

Un procedimiento efectivo para descomponer y modelar series temporales en agricultura

Francisco Aragón, Francisco Javier Baldán, Manuel Parra, José M. Benítez

Depto. Ciencias de la Computación e Inteligencia Artificial, DICITS, DaSCI

Universidad de Granada

Granada, España

{far, fjbaldan, manuelparra, J.M.Benitez}@decsai.ugr.es

Resumen—En este trabajo proponemos una forma innovadora de abordar casos reales de predicción de la producción de cultivos en una cooperativa. Nuestro enfoque consiste en la descomposición de la serie temporal original de los cultivos en sub-series temporales según una serie de factores, con el objetivo de generar un modelo predictivo del cultivo a partir de los modelos predictivos parciales de las sub-series. El ajuste de los modelos se realiza mediante un conjunto de técnicas estadísticas y de Aprendizaje Automático. Esta metodología se ha comparado con una metodología intuitiva que consiste en una predicción directa de las series temporales. Los resultados muestran que nuestro enfoque logra un mejor rendimiento de predicción que la manera directa, por lo que aplicar una metodología de descomposición es más adecuada para este problema que la no descomposición.

Palabras Clave—agricultura, predicción, modelos predictivos, series temporales, descomposición

I. INTRODUCCIÓN

La agricultura es una actividad económica que resulta de vital importancia en prácticamente todos los países del mundo. En los últimos años la mejora del rendimiento agrícola (cantidad de producción obtenida por área cultivada) había sido propiciada por avances en la maquinaria empleada, nuevas técnicas de siembra, así como mediante la mejora de semillas y agroquímicos o un mejor control de plagas y enfermedades. Pero ahora, la agricultura comercial se ha vuelto una actividad de alta tecnología, en la que los avances tecnológicos informáticos también tienen su aplicación ya que permiten generar información de alta calidad sobre los procesos productivos.

Por lo tanto, la agricultura puede beneficiarse del auge de las técnicas englobadas dentro del ámbito de la Inteligencia Artificial y del Análisis de Datos, para crear modelos predictivos que predigan situaciones futuras que sean de ayuda para mejorar tanto la productividad de las cosechas como la toma de decisiones en todo lo relativo a ellas. Dentro del ámbito de la agricultura una gran cantidad de procesos tales como la producción o la aparición de plagas entre otras, pueden modelarse como una colección de observaciones habitualmente ordenadas a lo largo de un periodo de tiempo, es decir, mediante una serie temporal. Por medio de la aplicación sobre las series temporales de un conjunto de técnicas estadísticas y de Aprendizaje Automático se pueden generar modelos

predictivos que sean capaces de extraer sus regularidades y de realizar predicciones.

Estos modelos predictivos resultantes de los análisis de los datos, pueden ser de gran importancia en las cooperativas y asociaciones de las que dependen los pequeños agricultores a la hora de la toma de decisiones en una gran cantidad de situaciones cotidianas relacionadas por ejemplo con la gestión de plagas, de los cultivos en el campo, o la gestión de la producción esperada ya sea para ofertarla en el mercado, para contratar el personal necesario para tratarla o para obtener el material necesario para su preparación y embalaje. La eficiencia en todas estas situaciones va a depender de la capacidad de transformar los datos brutos en información precisa que permita tomar la decisión acertada en cada una de ellas, siendo la gestión de la producción una de las actividades de mayor importancia. Por ello, disponer de información fiable sobre las expectativas de producción es crítico en el sector agrícola.

En este artículo se propone descomponer las series temporales del rendimiento de un cultivo en un conjunto de sub-series temporales en función de una serie de factores, siendo clave la semana de plantación, con la intención de que con la descomposición se obtengan sub-series con patrones más establecidos que la serie original general. El análisis y ajuste de los modelos predictivos de las sub-series temporales se realiza mediante una batería de métodos estadísticos como son los métodos ARIMA [3] y de Aprendizaje Automático como son las redes neuronales entre otras, para a partir de ellos, realizar una predicción general de los kilogramos del cultivo. La metodología propuesta se ha evaluado en varios casos de estudio reales, y se ha comparado con una metodología directa que consiste en una predicción de las series temporales sin descomponer. Los resultados muestran como se obtiene una mejora de la predicción al aplicar la metodología planteada con respecto a la serie sin descomponer.

El resto del artículo está organizado de la siguiente manera: la sección 2 describe el estado actual en que se encuentra la aplicación de modelos predictivos sobre problemas de predicción de cosechas modelados como series temporales. La sección 3 describe el problema, la forma en que ha llevado a cabo la descomposición, y la metodología empleada para resolverlo. En la sección 4 se detalla la experimentación realizada, se muestra el conjunto de datos y los resultados



obtenidos. En la sección 5 se discuten los resultados obtenidos. Por último, la sección 6 presenta las conclusiones obtenidas tras la realización del trabajo.

II. ESTADO DEL ARTE

En el ámbito de la agricultura, los investigadores han propuesto numerosos modelos y procedimientos para mejorar la predicción de cosechas. La mayoría de ellos han tenido un enfoque multivariante, en los que se trata de incorporar a la predicción el impacto que numerosas variables relativas a cultivos tales como la lluvia, la temperatura, los fertilizantes o la calidad del suelo tienen en la cosecha. Los métodos más ampliamente empleados para realizar este tipo de modelos multivariantes son las redes neuronales, más en concreto en su variante *feed-forward*. En [10] y [12] se desarrollan estudios sobre si redes neuronales con variables relacionadas al cultivo son adecuadas para realizar la predicciones futuras, realizan una evaluación de la capacidad predictiva de diferentes parametrizaciones de la red, y comparan los resultados con otros modelos de regresión. En otros trabajos como en [14] se ha realizado un estudio concreto de las variables óptimas para la predicción de cosecha de maíz. Existe un déficit de trabajos en cuanto a la predicción de cosechas que tengan un enfoque multivariante y que además empleen métodos distintos a redes neuronales. Algunos de esos trabajos son [6] y [9], que demuestran la capacidad de métodos de *Random Forests (RF)* para la estimación de cultivos, y en [4] para máquinas de soporte vectorial. En [16] se realiza una comparación entre diferentes metodos regresivos como redes neuronales, funciones de base radial o árboles de regresión.

En cuanto a un enfoque univariante, los métodos ARIMA son los más ampliamente usados en la predicción de cosechas. En [15] se emplean los métodos ARIMA para predecir el área y la producción de trigo en los próximos años. También los alisados exponenciales los acompañan en algunos trabajos como en [1]. En [5] se realiza una comparación sobre la capacidad predictiva entre los métodos ARIMA y los alisados exponenciales.

La mayoría de los trabajos que se han realizado en este ámbito se han realizado desde un punto de vista multivariante, centrando el estudio en la importancia del impacto de los factores endógenos sobre la variable a predecir. Las redes neuronales y los métodos ARIMA son los más empleados, siendo también aplicados aunque en menor medida las máquinas de soporte vectorial o los métodos *Random Forest*. Son menos los trabajos en los que se ha empleado un enfoque univariante que trabaje directamente sobre la serie temporal. Aún así, los trabajos tratan de una manera muy superficial a la serie temporal ya que no tratan de adaptarla o modificarla para obtener una serie más fácil de predecir, sino que se centran directamente en aplicar un método concreto, con diferentes entradas o diferentes parametrizaciones para obtener una mejor predicción. Además, los trabajos solo emplean un método para el análisis de la serie, obviando la posibilidad de combinar varios métodos sobre la misma serie mediante una subdivisión de la misma.

III. PROPUESTA

Para hacer una declaración más precisa se formaliza el problema de predicción abordado, por lo que a continuación se detalla nuestra propuesta.

III-A. Descripción del problema y formulación

El problema se ubica en el ámbito de una cooperativa de frutas y hortalizas. La cooperativa está compuesta de un conjunto de parcelas, pertenecientes a los agricultores asociados, y la producción se divide en campañas, que representan los años agrícolas, de septiembre a julio. Cada parcela planta un producto específico durante una campaña. La superficie de cultivo de las parcelas se mide en metros cuadrados y cada parcela tiene su propia superficie. Los productos que se plantan en la cooperativa son diferentes variedades de frutas y verduras y su producción se mide en kilogramos. La producción total de un producto está formada por la suma de la producción de cada parcela que planta el producto. Algunos de los productos se plantan durante un cierto periodo de tiempo, siguiendo un patrón de siembra similar en cada campaña, mientras que otros se plantan de forma irregular a lo largo de la campaña. La producción durante el ciclo de vida de un cultivo suele ser similar para el mismo producto, por lo que para los productos plantados en periodos cercanos la producción sigue un patrón similar.

El principal interés de la cooperativa reside en conocer la producción en unidades de peso (p.e. kilogramos) que producirá cada producto para la próxima semana, de manera que la cooperativa pueda manejar adecuadamente el volumen de producción que se tendrá. Por lo tanto, la mejor manera de medir la producción es de forma semanal, y el problema a tratar consiste en la predicción de la producción en kilogramos de producto para la próxima semana. Este problema se abordará descomponiendo las series temporales de producción en sub-series temporales de semanas de siembra significativas, y empleando sobre ellas técnicas de Inteligencia Artificial y métodos estadísticos para generar modelos predictivos cuyas predicciones se irán agregando para obtener finalmente la producción global.

III-B. Descomposición de las series temporales

La idea principal del enfoque de descomposición en sub-series temporales es capturar el comportamiento similar que debe tener la producción de un producto que ha sido plantado en el mismo corto periodo de tiempo, ya que su ciclo de vida será muy parecido. Por lo tanto, el principal criterio de descomposición está tomado de uno de los factores más influyentes en la producción: la fecha de siembra. Las fechas de plantación se agrupan por semanas. La descomposición se realiza en semanas de siembra significativas y habrá tantas sub-series temporales como semanas de plantación significativas se establezcan. Las semanas significativas son aquellas que presentan un gran número de parcelas en plantación y por tanto la mayor parte de la producción del producto. Para determinar el conjunto de semanas significativas, se computan

las frecuencias de las semanas de siembra durante las campañas anteriores para analizar el número de parcelas que han plantado el producto en cada semana, y por lo tanto elegir las semanas con una mayor frecuencia de parcelas en plantación. Las semanas cuyas frecuencias de siembra son pequeñas se agrupan en una o dos series temporales adicionales.

Para los casos particulares bajo estudio, la descomposición se hizo como se expresa en la Tabla I. La Tabla I revela la elección de semanas significativas para los productos, y por lo tanto el número de sub-series temporales. El producto 1 tiene tres importantes semanas consecutivas de plantación; 37, 38 y 39 que cubren el 70% de las parcelas. Para el producto 2 las semanas elegidas presentan cierta dispersión ya que tiene sus semanas significativas entre la 8 y la 11 y entre la 31 y la 38, representando el 67% del total de parcelas. En general, para ambos productos predomina la plantación en las semanas correspondientes a los meses de agosto y septiembre.

Tabla I
SEMANAS DE PLANTACIÓN SELECCIONADAS PARA CADA SUB-SERIE

Nº de la subserie	Semanas de plantación de la subserie	
	Producto 1	Producto 2
Sub1	37	8-9
Sub2	38	10-11
Sub3	39	31-32
Sub4	-	33-34
Sub5	-	35-36
Sub6	-	37-38
Sub7	-	Entre la 12 y la 30
Sub8	-	No entre la 12 y la 30

III-C. Metodología

Para resolver el problema, proponemos desarrollar modelos que predigan la producción de un producto para la semana siguiente. Esta producción se modela como una serie temporal del rendimiento histórico semanal del producto. El rendimiento del cultivo se utiliza como valor de la serie temporal en lugar de los kilogramos, ya que elimina el efecto de la influencia que tiene la superficie sobre el número total de kilogramos producidos por una parcela en un periodo. El rendimiento global de cada producto se obtiene dividiendo la suma total de los kilogramos producidos por las parcelas en producción en cada periodo de tiempo por la suma de las superficies de las parcelas que contribuyen en ese periodo de tiempo.

$$\text{Rendimiento de cultivo} = \frac{\sum kg}{\sum \text{superficie}} \quad (1)$$

La metodología propuesta para desarrollar los modelos predictivos se basa en la descomposición de las series temporales del rendimiento de los cultivos de un producto en sub-series temporales. En cada sub-serie temporal solo se tratan los cultivos que han sido plantados en un periodo de tiempo muy corto, por lo que como paso esencial para nuestra metodología se deben analizar y agrupar las semanas de plantación para aprovechar la homogeneidad en la duración y producción del cultivo y además cubrir el mayor número de parcelas posible.

Como resultado de esto, las sub-series temporales que se obtengan presentarán comportamientos más predecibles.

Una vez que se tienen las sub-series parciales de un producto, se ajusta un modelo predictivo para cada una de ellas. El proceso de construcción del modelo predictivo sigue los siguientes pasos:

1. Preprocesamiento: Se realiza una imputación de valores faltantes en las sub-series temporales en caso de que existan, para que todas las sub-series temporales tengan la misma longitud. Además, se realiza una normalización min-max entre el rango [0,1]:

$$\text{normalización } TS = \frac{TS - \min_{TS}}{\max_{TS} - \min_{TS}} \quad (2)$$

donde TS denota la sub-serie temporal original, y \min_{TS} y \max_{TS} los valores mínimo y máximo de la sub-serie temporal. Por último, las series parciales se dividen en datos de entrenamiento y datos de prueba.

2. Modelado: Se ajustan modelos predictivos de las sub-series temporales de entrenamiento utilizando un conjunto de técnicas estadísticas y de *machine learning* (ARIMA, alisados exponenciales, máquinas de soporte vectorial, *random forest*, redes neuronales, redes neuronales parcialmente recurrentes y modelos aditivos). Para cada una de las técnicas anteriores se generan un conjunto de modelos en función de las diferentes parametrizaciones empleadas.
3. Validación: El paso de validación determina en primer lugar cuál es el modelo más adecuado para cada técnica y, a partir de él, cuál es el mejor modelo global. El mejor modelo será aquel que obtenga una menor medida de error y por tanto una mayor capacidad predictiva. Para evaluar el error y la capacidad predictiva de cada uno de los modelos generados se aplica una validación cruzada *leave-one-out* [7] sobre las sub-series temporales de prueba. Para esta variante de la validación cruzada, existe un único valor de prueba para cada iteración, con la particularidad de que el conjunto de entrenamiento está formado por los valores que se producen temporalmente antes del valor de prueba. Así pues, no se usan valores futuros al valor de prueba en el entrenamiento del modelo. En cada iteración, el valor de prueba de la iteración anterior se incorpora al conjunto de entrenamiento, y el siguiente valor temporal que antes no ha sido usado se incorpora como valor de prueba.
4. Predicción: Después de completar los pasos anteriores, se tiene el modelo predictivo más adecuado para cada sub-serie temporal. A partir de estos modelos se genera una predicción para la semana siguiente del rendimiento de cultivo en cada sub-serie temporal. La predicción en kilogramos se obtiene multiplicando el rendimiento previsto obtenido de la cosecha y la superficie estimada de cultivo de las parcelas:

$$kg = \text{rendimiento} * \text{superficie} \quad (3)$$



El área de cultivo se estima a partir de las parcelas que han plantado en la presente campaña. Para ello se añade la superficie de todas las parcelas plantadas, menos la superficie de aquellas parcelas que han plantado y que no han podido producir al llevar plantadas menos días del tiempo que el producto necesita para empezar a producir, y aquellas parcelas que ya han terminado de producir.

5. Agregación: Para obtener la predicción global de kilogramos, se suman los valores pronosticados de kilogramos de cada sub-serie temporal.

$$KG = \sum_{i=1}^n kg_i \quad (4)$$

Dónde $i = \{1, 2, \dots, n\}$ es el número de sub-series temporales.

IV. ESTUDIO EMPÍRICO

En esta sección, se describe el estudio empírico que hemos diseñado para evaluar la validez y el desempeño de nuestra propuesta de predicción para el problema abordado. La metodología se ha aplicado a casos reales. Se detallan los conjuntos de datos, las medidas de rendimiento y los enfoques competitivos. Después, se presentan los resultados empíricos obtenidos tras la realización de los experimentos.

IV-A. Procedimiento

Para comprobar la validez, el enfoque propuesto en este trabajo se ha comparado con un enfoque de predicción directo del problema. Esta estrategia directa consiste en aplicar algunos métodos de predicción a la serie temporal completa sin ningún tipo de descomposición. Las diferentes técnicas de predicción empleadas son ampliamente utilizadas en el estado del arte actual. En particular, los modelos considerados son: ARIMA, redes neuronales, máquinas de soporte vectorial, alisados exponenciales, redes neuronales parcialmente recurrentes, *Random Forest* y el modelo aditivo *Facebook Prophet* [17]. Argumentamos que este enfoque directo obtiene una menor capacidad predictiva, y por lo tanto, un enfoque más efectivo es considerar la descomposición basada en la semana de siembra.

IV-B. Configuración

Las series temporales de los productos y las sub-series temporales son univariantes. El periodo de las series temporales y las sub-series temporales se ha establecido en 52. Cada observación de la serie temporal y las sub-series temporales representa un valor semanal. Los valores de las series temporales se dividen en conjunto de entrenamiento y conjunto de pruebas. El porcentaje de valores en entrenamiento para todas las series temporales es el primer 65% de los valores totales, dejando el 35% restante como conjunto de pruebas. La longitud de la serie temporal del producto 1 es de 248, 161 valores para entrenamiento y 87 para la prueba. Para el segundo producto la longitud es de 222, 144 valores para el entrenamiento y 78 para la prueba. Se ha utilizado el error

cuadrático medio (RMSE) y el error absoluto medio (MAE) como medida del rendimiento:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (o_t - \bar{p}e_t)^2} \quad (5)$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |o_t - \bar{p}e_t| \quad (6)$$

donde o_t denota el valor de rendimiento original de la sub-serie temporal, pe_t el valor de rendimiento predicho y T el número de valores de prueba. Por otro lado, todo el procesamiento se ha implementado en el lenguaje de programación R, utilizando gran cantidad de los paquetes disponibles en el repositorio CRAN [2], [8], [11], [13], [17].

IV-C. Conjunto de datos

Los datos de producción se han recogido de una cooperativa situada en el sur de España. Para este trabajo se han seleccionado dos cultivos:

- Producto 1. Vegetal de temporada. La producción de rendimiento para este producto está disponible desde septiembre de 2013 hasta abril de 2018, cubriendo las campañas 2013, 2014, 2015, 2016 y 2017. La plantación principal tiene lugar en el mes de septiembre, pero también se realizan pequeñas plantaciones en los meses de agosto y febrero.
- Producto 2. Fruta de temporada. La producción de rendimiento para este producto está disponible desde enero de 2014 hasta abril de 2018, cubriendo la mitad de la campaña de 2013 y las campañas completas de 2014, 2015 y 2016 y gran parte de la de 2017. La plantación principal ocurre en los meses de agosto-septiembre y en los meses de febrero-marzo.

IV-D. Resultados

La Tabla II muestra el RMSE en prueba obtenido con nuestra metodología propuesta y con el enfoque directo. La Tabla III muestra el MAE. En las filas de la primera columna de ambas tablas se muestra la metodología empleada, ya sea la metodología propuesta en este trabajo o la metodología directa. La segunda columna especifica el método de modelado empleado por la metodología, en el caso de la metodología propuesta se han empleado una combinación de todos los métodos para ajustar la serie, pero en el caso de la metodología directa se ha empleado solo un método en cada caso. La columna 3 se subdivide en dos columnas siendo cada una referente al error cometido en cada producto. Todos los valores RMSE y MAE están normalizados en el rango [0-1].

V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

Los resultados de la Tabla II y Tabla III muestran como nuestro enfoque logra una mejor capacidad predictiva que el enfoque directo para cada uno de los casos reales al obtener un error menor tanto en RMSE como en MAE.

Tabla II
RMSE NORMALIZADO OBTENIDO POR NUESTRA PROPUESTA Y DE MANERA DIRECTA SOBRE EL SUBCONJUNTO DE PRUEBA

Metodología	Modelo usado	Error cometido (RMSE)	
		Producto 1	Producto 2
Propuesta	Todos	0.0371	0.0368
Directa	ARIMA	0.0536	0.0458
Directa	Red Neuronal	0.0502	0.0446
Directa	SVM	0.0643	0.0685
Directa	Alisado Exponencial	0.0565	0.0475
Directa	Red Neuronal Parcialmente Recurrente	0.0457	0.0445
Directa	Random Forest	0.0658	0.0487
Directa	Modelo Aditivo	0.0600	0.0579

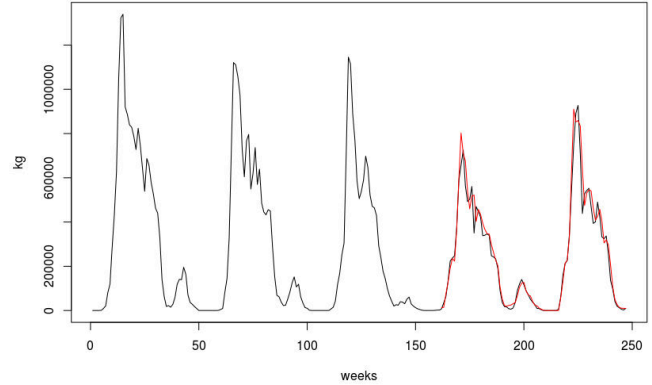
Tabla III
MAE NORMALIZADO OBTENIDO POR NUESTRA PROPUESTA Y DE MANERA DIRECTA SOBRE EL SUBCONJUNTO DE PRUEBA

Metodología	Modelo usado	Error cometido (MAE)	
		Producto 1	Producto 2
Propuesta	Todos	0.0214	0.0247
Directa	ARIMA	0.0346	0.0400
Directa	Red Neuronal	0.0331	0.0292
Directa	SVM	0.0485	0.0403
Directa	Alisado Exponencial	0.0384	0.0350
Directa	Red Neuronal Parcialmente Recurrente	0.0346	0.0302
Directa	Random Forest	0.0447	0.0343
Directa	Modelo Aditivo	0.0488	0.0488

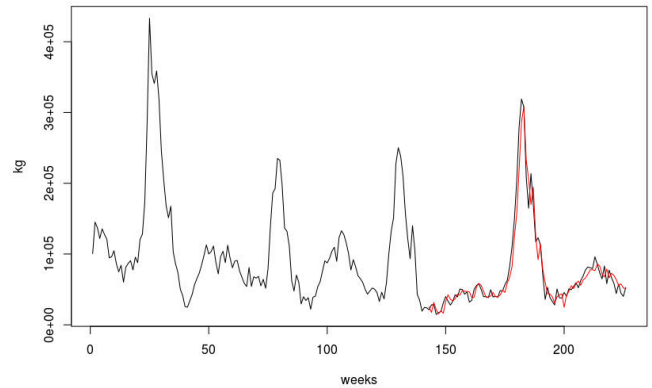
La mejora al emplear el enfoque con descomposición es considerable para el producto 1. El RMSE obtenido con nuestra metodología es de 0.0371, mientras que el menor RMSE obtenido con la metodología directa en cualquiera de sus variantes es de 0.0457 conseguido aplicando una red neuronal parcialmente recurrente. Para los ajustes con el resto de métodos en la metodología directa el RMSE es superior a 0.05 e incluso a 0.06. Por otro lado nuestra metodología obtiene un MAE más bajo de 0.0214, mientras que el menor MAE obtenido con la metodología directa es de 0.0331.

Mientras que para el producto 2 nuestra metodología ofrece también una mejor capacidad predictiva que la metodología directa. El RMSE más bajo obtenido con nuestro enfoque es del 0.0368 mientras que el menor de los RMSE obtenidos por el enfoque directo es de 0.0445, conseguido también con una red neuronal parcialmente recurrente. El resto de ajustes con el enfoque directo ofrecen un RMSE similar al 0.0445 conseguido por la red neuronal parcialmente recurrente, salvo varias excepciones. Con respecto al MAE el menor conseguido por nuestra metodología es de 0.0247, y el menor del enfoque directo es de 0.0292. El resto de variantes del enfoque directo obtienen un MAE superior al 0.03.

Así pues estos resultados muestran que tanto para el caso del producto 1 como el producto 2 los errores cometidos por nuestro enfoque son siempre menores que los errores obtenidos por el enfoque directo, siendo la mejora obtenida considerable.



(a) Producto 1



(b) Producto 2

Figura 1. Serie temporal de los kilogramos de producción real y predicción realizada por nuestra metodología (línea roja) en los casos de prueba para los dos productos analizados

VI. CONCLUSIONES

En este trabajo se ha presentado una forma innovadora de realizar el análisis de una serie temporal del rendimiento de la producción de productos reales de una cooperativa hortofrutícola mediante la descomposición de la serie temporal general en sub-series temporales según la semana de plantación, siendo el modelo predictivo final el resultado de la mezcla entre los mejores modelos ajustados para cada sub-serie a partir de un conjunto de métodos estadísticos y de Aprendizaje Automático. Además, esta metodología ha sido comparada con otra implementación posible del mismo problema.

Los resultados de los experimentos han mostrado que la metodología propuesta mejora la capacidad predictiva de la metodología comparada, ya que obtiene un error menor para ambos productos analizados. Además, siempre mejora el error obtenido en cada una de las comparaciones. Por lo que se puede concluir que una descomposición de la serie temporal según las semanas significativas de plantación se da lugar a



que se obtenga una mejora significativa en la predicción con respecto a la serie sin descomponer.

AGRADECIMIENTOS

Esta investigación ha sido parcialmente financiada por los Proyectos de Investigación Nacional TIN2013-47210-P, y TIN2016-81113-R y por el Proyecto de Investigación de Excelencia de la Junta de Andalucía P12-TIC-2958.

REFERENCIAS

- [1] Akram, M., Bhatti, I., Ashfaq, M., Khan, A.A. *Hierarchical Forecasts of Agronomy-Based Data*, American Journal of Mathematical and Management Sciences, 36(1), 49-65, 2017
- [2] Bergmeir C. and Benítez J.M. *Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS*. Journal of Statistical Software, 46(7), 1-26, 2012
- [3] Box G.E.P. and Jenkins G.M. *Time Series Analysis, Forecasting and Control*, 1970
- [4] Brdar S., Culibrk D., Marinkovic B., Crnobarac J., Crnojevic V. *Support Vector Machines with Features Contribution Analysis for Agricultural Yield Prediction*, Second International Workshop on Sensing Technologies in Agriculture, Forestry and Environment, 43-47, 2011
- [5] Choudhury, A. and Jones, J. *Crop yield prediction using time series models*, Journal of Economics and Economic Education Research., 15, 53-68, 2014.
- [6] Fukuda S., Spreer W., Yasunaga E., Yuge K., Sardud V. and Müller J. *Random Forests modelling for the estimation of mango (Mangifera indica L. cv. Chok Anan) fruit yields under different irrigation regimes*, Agricultural Water Management, 116(1), 142-150, 2013
- [7] Hastie T., Tibshirani R., Friedman, J. *The Elements of Statistical Learning*, Springer, New York, 2009.
- [8] Hyndman R.J. and Khandakar Y. *Automatic time series forecasting: the forecast package for R*. Journal of Statistical Software, 26(3), 1-22, 2008
- [9] Jeong J.H., Resop J., Mueller N., Fleisher D.H., Yun K., Butler E., Timlin D., Shim K., Gerber J., Reddy V., and Kim S.H. *Random Forests for Global and Regional Crop Yield Predictions*, PLoS One. 2016; 11(6): e0156571.
- [10] Ji, B., Sun Y., Yang S. and Wan J. *Artificial neural networks for rice yield prediction in mountainous regions*, Journal of Agricultural Science, 145: 249-26, 2007.
- [11] Karatzoglou A., Smola A., Hornik K. and Zeileis A. *kernlab - An S4 Package for Kernel Methods in R*. Journal of Statistical Software, 11(9), 1-20, 2004
- [12] Kaul M., Hill R.L., and Walthall C. *Artificial neural networks for corn and soybean yield prediction*, Agricultural Systems, 85(1): 1-18, 2005.
- [13] Liaw A. and Wiener M. *Classification and Regression by randomForest*. R News, 2(3), 18-22, 2002
- [14] Liu J., Goering C.E., and Tian L. *A neural network for setting target corn yields*, Transactions of the ASAE. Vol. 44(3): 705-713, 2007.
- [15] Najeeb I., Khuda B., Asif M. *Use of the ARIMA Model for Forecasting Wheat Area and Production in Pakistan*, Journal of Agriculture and Social Sciences, 1(2), 120-122, 2015
- [16] Ruß G. *Data Mining of Agricultural Yield Data: A Comparison of Regression Models*, In: Perner P. (eds) *Advances in Data Mining. Applications and Theoretical Aspects*, ICDM 2009. Lecture Notes in Computer Science, vol 5633.
- [17] Taylor S. and Letham B. *prophet: Automatic Forecasting Procedure. R package version 0.1*. 2017
- [18] Wuo W., Xue H. *An incorporative statistic and neural approach for crop yield modelling and forecasting*, Neural Computing and Applications, 21(1): 109-117, 2012.