



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

DOCTORADO EN CIENCIAS POR INVESTIGACIÓN

**MEJORA DE LA CALIDAD DE DATOS
DEL RENDIMIENTO DE MAÍZ EN EL DR
075 RÍO FUERTE, SIN., APLICANDO UN
FILTRO DE KALMAN SIMPLIFICADO**

FÉLIX ALBERTO LLERENA VILLALPANDO

T E S I S

PRESENTADA COMO REQUISITO PARCIAL

PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

MONTECILLO, TEXCOCO, ESTADO DE MÉXICO, MÉXICO

2022



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

La presente tesis titulada: **MEJORA DE LA CALIDAD DE DATOS DEL RENDIMIENTO DE MAÍZ EN EL DR 075 RÍO FUERTE, SIN., APLICANDO UN FILTRO DE KALMAN SIMPLIFICADO** realizada por el estudiante: **FÉLIX ALBERTO LLERENA VILLALPANDO** bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN CIENCIAS

CONSEJO PARTICULAR

CONSEJERO

DR. OSCAR LUIS PALACIOS VÉLEZ

ASESOR

DR. IRINEO LORENZO LÓPEZ CRUZ

ASESOR

DR. CARLOS ALBERTO ORTÍZ SOLORIO

ASESOR

DR. JORGE FLORES VELÁZQUEZ

ASESOR

DR. WALDO OJEDA BUSTAMANTE

Montecillo, Texcoco, Estado de México, noviembre de 2022

MEJORA DE LA CALIDAD DE DATOS DEL RENDIMIENTO DE MAÍZ EN EL DR 075 RÍO FUERTE, SIN., APLICANDO UN FILTRO DE KALMAN SIMPLIFICADO

Félix Alberto LLerena Villalpando, D.C.
Colegio de Postgraduados, 2022

RESUMEN

Los datos del rendimiento y de la producción de maíz de grano en las principales zonas productoras de México, son una base importante para llevar a cabo una adecuada planeación agrícola, así como para el establecimiento de políticas públicas. Sin embargo, generalmente dicha información no es del todo precisa debido a que hay fallas en los procedimientos que se siguen para obtenerla, por lo que, si se mejora, se coadyuvará a potenciar su uso. Con respecto a la predicción de rendimiento de cultivos, existe mucha literatura sobre los modelos que se utilizan y una gran cantidad de variables de entrada para generarlos, pero en la mayoría de los casos la información de dichas variables no está disponible para poderla utilizarla y aplicarla en grandes áreas. Por ello, se decidió utilizar variables climatológicas para generar los modelos predictivos de rendimiento, con la finalidad de desarrollar una metodología más práctica que utilice datos de variables accesibles, ya que los datos climáticos son los únicos que están disponibles en la mayoría de las regiones del país. También se presenta el uso del script *Climatol* de R para la obtención de datos climáticos faltantes, así como el uso de la herramienta *systemIdentification* de MATLAB para generar los modelos de predicción de rendimiento y la de GAM de R para obtener datos predictivos. Se aplicó un Filtro de Kalman simplificado (FKS), para mejorar la calidad de los datos existentes de los rendimientos de maíz en el DR 075 Río Fuerte, Sin., que es una variante de la herramienta matemático-estadística desarrollada en 1960 para mejorar la predicción de variables de estado, y que tiene numerosas aplicaciones en la ciencia y tecnología. La aplicación del FKS se hizo utilizando los datos obtenidos con un modelo seleccionado para la predicción del rendimiento de maíz de grano en dicho DR 075. La aportación del FKS se expresa en el hecho de que los valores mejorados siempre quedan entre los predichos y los medidos.

Palabras claves: Maíz, modelos de predicción; Modelos ARX, Modelos computarizados, rendimiento, Filtro de Kalman Simplificado.

IMPROVEMENT OF THE QUALITY OF CORN YIELD DATA IN DR 075 RÍO FUERTE, SIN., APPLYING A SIMPLIFIED KALMAN FILTER

Félix Alberto LLerena Villalpando, D.C.
Colegio de Postgraduados, 2022

ABSTRACT

Data on the yield and production of grain corn in the main producing areas of Mexico are an important basis for carrying out adequate agricultural planning, as well as for the establishment of public policies. However, this information is generally not entirely accurate because there are flaws in the procedures followed to obtain it, so if it is improved, it will help to promote its use. Regarding the prediction of crop yield, there is a lot of literature on the models that are used and a large number of input variables to generate them, but in most cases the information of these variables is not available to be able to use and apply it. in large areas. For this reason, it was decided to use climatological variables to generate the predictive performance models, with the aim of developing a more practical methodology that uses data from accessible variables, since climatic data are the only ones available in most regions of the world. country. The use of the R script *Climatol* to obtain missing climatic data is also presented, as well as the use of the MATLAB *systemIdentification* tool to generate the performance prediction models and the R GAM tool to obtain predictive data. A Simplified Kalman Filter (FKS) was applied to improve the quality of the existing data on maize yields in the DR 075 Río Fuerte, Sin., which is a variant of the mathematical-statistical tool developed in 1960 to improve the prediction of state variables, and which has numerous applications in science and technology. The application of the FKS was made using the data obtained with a selected model for the prediction of grain corn yield in said DR 075. The contribution of the FKS is expressed in the fact that the improved values are always between those predicted and those measured.

Keywords: Maize, prediction models; ARX models, Computer models, performance, Simplified Kalman Filter.

DEDICATORIA

Especialmente a mi queridísima esposa María del Carmen Bejarano Valenzuela

A mis padres que me formaron y a quienes les debo todo, David H. LLerena Lanzagorta, un distinguido y gran agrónomo chapinguero (QEPD) y Beatriz Villalpando Pazos (QEPD)

A mis hijos: Beatriz del Carmen, Alberto y Miguel Ángel

A mis 6 hermanos: David, Sara Beatriz, Enrique (QEPD), Jorge Alfredo, Silvia Elena y Ana Laura

A mis nueras: Tania y Tatiana

A mis nietos: Mónica, Alberto y Regina

A mi yerno: Juan Carlos Lastiri Trinidad

A todos mis amigos, especialmente a los de la generación 1965-1969 de la ENA

Al Ing. Ildfonso de la Peña de la Torre (QEPD)

Al Ing. Gerardo Cruickshank García (QEPD)

AGRADECIMIENTOS

A mi Escuela Nacional de Agricultura, hoy Universidad Autónoma Chapingo

Al Colegio de Postgraduados campus Montecillo, distinguidísima institución que ha contribuido grandemente a mi desarrollo profesional y a mi Doctorado por Investigación

Al Dr. Oscar Luis Palacios Vélez, extraordinario y visionario investigador y docente, por todo el apoyo otorgado como consejero para la realización de la presente tesis

Al Dr. Irineo Lorenzo López Cruz, por compartir y apoyar como asesor sus conocimientos de MATLAB y Climatol, entre otros, que coadyuvaron a complementar la presente tesis

A los Doctores y asesores Carlos Alberto Ortiz Solorio, Jorge Flores Velázquez y Waldo Ojeda Bustamante, por todas las sugerencias y recomendaciones recibidas

A los Ingenieros Enrique Caballero Obscura y Jorge Lomelí Osuna y al Dr. Ernesto Sifuentes Ibarra, por todo el apoyo recibido en mis visitas al DR 075 en Los Mochis, Sinaloa

A los Doctores Jaime Arturo Matus Gardea y Juan Antonio Leos Rodríguez, por su participación como sinodales interno y externo respectivamente y por la amistad que me han brindado hace más de 58 años

A Miguel Servín Palestina, por compartir sus conocimientos de Climatol

CONTENIDO

RESUMEN	iii
ABSTRACT	iv
DEDICATORIA	v
AGRADECIMIENTOS	vi
LISTA DE CUADROS	x
LISTA DE FIGURAS	xi
1. INTRODUCCIÓN	1
2. OBJETIVOS E HIPÓTESIS	3
2.1. Objetivos generales.....	3
2.2. Objetivos específicos.....	3
2.3. Hipótesis.....	4
3. REVISIÓN DE LITERATURA	5
3.1. El cultivo de maíz en México	5
3.2. Modelos para la estimación de rendimientos de cultivos y de maíz.	7
3.3. El paquete de R Climatol para complementar datos faltantes en un conjunto de series de variables climáticas.....	12
3.4. Algunas herramientas computacionales para la generación de pronósticos, predicciones y modelos predictivos.....	13
3.4.1. En el lenguaje de programación SAS (no es abierto).	13
3.4.2. En el lenguaje de programación PHYTON (es abierto).	14
3.4.3. En el lenguaje de programación IBM (no es abierto).	14
3.4.4. En el lenguaje de programación INFOSTAT (no es abierto).....	15
3.4.5. En el lenguaje de programación GRETL (Guía Gretl, 2005) (es abierto).	15
3.4.6. En el lenguaje de programación MATLAB (no es abierto).	16
3.4.7. En el lenguaje de programación R (es abierto).....	16
3.5. Sobre el Filtro de Kalman (FK).	18
3.6. Algunos lenguajes de programación que contienen funciones para aplicar el Filtro de Kalman.	23

4. MATERIALES Y MÉTODOS.....	25
4.1. Descripción de la zona de estudio.....	25
4.2. Recopilación de los datos climáticos.....	31
4.2.1. Selección de estaciones climáticas de la región.....	31
4.2.2. Recopilación en el CLICOM y en el Distrito de Riego 075 de los datos climáticos diarios de las estaciones seleccionadas.....	32
4.3. Obtención de los datos faltantes con Climatol de R.	32
4.4. Metodologías utilizadas para obtener los modelos dinámicos de predicción de rendimiento de maíz de grano.....	32
4.4.1. Uso de la herramienta <i>systemIdentification</i> de MATLAB, para determinar el mejor modelo matemático dinámico en tiempo discreto auto regresivo ARX lineal.....	34
4.4.2. Uso de Modelos aditivos generalizados (Generalized Additive Models, GAM) con el lenguaje de programación R.	35
4.5. Aplicación del Filtro de Kalman Simplificado (FKS) para mejorar los datos de rendimiento.....	36
4.5.1. El Filtro de Kalman (FK).....	36
4.5.2. Versión simplificada del Filtro de Kalman (FKS).....	38
4.5.3. Cálculo recursivo de la media y la varianza.....	40
4.5.4. Implementación del FKS en una hoja de cálculo.....	41
4.5.5. Programación del FKS en MATLAB.....	42
5. RESULTADOS Y DISCUSIÓN	43
5.1. Datos climáticos recopilados, completados y promediados utilizados para obtener los modelos de predicción del rendimiento de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa.	43
5.1.1. Descarga del CLICOM de los datos climáticos diarios de las estaciones seleccionadas de 1961 a 2012.....	43
5.1.2. Recopilación de la información climática del 2013 al 2021 en el Distrito de Riego.....	43
5.1.3. Obtención de los datos faltantes con Climatol.	44
5.1.4. Manejo y depuración de los datos climáticos.....	45

5.2. Resultados obtenidos con las dos metodologías utilizadas para obtener los modelos dinámicos de predicción de rendimiento de maíz de grano.	45
5.2.1. Con la herramienta systemIdentification de MATLAB, para determinar el mejor modelo ARX lineal.	45
5.2.2. Uso de Modelos aditivos generalizados (Generalized Additive Models, GAM) con el lenguaje de programación R.	53
5.3. Comparación de los datos predichos para seleccionar el mejor modelo y utilizar sus datos en la aplicación del Filtro de Kalman Simplificado.	53
5.4. Aplicación del Filtro de Kalman Simplificado (FKS) para mejorar la calidad de los datos del rendimiento.	55
6. CONCLUSIONES	58
7. LITERATURA CITADA	61
ANEXOS	67
Anexo 1. Metodología detallada que se siguió para completar los datos climáticos faltantes con el Script climatol del lenguaje de programación R	67
Anexo 2. Script general en r para aplicar las funciones ya establecidas de “GAM” para generar los modelos y predecir los datos	71
Anexo 3. Procedimiento detallado para la obtención de modelos ARX lineales para predicción del rendimiento probando diferentes escenarios con la herramienta System identification de MATLAB	72
Anexo 4. Pseudo-código elaborado en MATLAB por el Dr. Irineo López para verificar el algoritmo FKS	83
Anexo 5. Procedimiento detallado para la obtención de modelos ARX lineales para predicción del rendimiento probando diferentes escenarios con System identification de MATLAB	87

LISTA DE CUADROS

Cuadro 1. Balanza de oferta y demanda de maíz blanco en México.....	6
Cuadro 2. Modelos más implementados para predecir los rendimientos de maíz.....	10
Cuadro 3. Paquetes de programación para homogeneización de datos.	12
Cuadro 4. Superficie sembrada, rendimientos y producción obtenida de maíz de grano en el ciclo otoño-invierno en el Distrito de Riego 075 Río Fuerte, Sinaloa, de 1993 a 2021.....	29
Cuadro 5. Datos climáticos faltantes.	44
Cuadro 6. Ajustes que tuvieron las siete diferentes combinaciones de datos climáticos diarios y con 4 años anteriores.	47
Cuadro 7. Ajustes que tuvieron las diferentes combinaciones de datos climáticos mensuales y con 4 años anteriores.	48
Cuadro 8. Ajustes que tuvieron las diferentes combinaciones de datos climáticos anuales y con 4 años anteriores de los datos y otros.	49
Cuadro 9. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos diarios.....	50
Cuadro 10. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos mensuales.	51
Cuadro 11. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos anuales.....	52
Cuadro 12. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos anuales.....	53
Cuadro 13. Datos predichos de los 6 modelos seleccionados de MATLAB y de R.....	54
Cuadro 14. Resumen de operaciones realizadas para aplicar el FKS.	55
Cuadro 15. Ajustes que tuvieron las siete diferentes combinaciones de datos climáticos diarios y con 4 años anteriores.	90
Cuadro 16. Ajustes que tuvieron las diferentes combinaciones de datos climáticos mensuales y con 4 años anteriores.	90
Cuadro 17. Ajustes que tuvieron las diferentes combinaciones de datos climáticos anuales y con 4 años anteriores de los datos y otros.	91

LISTA DE FIGURAS

Figura 1. Superficie nacional sembrada de maíz en México.	5
Figura 2. Importaciones de maíz de México.	6
Figura 3. Diagrama de flujo del funcionamiento de Climatol y sus procesos iterativos.	13
Figura 4. Ubicación del Distrito de Riego 075 dentro del Organismo III de Cuenca Pacífico Norte de la CONAGUA.	26
Figura 5. Sistema de presas que abastecen al Distrito de Riego 075.	27
Figura 6. Almacenamiento histórico del sistema de presas que abastecen al Distrito de Riego 075.	27
Figura 7. Distrito de Riego 075.	28
Figura 8. Evolución de la superficie sembrada de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa, de 1993 a 2021	30
Figura 9. Evolución de los rendimientos de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa, de 1993 a 2021	30
Figura 10. Estaciones climatológicas del CLICOM en la zona de estudio.	31
Figura 11. Comparación en Excel de las predicciones de los modelos implementados.	56
Figura 12. Comparación en MATLAB de las predicciones de los modelos implementados.	57
Figura 13. Gráficas del comportamiento en el tiempo de las variables de entrada y de salida.	76
Figura 14. Ventana principal de systemIdentification de MATLAB.	76
Figura 15. Ventana de Import data	77
Figura 16. Ventana de “Import Data”	77
Figura 17. Ventana principal con figura Mydata1 y cuadro de diálogo “Operations”	78
Figura 18. Cuadros de diálogo de “Select Range”	78
Figura 19. Ventana principal con nuevas figuras “Mdata1e” y “Mydata1v”	79
Figura 20. Gráfica en donde se señalan los datos o rangos de datos de entrada utilizados para la estimación y validación del modelo.	79
Figura 21. Ventana principal en donde se selecciona “Polynomial Models”	80

Figura 22. Ventana de “Polynomial Models”.....	80
Figura 23. Ventana principal con el modelo generado en “Model Views”.....	81
Figura 24. Gráfica en donde se indica el porcentaje que explica el modelo generado	81
Figura 25. Ventana que muestra la ecuación matemática del modelo matemático generado.	82
Figura 26. Modelo matemático generado importado al ambiente de MATLAB en la ventana de “Command Windows”.....	82
Figura 27. Gráfica de la fluctuación diaria de las variables de entrada y salida en el ciclo de cultivo y en los 29 años considerados.	87
Figura 28. Gráfica con los datos climáticos mensuales promedio de los 29 años considerados.	88
Figura 29. Gráfica con los datos climáticos anuales promedio de los 29 años considerados.	88
Figura 30. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax.	92
Figura 31. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmin.	93
Figura 32. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variables Tmax y Tmin.....	93
Figura 33. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax.	94
Figura 34. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmin.	94

Figura 35. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax y pp.	95
Figura 36. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmin.	95
Figura 37. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable pp.	96
Figura 38. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax y pp.	96

1. INTRODUCCIÓN

El maíz es uno de los productos agrícolas más importantes en México, debido a que es parte esencial en la dieta de los mexicanos y, además, tiene una gran importancia estratégica y económica en la producción pecuaria y en el sector industrial (SIAP, 2022).

En el SIAP (2022), se reporta que en el año 2021 se sembraron 7'309,546 ha de maíz de grano en todo el país, de las cuáles 1'502,224 ha fueron de riego (20.6%) y 5'807,322 ha de temporal (79.4%). La producción total fue de 27'503,477 toneladas, de las cuales se obtuvieron 13'169,588 en riego (47.9%) y 14'333,889 en temporal (52.1%), con lo que se destaca la importancia de las áreas de riego en México.

Destaca el estado de Sinaloa debido a que, aunque la superficie que sembró en el 2021 fue de 489,457 ha (6.7% del total nacional), es el estado que tuvo la mayor producción con 5'535,561 toneladas (20.1% del total nacional), debido a que la mayor parte se siembra de riego y a que sus rendimientos son muy altos (11.32 ton/ha) (SIAP, 2022).

Dicho cultivo está sujeto a una serie de factores que causan volatilidad en sus precios y además, los productores no cuentan con la información suficiente para tomar las mejores y más adecuadas decisiones sobre su siembra, debido a que en México no se cuenta con una señal sobre el comportamiento de los precios futuros de maíz blanco como el Chicago Mercantile Exchange (CME) (REPOSITORIO, 2020).

En este contexto de alta volatilidad, contar con mejores datos de rendimiento y producción de maíz de grano permitiría a productores, comercializadores e industriales del sector, protegerse de variaciones en el precio del maíz.

Con respecto a los modelos que existen de predicción de rendimiento de cultivos, aunque existe mucha literatura sobre ellos y además, el número de variables que se utilizan para generarlos es muy amplio, lamentablemente en algunos casos la información requerida no está disponible o está incompleta para poder utilizarla en grandes áreas, lo que motiva que se deban buscar alternativas más prácticas para contrarrestar estas limitantes y que sean aplicables a cualquier cultivo y región del país.

La información que se genera sobre el rendimiento y la producción de maíz de grano en el país en general, es muy útil e indispensable y coadyuva a delinear y establecer una adecuada planeación y a la toma de mejores decisiones dentro del sector agroalimentario de México bajo diferentes escenarios.

Como ejemplos de lo anterior, está lo relacionado a la superficie de maíz que se debe sembrar y en donde hacerlo, así como promover políticas públicas regionales, estatales y nacionales con enfoque de seguridad alimentaria, que permitan determinar y definir las cuotas de exportaciones e importaciones de acuerdo con la oferta y la demanda, regular aspectos de precios, comercialización y mercado, generar y programar apoyos gubernamentales para aumentar la productividad, entre otros temas.

Sin embargo, la información que se reporta sobre los rendimientos de los cultivos y sobre su producción no es precisa debido a los procedimientos que se siguen para obtenerla, por lo que si se mejora, se coadyuvará a perfeccionar dichas políticas públicas.

Una forma de mejorar los datos existentes de los rendimientos de los cultivos y por lo tanto de la producción, es utilizar estrategias ya desarrolladas como las que proponen las diferentes versiones del Filtro de Kalman (FK), que es un algoritmo desarrollado en 1960 como una *herramienta matemático-estadística* para mejorar la predicción de variables de estado, el cual tiene numerosas aplicaciones en la ciencia y tecnología.

La aplicación y desarrollo del FKS en el presente trabajo para mejorar la calidad de los datos existentes de rendimiento de maíz de grano en el D.R. No. 75, Río Fuerte, Sin., se hizo utilizando los datos predictivos obtenidos con el mejor modelo desarrollado que se seleccionó entre cientos generados.

Para la generación de los modelos se utilizaron variables climatológicas, con la finalidad de desarrollar una metodología que utilice variables de entrada más accesibles, debido a que dichos datos son los únicos que están disponibles y actualizados en la mayoría de las regiones del país.

2. OBJETIVOS E HIPÓTESIS

2.1. Objetivos generales.

- 2.1.1.** Desarrollar una metodología práctica y simplificada para la generación de modelos para la predicción de rendimiento de maíz de grano en el Distrito de Riego 075 Río Fuerte del estado de Sinaloa, donde se utilicen variables de entrada disponibles y actualizadas para que sea factible de generalizar en cualquier región del país y para otros cultivos.
- 2.1.2.** La aplicación de una versión simplificada del Filtro de Kalman para mejorar la calidad de datos del rendimiento de maíz de grano en el Distrito de Riego 075 Río Fuerte del estado de Sinaloa.

2.2. Objetivos específicos

- 2.2.1.** Utilizar datos climáticos como variables de entrada para la generación de modelos para la predicción de rendimiento de maíz de grano en el DR 075 Río Fuerte del estado de Sinaloa
- 2.2.2.** Utilizar la función o paquete Climatol del lenguaje de programación de R para complementar los datos climatológicos faltantes de las estaciones seleccionadas de la zona de estudio.
- 2.2.3.** Utilizar herramientas modernas de programación, como MatLab y R, para generar modelos de predicción en el primer caso y datos predictivos en el segundo, del rendimiento del cultivo de maíz de grano en el DR 075 Río Fuerte del estado de Sinaloa.
- 2.2.4.** Analizar los niveles de ajuste de los diferentes modelos que se generen para seleccionar el que se utilizará para aplicar el FKS.
- 2.2.5.** Aplicar una versión simplificada del Filtro de Kalman (FKS) para la mejora de la calidad de los datos.

2.3. Hipótesis

- 2.3.1.** Es posible completar los datos climatológicos faltantes de las estaciones de la zona de estudio utilizando la función o paquete Climatol del lenguaje de programación R.
- 2.3.2.** Es posible utilizar herramientas modernas de programación, como MatLab y R, para generar modelos de predicción y predicciones respectivamente, del rendimiento del cultivo de maíz en el Distrito de Riego 075 Río Fuerte, Sinaloa.
- 2.3.3.** Con la aplicación de una versión simplificada del filtro de Kalman, se pueden mejorar la calidad de los datos existentes del rendimiento de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa.

3. REVISIÓN DE LITERATURA

3.1. El cultivo de maíz en México

El maíz es el cultivo más representativo de México por su importancia económica, social y cultural, además de que gran parte del territorio nacional es propicio para la producción de maíz de grano, por lo que se cultiva en los 32 Estados de la República Mexicana.

A nivel mundial, México ocupa el 8° lugar en producción de maíz y en 2017 exportó a 17 países, principalmente a Venezuela (58%), Kenia (33%) y Estados Unidos (4%), entre otros (6%), por lo que fue el 10° Exportador mundial de maíz de grano (ASERCA, 2021)

A continuación, se presenta la superficie sembrada en México tanto de maíz blanco como amarillo y su producción del año 2010 al 2020 (Figura 1), la balanza de oferta y demanda de maíz blanco (Cuadro 1) y las importaciones de maíz (Figura 2).

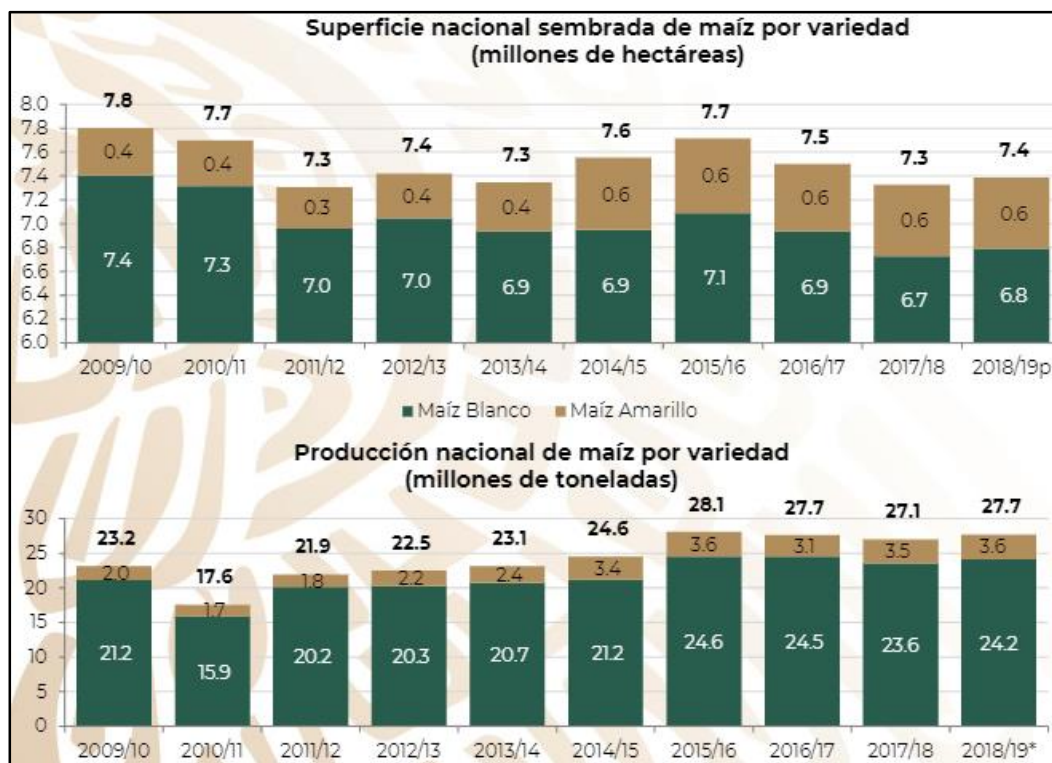


Figura 1. Superficie nacional sembrada de maíz en México.

Fuente: DOCPLAYER, 2022

Cuadro 1. Balanza de oferta y demanda de maíz blanco en México.

Balanza de Oferta y Demanda de Maíz Blanco				
(Miles de toneladas)				
Anual (oct/sep)				
		2017/18	2018/19	% Var (17/18 - 18/19)
Oferta	Oferta Total	27,519	27,673	0.6
	Inventario Inicial	2,106	2,485	18.0
	Producción	24,384	24,283	-0.4
	Importaciones	1,029	905	-12.1
Demanda	Demanda Total	25,034	25,091	0.2
	Exportaciones	909	585	-35.6
	Consumo Total	22,938	23,320	1.7
	Otros ¹	1,187	1,186	-0.1
Inventario Final		2,485	2,582	3.9

Fuente: DOCPLAYER, 2022



Figura 2. Importaciones de maíz de México.

Fuente: World Documents, 2022

En el año 2021, se sembraron en México un total de 7,139,620 ha de maíz de grano, de las cuales 5,666,163 fueron de temporal y 1,473,457 de riego (SIAP, 2022).

Los 8 estados de México que sembraron mayor superficie de maíz de grano fueron, Chiapas (9.44%), Veracruz (8.13%), Jalisco (8.04%), Oaxaca (7.04%), Puebla (7.02%), Guerrero (6.93%), Sinaloa (6.70%) y México (6.69%). Los estados que tuvieron los mayores valores totales de producción debido a sus más altos rendimientos por hectárea fueron, Sinaloa (22.14%), Jalisco (15.36%), México (8.0%), Guanajuato (7.78%), Michoacán (6.54%), Chihuahua (5.30%) y Guerrero (5.60%) (SIAP, 2022) y como se puede observar, el Estado de Sinaloa es el mayor productor de maíz en México.

3.2. Modelos para la estimación de rendimientos de cultivos y de maíz.

Convencionalmente los modelos para pronosticar o producir rendimientos de cultivos se pueden agrupar en dos conjuntos, modelos que calculan procesos y variables de estado a través de simulaciones sobre un área unitaria como biomasa, área foliar, desarrollo radical, etc., y modelos que calculan el resultado de los procesos que trabajan sobre una región, por ejemplo, los rendimientos y la producción de maíz.

Los modelos a desarrollar en este trabajo pertenecen al segundo grupo y en la literatura y en internet se reposta una grandísima cantidad de ellos. Dichos modelos, generalmente son sofisticados por el tipo de variables de entrada que utilizan, ya que están basados en una gran diversidad de factores que intervienen en los procesos productivos, como son las diferentes variables climáticas y vegetativas, aspectos fisiológicos y morfológicos de los cultivos, el tipo de suelo y los factores que participan en el manejo agronómico de los cultivos, principalmente, que y por lo tanto complicados para aplicarlos en grandes áreas.

Por lo anterior, solo se citarán algunos métodos y modelos con la finalidad de destacar el tipo de variables que utilizan. Por ejemplo, un método muy conocido y utilizado para la estimación de rendimientos de cultivos es el de la FAO, como lo menciona Ortiz et al (2005), que se basa en principios ecofisiológicos y en particular, en generar la cantidad total de biomasa neta producida por un cultivo y multiplicarla por un índice de cosecha, que es la parte proporcional de la biomasa útil para el hombre. Dicho autor menciona que en México, a pesar de que se conoce este método desde los años setenta, su aplicación ha sido muy restringida, principalmente por la falta de datos, tanto climáticos como de cultivos. Que, en el caso de los climáticos, la radiación global es la más significativa y

para el caso de los de cultivo, los menos conocidos pero más requeridos, son los índices de área foliar y de cosecha. Concluye en este trabajo, que la estimación de rendimientos de maíz bajo condiciones de riego en el ejido de San Salvador Atenco, con el Método FAO, resultó muy precisa al calibrar los índices de área foliar y de cosecha para cada una de las clases de tierras que se presentan en la zona.

Licona et al (2006), calibraron y validaron el modelo EPIC, para estimar rendimientos de maíz en el DDR 003-Texcoco, debido que el aplicar modelos de simulación calibrados con datos provenientes de experimentos establecidos en el campo, es una alternativa rápida y menos cara. Para calibrar el modelo, reunieron información de rendimiento, prácticas de manejo y la ubicación geográfica, pendiente y dimensiones de 56 parcelas cultivadas con maíz, así como información climática y de caracterización física y química de los suelos. Incorporaron por primera vez en México variables fisio-técnicas como relación biomasa-energía, índice de área foliar potencial e índice de cosecha. Para validar el modelo, compararon los rendimientos reportados en 65 puntos de evaluación. Concluyeron que en condición de riego, el modelo presentó un mejor ajuste que en condiciones de secano y que las variables fisio-técnicas empleadas en la investigación de relación biomasa-energía, índice de cosecha e índice de área foliar potencial, son confiables y deberían registrarse rutinariamente para complementar la información de campo.

Otros investigadores mexicanos han combinado y complementado algunos de los modelos mencionados, como es el caso de Soria et al (2007), del INIFAP, estableciendo el Programa Nacional de Estimaciones de Superficie, Producción y Rendimientos (PRONESPRES), con la participación del INIFAP, ASERCA y el SIAP. Estos autores describen la metodología y resultados del servicio de predicción de cosecha realizado por el INIFAP para el cultivo de maíz en el Estado de México (ciclo primavera-verano 2006), emitida por el Laboratorio de Geomática y el Laboratorio Nacional de Modelaje y Sensores Remotos del INIFAP, en donde la primera predicción de rendimiento se realizó con el modelo de simulación de crecimiento vegetal EPIC y el modelaje de clima y la segunda predicción, mediante el Modelo de Regresión Simple desarrollado por el INIFAP alimentado con datos de monitoreo del cultivo.

Gómez et al (2009) también utilizaron el método FAO en México en el cultivo de la cebada, y mencionan que el uso de ese método permite validar, calibrar y hacer el ajuste, de acuerdo con sus variables, mismas que pueden calcularse en campo, de manera experimental o por un procedimiento de muestreo que evalúe una superficie o región, para corroborar que el método estima el rendimiento de manera aceptable (alta precisión y bajo sesgo). Propone un factor de ajuste a la metodología de la FAO, para que la estimación se aproxime a la realidad, tanto por región como por genotipo.

Martínez (2013), reporta la evolución que el volumen de producción de maíz en México bajo riego ha tenido en los últimos 32 años, así como define a las variables que impactan sobre dicho volumen de producción. Analiza las variables de la superficie sembrada, siniestrada y precio medio rural del maíz haciendo uso de herramientas econométricas, específicamente del modelo de regresión lineal. Concluye que el modelo estimado es adecuado y que proporciona información que puede coadyuvar a los procesos de creación de conocimiento e investigación en cuanto al sector agrícola se refiere, ya que el modelo propuesto permitirá vislumbrar el comportamiento que la producción de maíz en cultivo de riego tendrá en los siguientes periodos.

Michel et al (2013), mencionan que los estudios de predicción se basan con frecuencia en las tendencias de rendimiento de los cultivos estimadas a partir de series temporales de rendimiento proporcionadas por organismos estadísticos nacionales y regionales. En su estudio presentan ocho modelos estadísticos para analizar series temporales de rendimiento y comparar su capacidad para predecir el rendimiento del trigo a escala nacional y regional, utilizando datos proporcionados por la Organización de las Naciones Unidas para la Agricultura y la Alimentación y por el Ministerio de Agricultura francés. Indican que los modelos lineales dinámicos y Holt-Winters tuvieron el mismo rendimiento, dando las predicciones más precisas del rendimiento del trigo, pero que los modelos lineales dinámicos tienen dos ventajas sobre los modelos Holt-Winters, que son que se pueden utilizar para reconstruir las tendencias de rendimiento pasadas retrospectivamente y para analizar la incertidumbre.

Delgadillo et al (2016), mencionan que el análisis de series de tiempo se utiliza para proyectar los datos a periodos futuros como pronósticos, por lo que esta metodología la

utilizaron en el estudio para comparar diferentes métodos de pronósticos en series de rendimiento de granos básicos de maíz, frijol, trigo y arroz en México, con el objetivo de predecir sus valores en el corto plazo. Los pronósticos se realizaron empleando los modelos Autoregressive Integrated Moving Average (ARIMA) para maíz, el de Brow para frijol, Suavización Exponencial Simple para trigo y el de Holt para arroz.

Noriega (2016), reporta algunos de los modelos más implementados para predecir el impacto en los rendimientos de maíz a nivel global, de 2006 a 2016, los cuáles se muestran en el Cuadro 2, en donde se indican las variables climáticas que toman en cuenta.

Cuadro 2. Modelos más implementados para predecir los rendimientos de maíz

MODELO	VARIABLES CLIMÁTICAS QUE UTILIZA
APSIM-Maize	Precipitación, temperaturas máximas y mínimas y radiación
AquaCrop	Precipitación, temperaturas máximas y mínimas y radiación
CropSyst	Precipitación, temperaturas máximas y mínimas, radiación, rocío y viento
CROPWAT	Precipitación, temperaturas máximas y mínimas, sol y viento
DSSAT-CERES-Maize	Precipitación, temperaturas máximas y mínimas y radiación
EPIC-Maize	Precipitación, temperaturas máximas y mínimas, radiación humedad relativa y viento
InfoCrop	Precipitación, temperaturas máximas y mínimas, radiación, vapor y viento
WOFOST	Temperaturas máximas y mínimas, radiación, vapor y viento

Fuente: Noriega, 2016

Dicho autor menciona que los más utilizados son DSSAT, CropSyst, APSIM y WOFOST y en México el AquaCrop y DSSAT.

Seffrin et al (2018), para determinar el modelo de regresión espacial más apropiado para explicar el rendimiento del maíz y sus variables predictivas, identificaron áreas que mostraron autocorrelación espacial de temperatura promedio del aire, lluvia, radiación solar, potencial agrícola del suelo y altitud. Probaron tres modelos para determinar cuál era el más adecuado para predecir el rendimiento de maíz, que fueron el modelo de regresión de mínimos cuadrados ordinarios (OLS), el modelo de retraso espacial (SAR) y el modelo de error espacial (CAR). El SAR lo recomendaron para todos los años de cultivo y el CAR solo para uno de los años de cosecha y concluyeron que las regresiones

espaciales SAR y CAR elegidas para predecir el rendimiento de maíz en los diversos años, tuvieron mejores resultados en comparación con el OLS que no incorpora la autocorrelación espacial de datos.

Ban et al (2019), mencionan que los modelos de crecimiento de cultivos y la teledetección son herramientas útiles para predecir el crecimiento y el rendimiento de los cultivos, pero cada herramienta tiene inconvenientes a escala regional. Para mejorar la precisión de las predicciones regionales de rendimiento de maíz, desarrollaron un enfoque simple para asimilar los productos del Espectro radiómetro de Imágenes de Resolución Moderada (MODIS) en un modelo de crecimiento de cultivos, y evaluaron la predicción de rendimiento regional en Illinois, US. Simularon el crecimiento y el rendimiento del maíz utilizando el Modelo de Síntesis de Recursos del Medio Ambiente del Cultivo (CERES), con insumos mínimos que comprenden la fecha de siembra, la cantidad de fertilizante, los coeficientes genéticos, el suelo y los datos climáticos. Concluyeron que la asimilación adicional de los factores de estrés hídrico derivados de los datos MODIS y LAI bajo la condición de riego automático, mostró la mayor precisión de predicción para la predicción anual de rendimiento de maíz, por lo que la estrategia actual para asimilar datos MODIS en un modelo de crecimiento de cultivos utilizando insumos mínimos fue exitosa para predecir los rendimientos regionales.

Noriega et al (2021), realizaron una revisión de estado del arte sobre los modelos de simulación del crecimiento y rendimiento de maíz, implementados para estimar el impacto bajo escenarios de cambio climático a nivel mundial, consultando un gran número de artículos publicados al respecto, de donde “seleccionaron los modelos APSIM-Maize (Agricultural Production Systems sIMulator), AquaCrop (modelo de crecimiento de cultivos para mejorar la productividad hídrica, desarrollado por la FAO), CropSyst (Cropping Systems simulator), CropWat (Crop Water Productivity Model), DSSAT-CERES-Maize (Decision Support System for Agrotechnology Transfer-Crop Environment Resource Synthesis), EPIC-Maize (Erosion Productivity Impact Calculator), InfoCrop (modelo dinámico de rendimiento de cultivos desarrollado por la Universidad de Wageningen en Holanda) y WOFOST (WOrld FOod STudies, desarrollado por la Universidad de Wageningen en Holanda)”.

3.3. El paquete de R Climatol para complementar datos faltantes en un conjunto de series de variables climáticas.

La gran mayoría de datos existentes sobre variables climáticas presentan deficiencias como el faltante de datos, por lo que, para atender este problema, se han desarrollado desde hace muchos años, metodologías de homogeneización de los datos que permitan eliminar o reducir en lo posible estas deficiencias indeseadas.

Algunos paquetes de programación que implementan estos métodos para que puedan ser usados por la comunidad climatológica, se presentan en el cuadro No. 3 (Climatol, 2020). Entre ellos está el paquete de R Climatol, que contiene funciones para el control de calidad, homogeneización y relleno de los datos faltantes en un conjunto de series de cualquier variable climática, el cual se utilizó en el presente trabajo.

Cuadro 3. Paquetes de programación para homogeneización de datos.

PAQUETE	VERSIÓN	LICENCIA	TIPO DE PROGRAMA	DISPONIBILIDAD
ACMANT	4	Software gratuito	Ejecutable	https://github.com/dpeterfree/ACMANT
AnClim ProClimDB	?	Software gratuito	Ejecutable	https://www.climahom.eu/
Climatol	3.0	Gpl	Paquete R	https://www.climatol.eu/index.html
GAHMIDI HOMAD	?	Gpl	R fuente R/Fortran	correo a andrea.toreti en giub.unibe.ch
GSIMCLI	0.0.1	Gpl	Pitón	https://filed.github.io/gsimcli/
HOMERO	2.6	Gpl	Fuente R	https://www.climatol.eu/pub/HOMER2_6.zip
MEZCLA	3.03	Software gratuito	Ejecutable	https://www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/
Prueba de redistribución	?	Software gratuito	Fuente R	correo a predrag.petrovic en hidmet.gov.rs
Pruebas de RH	4	Software gratuito	Fuente R	https://etcccdi.pacificclimate.org/software_e.shtml
USHCN	52i	Software gratuito	Fuente Fortran	ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/software/52i/phav52i.tar.gz

Fuente: Climatol, 2020

Además de normalizar los datos mediante división por sus valores medios, Climatol versión 3.1.1. también lo hace restando las medias o mediante una estandarización completa, de acuerdo con el procedimiento iterativo que se presenta en la Figura 3.

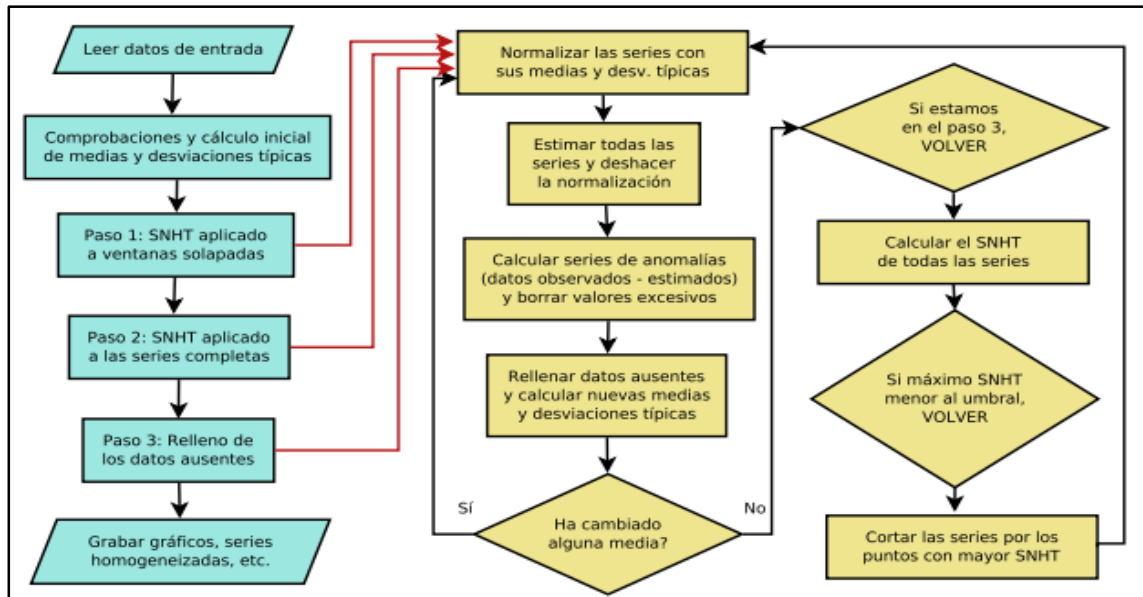


Figura 3. Diagrama de flujo del funcionamiento de Climatol y sus procesos iterativos.

*SNHT = Standard Normal Homogeneity Test

Fuente: (Gujjarro, J. 2018)

La versión 3.1.2. de Climatol se puede consultar en la página: <https://cran.r-project.org/web/packages/climatol/climatol.pdf>

3.4. Algunas herramientas computacionales para la generación de pronósticos, predicciones y modelos predictivos

Los avances tecnológicos y computacionales de los últimos años, han permitido que diferentes lenguajes de computación incluyan dentro de sus funciones, paquetes o herramientas para obtener pronósticos y predicciones, así como modelos predictivos.

Dentro de ellas se mencionarán algunos casos, destacando dos de ellos que se utilizaron en la presente tesis, el de la herramienta System Identification de MATLAB para determinar el mejor modelo ARX lineal y el uso de modelos aditivos generalizados (Generalized Additive Models, GAM) de R para generar datos predictivos.

3.4.1. En el lenguaje de programación SAS (no es abierto).

- a. **Procedimiento PLS en SAS/STAT.** Se utiliza para ajustar modelos mediante el uso de métodos predictivos lineales. Además, las técnicas utilizadas en el procedimiento PLS tienen en cuenta cualquier variación en los predictores. Las técnicas implementadas mediante el procedimiento PLS son, de regresión de

componentes principales, de regresión de rango reducido y de regresión de mínimos cuadrados parciales, donde se contabilizan tanto la variación de la respuesta como la variación del predictor (BUAP, 2017).

- b. En el SAS Visual Forecasting** (SAS, 2022). Algunas de sus principales ventajas son que permite preparar y segmentar datos de series temporales, crear modelos utilizando métodos de series temporales y redes neuronales disponibles, determina automáticamente los modelos de previsión más adecuados para los datos históricos, guiar a los usuarios a través de la selección de datos, modela y pronostica individualmente cada serie en la jerarquía, luego concilia los pronósticos y permite a los planificadores probar escenarios hipotéticos.

3.4.2. En el lenguaje de programación PHYTON (es abierto).

Tiene un modelo en particular llamado SARIMAX o Promedio Móvil Integrado Autorregresivo Estacional con Modelo de Variable Explicativa (Skforecast, 2022), que es estacional porque los datos tienen una tendencia estacional, es Autoregresivo porque depende de valores pasados, es Promedio móvil integrado debido a que toma en cuenta los errores de pronóstico pasados para corregir sus pronósticos futuros y es de Variable Explicativa, ya que considera que la evolución de la serie temporal de interés no solo depende de sí misma, sino también de variables externas. Puede crear y probar un modelo ForecasterAutoreg a partir de un regresor RandomForestRegressor y una ventana temporal que la define cuantos datos anteriores utiliza como predictores (anuales, mensuales, diarios, etc.). Usa modelos de regresión de Scikit-learn para realizar pronósticos sobre series temporales haciendo uso de Skforecast, una librería que contiene las clases y funciones necesarias para adaptar cualquier modelo de regresión de Scikit-learn a problemas de pronósticos.

3.4.3. En el lenguaje de programación IBM (no es abierto).

Una demostración de su librería para predicciones IBM SPSS Statistics, está disponible para todos los usuarios del software como descarga gratuita con todos los módulos al completo. Utiliza el procedimiento Modelizador de series temporales para estimar modelos de suavizado exponencial, modelos autorregresivos integrados de media móvil (ARIMA) univariante o modelos de función de transferencia multivariantes para series

temporales (IBM, 2022a). Su módulo adicional Predicciones proporciona varios procedimientos para llevar a cabo las tareas de creación de modelos y producción de predicciones, como modelizador de series temporales, aplicar modelos de series temporales y de causales temporales (IBM, 2022b).

3.4.4. En el lenguaje de programación INFOSTAT (no es abierto).

Es un software no abierto, pero con versión gratuita para análisis estadístico de aplicación general desarrollado bajo la plataforma Windows. Contiene un submenú Estimación, validación y pronósticos, en donde a su vez tiene una ventana con dos solapas (Infostat, 2008):

- a. La solapa Modelo.** Permite especificar la ecuación del modelo a estimar. Se puede seleccionar modelos estacionarios ARMA y modelos no estacionarios ARIMA y los modelos a estimar pueden ser No estacionales (sin tendencia) o estacionales.
- b. La solapa Estimación, validación, pronósticos, datos faltantes, otras herramientas.** Permite realizar controles sobre los procedimientos de estimación, indicar cuáles serán las herramientas utilizadas para el diagnóstico del modelo estimado y realizar pronósticos cuando el modelo ya ha sido validado.

3.4.5. En el lenguaje de programación GRETL (Guía Gretl, 2005) (es abierto).

Es un programa de código abierto de econometría y estadísticas que ofrece una línea de comandos y una interfaz gráfica basada en menús. Sus funciones principales están implementadas en C, pero también comprende un lenguaje de matrices y secuencias de comandos versátil llamado hansl que se puede utilizar para desarrollar fácilmente paquetes de funciones contribuidas. Tiene una colección de funciones FEP, para calcular diferentes tipos de estadísticas de evaluación de pronósticos, así como pruebas. Para facilitar el uso, se proporciona un marco de interfaz de secuencias de comandos común, que es lo suficientemente flexible como para adaptarse a futuras incorporaciones. También se puede acceder a la mayor parte de la funcionalidad a través de una interfaz de ventana de diálogo gráfica dentro de gretl. Esta documentación explica el uso y las

capacidades, además de proporcionar algunos antecedentes econométricos cuando sea necesario.

3.4.6. En el lenguaje de programación MATLAB (no es abierto).

Es una plataforma de programación y cálculo numérico utilizada por millones de ingenieros y científicos para analizar datos, desarrollar algoritmos y crear modelos. En la presente tesis se utilizó la herramienta *systemIdentification* de MATLAB para generar los determinar el mejor modelo de tipo ARX (Autoregressive with Exogeneous Input), que “es uno de los modelos lineales dinámicos más implementado en la actualidad, debido a que la gran mayoría de sistemas de procesos del mundo real se ajustan con precisión a este modelo” (Montes, 2020).

3.4.7. En el lenguaje de programación R (es abierto).

- a. Uso de Modelos aditivos generalizados (Generalized Additive Models, GAM) para generar datos predictivos.** Son modelos de regresión que están ganando gran popularidad en la práctica debido a su flexibilidad, su versatilidad, y su interpretabilidad en una gran variedad de estudios reales en diversas disciplinas (Cadarso, 2021).

La existencia de paquetes estadísticos potentes de uso libre implementando dichos modelos (fundamentalmente en R, SAS/STATA y XLSTAT), facilita enormemente que los GAM puedan transferirse con rapidez a la investigación aplicada. Son una extensión de los modelos tradicionales de regresión lineal para el análisis de los datos, que incorpora la no linealidad y la regresión no paramétrica. El modelo está construido por la suma de funciones suaves (splines) de las variables predictoras, pudiendo ser estas variables continuas, variables categóricas, número de casos y series de datos (DOCPLAYER, 2022).

Entre las bondades de estos modelos, está la no necesidad de probar que las variables son independientes y si tienen o no una distribución normal. Los algoritmos inclusive permiten introducir indicaciones respecto a la distribución (binomial, poisson, y algunas otras). Por lo tanto, las funciones de las relaciones que revelan los modelos permiten una comprensión mucho mayor que la de los

coeficientes beta es suponer la relación lineal de la regresión tradicional, ya que los efectos no son constantes dentro del rango de variación de las variables (DOCPLAYER, 2022).

- b. Paquetes u objetos clase forecast: forecast, ForecastComb, ForecastCombinatuions, forecastR, forecastHybrid, forecastMtL, forecastSNSStS y forecastTB.** Tienen información sobre el método de pronóstico, los datos utilizados, los pronósticos puntuales obtenidos, los intervalos de predicción, los residuos y los valores ajustados. Contiene métodos y herramientas para mostrar y analizar pronósticos de series de tiempo univariantes, que incluyen suavizado exponencial a través de modelos Espacio de Estados (State Space) y modelado automático ARIMA (metodología econométrica basada en modelos dinámicos que utiliza datos de series temporales). Los modelos ARIMA proporcionan métodos más sofisticados para crear modelos de los componentes de tendencia y estacionales que los modelos de suavizado exponencial y disponen de la ventaja añadida de poder incluir variables predictoras en el modelo. Los parámetros del modelo ARIMA se pueden usar como un modelo predictivo para hacer pronósticos de valores futuros de las series temporales, una vez que se selecciona el modelo más adecuado para los datos de la serie temporal. El valor d afecta a los intervalos de predicción; los intervalos de predicción aumentan de tamaño con valores más altos de « d ». Todos los intervalos de predicción serán esencialmente los mismos cuando $d = 0$ porque la desviación estándar de la previsión a largo plazo irá a la desviación estándar de los datos históricos. La función `forecast` en el paquete R de `forecast`, también se puede utilizar para pronosticar valores futuros de la serie temporal y se puede especificar el nivel de confianza para los intervalos de predicción utilizando el argumento de nivel (Mateos, 2020).
- c. Función predict.** Se usa para predicciones a partir de los resultados de varias funciones de ajuste de modelos y necesita un argumento `n.ahead` que especifica la cantidad de tiempo que se adelanta para predecir.

3.5. Sobre el Filtro de Kalman (FK).

Palacios (comunicación personal, 2021), define al FK como “una herramienta matemático-estadística para mejorar la predicción de variables de estado en modelos dinámicos lineales y no lineales, cuando se dispone de información de otra fuente (como mediciones), sobre la variable predicha. También menciona que otros autores aceptan sin problema que, como su nombre lo indica, el FK es un “filtro” que “deja pasar algo” y “retiene algo”. Para ellos, el FK es un algoritmo recursivo que puede correr en tiempo real usando únicamente las mediciones de entrada actuales, el estado calculado previamente y su matriz de incertidumbre y no requiere ninguna otra información adicional. Además, señala tres variantes del Filtro de Kalman: el modelo básico cuando las ecuaciones de las funciones de las variables del espacio de estados son lineales (KF); el modelo ampliado para modelos no-lineales diferenciables (Extended Kalman Filter, EKF) y el modelo denominado “sin aroma” (Unscented Kalman Filter, UKF). Su principal contribución fue explicar detalladamente todos aquellos artificios matemáticos y operaciones con matrices que se requieren en el proceso de desarrollar los diferentes algoritmos del FK, Así, Palacios (2021), concluye que el FK es una herramienta que utiliza información como mediciones u observaciones sobre la variable en estudio, para mejorar u optimizar predicciones de modelos dinámicos con formatos de ‘espacio de estados’ (formato en el cual el modelo predictivo al tiempo $k + 1$ hace una predicción tomando principalmente en cuenta el modelo al tiempo k). Que una característica fundamental del FK, es que considera que tanto las predicciones de los modelos, como las mediciones u observaciones en las que se basan las "correcciones", son imperfectas, es decir, tienen una cierta varianza.

Wan et al (2001), presentan el filtro alternativo Kalman sin aroma (UKF) con un rendimiento superior al del filtro Kalman extendido (EKF), debido a que en el EKF la distribución del estado se aproxima mediante variables aleatorias gaussianas (GRV), lo que puede introducir grandes errores en la verdadera media posterior y la covarianza del GRV transformado y conducir a un rendimiento subóptimo y a veces a la divergencia del filtro. El UKF aborda este problema utilizando un enfoque de muestreo determinista y su

rendimiento se demostró en varios dominios de aplicación, incluida la estimación del estado, la estimación dual y la estimación de parámetros.

La mejor explicación sobre lo que hace el FK, es la dada por Beven (2009), que consiste en lo siguiente: se tiene un modelo M1 que hace una predicción Y1 de una variable de interés, y se tiene otro modelo M2 que hace una predicción Y2 de la misma variable. Si no hay más información, una mejor estimación de la variable de estudio está dada por el promedio $(Y1 + Y2) / 2$. ¿Y qué pasa si se cuenta con información de las varianzas σ_1^2 y σ_2^2 de los dos modelos?. Entonces es posible mejorar los pronósticos de diferentes actividades utilizando simultáneamente dos o más modelos de predicción de manera “mixta” o complementaria. Así en este caso, una mejor estimación se logra con un estimador inversamente proporcional a las varianzas de los modelos y la solución que propone Beven (2009) con fórmulas, se utilizó en el presente trabajo y dichas fórmulas se presentan en el punto 4.5.

Morales et al (2014), mencionan que el Filtro de Kalman es un procedimiento matemático que opera por medio de un mecanismo de predicción y corrección, por lo que es un algoritmo que pronostica el nuevo estado a partir de su estimación previa, ya que va añadiendo un término de corrección proporcional al error de predicción, de tal forma que este último es minimizado estadísticamente. Mencionan que la solución es óptima en cuanto el filtro combina toda la información observada y el conocimiento previo acerca del comportamiento del sistema para producir una estimación del estado. Concluyen que el Filtro de Kalman es un algoritmo matemático que permite realizar pronósticos del comportamiento de un estado con características estocásticas, con resultados bastante aceptables en tiempo real y comprobaron que puede ser aplicado exitosamente al área de hidrología, como la predicción de caudales en tiempo real, mejorando los resultados obtenidos con otros modelos, destacando su utilidad como una herramienta de apoyo en el desarrollo de políticas de operación y control de los embalses.

De Wit et al (2007), mencionan que la incertidumbre en la distribución espacial y temporal de la lluvia en las simulaciones regionales de rendimiento de cultivos, comprende una fracción importante del error en los resultados de la simulación del modelo de cultivos. En su trabajo, utilizaron el Ensemble Kalman Filter (EnKF) para asimilar estimaciones de

humedad del suelo (SWI) derivadas de sensores de microondas satelitales de resolución gruesa para corregir errores en el balance hídrico del modelo de cultivo de estudios alimentarios mundiales (WOFOST). Llevaron a cabo simulaciones de modelos de cultivo con el EnKF para trigo de invierno y maíz en grano para España, Francia, Italia y Alemania durante el período 1992–2000. Sus resultados se evaluaron sobre la base de regresión con estadísticas conocidas de rendimiento de cultivos a nivel nacional y regional. Además, analizaron las innovaciones del filtro EnKF para ver si se podían encontrar tendencias sistemáticas que pudieran indicar deficiencias en el balance hídrico del modelo para crecimiento de cultivos WOFOST. Concluyen que los análisis de las innovaciones de filtro revelaron patrones espaciales y temporales, mientras que la distribución de innovaciones normalizadas no es gaussiana y tiene una media distinta de cero, lo que indica que el rendimiento de EnKF es subóptimo.

Sánchez et al, (2013), mencionan que se implementó el filtro de Kalman de ensambles (FKEN) y el filtro de Kalman de cuadratura (FKC) para estimar y pronosticar los estados de precipitación en Venezuela. Utilizaron la regla de cuadratura de Gauss - Hermite para aproximar la media a posterior mediante la integración de cuadratura numérica en el FKC y un modelo espacio temporal para la predicción de precipitación en periodo de tiempos cortos. Evaluaron el comportamiento de los algoritmos en términos del estadístico denominado tiempo promedio del error relativo, observándose que el FKEN tiene menos variabilidad en la estimación que el FKC.

Moira (2013), discute la aplicación de la idea de Filtro Federado con factor de intercambio de información adaptativo para la estimación de estado de una columna de destilación, así como su impacto en la tolerancia a fallos y rendimiento en el filtrado. Menciona que el desempeño de las tradicionales soluciones centralizadas es comparado con la estructura del filtro Federado, y su equilibrio entre la capacidad de tolerancia a fallas y la precisión en la estimación. En el artículo se muestra como el filtro extendido de Kalman adaptativo en una estructura de Filtro Federado es exitoso en la estimación de estado de una columna de destilación por lotes. Concluyen que la implementación y rendimiento de la solución estándar de EKF en comparación con el AEKF y con el FKF adaptativo son comparados. Menciona que los algoritmos de filtrado fueron usados para la estimación

de estado de una columna de destilación por lotes a partir de datos reales de la columna y que los resultados confirman que el FKF, con factores de intercambio de información adaptativo, es capaz de incrementar la robustez del filtro contra variaciones en la señal de medición mejorando de manera significativa la precisión de la estimación.

Vargas et al (2015), realizaron una revisión histórica mundial de modelación del proceso lluvia-escurrimiento, desde sus inicios mediados del siglo XIX con el método racional. Partieron de la teoría del hidrograma unitario y el empleo de las computadoras, hasta llegar a la aplicación de filtros de Kalman y las redes neuronales artificiales, para alcanzar mayor precisión numérica en diseño de estructuras hidráulicas y pronóstico de avenidas. Mencionan que en los países en desarrollo, como México, los modelos más usados son: 1) aquellos que estén disponibles de manera gratuita y con foros abiertos de usuarios compartiendo sus dudas, 2) aquellos con fortaleza teórica-práctica y 3) en los que estén disponibles los datos para alimentar y calibrar el modelo.

González et al (2015), propusieron la implementación del algoritmo del Filtro de Kalman Discreto (DKF) junto con un modelo autorregresivo con entrada exógena (ARX) para realizar el pronóstico de caudales a corto plazo con 24, 48, 72 y 96 horas de anticipación en la cuenca del río Turbio, localizada en el estado de Guanajuato y parte del estado de Jalisco. Entre sus principales conclusiones están que el pronóstico de caudales mediante el modelo ARX+DKF fue implementado de manera satisfactoria. Que los periodos previos de calibración para la metodología del modelo ARX+DKF son fundamentales para mejorar los pronósticos a través del tiempo, pues la función de respuesta en la cuenca puede ser variante según los periodos del ciclo hidrológico o debido a la presencia de eventos meteorológicos de consideración.

Sánchez et al (2017), presentan una metodología basada en geoestadística espacio-temporal para el diseño óptimo de redes de monitoreo automatizadas, empleando como criterio de optimización reducir la varianza del error en la estimación de los niveles del agua subterránea en ciertas posiciones espacio-temporales dada una frecuencia de monitoreo fija; la metodología se aplica para seleccionar sitios a instrumentar en un acuífero de Zacatecas, México. Sus principales conclusiones fueron que el FK resulta eficiente para reducir la varianza del error en la estimación por lo que lo convierte en una

herramienta muy útil para el diseño de redes de monitoreo. Que la aplicación de la metodología permitió asignar un orden de prioridad a cada uno de los pozos disponibles para ser instrumentados en función del nivel de representatividad que tienen sobre las estimaciones de los niveles del agua subterránea en una malla espacio-temporal definida para las zonas de interés.

Herrera et al (2020), analizaron la información de 205 de pozos de bombeo para uso agrícola o doméstico en las márgenes del Río Sinaloa. El objetivo de su trabajo fue encontrar la relación entre los datos de capacidad específica (Q_e) y transmisividad hidráulica (T) de la zona de estudio, usando la fórmula de Thiem, considerando un valor fijo del radio de influencia. Concluyeron que mediante el FK es posible eliminar o reducir tal componente, mejorando así la determinación de la relación T - Q_e de un valor r de 0.95 (sin filtro) a 0.97 (con filtro), para una relación lineal y exponencial.

Narváez et al, (2021), mencionan que la asimilación de datos integrada para el pronóstico de caudales puede brindar flexibilidad y reducción de errores sistemáticos en los modelos. Evaluaron la capacidad predictiva del FK discreto y el FK de conjuntos y su integración, utilizando registros horarios de caudal de dos estaciones ubicadas sobre el río Huaynamota. Concluyen que el ajuste entre la serie observada y las pronosticadas lo estimaron mediante el coeficiente de Nash-Sutcliffe y la raíz del cuadrado medio del error para determinar que el filtro de Kalman discreto alcanza mejor ajuste y actualización con base en el tiempo de retraso entre series. Que el filtro de Kalman de conjuntos genera un suavizado de la serie pronosticada, y al igual que la integración de filtros aumenta el efecto de desplazamiento de la serie pronosticada. Que el filtro de Kalman discreto alcanza ajuste superior a ARX y a la combinación ARX-DKF.

Becker (2022), menciona que la mayoría de los sistemas modernos están equipados con numerosos sensores que proporcionan una estimación de variables ocultas (desconocidas) basadas en series de mediciones, por lo que uno de los mayores desafíos del sistema de seguimiento y control es proporcionar una estimación exacta y precisa de las variables ocultas en presencia de incertidumbre. También menciona que el Filtro de Kalman es uno de los algoritmos de estimación más importantes y comunes y produce estimaciones de variables ocultas basadas en mediciones exactas e inciertas, por lo que

es una solución recursiva al problema de filtrado lineal de datos discretos. Además, que el filtro de Kalman proporciona una predicción del estado futuro del sistema, basado en estimaciones pasadas, por lo que hoy en día, el filtro Kalman se utiliza en objetivos de seguimiento (radar), sistemas de localización y navegación, sistemas de control, gráficos por computadora y mucho más.

3.6. Algunos lenguajes de programación que contienen funciones para aplicar el Filtro de Kalman.

En MathWorks (2022), se trata de diseño y utilización de filtros de Kalman en MATLAB y Simulink, y se utilizan las funciones `kalmanFilterForTracking` y `vision.KalmanFilterobjeto` y la configure `KalmanFilterfuncion`. Es un ejemplo de una función con su cuerpo principal en la parte superior y rutinas auxiliares en forma de funciones anidadas. Por ejemplo, el FK para el seguimiento o rastreo de objetos se enfoca en tres características importantes: predicción de la ubicación futura del objeto, reducción del ruido introducido por detecciones inexactas y facilitar el proceso de asociación de múltiples objetos a sus pistas. Específicamente MATLAB trata sobre el diseño de filtro de Kalman para estimación de estado

En Python (DelftStack, 2022), se usa la clase `KalmanFilter` de su biblioteca OpenCV para implementar dicho algoritmo y predecir estados como la trayectoria de un objeto dado recurriendo el estado anterior del objeto. Esta clase maneja una ecuación diferencial estocástica lineal, en donde una matriz de transición de estado vincula el estado actual con el estado anterior, para lo cual se requiere transformar el estado en algún dominio de medición que se logra utilizando una matriz de transformación: además, también es necesario que haya un vector de ruido de proceso con covarianza. La biblioteca OpenCV de Python utiliza tres atributos de un objeto para esta clase, que son los siguientes:

- `measurementMatrix` especifica la matriz de medición.
- `transitionMatrix` especifica la matriz de transición.
- `processNoiseCov` especifica la matriz de ruido del proceso con covarianza.

Ya utilizados los tres atributos se recurre a la función `predict` para hacer algunas predicciones. En dicha página se puede ver un ejemplo con un código en Python para la aplicación de la clase `KalmanFilter`.

4. MATERIALES Y MÉTODOS

El desarrollo de este trabajo se dividió en cinco etapas o secciones, cuyas finalidades fueron definir primeramente una metodología práctica y accesible para el desarrollo de modelos para la predicción del rendimiento de maíz de grano, para posteriormente aplicar el Filtro de Kalman Simplificado (FKS) para mejorar los resultados obtenidos con los modelos desarrollados.

- i. Descripción de la zona de estudio y de la situación del maíz de grano en el estado de Sinaloa y en el Distrito de Riego 075 Río Fuerte, Sinaloa.
- ii. Recopilación de los datos climáticos a utilizar como variables de entrada en los modelos de predicción.
- iii. Completar los datos climáticos faltantes con la herramienta de Climatol de R apoyada con MATLAB.
- iv. Aplicación de dos metodologías en MATLAB y en R para generar modelos de predicción de rendimiento de maíz de grano en el primer caso y solo predicciones de rendimiento de maíz en el segundo caso.
- v. Aplicación de un Filtro de Kalman Simplificado (FKS) con la finalidad de mejorar la calidad de los datos existentes sobre el rendimiento de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa, utilizando el mejor de los modelos generados.

4.1. Descripción de la zona de estudio

El Organismo de Cuenca Pacífico Norte de la CONAGUA que es el número III, es el que tiene la mayor superficie de Distritos de Riego de los XIII Organismos que existen en el país (aproximadamente el 30% del total nacional), en donde existen 10 Distritos de Riego, destacando dentro de ellos, los dos Distritos de Riego que tienen la mayor superficie en México, que son el 010 Culiacán y el 075 Río Fuerte (Figura 4).

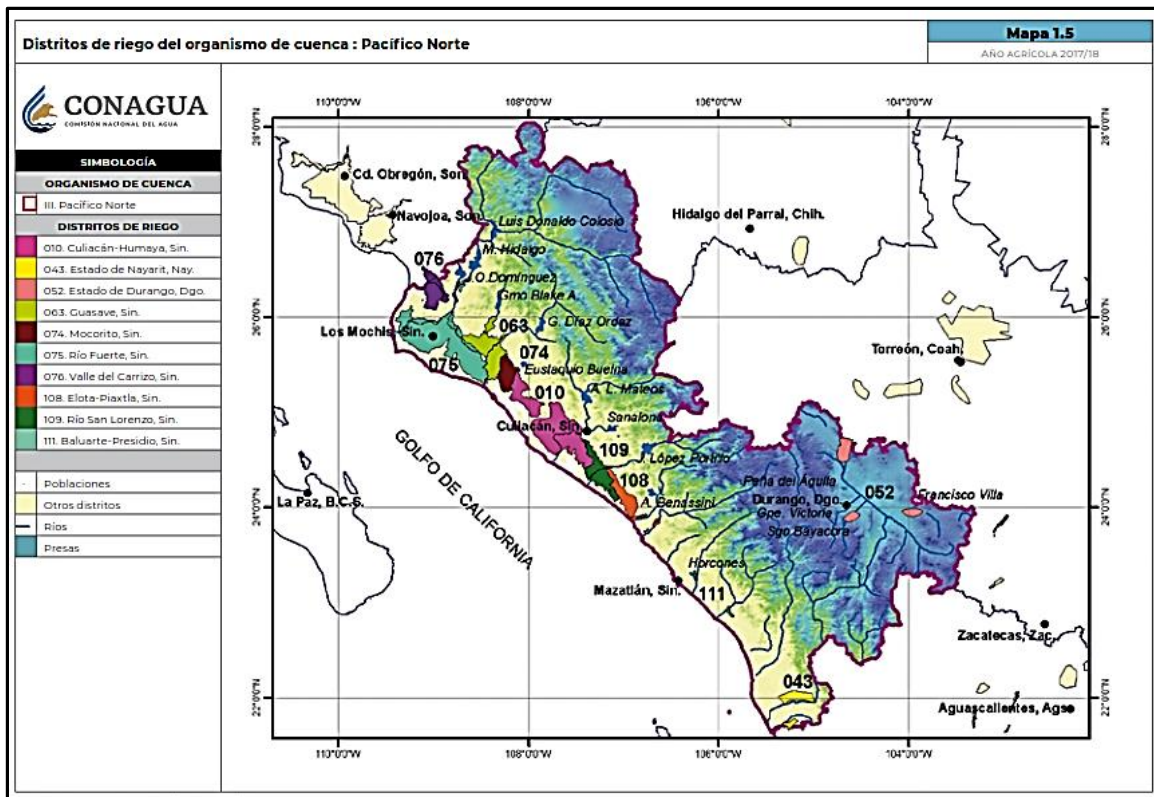


Figura 4. Ubicación del Distrito de Riego 075 dentro del Organismo III de Cuenca Pacífico Norte de la CONAGUA.

Fuente: CONAGUA, Estadísticas DR 2017-2019

El Distrito 075 Río Fuerte está ubicado al norte del estado de Sinaloa, dentro del cuadrante ubicado dentro de las coordenadas geográficas siguientes:

- 26°06'31" N y 109°09'23" O
- 25°42'21" N y 108°29'14" O
- 25°51'04" N y 109°22'18" O
- 25°27'31" N y 107°38'58" O

La superficie de riego que domina es de aproximadamente 230,000 ha y la que se siembra cada año que incluye segundos cultivos, depende de la disponibilidad de agua que existe en el sistema de 3 presas con las que cuenta para abastecerse (Miguel Hidalgo, Josefa Ortiz y Luis Donald Colosio o Huites, Figura 5), que se conoce con buena aproximación el primer día de octubre de cada año, y con exactitud en diciembre de cada año.



Figura 5. Sistema de presas que abastecen al Distrito de Riego 075.

Fuente: <https://redvalledelfuerte.org/>

El comportamiento que ha tenido este sistema de presas de 1995 a 2020, se presenta en la Figura 6.

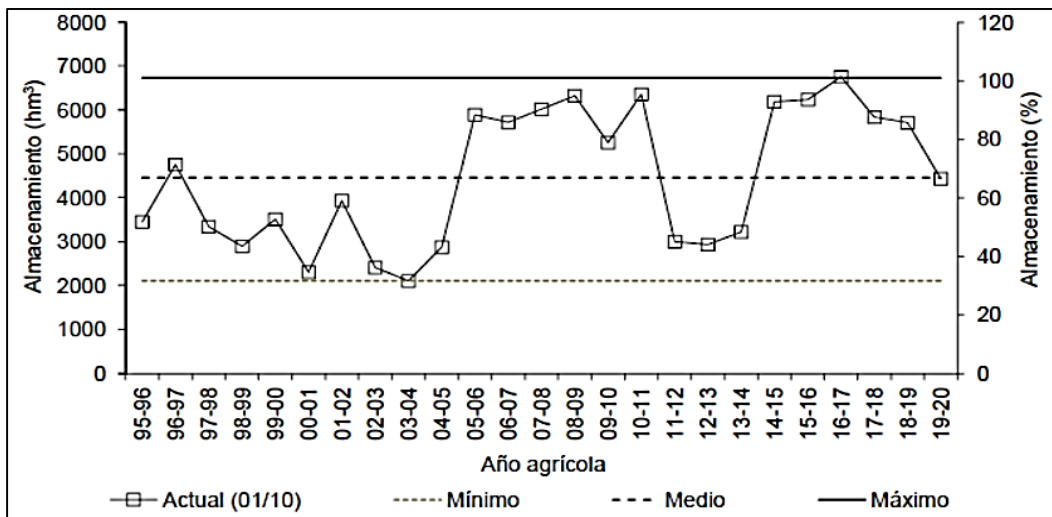


Figura 6. Almacenamiento histórico del sistema de presas que abastecen al Distrito de Riego 075.

Fuente: Sifuentes, 2020

Se observa que el comportamiento es irregular, en donde han existido períodos con escasez hídrica, destacando el de 1995 a 2005 que fueron 10 años continuos, y especialmente los años 2000/2001 y 2003/2004 que han sido los más críticos, así como otro periodo de escasez de 2011/2012 al 2013/2014.

Con respecto a la superficie total sembrada de maíz de grano en el estado de Sinaloa, en el año 2021 fueron 489,547 ha, de las cuales 459,835 de riego y 29,369 de temporal (SIAP, 2022), cosechándose 6'298,369 toneladas. Destaca el Distrito de Riego 075 Río Fuerte, en donde en el 2021 se cosecharon 131,356 ha (CONAGUA, 2022), por lo que es la región en donde se siembra la mayor superficie de maíz de grano además de tener de los rendimientos más altos, siendo la región maicera más importante del estado y del país, razón por la cual se seleccionó para realizar el presente trabajo (Figura 7).



Figura 7. Distrito de Riego 075.

Fuente: Red Mayor del Distrito de Riego 075.

Se recopilaron los datos que se reportan en las 29 estadísticas anuales que elabora la Gerencia de Distritos de Riego de la Comisión Nacional del Agua, del año 1993 al 2022, relacionada al maíz de grano que se siembra como 1er. cultivo en el ciclo otoño-invierno en el Distrito de Riego 075 Río Fuerte, Sinaloa, aclarando que también se siembra como segundo cultivo en el ciclo primavera-verano en superficies menores y con rendimientos más bajos.

Los datos de rendimiento fueron los utilizados para la generación de los diferentes modelos de predicción de rendimiento (Cuadro 4).

Cuadro 4. Superficie sembrada, rendimientos y producción obtenida de maíz de grano en el ciclo otoño-invierno en el Distrito de Riego 075 Río Fuerte, Sinaloa, de 1993 a 2021.

Año	Sup 1er cultivo (ha)	Rend (ton/ha)	Producc (ton)
1993	74,085	7.04	521,625
1994	98,955	7.77	768,819
1995	73,278	8.30	608,221
1996	80,374	8.21	659,825
1997	100,222	8.07	809,204
1998	84,617	8.40	710,783
1999	62,961	8.55	538,205
2000	90,002	9.12	821,159
2001	97,442	9.64	939,678
2002	91,175	9.17	836,042
2003	90,509	9.89	895,009
2004	119,569	8.71	1,041,856
2005	115,148	10.08	1,160,339
2006	108,442	10.06	1,091,427
2007	122,955	10.50	1,291,031
2008	121,843	10.19	1,242,142
2009	118,617	10.46	1,240,377
2010	110,630	10.77	1,191,361
2011	128,811*	10.52*	1,349,924*
2012	146,992	10.26	1,508,486
2013	152,608	10.47	1,598,514
2014	137,407	11.68	1,604,548
2015	165,346	11.86	1,961,642
2016	156,696	13.37	2,095,192
2017	142,384	13.37	1,904,350
2018	130,582	13.44	1,754,404
2019	160,807	13.54	2,178,061
2020	145,730	14.30	2,084,100
2021	131,357	13.62	1,788,682

* Datos promediados con los años 2010 y 2012 por evento extraordinario de helada en 2011.

Fuente: CONAGUA, Estadísticas anuales de los Distritos de Riego

En el cuadro 4 se detecta que la superficie de maíz de grano de riego en dicho Distrito se ha venido aumentando de 1993 a la fecha, debido principalmente a los apoyos

gubernamentales que se han otorgado, a los precios que ha tenido el maíz y a los incrementos en los rendimientos.

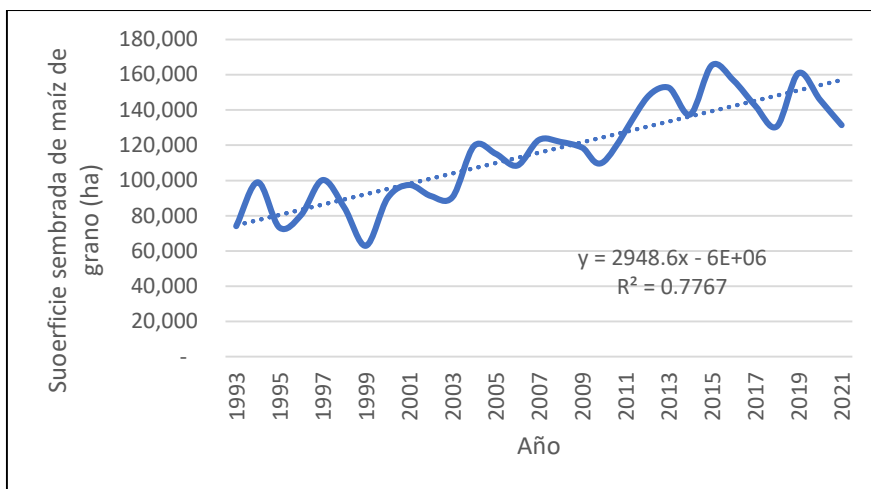


Figura 8. Evolución de la superficie sembrada de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa, de 1993 a 2021

Fuente: CONAGUA, Estadísticas de diferentes años de Distritos de Riego

Dichos incrementos en rendimiento por hectárea de maíz de grano, se deben tanto al desarrollo de nuevas variedades mejoradas que han venido impulsando el INIFAP y un gran número de empresas semilleras particulares nacionales e internacionales, como a un mejor conocimiento del proceso productivo por parte de los usuarios de riego (Figura 9).

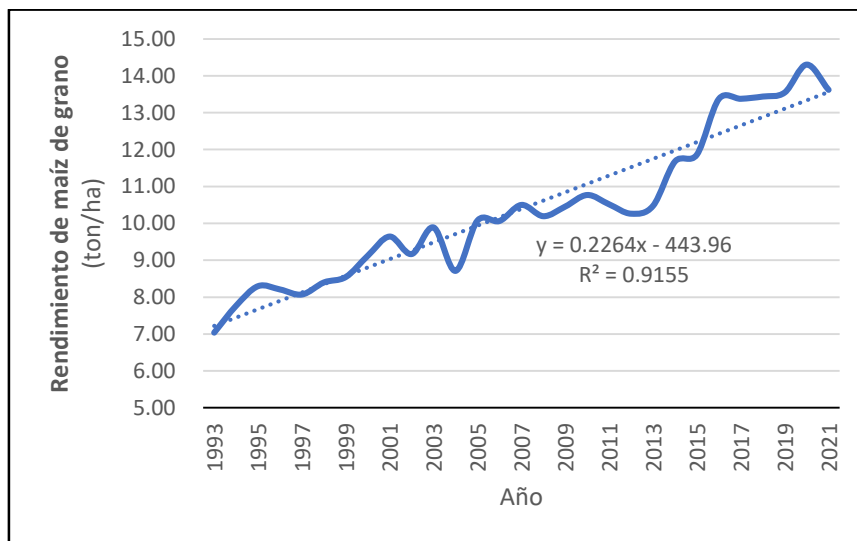


Figura 9. Evolución de los rendimientos de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa, de 1993 a 2021

Fuente: CONAGUA, Estadísticas de diferentes años de Distritos de Riego

4.2. Recopilación de los datos climáticos.

Existe una gran cantidad de variables de entrada para generar modelos de predicción de rendimiento de maíz de grano, información o datos que en su mayoría generalmente no está disponible, por lo que no es posible utilizarla para desarrollar modelos predictivos que sean prácticos y se puedan generalizar y aplicarlos a grandes superficies.

Por lo anterior, se decidió utilizar los datos de variables climáticas que estuvieran actualizados y disponibles, que en principio fueron temperatura máxima, temperatura mínima, precipitación y evaporación, debido a que existía información hasta 2013 en la base de datos climatológica nacional del sistema CLICOM (CLICOM, 2022).

4.2.1. Selección de estaciones climáticas de la región.

Se revisó en el CLICOM que estaciones meteorológicas están operando de las que se reportan en la zona de estudio, detectando que existían solo 5, que fueron 25003 "AHOME", 25042 "HIGUERA", 25008 "LOS MOCHIS", 25080 "RUIZ CORTINEZ" y 25038 "GUASAVE" (Figura 10).

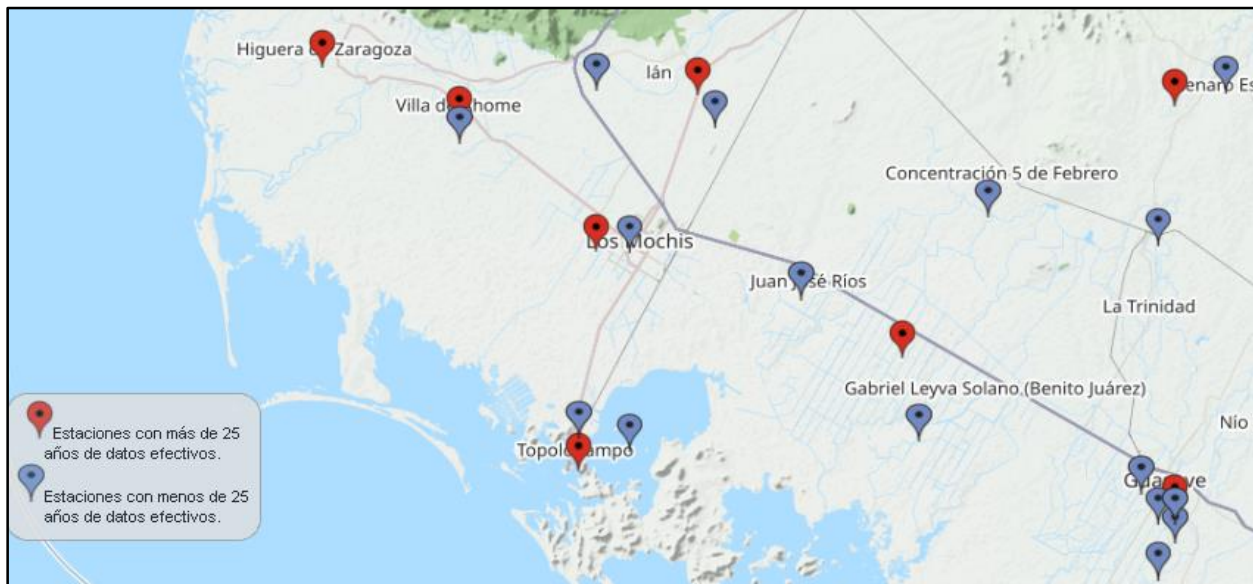


Figura 10. Estaciones climatológicas del CLICOM en la zona de estudio.

Fuente: CLICOM, 2022 <http://clicom-mex.cicese.mx/mapa.html>

4.2.2. Recopilación en el CLICOM y en el Distrito de Riego 075 de los datos climáticos diarios de las estaciones seleccionadas.

En general, la información diaria existente en el CLICOM de las variables climáticas de éstas cinco estaciones es de los años 1960's al 2013. Se procedió a descargar la información diaria del CLICOM de 4 variables que inicialmente se seleccionaron (temperatura máxima, temperatura mínima, precipitación y evaporación), obteniendo 4 archivos tipo csv de cada una de las 5 estaciones mencionadas.

Para completar y actualizar la información faltante en el CLICOM de los años 2013 al 2021 y algunos otros de las estaciones seleccionadas, en diciembre del año 2021 se hizo una segunda visita a Los Mochis, especialmente para recabar dicha información en el Distrito de Riego de la CONAGUA.

4.3. Obtención de los datos faltantes con Climatol de R.

Para completar u obtener los datos faltantes, se siguió un procedimiento existente que utiliza un script "Climatol" en el lenguaje de programación de R, apoyado por un paso previo de preparación de archivos que se tiene que realizar en MATLAB.

La metodología detallada que se siguió con un script Climatol adaptado al presente trabajo para completar los datos climáticos faltantes., se presenta en el Anexo 1.

4.4. Metodologías utilizadas para obtener los modelos dinámicos de predicción de rendimiento de maíz de grano.

Los modelos dinámicos son aquellos en los que el tiempo es la principal variable independiente, los cuáles se utilizan en prácticamente cualquier esfera de actividad humana, por lo que su uso ha venido en aumento acelerado en los últimos 40 o 50 años, a la par que se desarrollaban las tecnologías de la información.

En el caso de la predicción de rendimiento de cultivos, está comprobado que existen muchas complicaciones para desarrollarlos debido al gran número de factores de diferente tipo interrelacionados que participan, ya sean controlados o no, como se comprobó en el caso del presente trabajo, que propician que dicho rendimiento varíe

temporal y espacialmente, presentando desviaciones de un año a otro que puede conducir a resultados inexactos.

Sin embargo, se ha logrado desarrollar un sinnúmero de métodos más eficientes de modelos de predicción basados en datos de crecimiento y fisiología de cultivos, pero que difícilmente puede replicarse de cultivo a cultivo o de región a región, por el proceso que se requiere llevar a cabo para la medición y control de las variables de entrada que utilizan y poder utilizar y aplicar dichos modelos de manera general.

Así, los modelos más sencillos como los que se utilizaron en el presente trabajo basados en variables de entrada de datos climáticos, son más prácticos, económicos y fáciles de aplicar e implementar para diferentes regiones y distintos cultivos, debido a que no requieren de un conocimiento profundo sobre los mecanismos físicos que producen los datos, aunque sus resultados no son tan precisos como los basados en datos de crecimiento y fisiología de cultivos.

La variable dependiente para generar los modelos predictivos fue el rendimiento de maíz y las variables independientes utilizadas fueron los datos diarios, mensuales y anuales del ciclo de cultivo (noviembre a abril), de temperatura máxima, temperatura mínima y precipitación que se obtuvieron y promediaron de tres estaciones climatológicas de la zona de estudio.

Tomando en cuenta los comentarios anteriores, se buscaron alternativas y herramientas para la obtención de los modelos basados en datos y no en crecimiento de cultivos, considerando que la variable de interés rendimiento, puede ser calculada o predicha por más de un modelo.

Por otro lado, los responsables técnicos de los servicios meteorológicos, comisiones de cuenca, asociaciones de usuarios de riego, empresas agroindustriales, sector agrícola gubernamental, y algo similar en otras áreas de actividad, se preguntan si habrá manera de mejorar sus pronósticos utilizando simultáneamente dos o más modelos de predicción de manera “mixta” o complementaria. La respuesta es un contundente “sí” y el razonamiento que da sustento a esta aseveración, lo contemplan las diferentes versiones del Filtro de Kalman que se comentan más adelante.

Por ello, se consultó a varios investigadores para poder definir que estrategias y metodologías se pueden seguir para obtener modelos de predicción del rendimiento de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa, para lo cual se detectaron varias opciones y finalmente se decidió utilizar dos, que son las siguientes:

La primera es utilizar la herramienta *systemidentification* del ambiente de programación MATLAB para determinar el mejor modelo autoregresivo con variables externas (ARX), que permitió probar cientos de escenarios que se pueden generar, tanto 21 externos de MATLAB derivados de las 7 combinaciones de los datos climáticos y de las 3 periodicidades de los datos (diarios, mensuales y anuales), como internos dentro de System Identification por la gran cantidad de combinaciones que se derivan de las diferentes particiones que se prueban para generar y validar el modelo a obtener, como los años anteriores a tomar en cuenta en el modelo. En una primera etapa se generaron modelos utilizando 10 particiones y con datos de 4 años anteriores (que son los que tiene *systemidentification* predeterminados), para cada una de las 21 diferentes combinaciones de las variables climáticas.

Ya definidas las particiones más adecuadas para los modelos que presentan mayor ajuste con 4 años, en una segunda etapa se continuó afinando la búsqueda de modelos con diferentes combinaciones de años anteriores para seleccionar el mejor.

La segunda fue utilizar el lenguaje de programación R, que tiene un procedimiento de estadística no paramétrica conocida como Modelos aditivos generalizados (GAM).

Ambos procedimientos se resumen a continuación y se detallan en los Anexos 1 y 2 respectivamente.

4.4.1. Uso de la herramienta *systemidentification* de MATLAB, para determinar el mejor modelo matemático dinámico en tiempo discreto auto regresivo ARX lineal.

La herramienta *systemidentification* de MATLAB permite crear modelos lineales y modelos no lineales ARX o autoregresivos de sistemas dinámicos derivados de la teoría

de Identificación de Sistemas, que son modelos que no se pueden modelar fácilmente a partir de especificaciones o principios básicos.

Se basa en utilizar la ecuación en diferencias lineales usando datos de entrada y salida del dominio del tiempo y el dominio de la frecuencia para identificar funciones de transferencia de tiempo continuo y tiempo discreto, modelos de procesos y modelos de espacio de estados.

Asumen que los valores de la variable de salida dependen de los valores que haya tomado esa variable en el pasado y son modelos autoregresivos en los que el valor de salida no solo depende de los valores pasados de la misma variable, sino que también se ve influido por los valores pasados de un conjunto de variables exógenas de origen externo (Henares, 2018).

Este procedimiento se hizo para un gran número de combinaciones derivadas de los datos climáticos y sus periodicidades y de las que se pueden probar dentro de *systemIdentification*, que se originan con distintos valores de particiones y de los años a utilizar como parte del modelo, cuyo proceso detallado se presenta en el Anexo 2.

4.4.2. Uso de Modelos aditivos generalizados (Generalized Additive Models, GAM) con el lenguaje de programación R.

Los Modelos GAM, son modelos de regresión que están ganando gran popularidad en la práctica debido a su flexibilidad, su versatilidad y su interpretabilidad, por lo que es posible utilizarlos en una gran variedad de estudios reales en diversas disciplinas.

Por lo anterior, se decidió probar esta alternativa para generar los modelos requeridos de predicción, empleando únicamente los datos anuales promedio de las variables climáticas de entrada ya definidas, con la finalidad de evaluar sus resultados y difundir sus aplicaciones y ventajas.

El procedimiento seguido en R fue generar un guion (script) para aplicar las funciones ya establecidas de “gam”, tanto para generar los modelos como para predecir los datos, el cual se presenta de manera detallada en el Anexo 3.

4.5. Aplicación del Filtro de Kalman Simplificado (FKS) para mejorar los datos de rendimiento.

4.5.1. El Filtro de Kalman (FK)

En Wikipedia definen al Filtro de Kalman, como “un algoritmo desarrollado por Rudolf E. Kalman en 1960 que sirve para poder identificar el estado oculto (no medible) de un sistema dinámico lineal, al igual que el observador de Luenberger, pero sirve además cuando el sistema está sometido a ruido blanco aditivo. La diferencia entre ambos, es que en el observador de Luenberger la ganancia K de realimentación del error debe ser elegida “a mano”, mientras que el filtro de Kalman es capaz de escogerla de forma óptima cuando se conocen las varianzas de los ruidos que afectan al sistema (la definición anterior se consultó en Wikipedia https://es.wikipedia.org/wiki/Filtro_de_Kalman, 21 de abril de 2022).

Esta definición justifica ampliamente la necesidad de contar con una versión simplificada del FK, más al alcance de profesores, investigadores y estudiantes con sólo los conocimientos más esenciales de las matemáticas y de la estadística.

Por lo tanto, se retoma lo que comenta Beven (2009) tratado en el punto 3.5., de que si es posible mejorar los pronósticos de diferentes actividades utilizando simultáneamente dos o más modelos de predicción de manera “mixta” o complementaria.

Dicha aseveración se apoya en lo siguiente: si el modelo $M1$ para un cierto tiempo predice un valor $Y1$ y el modelo $M2$ para el mismo intervalo de tiempo predice un valor $Y2$ y, lo más importante: si los valores promedio de $Y1$ y $Y2$ son razonablemente “cercaños” (tienen la misma media poblacional), entonces un mejor estimador que $Y1$ o $Y2$ es $(Y1 + Y2) / 2$.

Obsérvese que este promedio sencillo le da el mismo peso ($1/2$) al valor predicho por $M1$ y al valor predicho por $M2$. Si no se cuenta con más información sobre los modelos, el promedio $(Y1 + Y2) / 2$ es lo mejor que se puede hacer.

Pero, ¿qué pasa si se conocen las varianzas de los modelos σ_1^2 y σ_2^2 ? Las varianzas son una medida de la dispersión con respecto a la media teórica y son también una

medida de la confiabilidad de los modelos. Entre mayor sea la varianza de las predicciones de un modelo, menor será la confianza que se le pueda tener. Si se conocen las varianzas, un mejor estimador que el promedio simple, es un promedio ponderado con ponderaciones inversamente proporcionales a las varianzas. Tal estimador Y^* se puede calcular con la siguiente fórmula (Beven, 2009, página 177; Pei et al., 2019):

$$Y^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} Y_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} Y_2 \quad (1)$$

Donde:

- Y^* = Valor de la variable dependiente mejorada;
- Y_1 = Valor de la variable dependiente obtenida con datos medidos u observados (primer modelo);
- Y_2 = Valor de la variable dependiente obtenida con datos predichos de rendimiento (segundo modelo);
- σ_1^2 = Varianza del valor predicho por el primer modelo
- σ_2^2 = Varianza de los valores medidos u observados
- $\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$ = Coeficiente de ponderación de Y_2
- $\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$ = Coeficiente de ponderación de Y_1

La fórmula (1) se puede expresar de manera totalmente equivalente, si renombraremos al Coeficiente de ponderación de Y_1 como coeficiente “K” o “Ganancia de Kalman”, como se indica a continuación:

$$K = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (2)$$

Si se utiliza esta fórmula para desarrollar la diferencia de $1 - K$, resulta:

$$1 - K = 1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2} - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^2 + \sigma_2^2 - \sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Sustituyendo en la fórmula (1) lo obtenido en el desarrollo anterior, resulta:

$$Y^* = (1 - K)Y_1 + KY_2 \quad (3)$$

O, de manera equivalente:

$$Y^* = Y_1 + K(Y_2 - Y_1) \quad (4)$$

Obsérvese que la ganancia de Kalman puede variar entre 0 y 1 y la variable dependiente mejorada Y^* , tomaría un valor cercano a Y_1 si $\sigma_1^2 \ll \sigma_2^2$. En caso contrario, Y^* tomaría un valor más cercano a Y_2 .

En los casos multidimensionales en lugar de varianzas se manejan matrices de varianzas y covarianzas. Cuando se presenta esa situación, en lugar de una división entre una suma de varianzas, se multiplica por la matriz inversa de la suma de las matrices de varianzas y covarianzas.

Estos conceptos son en esencia lo que hacen las distintas variantes del FK. A los lectores interesados en conocer con detalle alguna de las versiones del FK, como son la versión básica (FKB), la extendida (FKE), la “unscented” (FKU) u otras, se les recomienda consultar primero el artículo de Lim et. al. (2016), quienes presentan, con mucha claridad y detalle, la deducción de las ecuaciones del FKB, es decir para aplicación en modelos lineales y tiempo discretizado.

Otro documento que se recomienda ampliamente es el tutorial del Becker (2022), cuando se manejan simultáneamente 2 o más variables independientes y dependientes, cuyas varianzas hay que actualizar en cada paso del tiempo.

4.5.2. Versión simplificada del Filtro de Kalman (FKS).

Para la versión simplificada del FK, se consideran los dos modelos mencionados correspondiendo el primero a los datos medidos y el segundo a los datos obtenidos con el modelo predictivo, que a ciertos intervalos de tiempo predicen una misma variable dependiente. En el Kalman clásico, se tiene un primer modelo predictivo (de una o más variables de estado) y un segundo modelo que corresponden a las mediciones u observaciones de la misma variable, que para el presente trabajo se invirtió el orden.

En el FK lineal para el intervalo de tiempo “j”, se predicen las variables de estado (de estudio) con base en los valores encontrados en el intervalo de tiempo anterior “j-1”. Se conoce además la función que relaciona la función predictiva y las mediciones u observaciones al tiempo “j”.

Con las varianzas de los modelos, se calcula la ganancia de Kalman y con ésta se calculan las predicciones mejoradas utilizando las ecuaciones (2), (3) y (4).

Es normal que los lectores con conocimientos básicos sobre métodos de optimización, se pregunten si la ganancia de Kalman es un estimador óptimo, o si solamente cumple la condición que parecía “natural”, de que constituya un método matemático-estadístico para mejorar las predicciones de modelos dinámicos, calculando coeficientes de ponderación inversamente proporcionales a las varianzas de las variables de estudio.

En realidad, la ecuación (3) se obtiene de un proceso de optimización. Tal proceso no lo presenta Beven, 2009, pero su deducción es muy sencilla (Pei *et al.*, 2019), como se muestra a continuación:

Se parte de la ecuación (3):

$$Y^* = (1 - K)Y_1 + KY_2$$

Se toman las varianzas:

$$\sigma_*^2 = (1 - K)^2\sigma_1^2 + K^2\sigma_2^2 \quad (5)$$

Se calcula la primera derivada de (5) respecto de K:

$$\frac{d\sigma_*^2}{dK} = -2(1 - K)\sigma_1^2 + 2K\sigma_2^2 \quad (6)$$

Igualando a cero y despejando K se obtiene la fórmula (2).

$$K = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

El cálculo de la segunda derivada de la ecuación (9), permite verificar que esta es positiva ($2\sigma_1^2 + 2\sigma_2^2$), lo que significa que σ_*^2 alcanza un valor mínimo en el valor K .

Como puede verse, la ganancia de Kalman es un estimador de mínima varianza *cuando la varianza va variando de intervalo a intervalo de tiempo*. De acuerdo a pruebas numéricas realizadas con la ecuación (3), se puede imponer cualquier valor constante a la ganancia de Kalman a través de todo el tiempo de simulación, incluso ciertos valores negativos, que darán una secuencia de Y^* con menor varianza. Sin embargo, el valor de K calculado con la fórmula (2) es óptimo cuando las varianzas se van recalculando en cada intervalo de tiempo.

4.5.3. Cálculo recursivo de la media y la varianza

El FK corre en tiempo real, generalmente en forma discreta, por lo que cada vez que se avanza un intervalo de tiempo se tendrán nuevos valores de Y_1 y Y_2 , por lo que deberán actualizarse las varianzas con las que se calcula la ganancia de Kalman y los valores optimizados Y^* .

Para hacer estos cálculos, resulta muy conveniente utilizar las formas recursivas de la media y la varianza. La forma recursiva de la varianza para un intervalo de tiempo “ j ”, se expresa en función de la varianza y la media del intervalo de tiempo anterior, “ $j-1$ ”, así como del j -ésimo dato que se incorpora, como se observa a continuación (Mathematics, 2022):

$$S_j^2 = \frac{j-2}{j-1} S_{j-1}^2 + \frac{1}{j} (x_j - \bar{x}_{j-1})^2; \quad j = 3, 4, 5, \dots \quad (7)$$

Donde:

- S_j^2 es la expresión recursiva de las varianzas σ_1^2 y σ_2^2 , sólo que aquí el subíndice “ j ” se refiere al número de intervalo de tiempo:

$$\bar{x}_j = \frac{j-1}{j} \bar{x}_{j-1} + \frac{1}{j} x_j; \quad j = 2, 3, 4, \dots \quad (8)$$

- \bar{x}_j es la media recursiva de los primeros “ j ” valores predichos por los modelos (Kim and Huh, 2011).

Para arrancar los cálculos, se necesita conocer la media y la varianza de los modelos cuando $j=2$, lo que permite iniciar los cálculos recursivos a partir de $j = 3$:

Para el cálculo de la media:

$$\bar{x}_2 = \frac{x_1 + x_2}{2} \quad (9)$$

Para el cálculo de la varianza:

$$S_2^2 = \frac{1}{2}(x_1 - x_2)^2 \quad (10)$$

4.5.4. Implementación del FKS en una hoja de cálculo

La puesta en marcha del FKS en una hoja de cálculo es relativamente simple y el número y contenido de columnas requeridas se enlistan a continuación:

- i. En la primera columna se anota el número de intervalo de tiempo;
- ii. En la segunda columna se anota el año a que corresponden los datos.
- iii. En la tercera columna se anotan los datos medidos de rendimiento.
- iv. La cuarta columna corresponde a los datos predichos mejorados con el FKS, y se llena hasta después de haber calculado la columna correspondiente a la ganancia de Kalman (columna No. 10). Aunque este es el último cálculo, el propósito de colocarlo en la cuarta columna, es para poder observar que el valor mejorado por Kalman se encuentra comprendido entre los valores medidos (columna 3) y los datos predichos (columna 5).
- v. En la quinta columna se anotan las predicciones obtenidas con el modelo predictivo seleccionado.
- vi. En la sexta columna se anota el cálculo de la media recursiva obtenida con los datos medidos de rendimiento (fórmula (8)). Este dato sirve solamente para calcular la varianza σ_1^2 .
- vii. En la séptima columna se anota el cálculo de la varianza recursiva obtenida con los datos medidos de rendimiento (fórmula (7)).
- viii. En la octava columna se anota el cálculo con fórmula (8) de la media recursiva de las predicciones. Este dato sirve únicamente para calcular la varianza σ_2^2 .
- ix. En la novena columna se anota el resultado del cálculo de la varianza recursiva de las predicciones (fórmula (7)).
- x. En la décima columna se anota el cálculo de la ganancia de Kalman (fórmula (4)).

- xi.** Ya contando con la ganancia de Kalman, se regresa a la quinta columna que se había dejado temporalmente vacía y se calcula el valor optimizado de Y^* , utilizando cualquiera de las fórmulas (3) o (4).

4.5.5. Programación del FKS en MATLAB

Para verificar el algoritmo FKS, se elaboró en MATLAB un código de todo el proceso realizado en Excel que incluye la elaboración de la gráfica con los tres que se presenta en el Anexo 4 (López, 2022, comunicación personal).

5. RESULTADOS Y DISCUSIÓN

5.1. Datos climáticos recopilados, completados y promediados utilizados para obtener los modelos de predicción del rendimiento de maíz de grano en el Distrito de Riego 075 Río Fuerte, Sinaloa.

Se revisó en el CLICOM que estaciones meteorológicas están operando de las que reporta en la zona de estudio, detectando que existían solo 5, que fueron 25003 "AHOME", 25042 "HIGUERA", 25008 "LOS MOCHIS", 25080 "RUIZ CORTINEZ" y 25038 "GUASAVE".

5.1.1. Descarga del CLICOM de los datos climáticos diarios de las estaciones seleccionadas de 1961 a 2012.

Se procedió a descargar del CLICOM la información diaria disponible de las cuatro variables climáticas inicialmente seleccionadas (temperatura máxima, temperatura mínima, precipitación y evaporación), de las cinco estaciones mencionadas, obteniendo 20 archivos tipo csv (4 de cada una de las estaciones).

Al revisar la información existente de las 5 estaciones, se detectó que en general la información correspondía de los años 60's al 2013. Así mismo, se detectó que dos de ellas, la 25042 "HIGUERA" y la 25080 "RUIZ CORTINEZ", tenían varios periodos con mucha falta de información, por lo que se decidió descartarlas quedando las 3 restantes.

Al seguir analizando la información, se determinó iniciar a partir del año 1969 debido a que la estación 25038 "GUASAVE" reporta datos a partir de esa fecha, procediendo a continuación a complementar y actualizar la información faltante de 2013 a 2021.

5.1.2. Recopilación de la información climática del 2013 al 2021 en el Distrito de Riego.

La segunda visita realizada a Los Mochis en diciembre de 2021, tuvo la finalidad de obtener en el Distrito de Riego la información del 2013 al 2021 en el CLICOM de las 4 variables climáticas inicialmente seleccionadas, principalmente de los años 2013 al 2021, así como algunos otros de las 3 estaciones definidas.

Dado que durante la visita se comprobó que no contaban con todos los archivos concentrados y actualizados de toda la información requerida, fue necesario continuar posteriormente solicitando la faltante a través de llamadas telefónicas, correos y mensajes, la cual se pudo terminar y completar hasta junio de 2022.

Al revisar la información obtenida en el Distrito de Riego de los años 2013 al 2021, se detectó que desde los años 2013 en promedio ya no se contaban con datos de evaporación, por lo que se tuvo que descartar este factor climático como variable de entrada para la generación de los modelos de predicción, quedando las 3 restantes.

5.1.3. Obtención de los datos faltantes con Climatol.

Con la información completa diaria de los tres factores climáticos correspondientes al periodo del 1 de enero de 1969 al 31 de diciembre de 2021, se procedió a obtener los datos faltantes con el procedimiento que utiliza la aplicación del script Climatol de R apoyado de MATLAB descrito en el punto 5.6.3.2. y detallado en el Anexo 1.

Dichos datos faltantes en el periodo mencionado y para cada una de las tres estaciones, fueron los siguientes (Cuadro 5):

Cuadro 5. Datos climáticos faltantes.

FACTOR CLIMÁTICO	ESTACIÓN		
	AHOME	LOS MOCHIS	GUASAVE
Temperatura máxima	584	106	910
Temperatura mínima	558	129	943
Precipitación	854	99	3824

Fuente: Elaboración propia

Es importante destacar que, aunque la información climática que se utilizó para la generación de los modelos de predicción del rendimiento de maíz de grano en el Distrito de Riego 075 corresponde solamente al periodo del año 1993 al 2021 (29 años), para obtener los datos faltantes se utilizaron todos los datos recopilados de 1969 a 2021 (53 años), considerando que la literatura consultada que, a mayor periodo y mayor número de estaciones, mejores resultados en la complementación de los datos.

5.1.4. Manejo y depuración de los datos climáticos.

Para poder aplicar los datos climáticos ya completados en Climatol para la generación de los modelos de predicción del rendimiento de maíz, se conformaron tres escenarios de periodicidad de datos que se pueden utilizar en dicha generación, que son datos diarios de los 29 años y 6 meses del ciclo de cultivo, mensuales promedios de los 29 años de los seis meses del ciclo de cultivo y los 29 anuales promedios.

Para ello, primeramente se obtuvieron los datos promedio diarios de cada una de las tres variables climáticas de las tres estaciones consideradas, quedando un solo valor diario de cada factor climático. En seguida se seleccionaron los datos de los años 1992 a 2021, desechando los de 1969 a 1991, aclarando que se incluyeron los del año 1992 debido a que el ciclo del cultivo cosechado en 1993 inició en noviembre de 1992.

Estos datos diarios fueron depurados, quitando los datos de los meses de mayo a noviembre que no corresponden al ciclo del cultivo de maíz que es de los meses de noviembre abril, obteniendo el archivo correspondiente a “1993-2021-diario-aj.xls”. Finalmente, con estos datos se obtuvieron los promedios mensuales y anuales de las 3 variables climáticas.

5.2. Resultados obtenidos con las dos metodologías utilizadas para obtener los modelos dinámicos de predicción de rendimiento de maíz de grano.

5.2.1. Con la herramienta systemIdentification de MATLAB, para determinar el mejor modelo ARX lineal.

El procedimiento detallado seguido se presenta en el Anexo 5 y los resultados obtenidos con los cientos de escenarios probados con los que se llevó a cabo la generación de los modelos de predicción de rendimiento de maíz, son los siguientes

a. Primera etapa utilizando 21 escenarios de datos climáticos y 10 particiones de años anteriores a tomar en cuenta en el modelo que se genere.

Para la generación de los modelos, primero se trabajó con los 21 escenarios externos de datos climáticos que resultan de utilizar 3 periodicidades de datos climáticos y los 7 contextos que se obtienen al combinar las 3 variables climáticas.

- 5,256 datos diarios para cada factor climático de 29 años y de 6 meses del ciclo de cultivo de maíz (de noviembre a abril), de cada año.
- 181 datos mensuales promedio de 6 meses del ciclo de cultivo de maíz (de noviembre a abril) y de 29 años para cada factor climático.
- 29 datos anuales promedio de 29 años para cada factor climático.
- 7 combinaciones de las variables climáticas.
 - ✓ Temperatura máxima, Temperatura mínima y Precipitación.
 - ✓ Temperatura máxima y Temperatura mínima.
 - ✓ Temperatura máxima y Precipitación.
 - ✓ Temperatura mínima y Precipitación
 - ✓ Temperatura máxima.
 - ✓ Temperatura mínima.
 - ✓ Precipitación.

Ya definidos los 21 escenarios externos de MATLAB, se continuó con la definición dentro de MATLAB de los subconjuntos o particiones de datos para estimar y validar los modelos ARX que se probarían para cada una de las 21 combinaciones de datos climáticos, que fueron las 10 siguientes: 50:50%, 55:45%, 60:40%, 65:35%, 70:30%, 75:25%, 80:20%, 85:15%, 90:10% y 95:5% y en algunos casos se afinaron aún más.

Con estas 210 combinaciones se inició la generación de los modelos utilizando en esta primera etapa, los datos de 4 años anteriores que tiene *systemIdentification* predeterminados a tomar en cuenta en el modelo que se genere, para ya definida la partición del modelo que presenta mayor ajuste, probar varias alternativas de años anteriores.

Los resultados obtenidos de las particiones que presentaron los mejores ajustes para cada una de las tres periodicidades de los datos climáticos en esta primer

etapas en las que se llevó a cabo la generación de los modelos de predicción de rendimiento de maíz, fueron los siguientes.

- **Resultados con datos climáticos diarios.** En el Cuadro 6, se señalan con verde los 3 modelos obtenidos de las 70 generados que presentaron los mejores ajustes en esta primera etapa para el caso de datos climáticos diarios y con 4 años anteriores, y con amarillo las combinaciones que tuvieron los mejores resultados de ajuste para cada combinación de variables climáticas.

Cuadro 6. Ajustes que tuvieron las siete diferentes combinaciones de datos climáticos diarios y con 4 años anteriores.

DATOS CLIMATICOS	PARTICIONES (porcentajes/días)									
	50:50	55:45	60:40	65:35	70:30	75:25	80:20	85:15	90:100	95:100
	2628	2891	3154	3416	3679	3942	4205	4468	4730	5000
TMax	47.22	15.36	23.91	5.68	19.69	64.57	36.37	43.82	23.21	11.63
TMin	59.67	43.7	45.67	35.57	51.71	66.84	44.77	44.72	30.90	14.01
pp	20.02	15.14	24.59	25.37	45.48	22.45	37.38	38.41	19.92	12.39
[Tmax Tmin]	62.78	4.61	14.31	-93.71	13.24	62.36	22.14	44.76	30.44	11.54
[Tmax pp]	-15.14	-28.54	-5.39	-14.21	-2.86	57.66	45.53	44.97	26.62	15.17
[Tmin pp]	59.96	48.08	50.37	43.55	60.57	44.86	0.19	20.95	8.72	6.65
[Tmax Tmin pp]	-104.4	-112	-91.71	-93.71	-79.97	-31.15	51.65	37.27	34.16	16.28

Fuente: Elaboración propia

Con esta información, se decidió utilizar para buscar en la segunda etapa mejores modelos con diferentes años anteriores a tomar en cuenta, únicamente los 3 casos seleccionados en esta primera etapa, que fueron Tmax (64.57%), Tmin 66.84%) y [Tmax Tmin] (62.36%).

- **Resultados de la primera etapa con datos climáticos mensuales para generar modelos utilizando para cada una de las 7 diferentes combinaciones de las variables climáticas y 4 años anteriores a tomar en cuenta.** En el Cuadro 7, se señalan con verde los 3 modelos obtenidos que presentaron los mejores ajustes en esta primera etapa para el caso de datos climáticos diarios y con 4 años anteriores. Se presenta con amarillo el resto de las combinaciones que tuvieron los mejores resultados de ajuste.

Cuadro 7. Ajustes que tuvieron las diferentes combinaciones de datos climáticos mensuales y con 4 años anteriores.

DATOS CLIMÁTICOS	PARTICIONES (porcentajes/días)									
	50:50	55:45	60:40	65:35	70:30	75:25	80:20	85:15	90:100	95:100
	87	96	104	113	122	131	139	148	157	165
Tmax	-10.38	-24.77	-19.30	-34.01	-40.08	49.35	50.58	34.96	24.02	74.27
Tmin	38.81	24.46	28.28	18.95	24.72	54.69	51.96	38.05	28.30	73.16
pp	65.81	54.49	52.88	46.16	54.57	65.34	28.68	25.21	6.88	49.35
[Tmax Tmin]	-17.69	-33.56	-29.95	-47.60	-58.82	27.63	30.03	29.69	26.88	70.76
[Tmax pp]	-10.00	-25.03	-20.87	-177.6	-44.12	59.34	50.77	35.38	24.19	74.10
[Tmin pp]	41.77	26.62	29.84	68.76	25.03	42.59	55.98	38.09	28.31	71.20
Tmax Tmin pp]	-17.87	-33.89	-31.76	-50.44	-65.99	20.24	25.70	30.81	27.32	70.46

Fuente: Elaboración propia

Se observa que los modelos obtenidos en general, presentan mejor ajuste que en el caso del uso de datos climáticos diarios. Por lo tanto, se decidió probar para buscar mejores modelos con diferentes años anteriores a tomar en cuenta, los 3 casos de datos climáticos que tuvieron los mayores porcentajes de ajuste con la partición 95:100%. fueron 3 y son Tmax (74.27%), Tmin (73.16%) y [Tmax pp] (74.10%) en esta primera etapa.

- **Resultados de la primera etapa con datos climáticos anuales para generar modelos utilizando para cada una de las 7 diferentes combinaciones de las variables climáticas y 4 años anteriores de los datos a tomar en cuenta.** En el Cuadro 8 se señalan con verde los 3 modelos obtenidos que presentaron los mejores ajustes en esta primera etapa para el caso de datos climáticos diarios y con 4 años anteriores. Se presenta con amarillo el resto de las combinaciones que tuvieron los mejores resultados de ajuste.

Se aclara que MATLAB no permitió en varios de los casos la estimación de los modelos con 4 años anteriores de los datos, teniendo que utilizar 3 o 2 años anteriores, como se señala con anaranjado claro la línea superior correspondiente en el Cuadro 8.

Cuadro 8. Ajustes que tuvieron las diferentes combinaciones de datos climáticos anuales y con 4 años anteriores de los datos y otros.

DATOS CLIMÁTICOS	PARTICIONES (porcentajes/días)						
	50:50	55:45	60:40	65:35	70:30	75:25	80:20
	15	16	17	19	20	22	23
Tmax	34.91	46.90	22.42	35.37	30.37	31.49	15.11
Tmin	47.79	45.70	38.63	54.31	41.21	21.07	-23.28
pp	39.95	39.95	30.40	38.78	55.99	71.23	34.16
[Tmax Tmin]	[4 3 1]						
	33.89	36.91	17.54	25.99	-16.36	19.02	-48.88
[Tmax pp]		[3 3]					
	52.02	50.42	40.39	52.02	38.75	53.75	-40.33
[Tmin pp]		[3 3]					
	39.22	32.34	16.05	-12.88	19.46	42.59	-37.7
[Tmax Tmin pp]	[2 2 2]		[3 3 3]				

Fuente: Elaboración propia

En los casos cuando se usaron las particiones de 85:15%, 90:100% y 95:100%, dieron valores de 100% de ajuste, lo que no es lógico y no se tomaron en cuenta.

Para buscar mejores modelos y continuar con el proceso, se probaron diferentes años anteriores de los datos a tomar en cuenta, en los 3 casos de datos climáticos que tuvieron los mayores porcentajes de ajuste en esta primera etapa, que fueron 3 y diferentes particiones, que son Tmin (54.31%), pp (71.23%) y [Tmax pp] (53.75%).

b. Segunda etapa o final del proceso mediante la elección del criterio de ajuste con diferentes años anteriores a tomar en cuenta en el modelo que se genere.

- **Para el caso de datos climáticos diarios.** Se probaron diferentes combinaciones ARX [na nb nk] y en el Cuadro 9 se señalan los resultados obtenidos de los modelos que presentaron los mejores porcentajes de ajuste en esta segunda etapa con diferentes combinaciones de años anteriores. Se presentan con verde los mejores resultados de ajuste y con azul el seleccionado.

Cuadro 9. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos diarios.

DATOS CLIMATICOS	Porcentaje de ajuste para diferentes combinaciones ARX [na nb nk]									
Tmax	[5 5 5]	[5 5 1]	[4 4 4]	[3 3 1]	[6 6 1]	[4 4 1]	[6 6 6]	[4 4 3]	[5 4 3]	[3 3 2]
	67.57	67.52	67.33	66.24	65.44	64.57	64.46	51.13	52.66	36.94
Tmin	[2 2 2]	[2 2 1]	[3 3 1]	[4 4 1]	[5 5 5]	[5 5 1]	[6 6 6]	[3 3 3]	[6 6 1]	[3 3 2]
	67.61	67.39	67.36	67.07	66.84	64.90	64.36	63.71	63.54	60.44
[Tmax Tmin]	[5 5 1]	[3 3 1]	[4 4 3]	[2 2 2]	[4 4 1]	[5 5 5]	[6 6 1]	[3 3 4]	[4 4 10]	[3 3 3]
	66.02	65.63	62.87	62.36	61.1	58.72	56.4	50.15	44.99	40.04

Fuente: Elaboración propia

Los resultados obtenidos permitieron comprobar que solo se mejoró muy poco el porcentaje de ajuste para las diferentes combinaciones probadas de diferentes años anteriores a tomar en cuenta y una partición de 75:25%, en el modelo que se generó utilizando datos climáticos diarios.

Por lo anterior, se seleccionó como el mejor modelo ARX manejando datos diarios del clima y con una partición de 75:25%, al correspondiente del factor climático Tmin [2 2 2] con un porcentaje de ajuste de 67.61%, considerando también aspectos prácticos por la simplicidad de aplicación que es utilizar dos y tres años anteriores, fórmula que es la siguiente:

$$y(t) = 1.993 \times y(t - 1) - 0.9931 y(t - 2) + 1.397 \times 10^{-6} u_1(t - 1) - 7.459 \times 10^{-7} u_1(t - 2) \quad (11)$$

Donde:

- ⇒ $y(t)$ representa la variable dependiente rendimiento predicha en el tiempo (año) t.
- ⇒ $u_1(t)$, es la variable independiente temperatura mínima medida en el tiempo (año) t manejando datos diarios.
- ⇒ t representan los años anteriores a tomar en cuenta

- **Para el caso de datos climáticos mensuales** Se probaron combinaciones de diferentes años anteriores ARX [na nb nk] y diferentes particiones, por lo que en el Cuadro 10 se señalan los resultados obtenidos de los modelos que

presentaron mejores ajustes en esta segunda etapa para el caso de datos climáticos mensuales

Cuadro 10. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos mensuales.

DATOS CLIMÁTICOS	Porcentaje de ajuste para diferentes combinaciones ARX [na nb nk]								
Tmax	[8 8 1]	[7 7 1]	[6 6 1]	[5 5 1]	[5 5 5]	[4 4 1]	[4 4 4]	3 3 1	[3 3 3]
	99.20	96.05	89.49	85.26	76.01	74.27	67.81	48.15	40.44
Tmin	[6 6 1]	[7 7 7]	[6 6 6]	[5 6 4]	[5 5 1]	[4 4 4]	[4 4 1]	[3 3 3]	[2 2 2]
	93.00	91.92	91.45	90.05	84.86	77.3	76.16	52.36	23.47
[Tmax pp]	[2 2 1]	[4 4 1]	[3 3 1]	[4 2 2]	[6 6 1]	[3 2 2]	[2 2 2]	[5 5 2]	
	74.66	72.81	71.73	71.60	71.36	71.35	71.07	69.08	

Fuente: Elaboración propia

Además, en varios casos se obtuvieron valores mayores a 90% de ajuste, lo que no es posible y no se tomaron en cuenta.

Por lo anterior, se seleccionó como el mejor modelo ARX manejando datos mensuales del clima, al correspondiente del factor climático Tmax [6 6 1] con una partición de 75:25%, que tuvo un porcentaje de ajuste de 89.49%, cuya fórmula es la siguiente:

$$\begin{aligned}
 y(t) = & 1.977 \times y(t-1) - 1.161 y(t-2) + 0.1953 \times y(t-3) + \\
 & 0.03332 \times y(t-4) + 0.2621 \times y(t-5) + 0.2187 \times y(t-6) + \\
 & 0.0002037 u_1(t-1) - 0.001209 u_1(t-2) + 0.0006463 \times \\
 & u_1(t-3) + 0.0007045 \times u_1(t-4) + 0.00182 \times u_1(t-5) + \\
 & 0.001449 \times u_1(t-6)
 \end{aligned} \tag{12}$$

Donde:

- ⇒ $y(t)$ representa la variable dependiente rendimiento predicha en el tiempo (año) t .
- ⇒ $u_1(t)$, es la variable independiente temperatura máxima medida en el tiempo (año) t manejando datos mensuales.
- ⇒ t representan los años anteriores a tomar en cuenta

- **Para el caso de datos climáticos anuales.** Se probaron combinaciones de diferentes años anteriores ARX [na nb nk] y en el Cuadro 11 se señalan los resultados obtenidos de los modelos que presentaron mejores ajustes en esta segunda etapa para el caso de datos climáticos anuales.

Cuadro 11. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos anuales.

DATOS CLIMÁTICOS	Porcentaje de ajuste para diferentes combinaciones ARX [na nb nk]								
Tmin	[4 5 1]	[4 6 2]	[4 7 1]	[4 5 1]	[4 4 1]	[5 5 1]	[6 6 1]	[4 5 2]	[3 3 1]
	63.23	57.37	57.13	56.49	54.31	52.15	45.98	44.76	41.25
pp	[4 5 1]	[4 5 3]	[4 4 1]	[4 5 4]	[4 3 1]	[4 5 2]	[3 3 3]	[4 4 4]	[5 5 5]
	76.07	75.42	71.23	70.66	69.38	68.01	67.19	63.36	48.14
[Tmax pp]	[5 4 2]	[5 4 2]	[5 4 3]	[5 5 1]	[4 4 1]	[4 3 2]	[4 3 1]	[5 4 1]	[3 4 1]
	67.10	66.06	56.94	54.6	53.75	51.52	50.3	36.32	17.81

Fuente: Elaboración propia

Por lo anterior, se seleccionó como el mejor modelo ARX de MATLAB manejando datos anuales del clima, al correspondiente del factor climático precipitación con una partición de 75:25% y [4 5 1] años anteriores a tomar en cuenta, que tuvo un porcentaje de ajuste de 76.07%, cuya fórmula es la siguiente:

$$y(t) = -0.1714 \times y(t-1) + 0.3663 y(t-2) - 0.3194 \times y(t-3) + 0.5049 \times y(t-4) - 0.5897 \times u_1(t-1) + 0.8768 u_1(t-2) + 0.4648 \times u_1(t-3) + 0.5435 \times u_1(t-4) + 0.03579 \times u_1(t-5) \quad (13)$$

Donde:

- ⇒ $y(t)$ representa la variable dependiente rendimiento predicha en el tiempo (año) t .
- ⇒ $u_1(t)$ es la variable independiente precipitación medida en el tiempo (año) t manejando datos anuales.
- ⇒ t representan los años anteriores a tomar en cuenta

5.2.2. Uso de Modelos aditivos generalizados (Generalized Additive Models, GAM) con el lenguaje de programación R.

Este procedimiento fue mucho más sencillo que el utilizado con MATLAB, aunque tiene la desventaja que solo proporciona los datos predichos. Los resultados obtenidos para las siete combinaciones de los datos climáticos se presentan en el Cuadro 12, señalando con verde los 3 mejores resultados de ajuste.

Cuadro 12. Porcentaje de ajuste en orden decreciente para diferentes combinaciones ARX [na nb nk] utilizando datos climáticos anuales.

Variables	R-sq.(adj)	AIC
Tmax	0.667	99.448
Tmax Tmin P	0.658	100.543
Tmin P	0.625	102.540
Tmax Tmin	0.491	110.975
P	0.455	100.543
Tmax	0.354	100.543
Tmin	0.256	118.293

*AIC = Criterio de información de Akaike que mide la calidad relativa de un modelo estadístico para un conjunto dado de datos (a menor valor mejor calidad).

Fuente: Elaboración propia

Por lo tanto, se generaron los datos predichos para los tres primeros casos, los cuáles se compararon con los mejores obtenidos con MATLAB, para seleccionar el mejor a través de una prueba de desviación estándar contra los datos medidos, que fue el que se utilizó en la aplicación del Filtro de Kalman Simplificado.

Como un primer comentario, todas las medias de los valores predichos en R son iguales con este procedimiento y como un segundo comentario, con este procedimiento se hizo un ejercicio de incluir como variable de entrada a los años, obteniendo muchos mejores resultados de R-sq.(adj) y AIC.

5.3. Comparación de los datos predichos para seleccionar el mejor modelo y utilizar sus datos en la aplicación del Filtro de Kalman Simplificado.

Se obtuvieron los datos predichos de los 6 modelos seleccionados de MATLAB y de R, en los que solo se pueden aplicar utilizando los promedios de los datos anuales de las

variables climáticas, para calcular su desviación estándar con relación a los datos medidos (Cuadro 13):

Cuadro 13. Datos predichos de los 6 modelos seleccionados de MATLAB y de R

No	Año	Rend med	Datos predichos de rendimiento					
			Tmin	Tmax	pp	Tmax-Tmin-P	Tmax-P	Tmin-P
1	1993	7.04				8.20	7.76	8.95
2	1994	7.77				8.59	8.30	8.83
3	1995	8.30				8.34	8.12	8.54
4	1996	8.21	8.83			10.40	10.40	10.21
5	1997	8.07	8.12			8.60	8.67	8.54
6	1998	8.40	7.94		8.70	7.86	7.86	8.07
7	1999	8.55	8.72	8.61	8.62	8.36	8.24	9.02
8	2000	9.12	8.69	8.52	9.32	9.00	9.23	8.88
9	2001	9.64	9.69	9.68	8.96	9.09	9.07	9.85
10	2002	9.17	10.16	10.10	9.11	10.44	10.61	10.36
11	2003	9.89	8.70	8.57	9.33	10.14	10.29	9.69
12	2004	8.71	10.60	10.69	9.34	9.73	9.99	9.26
13	2005	10.08	7.55	7.33	9.81	8.66	9.06	8.19
14	2006	10.06	11.43	11.51	9.96	9.96	10.14	9.64
15	2007	10.50	10.05	9.93	10.73	10.06	10.17	10.03
16	2008	10.19	10.93	10.74	10.19	10.71	11.11	10.48
17	2009	10.46	9.89	10.14	10.64	10.19	9.95	10.33
18	2010	10.77	10.72	10.49	10.58	11.50	11.78	11.18
19	2011	10.52	11.08	11.05	10.57	9.44	9.43	9.47
20	2012	10.26	10.26	10.12	10.87	10.80	11.03	10.72
21	2013	10.47	10.01	10.16	10.76	11.65	11.65	11.65
22	2014	11.68	10.68	10.69	10.95	12.29	12.03	12.26
23	2015	11.86	12.87	12.74	10.61	13.05	12.37	13.60
24	2016	13.37	12.05	11.88	10.94	12.02	11.85	12.17
25	2017	13.37	14.87	14.90	11.29	12.94	13.08	12.41
26	2018	13.44	13.38	13.12	12.47	12.45	12.21	12.28
27	2019	13.54	13.49	13.27	13.18	11.70	11.50	12.00
28	2020	14.30	13.65	13.68	13.77	13.42	13.48	13.36
29	2021	13.62	15.05	14.71	13.45	11.73	11.99	11.41
	Prom	10.91	10.74	10.98	10.58	10.39	10.39	10.39
	Desv E	1.95	1.74	1.28	1.60	1.60	1.54	

Fuente: Elaboración propia

El modelo ARX de MATLAB que obtuvo la desviación más baja fue cuando se utilizó como variable de entrada a la precipitación manejando datos anuales del clima, con una

partición de 75:25% y [4 5 1] años anteriores a tomar en cuenta, por lo que los datos predichos inician en 1998 y el modelo matemático se presentó en la ecuación (13).

5.4. Aplicación del Filtro de Kalman Simplificado (FKS) para mejorar la calidad de los datos del rendimiento.

En el Cuadro 14 se presenta el resumen de operaciones para aplicar el FKS en Excel, en donde estrictamente el usuario sólo tiene que llenar la segunda, tercera y cuarta columnas y la hoja de cálculo hace el resto de las operaciones.

Cuadro 14. Resumen de operaciones realizadas para aplicar el FKS.

Número de intervalo de tiempo	Año	Mediciones rendimiento	Predicciones optimizadas por el FKS	Predicciones rendimiento	Media recursiva predicciones	Varianza recursiva predicciones	Media recursiva mediciones	Varianza recursiva mediciones	Ganancia de Kalman
1	1998	8.400	8.550	8.701					
2	1999	8.548	8.565	8.619	8.660	0.003	8.474	0.011	0.768
3	2000	9.124	9.219	9.316	8.878	0.145	8.691	0.146	0.502
4	2001	9.643	9.485	8.961	8.899	0.098	8.929	0.324	0.767
5	2002	9.170	9.155	9.110	8.941	0.083	8.977	0.255	0.755
6	2003	9.889	9.771	9.326	9.005	0.091	9.129	0.342	0.790
7	2004	8.713	8.856	9.339	9.053	0.092	9.070	0.310	0.772
8	2005	10.077	10.003	9.811	9.148	0.150	9.196	0.393	0.723
9	2006	10.065	10.031	9.962	9.238	0.205	9.292	0.427	0.676
10	2007	10.500	10.602	10.733	9.388	0.406	9.413	0.526	0.564
11	2008	10.195	10.192	10.189	9.461	0.424	9.484	0.529	0.555
12	2009	10.457	10.545	10.643	9.559	0.502	9.565	0.560	0.527
13	2010	10.769	10.683	10.584	9.638	0.541	9.658	0.624	0.536
14	2011	10.516	10.542	10.571	9.705	0.561	9.719	0.629	0.528
15	2012	10.262	10.567	10.867	9.782	0.611	9.755	0.604	0.497
16	2013	10.475	10.624	10.764	9.844	0.631	9.800	0.596	0.486
17	2014	11.677	11.339	10.948	9.909	0.663	9.911	0.766	0.536
18	2015	11.864	11.348	10.610	9.948	0.652	10.019	0.933	0.589
19	2016	13.371	12.612	10.938	10.000	0.667	10.195	1.472	0.688
20	2017	13.375	12.804	11.285	10.064	0.715	10.354	1.900	0.727
21	2018	13.435	13.148	12.467	10.178	0.954	10.501	2.257	0.703
22	2019	13.545	13.420	13.176	10.315	1.317	10.639	2.571	0.661
23	2020	14.301	14.106	13.773	10.465	1.777	10.799	3.037	0.631
24	2021	13.617	13.550	13.445	10.589	2.070	10.916	3.236	0.610
Medias		10.916	10.822	10.589					
Varianzas		3.236	2.715	2.070					

Fuente: Elaboración propia

Al observar el Cuadro 14, se detecta claramente que el valor mejorado de la predicción de rendimiento de maíz por el FKS (columna 4), siempre se encuentra comprendido entre los valores medidos (columna 3) y los datos predichos (columna 5). Así mismo, que en

cada paso de tiempo, dicho valor mejorado se aproxima más al valor que tiene menor varianza.

En la Figura 11 se muestran gráficamente que los valores calculados con el FKS siempre están comprendidos entre los valores calculados por los modelos 1 y 2, pues esto hace que las 3 medias sean prácticamente iguales. los valores de las mediciones (primer modelo), de las predicciones obtenidas con el segundo modelo y de las predicciones mejoradas del segundo modelo por el FKS.

Llama la atención que la media de las predicciones mejoradas por el FKS es muy cercana a la de los valores medidos.

Las Figuras 11 y 12 muestran el resultado de la programación del FKS en Excel y MATLAB respectivamente, y su comparación con los modelos 1 y 2. Ambos programas en Excel y MATLAB produjeron resultados idénticos, lo cual queda claro al comparar las Figuras 11 y 12 y también los valores de las medias y varianzas de los tres modelos, los cuales en ambas implementaciones resultaron ser los mismos.

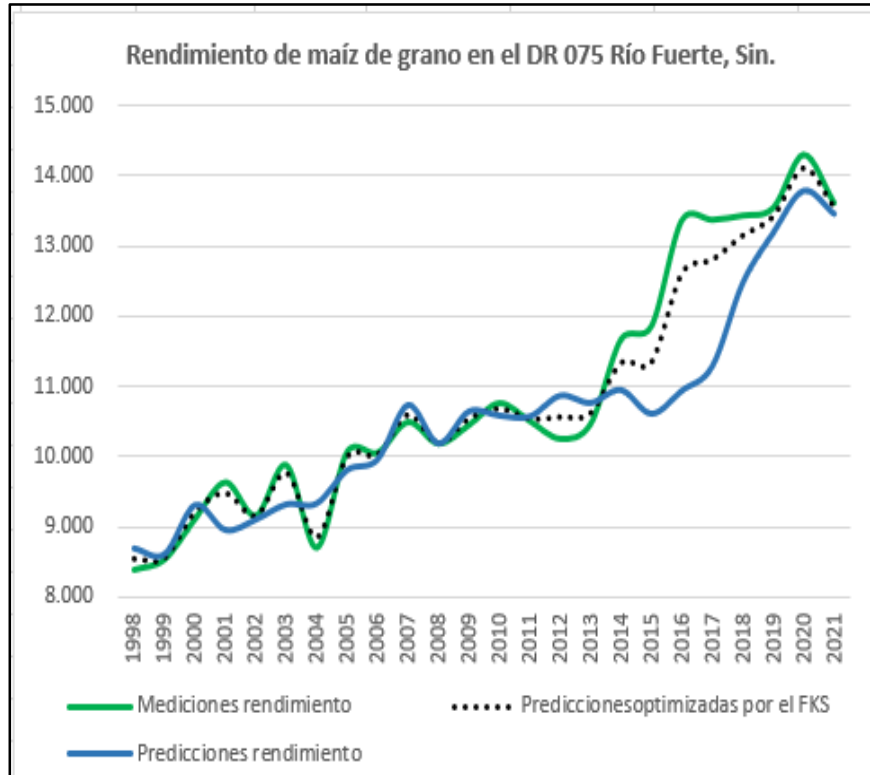


Figura 9. Comparación en Excel de las predicciones de los modelos implementados.

Fuente: Elaboración propia

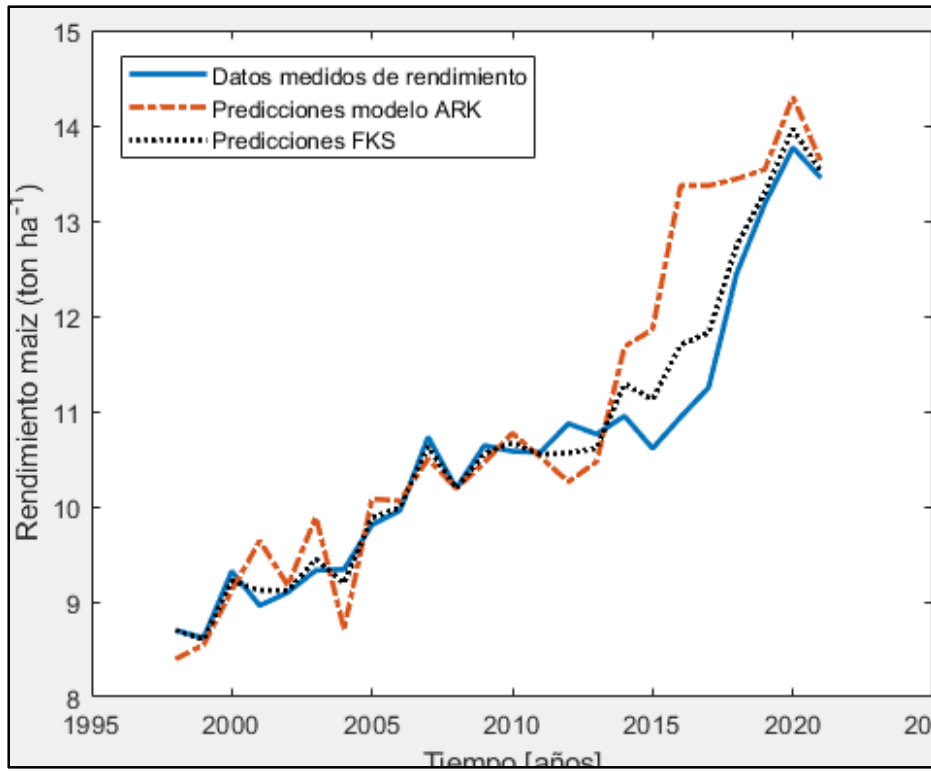


Figura 10. Comparación en MATLAB de las predicciones de los modelos implementados.
Fuente: Elaboración propia

6. CONCLUSIONES

- 6.1. Con el script *Climatol* de R fue posible completar los datos climáticos faltantes.
- 6.2. Con la herramienta *systemIdentification* del ambiente de programación MATLAB, se generaron un gran número de modelos ARX para la predicción de rendimiento de maíz, entre cientos de escenarios que se pueden manejar internamente, utilizando como variables de entrada las 7 combinaciones de datos climáticos, que fueron temperatura máxima (Tmax), temperatura mínima (Tmin) y precipitación (pp) y ciento de combinaciones de particiones y de años a tomar en cuenta en la generación del modelo.
- 6.3. El procedimiento estadístico no paramétrico conocido como Modelos aditivos generalizados (GAM) en el lenguaje de programación R, permitió con una relativa facilidad, obtener modelos y datos predictivos de rendimiento de maíz de grano con datos climáticos anuales.
- 6.4. Los tres modelos generados con MATLAB después de la segunda etapa que tuvieron mayor ajuste para cada una de las tres periodicidades de datos, fueron:
 - ⇒ Con datos diarios, cuando se utilizó la variable temperatura mínima (Tmin), con la partición 75:25% y [2 2 2] de años anteriores, que tuvo un porcentaje de ajuste de 67.61% y se puede observar en la Fórmula (11).
 - ⇒ Con datos mensuales, cuando se utilizó la variable temperatura máxima (Tmax), con la partición 95:100% y [6 6 4] de años anteriores, que tuvo un porcentaje de ajuste de 89.49% y se puede observar en la Fórmula (12).
 - ⇒ Con datos anuales, cuando se utilizó la variable precipitación (pp), con la partición 75:25% y [4 5 1] de años anteriores, que tuvo un porcentaje de ajuste de 76.07% y se puede observar en la Fórmula (13).
- 6.4. Se obtuvieron con R diferentes modelos GAM de rendimiento de maíz sin contar con las respectivas ecuaciones matemáticas, utilizando como variables de entrada, las distintas combinaciones de datos climáticos que fueron temperatura máxima (Tmax), temperatura mínima (Tmin) y la precipitación (pp).

6.5. Los tres mejores modelos GAM generados con R para el caso de la periodicidad de dato anuales fueron:

- ⇒ Cuando se utilizó la variable temperatura máxima (Tmax), que tuvo una R-sq.(adj) de 0.667%.
- ⇒ Cuando se utilizaron las variables temperatura máxima (Tmax), temperatura mínima (Tmin) y precipitación (P), que tuvo una R-sq.(adj) de 0.658%.
- ⇒ Cuando se temperatura mínima (Tmin) y precipitación (P), que tuvo una R-sq.(adj) de 0.625%.

6.6. Para seleccionar el mejor modelo de todos los generados y utilizar sus predicciones para aplicar el FKS, se utilizaron los 6 mejores modelos obtenidos con MATLAB y R para obtener los datos predictivos y evaluarlo, para lo cual se calculó su desviación estándar con relación a los datos medidos. El que tuvo la menor desviación fue el correspondiente al factor climático precipitación, con una partición de 75:25% y [4 5 1]) años anteriores a tomar en cuenta, que tuvo un porcentaje de ajuste de 76.07%. Su ecuación es la siguiente:

$$y(t) = -0.1714 \times y(t-1) + 0.3663 y(t-2) - 0.3194 \times y(t-3) + 0.5049 \times y(t-4) - 0.5897 \times u_1(t-1) + 0.8768 u_1(t-2) + 0.4648 \times u_2(t-3) + 0.5435 \times u_3(t-4) + 0.03579 \times u_5(t-5) \quad (13)$$

Donde:

- $y(t)$ representa la variable dependiente rendimiento predicha en el tiempo (año) t.
- $u_1(t)$ es la variable independiente precipitación medida en el tiempo (año) t manejando datos anuales.
- t representan los años anteriores a tomar en cuenta.

6.7. Se desarrolló una versión simplificada del Filtro de Kalman, que permite mejorar la calidad de los datos existentes de los rendimientos de maíz de grano en el DR 75, Río Fuerte, Sin., a partir de una serie de mediciones y un modelo predictivo,

interpolando con factores de ponderación inversamente proporcionales a las varianzas, las que se actualizan de manera recursiva en cada intervalo de tiempo.

- 6.8. Se muestra que la “ganancia de Kalman”, fórmula (2), que es el parámetro central del filtro, es resultado de un proceso de optimización o minimización de las varianzas.
- 6.9. La aportación del FKS se expresa en el hecho de que la secuencia de valores mejorados por el uso del filtro, siempre queda comprendida entre los valores predichos y los medidos, acercándose más los valores mejorados al valor de los predichos y de los medidos que tenga la menor varianza.

7. LITERATURA CITADA

- ASERCA, 2019. Obtenido de: <https://www.gob.mx/aserca/articulos/maiz-grano-cultivo-representativo-de-mexico?idiom=es#:~:text=El%20ma%C3%ADz%2C%20es%20uno%20de,importancia%20econ%C3%B3mica%2C%20social%20y%20cultural>
- ASERCA, 2021. Obtenido de: <https://www.gob.mx/aserca/articulos/maiz-grano-cultivo-representativo-de-mexico?idiom=es>
- Ban H-Y, Ahn J-B, Lee B-W. 2019. Assimilating MODIS data-derived minimum input data set and water stress factors into CERES-Maize model improves regional corn yield predictions. PLoS ONE 14(2): e0211874. Obtenido de:
<https://doi.org/10.1371/journal.pone.0211874>
- Beven, K. 2009. Environmental Modelling: An Uncertain Future? New York, USA: Routledge.
- Becker, A. Kalman Filter Tutorial. 2022. Obtenido de:
<https://www.kalmanfilter.net/default.aspx>
- BUAP, 2017. Obtenido de:
<https://repositorioinstitucional.buap.mx/handle/20.500.12371/488>
- Cadarso S. M. C. 2021. Obtenido de:
[https://mesioupcub.masters.upc.edu/en/xiv-summer-school-2021/courses/modelos-aditivos-generalizados-y-extensiones-aplicaciones-en-biomedicina-july-5th-to-july-9th#:~:text=Los%20Modelos%20Aditivos%20Generalizados%20\(Generalized,%2C%20Econom%C3%ADa%2C%E2%80%A6%2C%20La](https://mesioupcub.masters.upc.edu/en/xiv-summer-school-2021/courses/modelos-aditivos-generalizados-y-extensiones-aplicaciones-en-biomedicina-july-5th-to-july-9th#:~:text=Los%20Modelos%20Aditivos%20Generalizados%20(Generalized,%2C%20Econom%C3%ADa%2C%E2%80%A6%2C%20La)
- CLICOM, 2022. Obtenido de: <http://clicom-mex.cicese.mx/mapa.html>
- Climatol 2019. Obtenido de: <https://CRAN.R-project.org/package=climatol>
- Climatol 2020. Obtenido de: <http://www.climatol.eu/tt-hom/index.html>
- CONAGUA. Estadísticas DR 2017-2018
- CONAGUA, 2022, ESTADÍSTICAS 2021,
- CONAGUA, 1992-2021. Estadísticas de diferentes años de Distritos de Riego
- Cottrell A. y Lucchetti R. J. 2005. Guía del usuario de Gretl. Obtenido de:
https://ocw.uniovi.es/pluginfile.php/2849/mod_resource/content/1/T_1C%2CA_668/Gretl/Guia_Gretl.pdf

Delgadillo R. O., Ramírez M. P. P., Leos R. J. A., Salas G., J. M. y Valdez.C. R. D. 2016. Pronósticos y series de tiempo de rendimientos de granos básicos en México. Acta Universitaria, 26(3), 23-32. doi: 10.15174/au.2016.882. Obtenido de:

https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-62662016000300023

De Wit A.J.W. and van Diepen C.A. 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. Wageningen University & Research, Agricultural and Forest Meteorology. Volume 146, Issues 1–2, 11 September 2007, Pages 38-56. Obtenido de:

<https://www.sciencedirect.com/science/article/abs/pii/S0168192307001402?via%3Dihub>

DelftStack, 2022. Obtenido de:

<https://www.delftstack.com/es/howto/python/opencv-kalman-filter/>

DOCPLAYER, 2022a. Obtenido de <https://docplayer.es/153629699-Reporte-del-mercado-de-maiz-agosto-2019.html>

DOCPLAYER, 2022b. Obtenido de: <https://docplayer.es/36813202-Modelo-aditivo-generalizado-gam.html>

Gómez M. R., Ortiz S. C., Zamora D. M., Soria R. J., Trinidad S. A. y Carballo C. A. 2009. Estimación del rendimiento de cebada (*Hordeum vulgare* L. maltera con el método FAO. Agricultura Técnica en México Vol. 35 Núm. 1 p. 13-23, México. Obtenido de:

<https://www.scielo.org.mx/pdf/agritm/v35n1/v35n1a2.pdf>

González L. F., Ibáñez C. L. A., Valdés, J. B., Vázquez P. M. A. y Ruiz G. A. 2015. Pronóstico de caudales con Filtro de Kalman Discreto en el río Turbio. Tecnología y Ciencias del Agua, 6(4), 5-24. Obtenido de:

<http://www.revistatyca.org.mx/ojs/index.php/tyca/article/view/1176/1332>

Guijarro, J. 2018_ Obtenido de: https://www.climatol.eu/homog_climatol-es.pdf.

Herrera B. J., Peinado G. H. J., Campos G. J., Cruz E. A., Peinado G., V. M., Ladrón de Guevara T. M. Belmonte J. S. I. y Herrera, L. E. 2021. Determinación de la transmisibilidad hidráulica en un acuífero costero mediante estimación óptima de la relación Qe-T usando e lfiltro de Kalman *HIDROBIOLÓGICA*, 30(3). Obtenido de:

<https://doi.org/10.24275/uam/izt/dcbs/hidro/2020v30n3/Herrera>

[Vista de Vol. 30 Núm. 3 \(2020 \(uam.mx\)](#)

IBM, 2022a. Obtenido de:

<https://www.ibm.com/docs/es/spss-statistics/saas?topic=forecasting-time-series-modeler>

Mathematics, 2022. Obtenido de:

<https://math.stackexchange.com/questions/2837813/sample-variance-recursive-relation-n-s-n12-n-1s-n2-big-fracnn>

IBM, 2022b. Obtenido de:

<https://www.ibm.com/docs/es/spss-statistics/saas?topic=series-building-models-producing-forecasts>

Infostat: manual del usuario 2008. Obtenido de:

https://repositorio.catie.ac.cr/bitstream/handle/11554/10346/Manual_INFOSTAT_2008.pdf?sequence=1 y <https://www.infostat.com.ar/>

Kim P and Huh L. 2011. Kalman Filter for Beginners with MATLAB Examples.

Licona S., J.; Martínez M., Mario R.; Mendoza O., L. E.; Figueroa S., B.; Fernández R., D. S. 2006. Uso del modelo EPIC para estimar rendimientos de maíz con base en variables fisiotécnicas en el oriente del Estado de México. SMCS. Terra Latinoamericana, vol. 24, núm. 2, abril-junio, 2006, pp. 283-291, Chapingo, México. Obtenido de:

<https://www.redalyc.org/pdf/573/57311108016.pdf>

López C. L. I. 2022. Elaboración en MATLAB de un pseudo-código de todo el proceso realizado en Excel que incluye la elaboración de la gráfica (comunicación personal).

Martínez S. L. E. 2013. Modelo econométrico para el volumen de producción de maíz en cultivo de riego (1980-2011). Tiempo Económico. Núm. 23, vol. VIII, Primer cuatrimestre de 2013. Obtenido de:

<http://tiempoeconomico.azc.uam.mx/wp-content/uploads/2017/08/23te5.pdf>

Mateos M. M. 2020. Clasificación de paquetes de series temporales en R. Ejemplo de modelización ARIMA con datos de contaminación de Madrid. Obtenido de:

<https://uvadoc.uva.es/bitstream/handle/10324/43849/TFG-G4617.pdf?sequence=1&isAllowed=y>.

MathWorks, 2022. [Filtros Kalman - MATLAB & Simulink \(mathworks.com\)](https://www.mathworks.com/help/matlab/kalman_filters.html)

Michel L. and Makowski D. 2013. Comparison of Statistical Models for Analyzing Wheat Yield Time Series. PLoS ONE 8(10): e78615.

Moira M. D. E. 2013. Filtro de Kalman Federado con fusión adaptativa: Aplicación a la destilación por lotes. Congreso Nacional de Control Automático, Ensenada, Baja California, Mexico, Octubre 16-18, 2013

<http://amca.mx/memorias/amca2013/papers/0057.pdf>

Montes G. J. 2020. Control predictivo basado en Machine Learning de una planta de laboratorio. Trabajo de fin de grado en Ingeniería Electrónica, Robótica y Mecatrónica. Escuela Técnica Superior de Ingeniería, Universidad de Sevilla, España.

<https://biblus.us.es/bibing/proyectos/abreproy/93346/fichero/TFG-3346+MONTES+GALLARDO%2C+A.J..pdf>

Morales V., M. I. y. Aparicio M. F. J. 2014. Pronóstico de avenidas usando el filtro de Kalman discreto. Tecnol. cienc. agua vol.5 no.2 Jiutepec mar./abr.

https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-24222014000200006#:~:text=La%20aplicaci%C3%B3n%20del%20filtro%20de%20Kalman%20al%20pron%C3%B3stico%20de%20avenidas,la%20operaci%C3%B3n%20de%20grandes%20embalses.

Narvárez O. I., Ibáñez C., L. A., Arteaga R., R., Vázquez P. M., y Cántora G. C. 2021. Comparación del filtro de Kalman discreto con el filtro de conjuntos para pronóstico de caudales horarios en el río Huaynamota, Nayarit, México. Tecnología Y Ciencias Del Agua, 12(6), 239-281. <https://doi.org/10.24850/j-tyca-2021-06-06>
Obtenido de:

<http://revistatyca.org.mx/index.php/tyca/article/view/2507>

Noriega N. J. L. 2016. Utilización de modelos de simulación de crecimiento para maíz bajo escenarios de cambio climático. 6° Congreso Nacional de investigación en cambio climático. México. Obtenido de:

<https://cienciasagricolas.inifap.gob.mx/index.php/agricolas/article/view/2552/3859>

Noriega N. J. L., Salazar M. R., López C. I. L. 2021. Revisión: modelos de crecimiento y rendimiento de maíz en escenarios de cambio climático. Revista Mexicana de Ciencias Agrícolas volumen 12 número 1 01 de enero - 14 de febrero, 2021.
Obtenido de:

https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-09342021000100127

Ortiz S. C. A., Gutiérrez C. M. C. y Nieves F. J. 2005. Estimación de rendimientos de maíz con el método FAO en el ejido de Atenco, Estado de México. Revista de Geografía Agrícola, núm. 35, pp. 57-65, Universidad Autónoma Chapingo.
Obtenido de: <https://www.redalyc.org/articulo.oa?id=75703503>

Palacios V. L. O. 2021. Comunicación personal sobre información del Filtro de Kalman.

Pei Y, Biswas S, Fussell DS and Pingali k. 2019. An elementary introduction to Kalman. Filtering. Communications of the ACM 62 (11): 122-133. Obtenido de:

<https://arxiv.org/pdf/1710.04055.pdf>

REPOSITORIO, 2020. Obtenido de:

<https://repositorio.chapingo.edu.mx/handle/20.500.12098/634>

Red Valle del Fuerte, 2021. Obtenido de: <https://redvalledelfuerte.org/>

Sánchez H. G. A., Júnez F. H. E., Alvarez R. W., González T. J., Bautista C. C. y Herrera Z. G. S. 2017 Optimización del monitoreo del nivel del agua subterránea para una frecuencia fija. IMTA. Obtenido de:

https://www.scielo.org.mx/scielo.php?pid=S2007-24222017000500019&script=sci_arttext

Sánchez L. y Saba I. L. S. 2012. El filtro de Kalman de los ensambles y filtro de Kalman de cuadratura para predecir precipitación en Venezuela. Revista de Climatología, Vol. 12 (2012: 33-48. Obtenido de:

https://www.researchgate.net/publication/312660256_FILTRO_DE_KALMAN_DE_LOS_ENSAMBLAS_Y_FILTRO_DE_KALMAN_DE_CUADRATURA_PARA_PREDICIR_PRECIPITACION_EN_VENEZUELA

SAS, 2022. SAS Visual Forecasting | SAS Support. Obtenido de:

<https://support.sas.com/en/software/visual-forecasting-support.html>

SIAP, 2022. Obtenido de <https://nube.siap.gob.mx/cierreagricola/>

Sifuentes I. E. 2020. Modelo de gestión del servicio de riego para grandes zonas agrícolas, apoyado con percepción remota e información climática. Tesis Doctoral Instituto Mexicano de Tecnología del Agua (IMTA).

Skforecast, 2022. Forecasting series temporales con Python y Scikit-Seffrin R., Coimbra de Araújo E. and Leones B. C. 2018. Regression models for prediction of corn yield in the state of Paraná (Brazil from 2012 to 2014 Biometry, Modeling and Statistics. Acta Scientiarum. Agronomy, v. 40, e36494, **learn**. Obtenido de:

<https://www.cienciadedatos.net/documentos/py27-forecasting-series-temporales-python-scikitlearn.html>

Soria R. J., Báez G. A. D., Rodríguez M. V. M., Ramos G. J. L., Reyes M. L., González G. M. A. 2007. Publicación Especial producida por el Laboratorio Nacional de Modelaje y Sensores Remotos y el Laboratorio de Geomática del Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. ISBN 978-970-43-0157-6. Obtenido de:

https://www.scielo.org.mx/scielo.php?pid=S2007-24222017000500019&script=sci_arttext

Vargas C.G., Ibáñez C., L. A., y Arteaga R. R. 2015. Development, classification and trends in rainfall-runoff modeling. Ingeniería Agrícola y Biosistemas, 7(1), 5-21. doi: 10.5154/r.inagbi.2015.03.002. Obtenido de:

<https://revistas.chapingo.mx/inagbi/revista/articulos/r.inagbi.2015.03.002.pdf>

Wikipedia https://es.wikipedia.org/wiki/Filtro_de_Kalman, 21 de abril de 2022

Wan E. A. and Van Der Merwe R. 2001. The unscented Kalman Filter, Chapter 5. Obtenido de:

https://www.pdx.edu/biomedical-signal-processing-lab/sites/www.pdx.edu/biomedical-signal-processing-lab/files/ukf.wan_chapt7.pdf

World Documents, 2022. Reporte del mercado de maíz enero 2020, Obtenido de <https://vdocuments.net/reportes-del-mercado-de-maz-a-el-maz-junto-con-el-trigo-y-el-arroz-es-uno.html?page=6>

ANEXOS

Anexo 1. Metodología detallada que se siguió para completar los datos climáticos faltantes con el Script climatol del lenguaje de programación R

1. **Preparación de todos los datos climáticos obtenidos.** Para poder trabajar Climatol en R, se requieren dos archivos de entrada, uno tipo *dat* con todos los datos en orden cronológico desde la primera estación hasta la última, el cual se genera en MATLAB, y otro tipo *txt* con información de las estaciones.

Los pasos que se siguieron para elaborar los dos archivos requeridos fueron los siguientes.

- i. Crear un archivo en Excel con todos los datos diarios de las 3 variables de las 3 estaciones.** Para ello, se concentró en un archivo general de Excel, todos los datos climáticos diarios de cada estación (9 archivos *csv*), colocando en tres hojas diferente del archivo los 19,358 datos diarios de 1969 a 2021 de cada una de las 3 variables climáticas de las 3 estaciones. Se verificó que la fecha de inicio fuera el 1 de enero de 1969 y la de término el 31 de diciembre de 2021, que es un requisito de la metodología. Así, los datos totales para cada variable fueron 19,358 en su respectiva columna y un total de 58,074 en las tres columnas de las 3 estaciones, información que se requiere indicar en el script de MATLAB.
- ii. Reemplazo de datos faltantes por NaN.** En dicho archivo se deben sustituir todos los datos vacíos por NaN, para que Climatol los pueda identificar.
- iii. Nombre del archivo concentrado de Excel.** Se nombró "DATOS_ALL_1969-2021.xls".
- iv. Procesamiento en MATLAB del archivo de Excel.** Este archivo fue procesado en MATLAB para cada variable climática, utilizando un script "GenArc.m" que proporcionó el alumno Miguel Servín de postgrado de la UACH, del Dr. Irineo López y que se adaptó por el suscrito para cada una, siendo los tres scripts los siguientes:
 - "GenArc_Tmax_69_21"
 - "GenArc_Tmin_69_21"

- “GenArc_pp_69_21”.

Como ejemplo de los scripts de MATLAB, para el caso de temperatura máxima (Tmax) fue el siguiente:

```
clear; clc; close all;
Var=xlsread('Datos_1969_2021.xlsx', 'Tmax', 'B3:D19360');
L=size(Var,1);
Tmax1=Var(1:L,1);
Tmax2=Var(1:L,2);
Tmax3=Var(1:L,3);
%Vector de datos
Tmax=[Tmax1', Tmax2', Tmax3'];
n=length(Tmax);
%Abre y genera archivo de datos
Prueba= fopen('Tmax_1969-2021.txt','w');
for i=1:size(Tmax,2)
    fprintf(Prueba, '%f ',Tmax(i));
```

MATLAB generó tres archivos de datos en formato *txt*, que fueron “Tmax_1969-2021.txt”, “Tmin_1969-2021.txt” y “PP_1969-2021.txt”.

- v. Cambio manual del nombre del archivo con terminación **.txt* a terminación **.dat*.** A los tres archivos tipo *txt* generado por MATLAB, se les cambió de manera manual el formato a *dat*, para que Climatol en R los pueda leer, utilizando el explorador en donde al seleccionar el archivo, con el botón derecho se les da editar y se les agrega la terminación *dat*.
- vi. Elaboración de los archivos con información de las estaciones.** Son archivos que se generan en el bloc de notas con terminación *txt* para cada variable, con una lista o información de las 3 estaciones, en donde viene los datos de cada una de ellas de longitud, latitud, msnm, clave y nombre de las 3 estaciones, quedando como sigue.

```
-109.183 25.917 10 "25003" "AHOME"
-109.030 25.800 11 "25008" "LOS MOCHIS"
-108.500 25.600 22 "25038" "GUASAVE"
```

A los tres archivos tipo *txt* generado por MATLAB, se les cambió de manera manual el formato a *est*, para que Climatol en R los pueda leer, utilizando el explorador que

al seleccionar el archivo, con el botón derecho se les da editar y se les agrega la terminación est, "Tmax_1969-2021.est", "Tmin_1969-2021.est" y "PP_1969-2021.est".

2. **Aplicación del script Climatol de R para completar los datos faltantes.** Para utilizar el script, se apoyó de una guía del usuario de Climatol de R proporcionada por el citado Miguel Servín, que utiliza datos de varias estaciones para estimar datos faltantes en función de datos aledaños alrededor de una estación, y contiene funciones para el control de calidad, homogeneización y relleno de datos faltantes en un conjunto de series de cualquier variable climática (Guijarro, J. 2018).

Se adaptó para el presente trabajo el script de Climatol proporcionado por Miguel Servín, estudiante de posgrado del Dr. Irineo López, obteniendo tres, uno para cada variable climatológica, los cuáles se ejemplifican con el caso del utilizado para la variable temperatura máxima (Tmax).

```
# Script de Climatol para temperatura máxima Tesis A LLerena

setwd("~/_____2022/____DOCTORADO/___EXAMEN DE
GRADO/_TESIS/_____2022/CLIMATOL")
library(climatol)

### Análisis exploratorio de los datos
homogen('Tmax', 1969, 2021, dz.max=9, snht1=60, snht2=70)

# Calcular totales mensuales en lugar de valores medios)
dd2m('Tmax', 1969, 2021) #Tmax, Tmin, Eva

#Homogeneización de las series mensuales
homogen('Tmax-m', 1969, 2021)
# Ajuste de las series diarias con los puntos de corte mensuales
homogen('Tmax', 1969, 2021, dz.max=7, metad=TRUE)

# Obtención de productos con los datos homogeneizados
load('Tmax_1969-2021.rda')

# Calcular totales mensuales en lugar de valores medios)
dd2m('Tmax', 1969, 2021, homog=TRUE)
# Para precipitaciones, añadir el parámetro valm=1 para

# calcular totales mensuales en lugar de valores medios
```

```
# Para precipitacion dd2m('pp', 1969, 2021, valm=1)

# Series homogeneizadas y resúmenes estadísticos
dahstat('Tmax', 1969, 2021, stat='series')

# Para obtener los resúmenes estadísticos si se quieren
dahstat('Tmax',1969,2021) #medias de las series diarias
dahstat('Tmax',1969,2021, mh=TRUE) #medias de los valores mensuales
dahstat('Tmax',1969,2021, mh=TRUE,stat='tnd') #tendencias y p-valores
dahstat('Tmax',1969,2021, stat='q',prob=.2) #primer quintil (diarios)
```

El tiempo que utilizó R para cada caso fue de más de 30 minutos por el gran volumen de datos procesados.

Los resultados o salida de archivos con datos completados obtenidos en R con Climatol de cada una de las variables climáticas, son archivos individuales con terminación *_series.csv*, siendo los 3 archivos obtenidos los siguientes:

- Tmax_1969-2021_series.csv
- Tmin_1969-2021_series.csv
- pp_1969-2021_series.csv

Si se requiere más información relacionada con la interpretación de la homogenización de los datos, esta se puede consultar en:

<https://cran.r-project.org/web/packages/climatol/climatol.pdf>

Con la información obtenida de los datos climáticos diarios ya completados, primeramente se procedió a seleccionar únicamente los datos de los años de 1992 a 2021, para después separar los datos de los meses de noviembre a abril para obtener los modelos, como se ampliará en Resultados y discusión.

Anexo 2. Script general en r para aplicar las funciones ya establecidas de “GAM” para generar los modelos y predecir los datos

Se utilizó el siguiente script desarrollado por el Dr. Humberto Vaquera para el caso de los datos climáticos anuales

```
# Obtención de modelos GAM con datos climáticos promedio anuales del ciclo de
# cultivo

library(readr)
setwd("~/_____2022/____DOCTORADO/___EXAMEN DE
GRADO/_TESIS/ _____2022/GAM R")

# lectura de datos
DATOS<-read_csv("_____2022/____DOCTORADO/
__6oCUATRIMESTRE/
EST526/ TAREAS/TAREA FINAL/TESIS/_____2022/GAM R/DATOS.csv")
library(mgcv)
head(DATOS)

modx=gam(Rend~s(variable x), data=DATOS)
summary(mod) # Para desplegar el ANOVA del modelo y verificar R-sq.(adj)
AIC(modx) # Para desplegar el AIC del modelo, entre mas pequeño mejor ajuste
plot(mod) # Para graficar el modelo

plot(mod_ano) # Para graficar el modelo

predicciones=data.frame(predict(modx)) # Para que calcule los valores de las
# predicciones de rendimiento del maíz en los años estudiados
write.csv(predicciones,"mod.csv") # Para que exporte los valores

# Para calcular las predicciones de años siguientes
a_2022_2025=data.frame(Ano=c(2022, 2023, 2024, 2025))
predic_2022_2025=predict(mod_ano,a_2022_2025)
predic_2022_2025
write.csv(predic_2022_2025,"predic_2022_2025.csv") #para exportar los valores
```

Anexo 3. Procedimiento detallado para la obtención de modelos ARX lineales para predicción del rendimiento probando diferentes escenarios con la herramienta System identification de MATLAB

System identification de MATLAB proporciona técnicas de identificación tales como máxima verosimilitud, minimización de errores de predicción e identificación de sistemas subespaciales. También soporta el modelado de datos de series temporales y la predicción de series temporales.

Un modelo dinámico en tiempo discreto, auto regresivo (ARX) con variables exógenas para un sistema con una variable de entrada y una variable de salida, está dado por una ecuación en diferencias lineal (Ljung, 1999):

$$y(t) + a_1y(t - 1) + \dots + a_{na}y(t - na) = b_1u(t - nk) + \dots + b_{nb}u(t - nk - nb + 1) + e(t) \quad (15)$$

Donde:

- $y(t)$ es la variable de salida del modelo para los tiempos $t=t, t-1, \dots, t-na$;
- $u(t)$ es la variable de entrada del modelo para los instantes de tiempo $t=t-nk, t-nk-1, \dots, t-nk-nb+1$;
- na es el número de pasos en el tiempo para la salida en el pasado;
- nb es el número de pasos en el tiempo para la entrada en el pasado;
- nk es el retraso existente entre la entrada y la salida;
- $e(t)$ es el ruido blanco asociado con la variable de salida.

La estructura de un modelo ARX lineal, se determina estimando los coeficientes a_1, \dots, a_{na} y b_1, \dots, b_{nb} , así como el orden del modelo dado por los valores de los parámetros na, nb y nk , a partir de las mediciones (datos) de la variable de entrada (independiente) y variable de salida (dependiente). Es claro que un sistema puede tener varias variables independientes y una o muchas variables dependientes.

Esta herramienta permitió probar cientos de escenarios que se pueden derivar, tanto externamente con las combinaciones que se derivan de los datos de entrada a introducir en el modelo, como internamente en la herramienta de systemIdentification de MATLAB con las particiones que se pueden dar para generar y validar el modelo a obtener, como los años anteriores a tomar en cuenta en el modelo.

Para el caso de la predicción de la producción de maíz, la variable de salida es el rendimiento y las variables climáticas de entrada fueron la temperatura máxima, la temperatura mínima y la precipitación.

Los pasos seguidos para la obtención de modelos ARX lineales para predicción del rendimiento probando diferentes escenarios. fueron los siguientes.

1. **Obtención de la variable de salida y las de entrada.** La variable dependiente fue el rendimiento de maíz y las variables independientes probadas, fueron los datos diarios, mensuales y anuales del ciclo de cultivo (noviembre a abril), de temperatura máxima (Tmax), temperatura mínima (Tmin) y precipitación (pp), que se promediaron de tres estaciones climatológicas de la zona de estudio. Las variables de entrada que se utilizaron que fueron los datos climáticos, propiciaron 21 escenarios que se derivaron de ellas, de acuerdo a las 3 periodicidades que se utilizaron (datos diarios, mensuales y anuales) y a las 7 combinaciones que se pueden formar al introducirlas al modelo.
2. **Selección de la estructura del modelo ARX lineal.** Se probaron solamente modelos auto regresivos con variables exógenas (ARX) lineales.
3. **Elección del mejor modelo ARX o mejor ajuste.** Se usó el criterio del error de predicción final (FPE) de Akaike (Ljung, 1999), definido como:

$$FPE = \left(\frac{1+d/N}{1-d/N} \right) V \quad (16)$$

Donde:

- V es la función que minimiza el cuadrado de las diferencias entre predicciones y mediciones;

- d es el número de parámetros estimados;
- N es el número de mediciones.

4. **Selección de los subconjuntos de datos para estimar y validar los modelos.**

En este caso hay dos conceptos internos dentro de systemIdentification que se deben definir antes de estimar el modelo de predicción, las particiones o rangos que se pueden asignar para generar y validar el modelo a obtener y los años anteriores que se deben considerar en el modelo de las variables de entrada y de salida, por lo que este proceso se realizó en dos etapas, primero se seleccionó para cada caso la partición con mayor ajuste para cada una de las siete combinaciones de los datos climáticos, para posteriormente continuar con la definición de cuantos años anteriores se deben incluir en el modelo.

- **Combinaciones internas de systemIdentification de MATLAB. Particiones probadas para generar y validar el modelo a obtener.** Se inició probando para cada una de las siete combinaciones de los datos climáticos, las particiones 50:50%, 55:45%, 60:40%, 65:35%, 70:30%, 75:25%, 80:20%, 85:15%, 90:10% y 95:5%, aunque en algunos casos se afinaron aún más.
- **Consideraciones de años anteriores a tomar en cuenta.** Los años que tiene systemIdentification predeterminados son 4, que fueron los que se utilizaron en la etapa anterior, para ya definida la partición del modelo que presenta mayor ajuste, probar varias alternativas.

5. **Evaluación de los modelos.** Se evaluaron hasta un total de 10,000 modelos ARX, ya que valor máximo para na (una variable de salida) y para nb (tres variables de entrada), fue de 10 y $nk = 1$, por lo que el número de combinaciones (número de modelos evaluados) resultante es de 10^4 .

6. **Nivel de ajuste.** Se utilizó el nivel de ajuste (%) entre los rendimientos estimados y medidos para seleccionar los modelos utilizados en el presente trabajo.

7. **Metodología.** El procedimiento seguido fue el siguiente:

- a. Se utilizó el siguiente script desarrollado por el Dr. Irineo López para el caso de los datos climáticos diarios:

```
% CODIGO GENERAL arxmodelrendmaizdiario2
% SE UTILIZAN DATOS CLIMATICOS DIARIOS
% DE 29 AÑOS DEL 1 DE NOVIEMBRE DE 1992 AL 30 DE ABRIL DE 2021
% PARA UN TOTAL DE 5256 DATOS DE CADA VARIABLE CLIMATICA
% LOS DATOS DE RENDIMIENTO SON DE 1993 A 2021

clear
close all
Entradas= xlsread('1993-2021-diario-aj.xlsx',2,'B2:D5257');
Salida=xlsread('1993-2021-diario-aj.xlsx',1,'B3:B31');
[renglones_entradas,columnas_entradas]=size(Entradas);
[renglones_salida,columnas_salida]=size(Salida);
Tmax=Entradas(1:renglones_entradas,1);
Tmin=Entradas(1:renglones_entradas,2);
pp=Entradas(1:renglones_entradas,3);
% tiempos de medicion de la variable rendimiento
t1=[1:1:renglones_salida];
% Tiempo de medición de las variables de entrada
t=[1:1/188:29-9/188];
% Se interpolan los valores de rendimiento para que se tengan la misma
% cantidad de datos para las variables de entrada y la variable de salida
Rendimiento=interp1(t1,Salida,t);

figure(1);
subplot(4,1,1);
plot(t,Tmax)
ylabel('T_{max}')
subplot(4,1,2);
plot(t,Tmin)
ylabel('T_{min}')
subplot(4,1,3);
plot(t,pp)
ylabel('pp')

subplot(4,1,4);
plot(t,Rendimiento)
ylabel('Rendimiento')

systemIdentification
```


- b. Al finalizar de correr el código aparecen dos ventanas, una con la gráfica del comportamiento de las variables de entrada y de salida (Figura 13) y la ventana principal de System identification en la que se hacen los diferentes procesos (Figura 14):

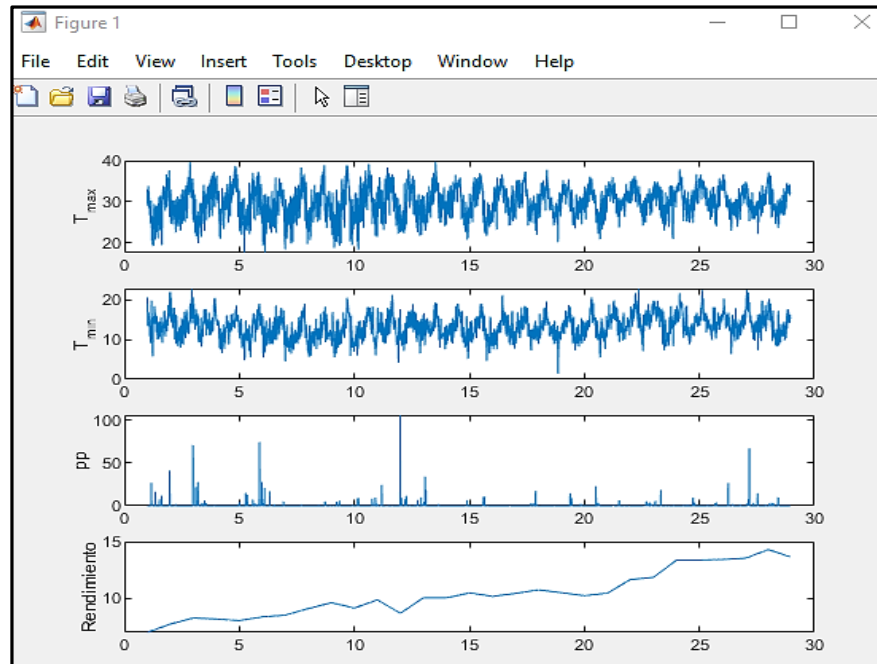


Figura 11. Gráficas del comportamiento en el tiempo de las variables de entrada y de salida.

Fuente: Generado por systemIdentification de MATLAB.

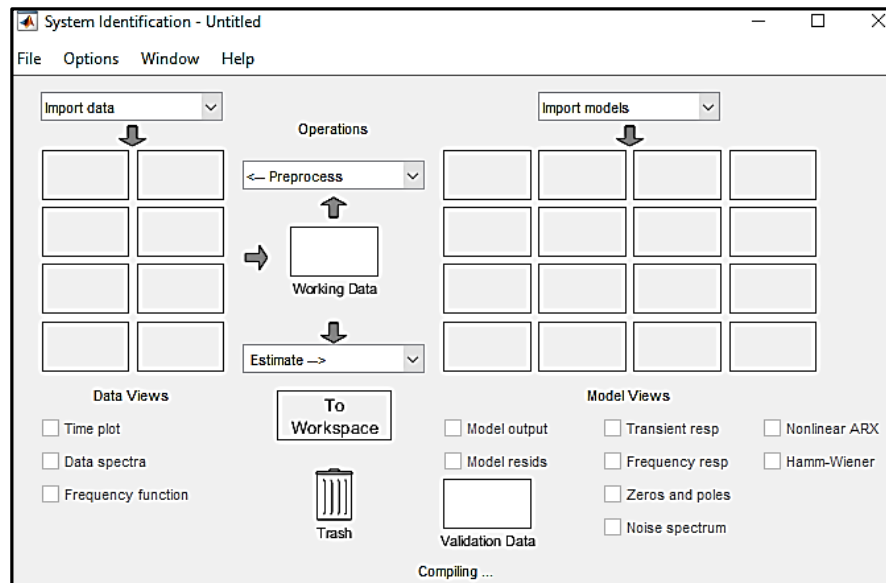


Figura 12. Ventana principal de systemIdentification de MATLAB.

Fuente: Generado por systemIdentification de MATLAB.

- c. En esta ventana hay que abrir el cuadro de diálogo que dice “Import data” y seleccionar la pestaña “Time domain data...” (Figura 15).

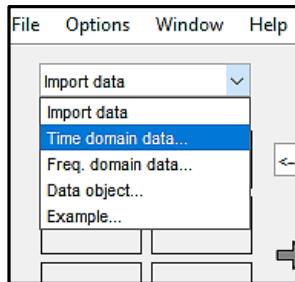


Figura 13. Ventana de Import data

Fuente: Generado por systemIdentification de MATLAB.

- d. Esto abre otra ventana de “Import Data”, en la cual se deben indicar las variables de entrada y de salida que participan en el modelo, así como el tiempo de inicio y el tiempo de muestra o ciclo. Así, en la parte de “Workspace Variable”, se colocan las combinaciones de las variables de entrada (Tmax, Tmin y pp) y en “Output” la variable de salida (Rendimiento). En la parte de “Data Information” en “Data name” se coloca el nombre del archivo de salida (“mydata1”); en “Start Time” se escribe el inicio del ciclo que generalmente es 0; en “Sample time” se escribe el dato del ciclo de cultivo que para el caso de datos diarios es 1/181, para los datos mensuales es 1/29 y para los datos anuales es 1/1. Para finalizar se selecciona “Import” (Figura 16).

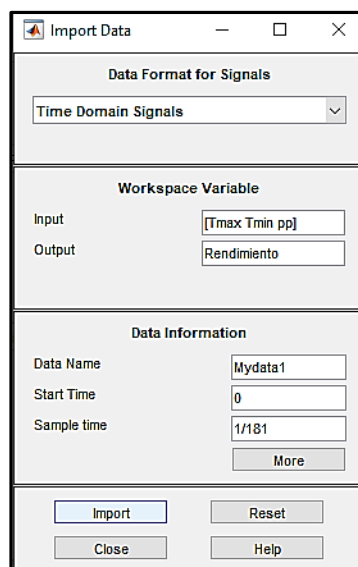


Figura 14. Ventana de “Import Data”.

Fuente: Generado por systemIdentification de MATLAB.

- e. Con el procedimiento anterior, aparece una figura en el cuadro de diálogo “Data Views” de la ventana principal de systemIdentification con el nombre asignado más 1 (Mydata1). Se selecciona y abre la pestaña o cuadro de diálogo “Operations” y se selecciona “Select range...” (Figura 17).

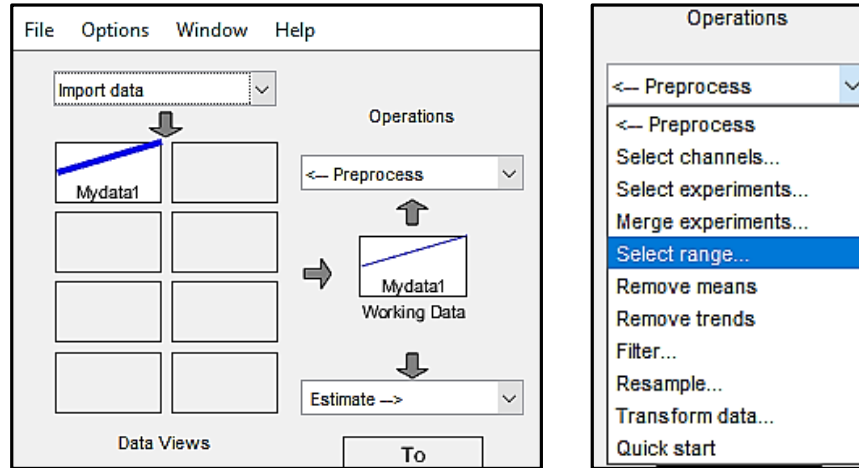


Figura 15. Ventana principal con figura Mydata1 y cuadro de diálogo “Operations”.

Fuente: Generado por systemIdentification de MATLAB.

- f. Se abre una nueva ventana de “Select Range”, en donde en dos pasos se deben colocar los dos valores de la partición en “Samples”; en un primer paso estimar los modelos y se selecciona “Insert”, y en un segundo paso se ponen los datos para validarlos y se vuelve a seleccionar “Insert”, probando y evaluando diferentes particiones en el proceso (Figura 18).

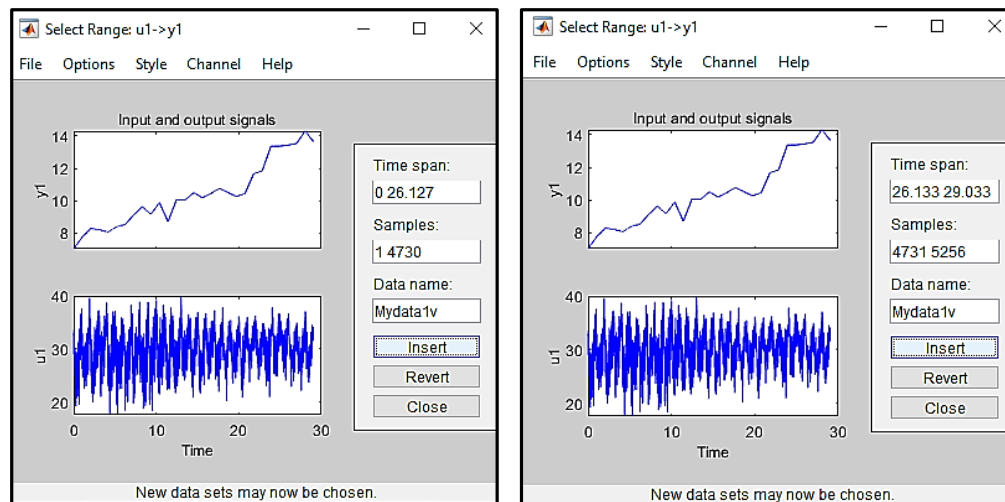


Figura 16. Cuadros de diálogo de “Select Range”.

Fuente: Generado por systemIdentification de MATLAB.

- g. Cada vez que se seleccionó “Insert”, originó que en la ventana principal apareciera una nueva figura en el cuadro de “Data Views”, “Mdata1e” de color verde y “Mydata1v” de color rojo (Figura 19).

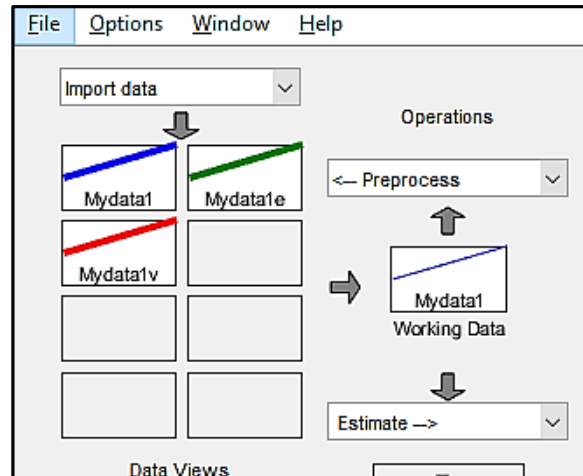


Figura 17. Ventana principal con nuevas figuras “Mdata1e” y “Mydata1v”.
Fuente: Generado por systemIdentification de MATLAB.

- h. Se procede a arrastrar con el ratón la figura del cuadro “Mydata1e” al cuadro de “Working Data” y la figura del cuadro “Mydata1v” al cuadro de “Validation Data” y aparece otro cuadro con una gráfica en donde se especifican con color verde las particiones para la estimación del modelo y con color azul para la validación del modelo (Figura 20).

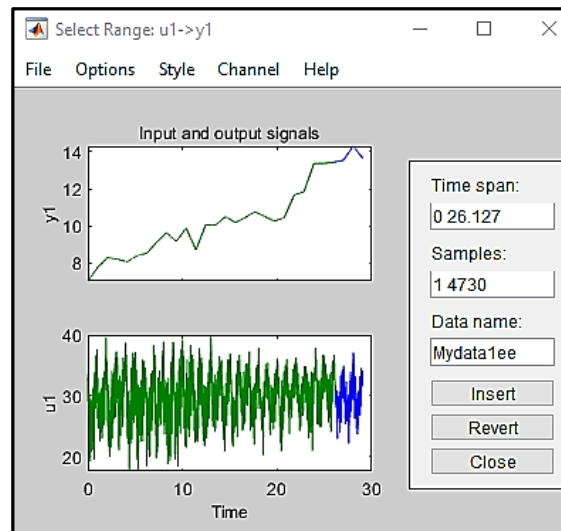


Figura 18. Gráfica en donde se señalan los datos o rangos de datos de entrada utilizados para la estimación y validación del modelo.

Fuente: Generado por systemIdentification de MATLAB.

- i. Se inicia el proceso de elaboración del modelo seleccionando en la ventana principal la pestaña de “Estimate” y dentro de ella seleccionar “Polynomial Models”, lo que abre otra ventana (Figura 21).

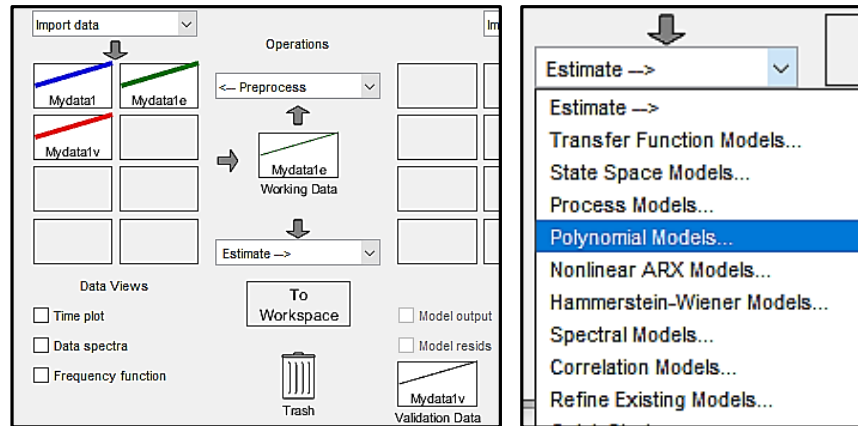


Figura 19. Ventana principal en donde se selecciona “Polynomial Models”.
Fuente: Generado por systemIdentification de MATLAB.

- j. Lo anterior abre la ventana de “Polynomial Models”, en donde en la pestaña “Orders” se seleccionan los años hacia atrás de información que utilizará el modelo para hacer las predicciones, que por default para el caso de tres variables son 4 años representados por “Orders” [4 4 4] [1 1 1]]. Para el presente trabajo se probaron diferentes alternativas con las distintas combinaciones de variables de entrada (Figura 22).

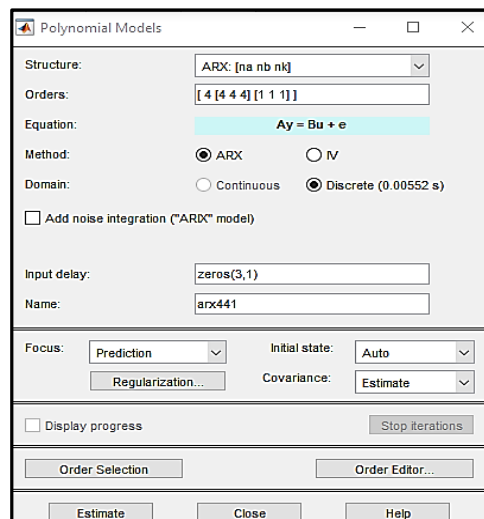


Figura 20. Ventana de “Polynomial Models”.
Fuente: Generado por systemIdentification de MATLAB.

- k. Ya asignado en “Orders” lo anterior, se abre la pestaña “Estimate”, lo que genera en la ventana principal de “System Identification User”, un primer modelo arx441 que aparece en el primer cuadro de “Models View” (Figura 23).

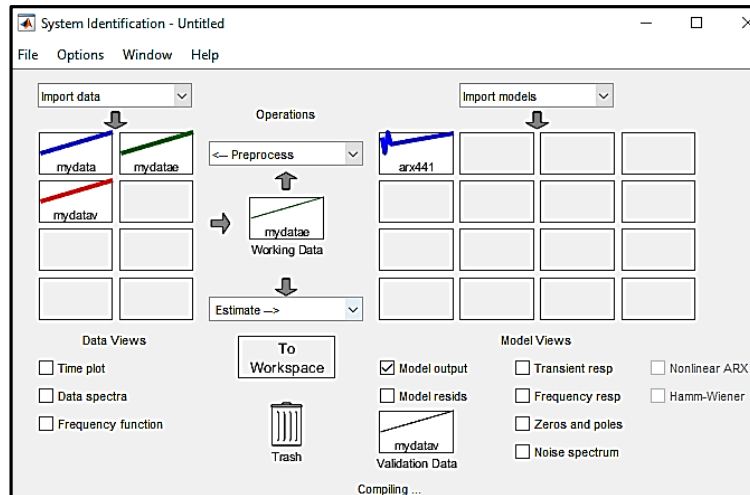


Figura 21. Ventana principal con el modelo generado en “Model Views”.
Fuente: Generado por systemIdentification de MATLAB.

- l. Para revisar los resultados, se selecciona la pestaña “Model output”, abriéndose una nueva ventana en donde se indica que porcentaje se explica con el modelo elaborado, que para este caso fue de 71.8% (Figura 24).

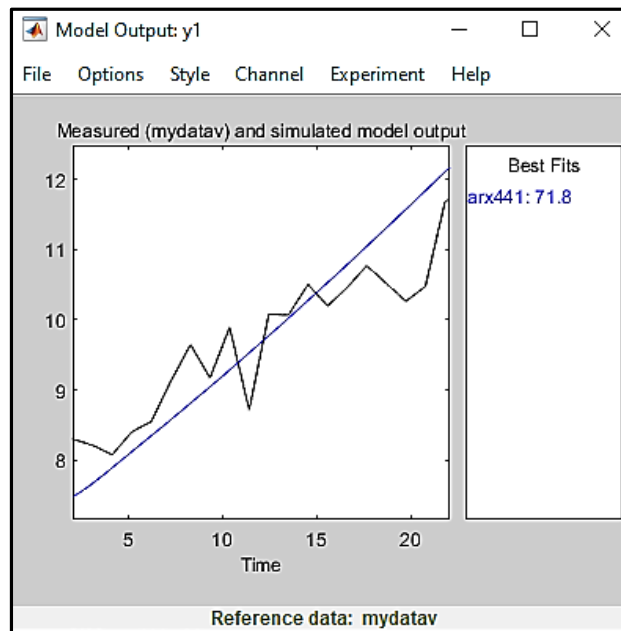


Figura 22. Gráfica en donde se indica el porcentaje que explica el modelo generado

Fuente: Generado por systemIdentification de MATLAB.

- m. Para visualizar el modelo matemático desarrollado, se pica en el modelo arx441 que apareció y se muestra una nueva ventana con información en donde se puede observar la ecuación matemática de dicho modelo (Figura 25).

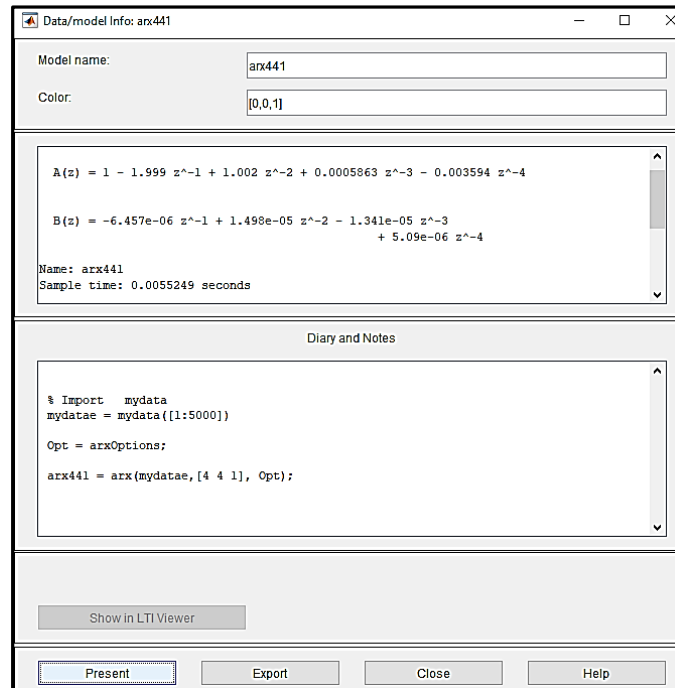


Figura 23. Ventana que muestra la ecuación matemática del modelo matemático generado.

Fuente: Generado por systemIdentification de MATLAB.

- n. Para importar el modelo matemático al ambiente de MATLAB, se selecciona “Present” y aparece en la ventana de “Command Windows” (Figura 26).

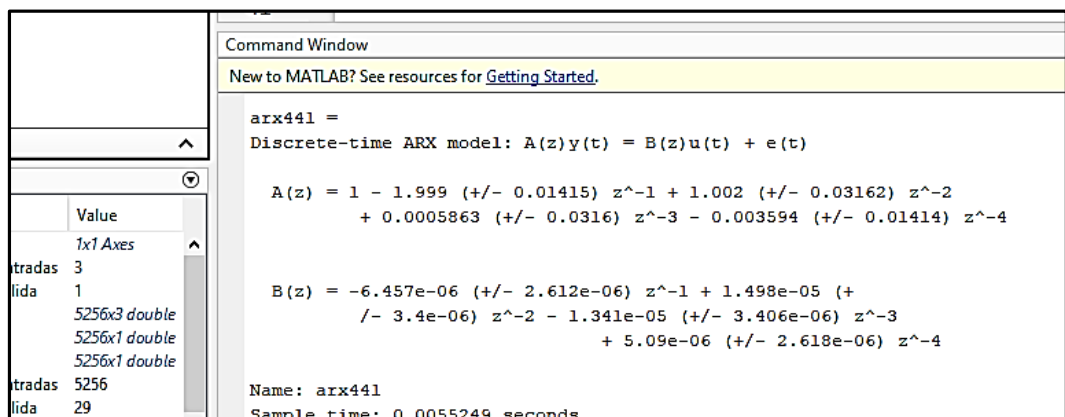


Figura 24. Modelo matemático generado importado al ambiente de MATLAB en la ventana de “Command Windows”.

Fuente: Generado por systemIdentification de MATLAB.

Anexo 4. Pseudo-código elaborado en MATLAB por el Dr. Irineo López para verificar el algoritmo FKS

Procedimiento FKS

- Leer los valores predichos por los modelos
- Inicializar las variables
- Realizar los cálculos para el primer y segundo instantes de tiempo ($J = 1,2$)
- Para $j=3$ hasta el último instante de tiempo
- Cálculo de medias recursivas de los dos modelos (Ecuación 10)
- Cálculo de varianzas recursivas de los dos modelos (Ecuación 9)
- Cálculo de la ganancia de Kalman (Ecuación 4)
- Cálculo del valor del segundo modelo mejorado por el FKS (Ecuación 2)
- Termina iteración
- Termina procedimiento FKS

% programa del algoritmo de Filtro de Kalman simple desarrollado por el Dr.

% Oscar Palacios Velez

clear

close all

% primer columna año, segunda modelo1, tercera modelo 2

datos=[1998 8.70 8.40

1999 8.62 8.55

2000 9.32 9.12

2001 8.96 9.64

2002 9.10 9.17

2003 9.33 9.89

2004 9.34 8.71

2005 9.81 10.08

2006 9.96 10.06

2007 10.73 10.50

2008 10.19 10.19

2009 10.64 10.46

2010 10.58 10.77

2011 10.57 10.52

2012 10.87 10.26

2013 10.76 10.47

2014 10.95 11.68

2015 10.61 11.86

2016 10.94 13.37

2017 11.25 13.37


```

2018 12.45 13.44
2019 13.18 13.54
2020 13.77 14.30
2021 13.45 13.62];
t=datos(:,1);
mp1=datos(:,2);
mp2=datos(:,3);

avgm1=zeros(size(t,1),1);
avgm2=zeros(size(t,1),1);
varm1=zeros(size(t,1),1);
varm2=zeros(size(t,1),1);
K=zeros(size(t,1),1);
mp1m=zeros(size(t,1),1);
% La media es inicialmente al primer valor predicho por los modelos
avgm1(1)=mp1(1);
avgm2(1)=mp2(1);
% La primer varianza debe ser cero al aplicar la definición
varm1(1)=0;
varm2(1)=0;
K(1)=1; % Este valor no se usa
mp1m(1)=mp1(1);
for k=2:size(t,1)
% Cálculo de las medias en forma recursiva
avgm1(k)=(k-1)/k*avgm1(k-1)+1/k*mp1(k);
avgm2(k)=(k-1)/k*avgm2(k-1)+1/k*mp2(k);
% Cálculo de las varianzas en forma recursiva
varm1(k)=varm1(k-1)+k/(k-1)*(avgm1(k)-mp1(k))^2;
varm2(k)=varm2(k-1)+k/(k-1)*(avgm2(k)-mp2(k))^2;
% Cálculo de la ganancia de Kalman
K(k)=varm1(k)/(varm1(k)+varm2(k));
% Uso de la ecuación  $Y^*=Y1+K(Y2-Y1)$ 
mp1m(k)=mp1(k)+K(k)*(mp2(k)-mp1(k));
end

figure(1)
plot(t,mp1,t,mp2,'-',t,mp1m,'k','LineWidth',2.0)
xlabel('Tiempo [años]');
ylabel('Rendimiento maiz (ton ha^{-1})')
legend('Datos medidos de rendimiento (primer modelo)', ...
'Predicciones modelo ARK (segundo modelo)', 'Predicciones FKS')
figure(2)
plot(t,K)
title('Ganancia de Kalman')
xlabel('tiempo [años]');

TotAvgm1=mean(mp1)

```

```

TotAvgm2=mean(mp2)
TotAvgFK=mean(mp1m)
Totvarm1=var(mp1)
Totvarm2=var(mp2)
TotvarFK=var(mp1m)

% programa del algoritmo de Filtro de Kalman simple desarrollado por el Dr.
% Oscar Palacios Velez
clear
close all
% primer columna año, segunda modelo1, tercera modelo 2
datos=[1998 8.70 8.40
1999 8.62 8.55
2000 9.32 9.12
2001 8.96 9.64
2002 9.10 9.17
2003 9.33 9.89
2004 9.34 8.71
2005 9.81 10.08
2006 9.96 10.06
2007 10.73 10.50
2008 10.19 10.19
2009 10.64 10.46
2010 10.58 10.77
2011 10.57 10.52
2012 10.87 10.26
2013 10.76 10.47
2014 10.95 11.68
2015 10.61 11.86
2016 10.94 13.37
2017 11.25 13.37
2018 12.45 13.44
2019 13.18 13.54
2020 13.77 14.30
2021 13.45 13.62];
t=datos(:,1);
mp1=datos(:,2);
mp2=datos(:,3);

avgm1=zeros(size(t,1),1);
avgm2=zeros(size(t,1),1);
varm1=zeros(size(t,1),1);
varm2=zeros(size(t,1),1);
K=zeros(size(t,1),1);
mp1m=zeros(size(t,1),1);
% La media es inicialmente al primer valor predicho por los modelos
avgm1(1)=mp1(1);

```

```

avgm2(1)=mp2(1);
% La primer varianza debe ser cero al aplicar la definición
varm1(1)=0;
varm2(1)=0;
K(1)=1; % Este valor no se usa
mp1m(1)=mp1(1);
for k=2:size(t,1)
% Cálculo de la medias en forma recursiva
avgm1(k)=(k-1)/k*avgm1(k-1)+1/k*mp1(k);
avgm2(k)=(k-1)/k*avgm2(k-1)+1/k*mp2(k);
% Cálculo de las varianzas en forma recursiva
varm1(k)=varm1(k-1)+k/(k-1)*(avgm1(k)-mp1(k))^2;
varm2(k)=varm2(k-1)+k/(k-1)*(avgm2(k)-mp2(k))^2;
% Cálculo de la ganancia de Kalman
K(k)=varm1(k)/(varm1(k)+varm2(k));
% Uso de la ecuación  $Y^*=Y1+K(Y2-Y1)$ 
mp1m(k)=mp1(k)+K(k)*(mp2(k)-mp1(k));
end

figure(1)
plot(t,mp1,t,mp2,'-',t,mp1m,'k:','LineWidth',2.0)
xlabel('Tiempo [años]');
ylabel('Rendimiento maiz (ton ha^{-1})')
legend('Datos medidos de rendimiento ', ...
'Predicciones modelo ARK ', 'Predicciones FKS')
figure(2)
plot(t,K)
title('Ganancia de Kalman')
xlabel('tiempo [años]');

TotAvgm1=mean(mp1)
TotAvgm2=mean(mp2)
TotAvgFK=mean(mp1m)
Totvarm1=var(mp1)
Totvarm2=var(mp2)
TotvarFK=var(mp1m)

```

Anexo 5. Procedimiento detallado para la obtención de modelos ARX lineales para predicción del rendimiento probando diferentes escenarios con System identification de MATLAB

1. PRIMERA etapa utilizando 21 escenarios de datos climáticos y 10 particiones de años anteriores a tomar en cuenta en el modelo que se genere.

Para la generación de los modelos, se trabajó con los 21 escenarios externos de datos climáticos que resultan de utilizar 3 periodicidades de datos climáticos y los 7 contextos que se obtienen al combinar las 3 variables climáticas.

Los datos diarios son 5,256 para cada factor climático de 29 años y de 6 meses del ciclo de cultivo de maíz, noviembre a abril, de cada año (Figura 27).

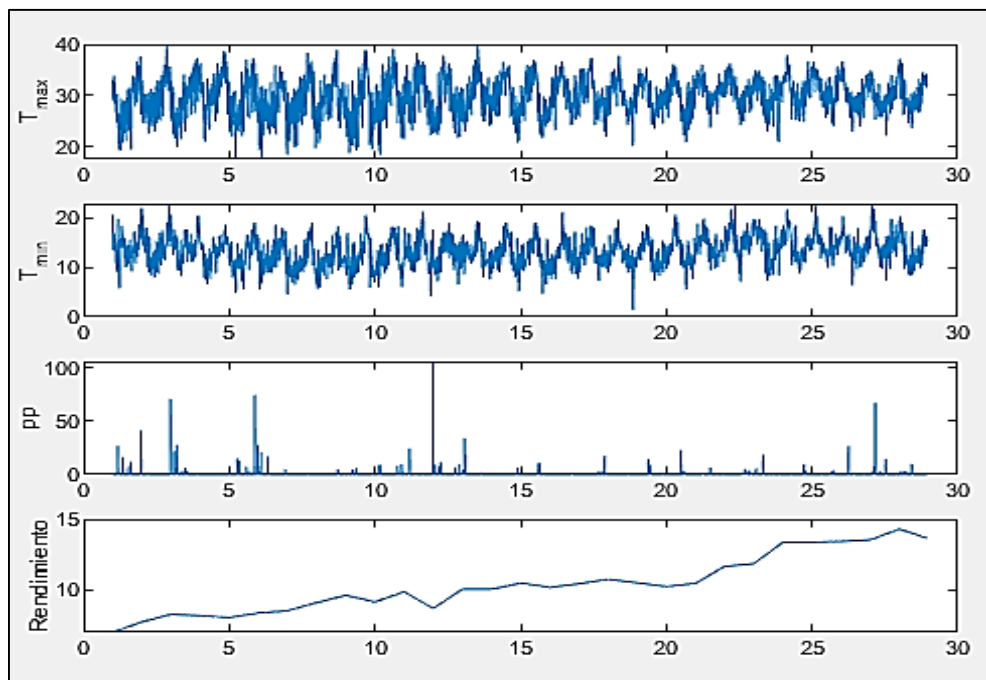


Figura 25. Gráfica de la fluctuación diaria de las variables de entrada y salida en el ciclo de cultivo y en los 29 años considerados.

Fuente: Generado por systemIdentification de MATLAB.

Los datos mensuales promedio son 181 para cada factor climático de 6 meses del ciclo de cultivo de maíz, noviembre a abril y de 29 años (Figura 28).

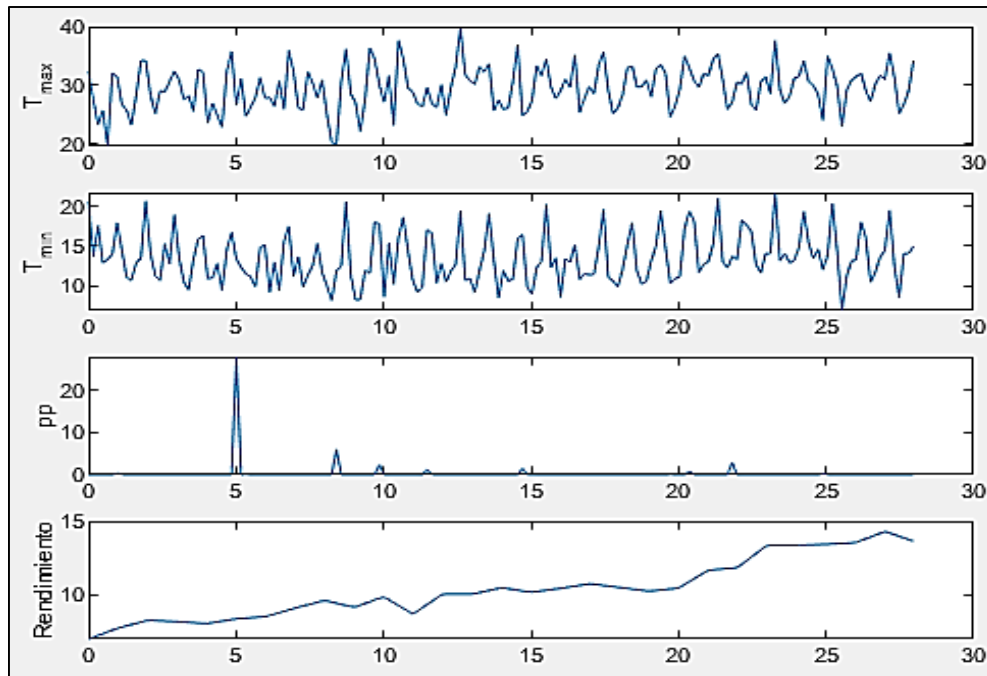


Figura 26. Gráfica con los datos climáticos mensuales promedio de los 29 años considerados.

Fuente: Generado por systemIdentification de MATLAB.

Los datos anuales promedio son de 29 años para cada factor climático (Figura 29).

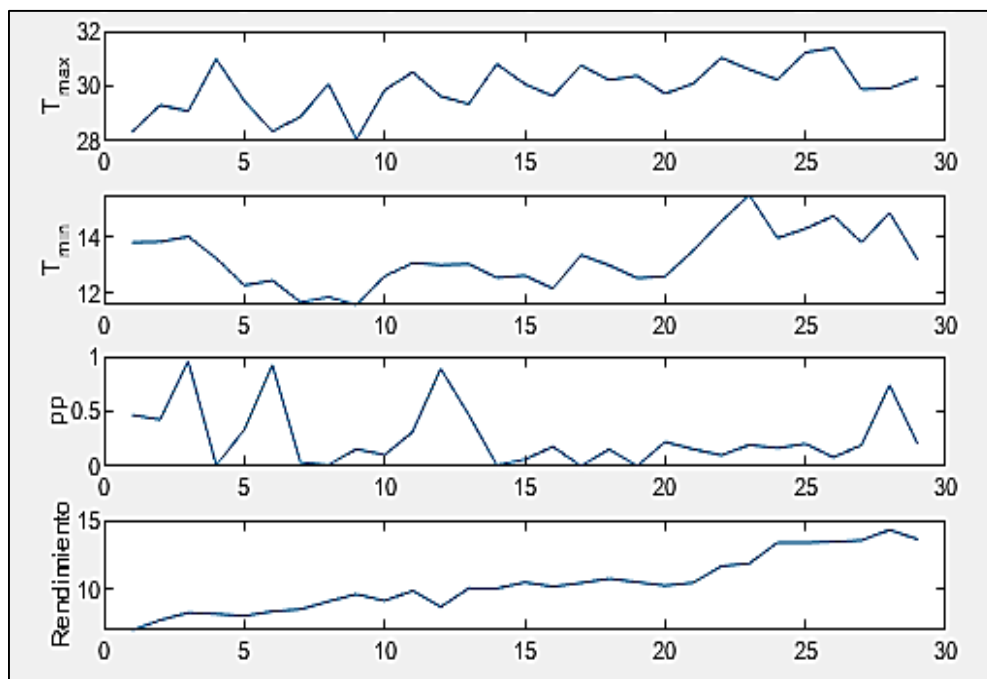


Figura 27. Gráfica con los datos climáticos anuales promedio de los 29 años considerados.

Fuente: Generado por systemIdentification de MATLAB.

Las 7 combinaciones de las variables climáticas fueron las siguientes.

- ✓ Temperatura máxima, Temperatura mínima y Precipitación
- ✓ Temperatura máxima y Temperatura mínima.
- ✓ Temperatura máxima y Precipitación.
- ✓ Temperatura mínima y Precipitación
- ✓ Temperatura máxima.
- ✓ Temperatura mínima.
- ✓ Precipitación.

Con las 21 combinaciones de datos climáticos externos de MATLAB, se continuó con la definición interna de los subconjuntos o particiones de datos para estimar y validar los modelos ARX que se probarían para cada una, que fueron las 10 siguientes: 50:50%, 55:45%, 60:40%, 65:35%, 70:30%, 75:25%, 80:20%, 85:15%, 90:10% y 95:5% y en algunos casos se afinaron aún más.

Con estas 210 combinaciones, se inició la generación de los modelos utilizando los datos de 4 años anteriores que tiene *systemIdentification* predeterminados a tomar en cuenta en el modelo que se genere, para ya definida la partición del modelo que presenta mayor ajuste, probar varias alternativas de años anteriores para buscar mejorar el ajuste

Los resultados obtenidos se las particiones que presentaron los mejores ajustes para cada una de las tres periodicidades de los datos climáticos en cada una de las dos etapas en las que se llevó a cabo la generación de los modelos de predicción de rendimiento de maíz, fueron los siguientes.

- a. **Resultados de la primera etapa con datos climáticos diarios.** En el Cuadro 15 se señalan con amarillo los resultados obtenidos de los modelos que presentaron mejores ajustes en esta primera etapa para el caso de datos climáticos diarios, las siete combinaciones y con ajuste de 4 años anteriores. Se presentan con amarillo fuertes los 3 mejores resultados de ajuste y con amarillo débil el resto.

Cuadro 15. Ajustes que tuvieron las siete diferentes combinaciones de datos climáticos diarios y con 4 años anteriores.

DATOS CLIMÁTICOS	PARTICIONES (porcentajes y días)									
	50:50%	55:45%	60:40%	65:35%	70:30%	75:25%	80:20%	85:15%	90:10%	95:5%
	2628	2891	3154	3416	3679	3942	4205	4468	4730	5000
Tmax	47.22	15.36	23.91	5.68	19.69	64.57	36.37	43.82	23.21	11.63
Tmin	59.67	43.7	45.67	35.57	51.71	66.84	44.77	44.72	30.9	14.01
pp	20.02	15.14	24.59	25.37	45.48	22.45	37.38	38.41	19.92	12.39
[Tmax Tmin]	62.78	4.61	14.31	-93.71	13.24	62.36	22.14	44.76	30.44	11.54
[Tmax pp]	-15.14	-28.54	-5.391	-14.21	-2.882	57.66	45.53	44.97	26.62	15.17
[Tmin pp]	59.96	48.08	50.37	43.55	60.57	44.88	0.197	20.95	8.72	6.65
[Tmax Tmin pp]	-104.4	-112	-91.71	-93.71	-79.97	-31.15	51.65	37.27	34.16	16.28

Fuente: Elaboración propia

Así, se utilizaron los 3 mejores modelos de los 7 datos climáticos que tuvieron los mayores porcentajes de ajuste en esta primera etapa, que son Tmax (64.57%), Tmin 66.84\$) y [Tmax Tmin] (62.36%), para buscar mayor ajuste con diferentes años anteriores a tomar en cuenta.

- b. Resultados de la primera etapa con datos climáticos mensuales.** En el Cuadro 16 se señalan con verde los resultados obtenidos de los modelos que presentaron mejores ajustes en esta primera etapa para el caso de datos climáticos diarios y con 4 años anteriores y con amarillo los 4 restantes de las 7 combinaciones de los datos climáticos.

Cuadro 16. Ajustes que tuvieron las diferentes combinaciones de datos climáticos mensuales y con 4 años anteriores.

DATOS CLIMÁTICOS	PARTICIONES (porcentajes y días)									
	50:50%	55:45%	60:40%	65:35%	70:30%	75:25%	80:20%	85:15%	90:10%	95:5%
	87	96	104	113	122	131	139	148	157	165
Tmax	-10.38	-24.77	-19.30	-34.01	-40.08	49.35	50.58	34.96	24.02	74.27
Tmin	38.81	24.46	28.28	18.95	24.72	54.69	51.96	38.05	28.30	76.16
pp	65.81	54.49	52.88	46.16	54.57	65.34	28.68	25.21	6.88	49.35
[Tmax Tmin]	-17.69	-33.56	-29.95	-47.60	-58.82	27.63	30.03	29.69	26.88	70.76
[Tmax pp]	-10.00	-25.03	-20.87	-177.6	-44.12	59.34	50.77	35.38	24.19	74.10
[Tmin pp]	41.77	26.62	29.84	68.78	25.03	42.59	55.98	38.09	28.31	71.20
[Tmax Tmin pp]	-17.87	-33.89	-31.76	-50.44	-65.99	20.24	25.70	30.81	27.32	70.46

Fuente: Elaboración propia

Se observa que los modelos obtenidos en general, presentan mejor ajuste que en el caso del uso de datos climáticos diarios. Así, se utilizaron los 3 mejores modelos de los 7 datos climáticos que tuvieron los mayores porcentajes de ajuste en esta primera etapa, que son Tmax (74.27%), Tmin (76.16%) y [Tmax pp] (74.10%), para buscar mayor ajuste con diferentes años anteriores a tomar en cuenta.

- c. **Resultados de la primera etapa con datos climáticos anuales.** En el Cuadro 17 se señalan con amarillo los resultados obtenidos de los modelos que presentaron mejores ajustes en esta primera etapa para el caso de datos climáticos anuales y con 4 años anteriores de los datos. Se presentan con amarillo fuertes los 3 mejores resultados de ajuste que fueron los que se probaron con varios años anteriores **de los datos** a tomar en cuenta y con amarillo débil el resto que ya no se probaron.

Cuadro 17. Ajustes que tuvieron las diferentes combinaciones de datos climáticos anuales y con 4 años anteriores de los datos y otros.

DATOS CLIMÁTICOS	PARTICIONES (porcentajes y días)						
	50:50%	55:45%	60:40%	65:35%	70:30%	75:25%	80:20%
	15	16	17	19	20	22	23
Tmax	34.91	46.90	22.42	35.37	30.37	31.49	15.11
Tmin	47.79	45.70	38.63	54.31	41.21	21.07	-23.28
pp	39.95	39.95	30.40	38.78	55.99	71.23	34.16
	[4 3 1]						
[Tmax Tmin]	33.89	36.91	17.54	25.99	-16.36	19.02	-48.88
		[3 3]					
[Tmax pp]	52.02	50.42	40.39	52.02	38.75	53.75	-40.33
	[3 3]						
[Tmin pp]	39.22	32.34	16.05	-12.88	19.46	42.59	-37.7
	[2 2 2]		[3 3 3]				
[Tmax Tmin pp]	-75.85	-74.94	-31.01	18.93	22.48	-43.37	-16.71

Fuente: Elaboración propia

Se aclara que MATLAB no permitió en varios de los casos la estimación de los modelos con 4 años anteriores de los datos, teniendo que utilizar 3 o 2 años anteriores, como se señala con anaranjado claro la línea superior correspondiente en el Cuadro 17.

Además, en los 7 casos cuando se usaron las particiones de 85:15%, 90:100% y 95:100%, dio valores de 100% de ajuste, lo que no es posible y no se tomaron en cuenta.

Así, se utilizaron los 3 mejores modelos de los 7 datos climáticos que tuvieron los mayores porcentajes de ajuste en esta primera etapa, que son Tmin (54.31%), pp (71.23%) y [Tmax pp] (53.75%), para buscar mayor ajuste con diferentes años anteriores a tomar en cuenta.

2. Segunda etapa o final del proceso mediante la elección del criterio de ajuste con diferentes años anteriores a tomar en cuenta en el modelo que se genere.

a. Para el caso de datos climáticos diarios. Se probaron diferentes combinaciones ARX [na nb nk] y algunos ejemplos de los ajustes obtenidos para las diferentes variables climáticas, se presentan en las Figuras 30, 31 y 32,

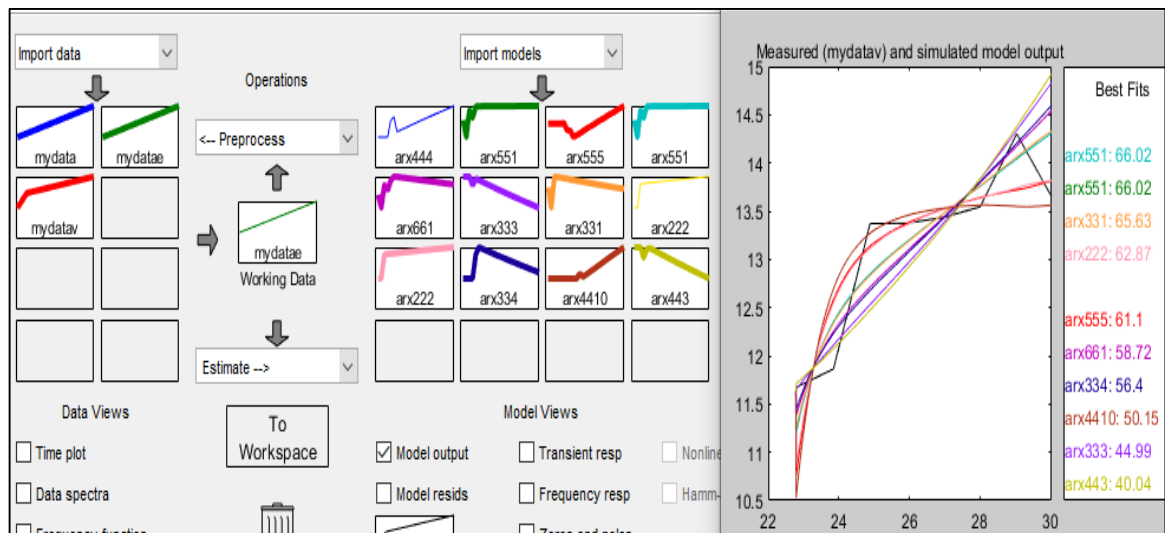


Figura 28. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax.

Fuente: Generado por systemIdentification de MATLAB.

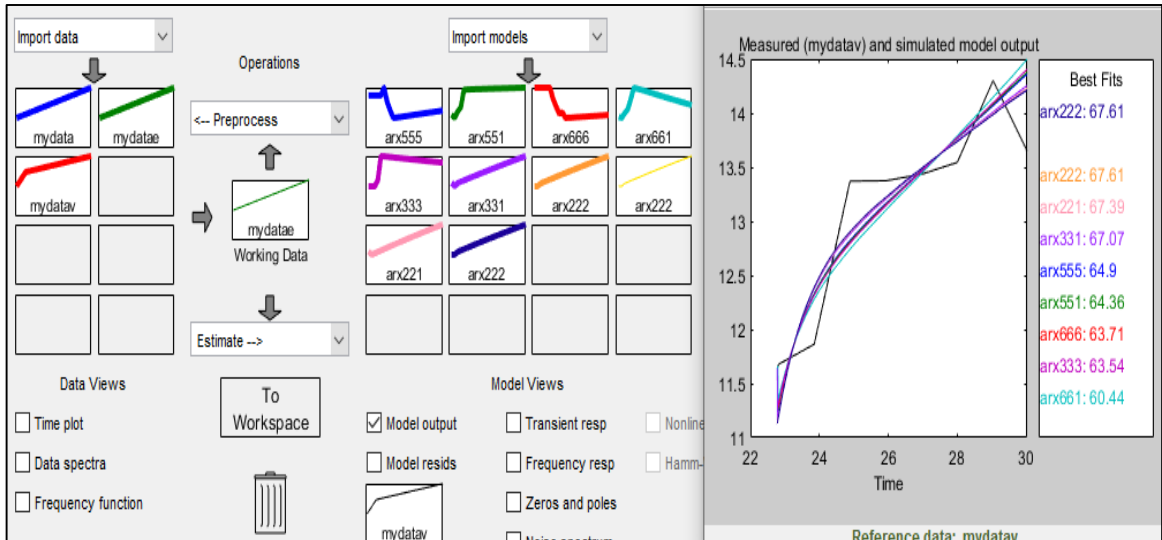


Figura 29. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmin.

Fuente: Generado por systemIdentification de MATLAB.

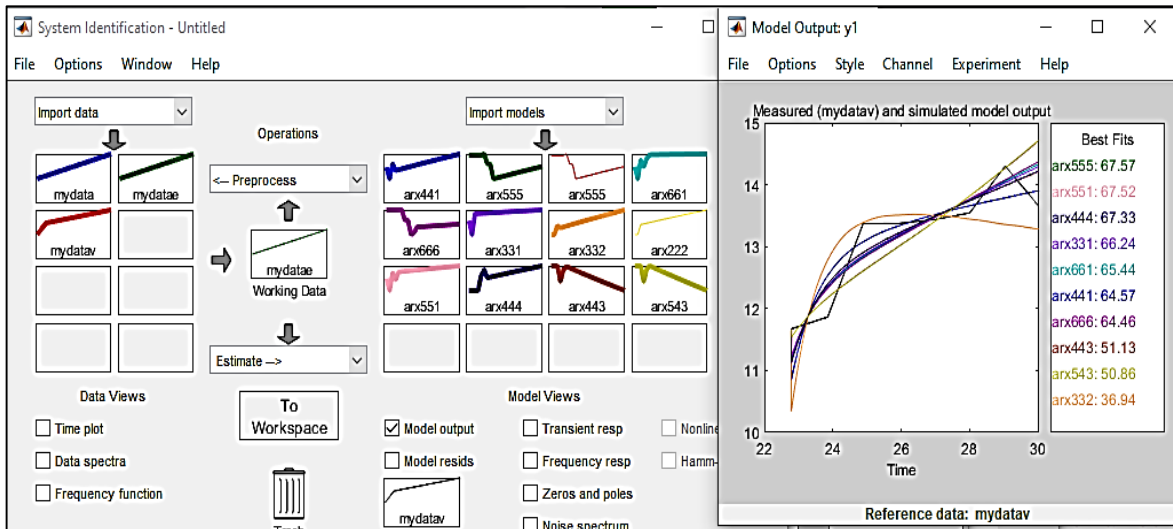


Figura 30. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para las variables Tmax y Tmin.

Fuente: Generado por systemIdentification de MATLAB.

- b. Para el caso de datos climáticos mensuales. Se probaron combinaciones de diferentes años anteriores ARX [na nb nk] y algunos ejemplos de los ajustes obtenidos para las diferentes variables climáticas, se presentan en las Figuras 33, 34 y 35.

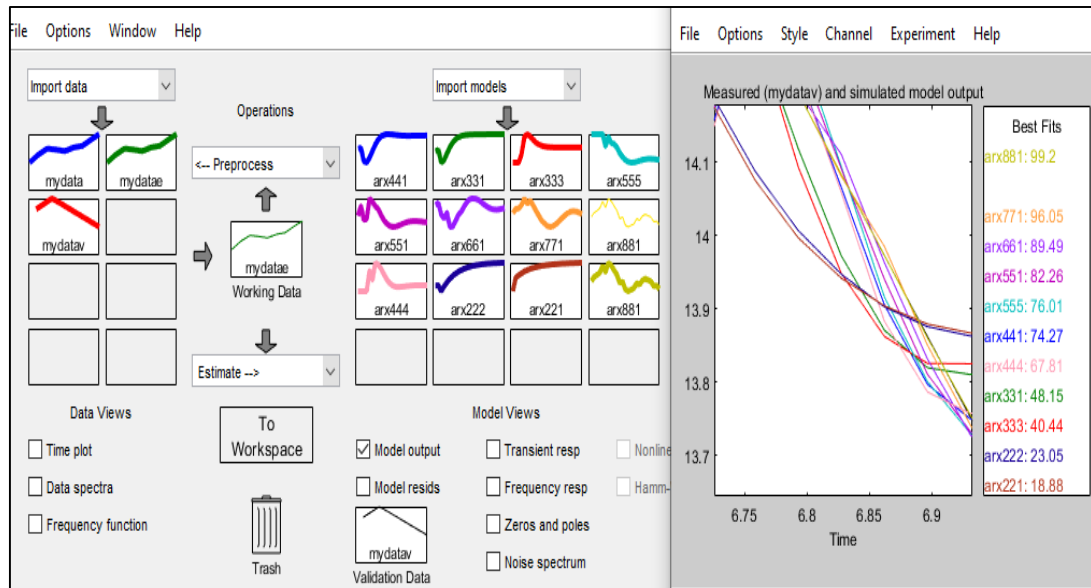


Figura 31. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax.

Fuente: Generado por systemIdentification de MATLAB.

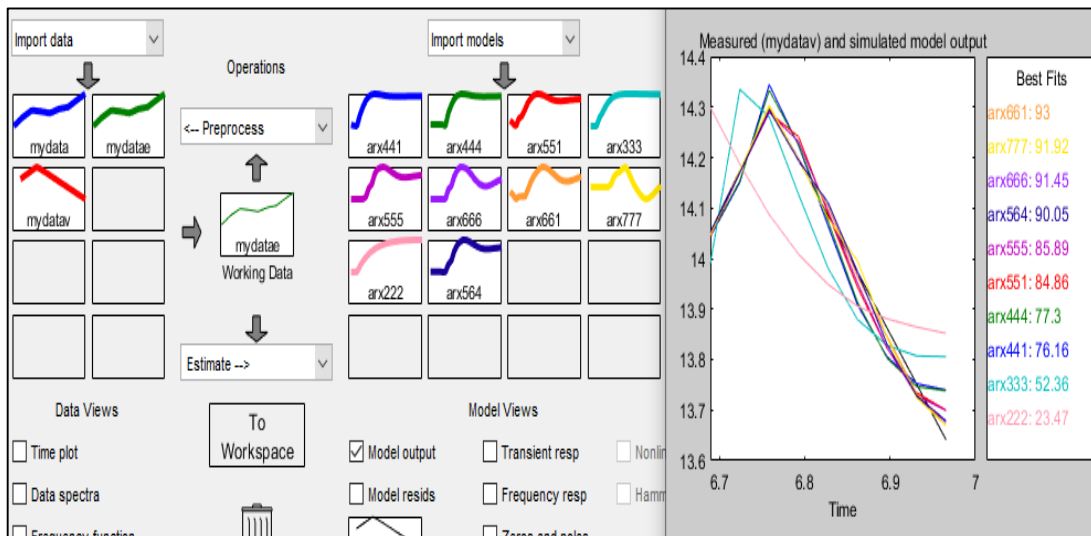


Figura 32. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmin.

Fuente: Generado por systemIdentification de MATLAB.

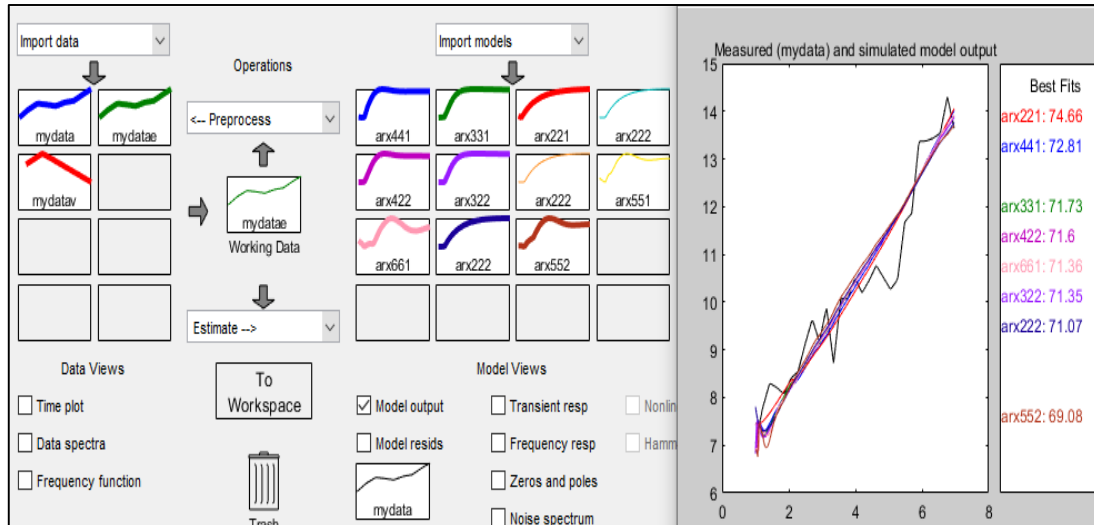


Figura 33. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax y pp.

Fuente: Generado por systemIdentification de MATLAB.

- c. **Segunda etapa de elección del criterio de ajuste con diferentes años anteriores de los datos a tomar en cuenta en el modelo que se genere.** Se probaron combinaciones de diferentes años anteriores ARX [na nb nk] y algunos ejemplos de los ajustes obtenidos para las diferentes variables climáticas se presentarán en las Figuras 36, 37 y 38.

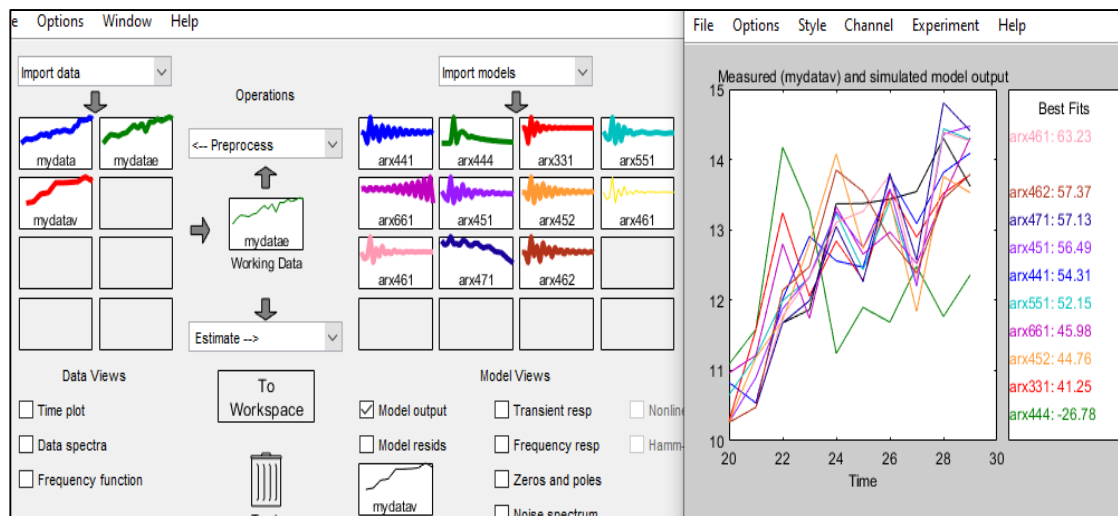


Figura 34. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmin.

Fuente: Generado por systemIdentification de MATLAB.

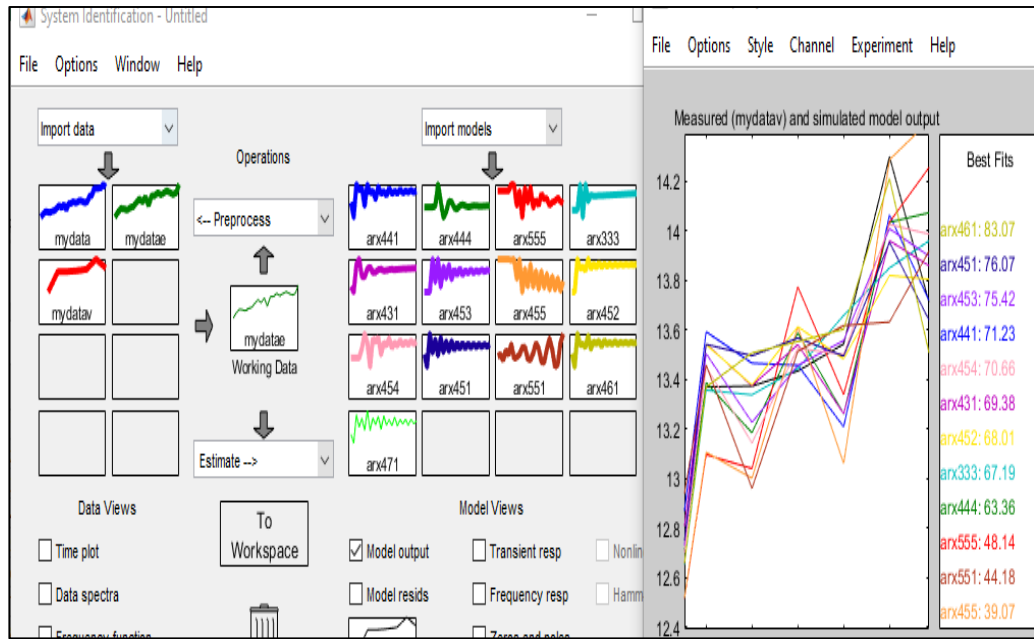


Figura 35. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable pp.

Fuente: Generado por systemIdentification de MATLAB.

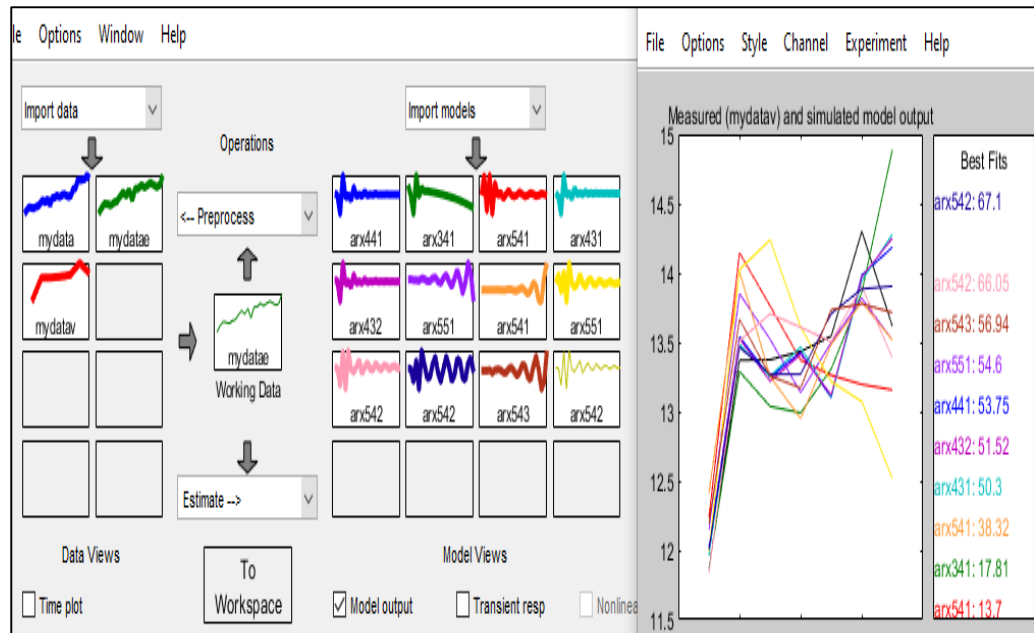


Figura 36. Ventana principal de systemIdentification y gráfica con los resultados de porcentajes de ajuste para los diferentes modelos probados para la variable Tmax y pp.

Fuente: Generado por systemIdentification de MATLAB.