

# **EXPLORACIÓN DEL FILTRO DE KALMAN COMO HERRAMIENTA DE PREDICCIÓN EN LA COSECHA DE BRÓCOLI**

## *KALMAN FILTER AND THE BROCCOLI CROP YIELD PREDICTION FOR A LOCAL FARM*

**Oscar Francisco González Ramírez**

Tecnológico Nacional de México / ITS Lerdo, México  
*mm2210886@itslerdo.edu.mx*

**Raymundo Juárez Del Toro**

Universidad Autónoma de Coahuila / FCA Torreón, México  
*r.juarez@uadec.edu.mx*

**Francisco Ruvalcaba Granados**

Tecnológico Nacional de México / ITS Lerdo, México  
*francisco.rg@itslerdo.edu.mx*

**José Cruz Olvera Ávila**

Tecnológico Nacional de México / ITS Lerdo, México  
*mm2210984@itslerdo.edu.mx*

**Recepción:** 27/febrero/2023

**Aceptación:** 12/octubre/2023

### **Resumen**

En este trabajo se propone la aplicación del Filtro de Kalman (FK) en Python, como herramienta de pronóstico en la estimación del rendimiento de producción de brócoli, en una granja agrícola local en Aguascalientes, México. La aplicación del FK se realiza en dos etapas recursivas, estimación y actualización, a partir de la existencia de un modelo lineal generalizado que describe el rendimiento de brócoli, obtenido previamente. El modelo utilizado para tal efecto es un modelo econométrico del tipo Vectorial Autorregresivo (VAR) y las variables que lo describen son meteorológicas. El brócoli convencional representa el principal producto de la finca agrícola con el 39% de la producción total de la granja agrícola en Aguascalientes, México. El objetivo de la previsión es mejorar la planificación, logística y administración de los recursos, humanos y materiales, disponibles para llevar a cabo esta tarea.

El FK es una herramienta matemática para la estimación estocástica a partir de observaciones con ruido. Esta herramienta recursiva, recibió su nombre de Rudolph E. Kalman, quien en 1960 describió la solución recursiva al problema del filtrado lineal de datos discretos.

El FK está compuesto por un sistema de ecuaciones que describen un estimador del tipo estimación-corrección. Este estimador es óptimo ya que minimiza la covarianza del error estimado, siempre que cumpla con ciertas condiciones específicas. Desde su aparición, el FK ha sido objeto de múltiples investigaciones y aplicaciones.

**Palabras Clave:** Agricultura de precisión, Filtro de Kalman, Producción agrícola.

### **Abstract**

*In this work, the application of the Kalman Filter (KF) in the Python programming language is proposed, as a forecasting tool in the estimation of broccoli production yield, in a local agricultural farm in Aguascalientes, Mexico. The application of the KF is carried out in two recursive stages, estimation and updating, based on the existence of a generalized linear model that describes the yield of broccoli, previously obtained. The model used for this purpose is an econometric model of the Vector Autoregressive (VAR) type and the variables that describe it are meteorological. Conventional broccoli represents the main product of the agricultural farm with 39% of the total production. The objective of the forecast is to improve the planning, logistic and administration of the resources, human and material, available to carry out this task.*

*The Kalman Filter is a mathematical tool for stochastic estimation from observations obtained from a noisy sensor. This recursive tool was named after Rudolph E. Kalman, who in 1960 described the recursive solution to the problem of linear filtering of discrete data. The FK is made up of a system of equations that describe an estimator of the estimation-correction type. This estimator is optimal since it minimizes the covariance of the estimated error, as long as it meets certain specific conditions.*

**Keywords:** Kalman Filter; Crop Yield; Agriculture automation.

## **1. Introducción**

Estimar el rendimiento de cosecha en fechas precisas mientras se identifican y seleccionan los mejores predictores meteorológico significativos es la idea básica de esta contribución. El modelo lineal de estimación de rendimiento elegido es un modelo econométrico vectorial autorregresivo (VAR). Este modelo lineal permite una detección y gestión temprana de problemas limitantes en la producción. La estimación de la producción de cosecha puede ayudar a aumentar los rendimientos en las fechas precisas y así obtener más ingresos para una finca local en Aguascalientes, México. Rancho Medio Kilo, en Aguascalientes México, es una finca social y ecológicamente responsable, que produce, procesa y distribuye vegetales como brócoli, coliflor, zanahoria, calabaza, acelga, espinaca, cebolla, lechuga, tomate, chile, etc. También es posible mejorar la rentabilidad de la explotación, así como equilibrar la cantidad de medios de producción utilizados, como mano de obra humana, maquinaria, recursos materiales y también para reaccionar frente a plagas o condiciones climáticas inesperadas. Además, un monitoreo temprano y confiable del rendimiento de los cultivos puede ayudar a los formuladores de políticas y las agencias de mercadeo agrícola a planificar las importaciones y exportaciones. El rendimiento de la cosecha aumenta las ganancias de las plantas de procesamiento. Las herramientas de pronóstico también brindan información valiosa a partir de la cual es posible estimar la cantidad de materia prima que la empresa podrá procesar. Los modelos estadísticos son las técnicas más comunes utilizadas para pronosticar el rendimiento de los cultivos, pero también lo son los modelos econométricos, que permiten la estimación del rendimiento durante la temporada de crecimiento a través del registro de los datos de cada variable en fechas específicas y que constituyen las llamadas series de tiempo. La predicción de cultivos para la temporada de crecimiento actual es una tarea relativamente difícil debido a la complejidad de la relación entre el proceso de crecimiento de la planta y los factores ambientales como el clima.

Una serie temporal es una serie de datos indexados en orden temporal. El pronóstico de series de tiempo se enfoca en la influencia de las reglas de un fenómeno dado para predecir tendencias futuras basadas en datos históricos [1]. El

modelo VAR es un modelo estadístico que se utiliza para capturar la relación entre múltiples cantidades a medida que cambian con el tiempo. El VAR es un tipo de modelo de proceso estocástico. Los modelos VAR generalizan el modelo autorregresivo de una sola variable (univariante) al permitir series temporales multivariantes. Los modelos VAR se utilizan a menudo en economía y ciencias naturales. En el modelo VAR, cada variable tiene una ecuación que modela su evolución en el tiempo. Esta ecuación incluye los valores pasados de la variable, los valores pasados de las demás variables del modelo y un término de error. Los modelos VAR solo requieren del conocimiento previo de una lista de variables de las que se puede suponer que se afectan entre sí a lo largo del tiempo.

En los últimos años, se ha sugerido el modelo de espacio de estado que integra modelos VAR en el filtro de Kalman [2]. Estos algoritmos de filtrado de Kalman pueden implementar la estimación óptima de los estados del sistema. Construye estimaciones de estado óptimo mediante métodos como la estimación de máxima verosimilitud o la solución de mínimos cuadrados y es de naturaleza autoadaptativa [3]. Esto mejora la precisión y la eficiencia de la predicción de series temporales. El filtro de Kalman es un estimador de estado estadístico. El filtro es extremadamente útil para estimar los estados inciertos de un sistema cuando la incertidumbre surge debido a entradas impredecibles como el ruido en la dinámica del sistema o en el sistema de medición. El filtro toma la forma de un algoritmo de corrección de predicción y procesa los datos secuencialmente. Es decir, se predice una estimación del estado del sistema utilizando la dinámica del modelo y la estimación del estado anterior. Este estado pronosticado luego se actualiza usando una corrección ponderada del error de predicción calculado usando la siguiente observación o punto de datos. La ponderación tiene en cuenta la cantidad de corrupción de ruido que se supone presente en la dinámica y en las observaciones.

El filtro de Kalman es un estimador de estado óptimo para sistemas lineales y suposiciones de corrupción de ruido gaussiana aditiva. El filtro de Kalman extendido es una versión subóptima del filtro de Kalman diseñado para tratar las no linealidades en la dinámica del sistema y en las funciones de medición. En los últimos tiempos, se ha acumulado un cuerpo de trabajo sobre el uso del algoritmo

de filtro de Kalman como una alternativa al entrenamiento estándar de descenso más pronunciado de funciones no lineales para aplicaciones como la predicción de series temporales. El FK también se ha utilizado para ayudar a resolver problemas en economía y agricultura. Ver [4] para más aplicaciones desde la mecatrónica en la agricultura y el trabajo de G. Maldonado sobre Agricultura Inteligente [5].

El brócoli es un cultivo que ha recibido una atención considerable con respecto a la predicción de la fecha de cosecha como en Marshall [6] y Pearson [7]. Wurr en [8] estableció una relación cuadrática entre el logaritmo natural del diámetro de la cabeza y los grados diarios efectivos acumulados desde el inicio de la cabeza, que podría usarse para predecir cuándo las cabezas alcanzan un tamaño específico. Se desarrolló un modelo informático para predecir las fechas de cosecha después de tomar muestras de la cabeza en una etapa joven. En [9] se desarrollaron modelos basados en aprendizaje profundo para evaluar cómo funcionan los algoritmos del aprendizaje automático (ML) con respecto a diferentes criterios de rendimiento. En esta contribución exploramos la técnica del Filtro de Kalman para pronosticar el rendimiento de los cultivos.

La presente contribución tiene la siguiente estructura, la sección 2 está dedicada a revisar los aspectos generales sobre la descripción de datos y variables que incluyen la vida de la planta del brócoli y los ciclos de cultivo. En la sección 3 presentamos métodos y ecuaciones significativas. En la sección 4 presentamos tanto el modelo VAR como la aplicación del FK. La sección 5 resume la contribución y finalmente se presentan las referencias importantes.

## **2. Descripción de los datos**

En el Rancho Medio Kilo, el brócoli convencional representa el 39% de la producción total de la finca. Solo se utilizan dos variedades de brócoli en la granja, etiquetadas como V106 y V194. La principal diferencia entre ellos es la capacidad de lidiar con el clima frío del invierno (V194), pero no es el centro de esta contribución. En periodos de cosecha la finca debe contratar personal, y debe preparar maquinaria, equipo y recursos adicionales. La estimación de la producción total (rendimiento) y la identificación de los mejores predictores meteorológicos

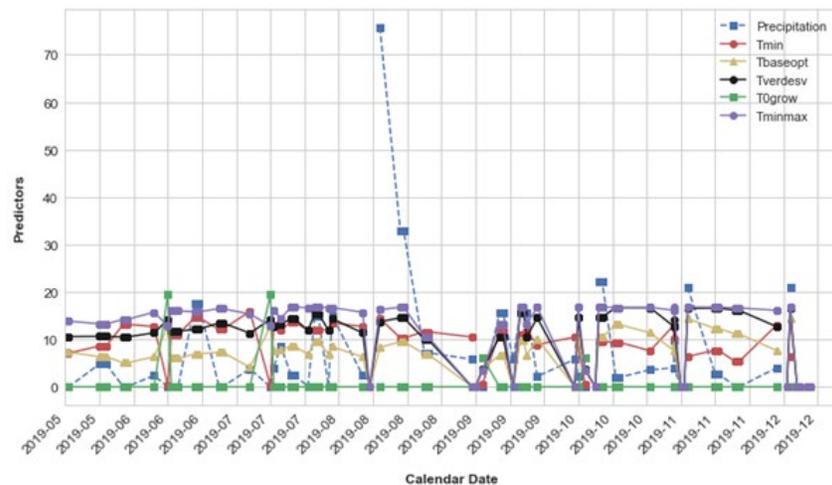
significativos son objetivos básicos de la producción agrícola [10]. En este trabajo se realiza el pronóstico de rendimiento de brócoli para tomar decisiones en cuanto a la planificación estacional del cultivo, y para equilibrar la cantidad de medios de producción utilizados, como mano de obra humana, maquinaria, anticipación de recursos, y también para estar preparados para reaccionar contra plagas o condiciones climáticas inesperadas. El rendimiento del cultivo se mide en kilogramos por hectárea (kg/Ha). La evaluación del pronóstico del rendimiento del cultivo se mide con respecto a su valor real en el siguiente intervalo de variación de  $[-15\%, 15\%]$ . El modelo estadístico utilizado para describir la producción total del brócoli es un modelo VAR y a este modelo se le aplica el FK para la estimación de los datos.

Los datos meteorológicos se recolectaron en la finca de mayo a diciembre de 2019. Los datos del clima fueron generados por una de las estaciones meteorológicas operadas por la oficina de agricultura mexicana llamada Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP), en Aguascalientes, México. Los datos meteorológicos tienen nueve variables que incluyen precipitación, radiación y temperaturas, como mínimo-máximo, base óptima, vernalización-desvernalización, crecimiento cero y crecimiento óptimo. La tabla 1 presenta las abreviaturas utilizadas aquí para las variables meteorológicas y la figura 1 ilustra el gráfico de estos predictores. La precipitación es cualquier producto de la condensación del vapor de agua atmosférico que cae bajo la atracción gravitacional de las nubes.

Tabla 1 Predictores meteorológicos para la producción total del brócoli.

<b>Variable</b>	<b>descripción</b>
Precipitation	Precipitación
Radiation	Radiación solar
Tmin	Temperatura mínima
Tmax	Temperatura máxima
Tbaseopt	Temperatura base óptima
Tverdesv	Temperatura Vernalización-Desvernalización
T0grow	Temperatura cero de crecimiento
Tminmax	Temperatura Mínima-Máxima
Toptgrow	Temperatura de crecimiento óptima

*Fuente: Elaboración propia.*



Fuente: Elaboración propia en Python

Figura 1 Predictores meteorológicos de la producción del brócoli.

Las principales formas de precipitación incluyen llovizna, lluvia, aguanieve, nieve, gránulos de hielo, granizo y granizo. La radiación solar es la energía radiante emitida por el sol que crea energía electromagnética. Las unidades de medida son vatios por metro cuadrado. La temperatura influye en la mayoría de los procesos de las plantas, incluida la fotosíntesis, la transpiración, la respiración, la germinación y la floración. A medida que la temperatura aumenta hasta cierto punto, aumentan la fotosíntesis, la transpiración y la respiración. Cuando se combina con la duración del día, la temperatura también afecta el cambio de crecimiento vegetativo a reproductivo. Según la situación y la planta específica, el efecto de la temperatura puede acelerar o ralentizar esta transición. Las plantas crecen mejor cuando la temperatura del día es de 10 a 15 grados más alta que la temperatura de la noche. Esta temperatura se denomina temperatura cero y temperatura óptima de crecimiento. La vernalización es la inducción del proceso de floración de una planta por exposición al frío prolongado del invierno, o por un equivalente artificial. Después de la vernalización, las plantas han adquirido la capacidad de florecer, pero pueden requerir señales estacionales adicionales o semanas de crecimiento antes de que realmente florezcan. Las temperaturas típicas de vernalización están entre 1 y 7 °C. Es posible desvernalizar una planta exponiéndola a veces a temperaturas altas y bajas posteriores a la vernalización. La temperatura base es la temperatura fresca a la que una planta no se desarrolla. En la temperatura base, o por debajo

de esta, las plantas ya no desarrollan hojas y cesa el progreso hacia la floración. A medida que la temperatura aumenta por encima de la temperatura base, las plantas se desarrollan progresivamente más rápido. El valor de temperatura base de un cultivo se puede estimar experimentalmente cultivando plantas en un rango de temperaturas, registrando el tiempo hasta la floración y luego extrapolando de los datos la temperatura en la que el tiempo para florecer es infinito. El brócoli convencional es el cultivo objetivo en este trabajo. El rendimiento del cultivo es el principal objetivo de pronóstico.

### 3. Métodos y técnicas

El modelo autorregresivo, que describe un proceso variable en el tiempo, especifica que la variable estimada solo depende linealmente de sus propios datos previos y de un término estocástico [11]. La notación  $AR(p)$  se define como se muestra en la ecuación 1.

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \epsilon_t \quad (1)$$

Donde  $y_t$  es la serie de tiempo bajo investigación,  $\alpha_i$  denota el coeficiente de autorregresión,  $p$  representa el orden del modelo que generalmente es mucho más pequeño que la longitud de la serie, y  $\epsilon_t$  siempre se supone que es el ruido blanco gaussiano. El modelo autorregresivo asume que la serie  $x_t$  es estacionaria, lo que significa que la distribución de probabilidad conjunta del proceso estocástico no cambia cuando se desplaza en el tiempo. Sin embargo, si consideramos la serie como un todo, el orden de toda la serie puede no ser el más apropiado para cada paso de tiempo porque la serie puede ser no estacionaria.

En el filtrado clásico y la predicción de sistemas dinámicos por medio de la transición de estado, uno está sujeto a muchas limitaciones que pueden tener un impacto adverso en sus aplicaciones prácticas [12]. Para eludir las dificultades que reducirán la utilidad en el mundo real, el filtro de Kalman proporciona un medio recursivo eficiente para estimar el espacio de estado y es muy poderoso para resolver problemas de filtrado lineal [13]. Los filtros de Kalman, basados en un sistema

dinámico lineal discretizado en el dominio del tiempo, pueden estimar el estado actual con el estado estimado del paso de tiempo anterior y la medición actual. En otras palabras, este modelo asume que el estado de un sistema en el momento  $t$  se basa en el estado anterior en el momento  $t - 1$  y obtiene una observación  $y_t$  del estado verdadero  $x_t$  en el momento  $t$  de acuerdo con el modelo de espacio de estado mostrado en las ecuaciones 2 y 3.

$$x_t = A_t x_{t-1} + w_t \quad (2)$$

$$y_t = H_t x_t + v_t \quad (3)$$

Donde  $x_t$  denota el vector de estado en el paso de tiempo  $t$ ,  $A_t$  representa la matriz de transición de estado basada en el parámetro de estado del sistema en el paso de tiempo  $t - 1$ ,  $w_t$  es el vector de ruido del proceso en el paso de tiempo  $t$ , es decir,  $p(w_t) \sim N(0, Q_t)$ ;  $y_t$  denota el vector de mediciones,  $H_t$  representa la matriz de transformación que asigna los parámetros del vector de estado al dominio de medición,  $v_t$  es el término de ruido de medición en el paso de tiempo  $t$ , es decir,  $p(v_t) \sim N(0, R_t)$ . El filtro de Kalman consta de dos etapas: predicción y actualización de la medición. Las ecuaciones relacionadas con la etapa de predicción son las ecuaciones 4 y 5.

$$\hat{x}_{t|t-1} = A_t \hat{x}_{t-1|t-1} \quad (4)$$

$$P_{t|t-1} = A_t P_{t-1|t-1} A_t^T + Q_t \quad (5)$$

Donde  $Q_t$  denota la matriz de covarianza del ruido del proceso asociada con las entradas de control ruidosas,  $\hat{x}_{t|t-1}$  es la media del estado predicho,  $P_{t|t-1}$  representa la matriz de covarianza del error. Se realiza la estimación previa  $\hat{x}_{t|t-1}$  y  $P_{t|t-1}$  para la actualización de la medición. Como  $y_t$  se ha obtenido en el tiempo  $t$ , este modelo puede actualizar la media condicional y la matriz de covarianza de las estimaciones según las ecuaciones 6,7 y 8.

$$K_t = P_{t|t-1} H_t^T [H_t P_{t|t-1} H_t^T + R_t]^{-1} \quad (6)$$

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t [y_t - H_t \hat{x}_{t|t-1}] \quad (7)$$

$$P_{t|t} = P_{t|t-1} - K_t H_t P_{t|t-1} \quad (8)$$

Donde  $K_t$  es la matriz de ganancia de Kalman que minimiza la covarianza del error a posteriori,  $p_{t|t}$  denota la matriz de covarianza del error actualizada. Una vez que este modelo obtiene el resultado de la siguiente medición, estas estimaciones se actualizan en base a un promedio ponderado. El algoritmo es recursivo y no requiere ninguna información pasada adicional, solo tiene en cuenta las medidas de entrada actuales y el estado calculado previamente y su matriz de incertidumbre. Para resolver el problema de que el modelo  $AR(p)$  clásico solo se puede ajustar utilizando datos estacionarios, el modelo  $AR(p)$  se integra en el filtro de Kalman, ya que el FK puede aceptar datos no estacionarios. Se supone que la covarianza del ruido del proceso  $Q_t$  y los términos del ruido de medición  $R_t$  son constantes desconocidas,  $Q$  y  $R$ , respectivamente, lo que puede simplificar el FK.

#### 4. Resultados

El análisis de correlación es el análisis estadístico para realizar la mejor selección de predictores. El factor de correlación denominado coeficiente de Pearson, en el intervalo  $[-1,1]$ , determinará el orden de importancia respecto a la previsión. El análisis de correlación permitió seleccionar los principales predictores significativos. Una correlación perfecta también es un efecto indeseable en las relaciones variables y debe evitarse. Aquí se realiza un análisis de correlación de 99% entre la variable de rendimiento de cultivo y todo el grupo Clima, y todas las observaciones en todo el período de tiempo en la base de datos, de mayo a diciembre de 2019. La tabla 2 presenta los predictores más significativos de Rendimiento de cultivo variables.

Tabla 2 Coeficientes de correlación de Pearson de los predictores meteorológicos de la producción del brócoli.

Variable	Coficiente de Pearson
Precipitación	0.239
Tmin	0.209
Tbaseopt	-0.406
Tver/desver	-0.28
Tminmax	-0.128
T0grow	0.139

Fuente: Elaboración propia.

Está claro que el número de predictores significativos del rendimiento del cultivo es limitado. De todo el grupo meteorológico, solo el 67% de los predictores son significativos. Dado que los coeficientes de Pearson son cercanos a cero, la influencia del descriptor y la eficiencia de predicción también son limitadas. Estos hechos limitan fuertemente la calidad de la predicción. Los valores absolutos de los coeficientes de Pearson están en el intervalo [0.142, 0.406]. Para mejorar la calidad de la predicción es necesario ampliar el número de variables meteorológicas y el periodo de recogida de datos. Se realizó una escala de normalización a algunas variables meteorológicas con el fin de equilibrar y mejorar el análisis de correlación. Los predictores significativos de la tabla 2 son la base del modelo VAR y el análisis KF de la siguiente sección.

El modelo VAR describe la evolución de un conjunto de  $k$  variables, denominadas variables endógenas, a lo largo del tiempo. Cada período de tiempo está numerado,  $t = 1, \dots, T$ . Las variables se recogen en un vector,  $x_t$ , que es de longitud  $k$ . De manera equivalente, este vector podría describirse como una matriz. El vector se modela como una función lineal de su valor anterior. Los componentes del vector se denominan  $x_{it}$ , lo que significa la observación en el tiempo  $t$  de la  $i$ -ésima variable. Por ejemplo, si la primera variable del modelo mide el precio del trigo a lo largo del tiempo,  $x_{1,1998}$  indicaría el precio del trigo en el año 1998.

Los modelos VAR se caracterizan por su orden, que se refiere a la cantidad de períodos de tiempo anteriores que utilizará el modelo. Un retraso es el valor de una variable en un período de tiempo anterior. Entonces, en general, un VAR de  $p$ -ésimo orden se refiere a un modelo de VAR que incluye retrasos para los últimos  $p$  períodos de tiempo. Un VAR de  $p$ -ésimo orden se denota VAR( $p$ ) y, a veces, se denomina VAR con rezagos  $p$ . En este trabajo consideramos solo el caso de un modelo VAR de primer orden que para este caso es la ecuación del sistema 9.

$$x_t = A_t x_{t-1} + c + e_t \quad (9)$$

Las variables de la forma  $y_{t-i}$  indican el valor de esa variable en periodos de tiempo anteriores y se denominan retraso  $i$ -ésimo de  $y_t$ . La variable  $c$  es un vector  $k$  de constantes que sirve como intersección del modelo.  $A_j$  es una matriz invariante en

el tiempo y  $e_t$  es un vector  $k$  de términos de error. Los términos de error deben cumplir tres condiciones:  $E(e_t) = 0$ , eso significa que todo término de error tiene una media de cero;  $E(e_t, e'_t) = 0$ , lo que significa que la matriz de covarianza contemporánea de los términos de error es una matriz semidefinida positiva denominada  $\Omega$ ;  $E(e_t, e'_{t-k}) = 0$ , para cualquier  $k$  distinto de cero, lo que significa que no hay correlación a lo largo del tiempo. En particular, no existe una correlación serial en términos de error individuales [1].

El modelo VAR se obtuvo como modelo 1-VAR sin vector constante  $c$ . Esta se obtuvo en el software econométrico Eviews® 12 utilizando las variables climáticas: PRECIPITATION TMIN, TBASEOPT, TVERDESV, T0GROW, TMINMAX. Luego, se formó el vector del sistema con el PRODUCCIONTOTAL y las variables anteriores,  $x_t = [\text{PRODUCCIONTOTAL}, \text{PRECIPITATION}, \text{TMIN}, \text{TBASEOPT}, \text{TVERDESV}, \text{T0GROW}, \text{TMINMAX}]$ . Solo se utilizaron 449 observaciones de 2450. Los valores  $R$ -cuadrado y  $R$ -cuadrado ajustado fueron 0.927 y 0.926, respectivamente. Y la matriz  $A_t$  tiene los siguientes valores:

$$A_t = \begin{bmatrix} 0.966133 & 23.26474 & 2.570248 & 19.88996 & -24.12823 & -4.753366 & 10.6537 \\ 1.83E-6 & 0.979752 & -0.008785 & -0.02755 & -0.004803 & 0.012613 & 0.006146 \\ -7.17E-7 & 0.012533 & 0.965301 & 0.002909 & -0.007855 & -0.006131 & 0.006935 \\ 3.57E-5 & -0.241020 & 0.567001 & 0.860126 & 0.1789 & 0.257049 & -0.112196 \\ 3.74E-5 & -0.258291 & 0.71995 & -0.162485 & 1.192083 & 0.368286 & -0.123863 \\ 4.47E-6 & -0.039351 & 0.072001 & -0.007102 & 0.023956 & 0.989412 & -0.019757 \\ 2.91E-5 & -0.167122 & 0.639954 & -0.167704 & 0.188575 & 0.387141 & 0.884157 \end{bmatrix}$$

La ecuación del sistema se completa en la forma:  $x_t = A_t x_{t-1} + w_t$ . Con este modelo VAR al implementar el algoritmo FK, presenta los resultados mostrados en figura 2.

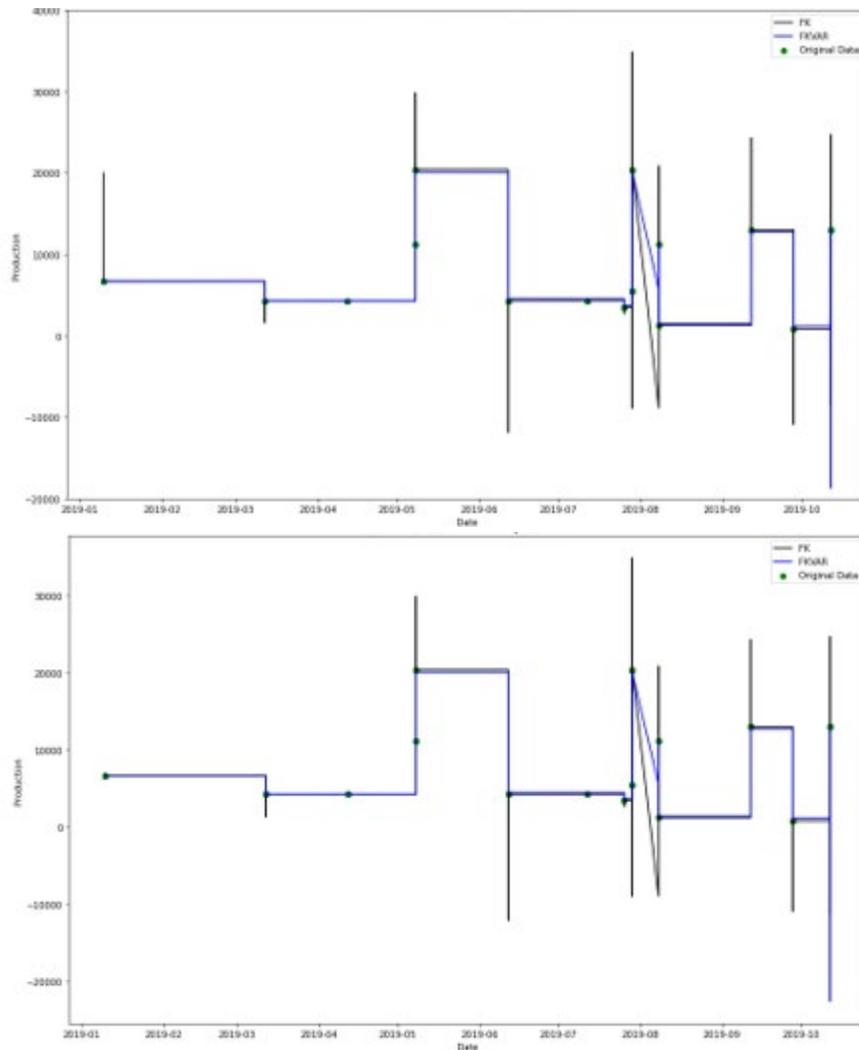
R-squared	0.927227	0.960418	0.995024	0.947066	0.951126	0.983787	0.969414
Adj. R-squared	0.926240	0.959880	0.994956	0.946348	0.950462	0.983567	0.968999
Sum sq. resids	1.22E+09	601.6336	63.57285	24532.97	40037.15	380.9692	32634.48
S.E. equation	1659.286	1.166689	0.379249	7.450132	9.517446	0.928397	8.592652
F-statistic	938.6191	1787.425	14730.64	1318.011	1433.599	4469.942	2334.856
Log likelihood	-3962.526	-702.7979	-198.2426	-1535.272	-1645.230	-600.2171	-1599.333
Akaike AIC	17.68163	3.161683	0.914221	6.869808	7.359598	2.704753	7.151519
Schwarz SC	17.74566	3.225713	0.978250	6.933837	7.423627	2.768782	7.219189
Mean dependent	8299.407	3.440980	7.670379	81.37082	116.8820	5.211581	132.0535
S.D. dependent	6109.553	5.824725	5.340173	32.16399	42.76140	7.242206	48.80217
Determinant resid covariance (dof adj.)		3216713.					
Determinant resid covariance		2881667.					
Log likelihood		-7798.910					
Akaike information criterion		34.95728					
Schwarz criterion		35.40549					
Number of coefficients		49					

Vector Autoregression Estimates  
Date: 02/05/23 Time: 18:16  
Sample (adjusted): 2 450  
Included observations: 449 after adjustments  
Standard errors in ( ) & t-statistics in [ ]

Fuente: Elaboración propia en Eviews® 12.

Figura 2 Coeficientes del análisis de estimación del modelo VAR para los predictores meteorológicos.

La figura 3 presenta las observaciones del Rendimiento del Cultivo en 2019; el modelo VAR evaluado sobre las observaciones; y la estimación del FK. El primer gráfico de la Figura 6 tiene también la contribución del ruido gaussiano y una condición inicial alejada de la observación inicial. Como el lector puede ver, el FK constituye una buena estimación para la producción total del cultivo en ambos casos. A pesar del error de estimación, el FK es una herramienta numérica apropiada para pronosticar la producción total.



Fuente: Elaboración propia en Python.

Figura 3 Comportamiento del FK en la estimación de producción total del brócoli en 2019.

La figura 3 ilustra el error cuadrático medio para la estimación del FK con respecto al rendimiento real del cultivo en 2019. Aquí, los 12 últimos puntos del gráfico

representan 12 puntos de pronóstico puro sin el conocimiento de las observaciones. En lugar de observaciones, los últimos 12 puntos de la Figura 3 provienen de los últimos 12 puntos pronosticados del sistema proporcionado por el algoritmo del FK.

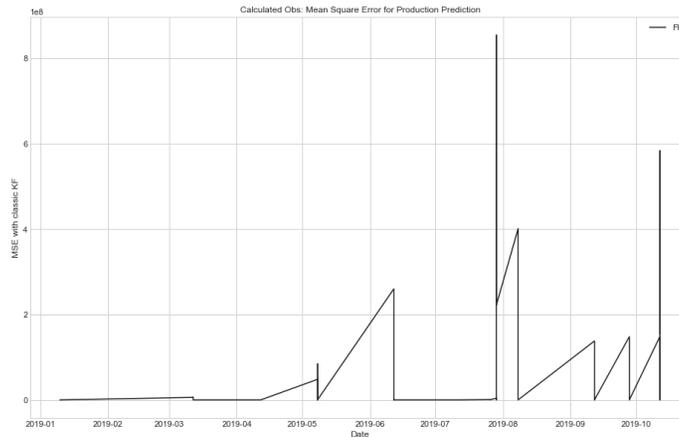


Figure 4 Error medio cuadrático de las estimaciones del FK.

## 5. Conclusiones

En este trabajo se exploró la aplicación del Filtro de Kalman (FK) clásico en Python, como herramienta de pronóstico en la estimación del rendimiento de producción de brócoli, en una granja agrícola local en Aguascalientes, México. La aplicación del FK se realiza en dos etapas recursivas, estimación y actualización, a partir de la existencia de un modelo lineal generalizado que describe el rendimiento de brócoli, obtenido previamente. El modelo utilizado para tal efecto es un modelo econométrico del tipo Vectorial Autorregresivo (VAR) y las variables meteorológicas que lo describieron fueron las siguientes: PRECIPITATION TMIN, TBASEOPT, TVERDESV, T0GROW, TMINMAX. El brócoli convencional representa el principal producto de la finca agrícola con el 39% de la producción total de la granja agrícola en Aguascalientes, México. Se compararon en una sola gráfica las observaciones de la producción total del brócoli, el modelo VAR evaluado en los puntos de observación y la estimación del FK evaluado también en los puntos de observación. Sin embargo, los 12 últimos puntos de la gráfica en estas tres variables descritas anteriormente, ilustran que: El modelo VAR evaluado en otros puntos diferentes a las observaciones presenta diferencias significativas respecto a los 12

observaciones reales; el FK evaluado en las predicciones propias y previas del algoritmo predicen con significativa ventaja a los observaciones, esto respecto a tres aspectos, el comportamiento del modelo VAR, las observaciones reales y al hecho de que el FK es de los pocos algoritmos que puede hacer predicciones fuera del rango de datos. Esta última conclusión se confirma aún bajo la presencia de ruido gaussiano en las observaciones reales. Finalmente, se espera que la predicción obtenida en este trabajo mejore la planificación, logística y administración de los recursos, humanos y materiales, disponibles para llevar a cabo esta tarea en la finca.

## **6. Bibliografía y Referencias**

- [1] S. Kendall and K. Ord, *Time Series*. Oxford University Press, 1990.
- [2] I. Arasaratnam and S. Haykin, "Cubature kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, 2009.
- [3] G. Kitagawa and W. Gersch, "A smoothness priors state space modeling of time series with trend and seasonality," *Journal of the American Statistical Association*, vol. 79, no. 386, pp. 378–389, 1984.
- [4] Negrete, J.C. (2016). *Mechatronics in Mexican Agriculture Current Status and Perspectives*.
- [5] Maldonado-Guzman, G., Garza-Reyes, J. A., & Solano-Romo, L. I. (2019). *Intelligent Agriculture: Developing a System for Monitoring and Controlling Production*. Emerald Group Publishing.
- [6] Marshall, B., Thompson, R., 1987. Applications of a model to predict the time to maturity of calabrese brassica oleracea. *Annals of Botany* 60, 521–529.
- [7] Pearson, S., Hadley, P., Wheldon, A., 1993. A reanalysis of the effects of temperature and irradiance on time to flowering in chrysanthemum (*dendranthema grandiflora*). *Journal of horticultural science* 68, 89–97.
- [8] Oikonomidis, A., Catal, C., Kassahun, A., 2022. Hybrid deep learning-based models for crop yield prediction. *Applied artificial intelligence*, 1–18.
- [9] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

- [10] Wurr, D., Fellows, J.R., Hambidge, A.J., 1992. The effect of plant density on calabrese head growth and its use in a predictive model. *Journal of Horticultural Science* 67, 77–85.
- [11] Hara, P., Piekutowska, M., Niedbala, G., 2021. Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data. *Land* 10, 609.
- [12] Jaramillo, N., Díaz, D., et al., 2006. El cultivo de las crucíferas: Brócoli, coliflor, repollo, col china. Technical Report. Corporacion Colombiana de Investigación Agropecuaria - AGROSAVIA.
- [13] B. Gary and W. Greg, “An introduction to the kalman filter,” *Proc of SIGGRAPH, Course*, vol. 8, no. 27599–23175, p. 41, 2001.