



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación
y Desarrollo Tecnológico

Tesis de Doctorado

Modelo predictivo de evapotranspiración de
referencia para invernaderos utilizando cómputo
evolutivo

presentada por

Francisco Javier Ruiz Ortega

como requisito para la obtención del grado de
Doctor en Ciencias de la Computación

Director de tesis

Dra. Alicia Martínez Rebollar

Codirector de tesis

Dr. José Jassón Flores Prieto

Dr. Eddie E. Clemente Torres

Cuernavaca, Morelos, México. Agosto de 2024.



Cuernavaca, Mor.,
No. De Oficio:
Asunto:

08/agosto/2024
SAC/224/2024
Autorización de
impresión de tesis

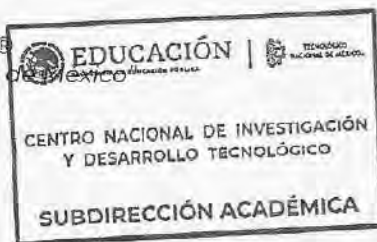
FRANCISCO JAVIER RUIZ ORTEGA
CANDIDATO AL GRADO DE DOCTOR
EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **“MODELO PREDICTIVO DE EVAPOTRANSPIRACIÓN DE REFERENCIA PARA INVERNADEROS UTILIZANDO CÓMPUTO EVOLUTIVO”**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®
“Conocimiento y tecnología al servicio de México”




CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO

C. c. p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



 Centro Nacional de Investigación y Desarrollo Tecnológico	ACEPTACIÓN DE IMPRESIÓN DEL DOCUMENTO DE TESIS DOCTORAL	Código: CENIDET-AC-006-D20
		Revisión: 0
	Referencia a la Norma ISO 9001:2008 7.1, 7.2.1, 7.5.1, 7.6, 8.1, 8.2.4	Página 1 de 1

Cuernavaca, Mor., a 26 de junio de 2024

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE


AT'n: JUAN GABRIEL GONZÁLEZ SERNA
PRESIDENTE DEL CLAUSTRO DOCTORAL


Los abajo firmantes, miembros del Comité Tutorial de la estudiante **FRANCISCO JAVIER RUIZ ORTEGA** manifiestan que después de haber revisado el documento de tesis titulado "MODELO PREDICTIVO DE EVAPOTRANSPIRACIÓN DE REFERENCIA PARA INVERNADEROS UTILIZANDO CÓMPUTO EVOLUTIVO", realizado bajo la dirección de Alicia Martínez Rebollar y codirección de José Jassón Flores Prieto y Eddie Helbert Clemente Torres, el trabajo se **ACEPTA** para proceder a su impresión.

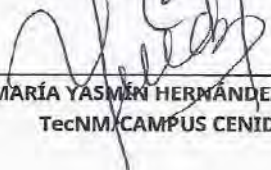
ATENTAMENTE


 ALICIA MARTÍNEZ REBOLLAR
 TecNM/CAMPUS CENIDET


 JOSÉ JASSÓN FLORES PRIETO
 TecNM/CAMPUS CENIDET


 EDDIE HELBERT CLEMENTE TORRES
 TecNM/I.T. ENSENADA


 HUGO ESTRADA ESQUIVEL
 TecNM/CAMPUS CENIDET


 MARÍA YASMÍN HERNÁNDEZ PÉREZ
 TecNM/CAMPUS CENIDET


 DANTE MÚJICA VARGAS
 TecNM/CAMPUS CENIDET


 CRISTINA RAMÍREZ FERNÁNDEZ
 TecNM/I.T. ENSENADA

c.c.p: María Elena Gómez Torres / Jefa del Departamento de Servicios Escolares
 c.c.p: Noé Alejandro Castro Sánchez / Jefe del Departamento C. Computacionales
 c.c.p: Expediente

Dedicatoria

A Dios, por ser la guía y fortaleza en cada paso de este arduo camino.

A mi amada esposa Lupita, cuyo amor incondicional y apoyo constante han sido el pilar sobre el cual se ha construido este logro.

A mis hijos, Avril María y Javier, quienes con su ternura y alegría me han dado la inspiración y motivación necesarias para seguir adelante.

A mis queridos padres, quienes desde el cielo continúan guiándome con su amor y enseñanzas. A mis hermanos, María de Jesús, Judith y David, por su cariño y respaldo inquebrantable en todo momento. Y muy especialmente, a la memoria de mi querido hermano Amalio, cuyo recuerdo y enseñanzas permanecerán siempre vivos en mi corazón. Su partida ha dejado un vacío inmenso, pero su espíritu sigue siendo una fuente de fuerza y perseverancia.

A todos ustedes, esta tesis es testimonio de la gratitud y el profundo respeto que les guardo.

Agradecimientos

La ética y la ciencia deben caminar juntas para asegurar que el progreso tecnológico esté alineado con el bienestar humano y el medio ambiente.
"Stephen John"

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por el apoyo económico que me brindó para realizar mis estudios de doctorado en el Tecnológico Nacional de México campus Centro Nacional de Investigación y Desarrollo Tecnológico (TecNM/CENIDET).

Al Tecnológico Nacional de México campus Tecnológico de Torreón (TecNM/I.T. Torreón) por el apoyo brindado para realizar mis estudios de doctorado.

Al Tecnológico Nacional de México campus Centro Nacional de Investigación y Desarrollo Tecnológico (TecNM/CENIDET) por brindarme la oportunidad de superarme académicamente en el programa de Doctorado en Ciencias de la Computación.

Quiero expresar mi más profundo agradecimiento a mi asesora de tesis, la Dra. Alicia Martínez Rebollar, y a mi revisor el Dr. Hugo Estrada Esquivel, quienes han sido pilares fundamentales en mi desarrollo académico y personal. Su guía, enseñanzas sustanciales y apoyo incondicional me han inspirado a alcanzar altos estándares de excelencia. Aprecio profundamente la amistad que hemos cultivado, su sentido humano, y la humildad y el valor con los que abordan cada desafío. Gracias por compartir conmigo su vasto conocimiento, sus valores y su amistad, y por ser ejemplo excepcional de integridad y compromiso.

Al Dr. José Jassón Flores Prieto, mi codirector de tesis, por su constante apoyo, sus valiosos consejos y la confianza que me ha brindado. Su disposición para compartir sus experiencias ha sido una fuente de inspiración y aprendizaje continuo. Gracias, Dr. Jasson, por creer en mí y por ser un mentor excepcional que ha dejado una huella indeleble en mi desarrollo personal y profesional.

Mi más sincero agradecimiento al Dr. Eddie E. Clemente Torres, mi codirector de tesis, por su invaluable apoyo a lo largo de este proceso. Sus enseñanzas han sido fundamentales para mi desarrollo académico, y su paciencia inquebrantable me ha permitido superar cada reto con confianza y determinación. Aprecio profundamente los desafíos que me ha planteado, ya que me han impulsado a alcanzar nuevos niveles de excelencia y a crecer como investigador. Además de su rol como mentor académico, valoro enormemente la gran amistad que hemos cultivado. Su disposición para ofrecer consejos y su apoyo constante han sido una fuente de fortaleza y motivación. Gracias, Dr. Eddie, por ser un mentor ejemplar y por compartir conmigo su conocimiento, sabiduría y amistad. Estoy profundamente agradecido por haber tenido la oportunidad de aprender y crecer bajo su guía y apoyo.

Quiero expresar mi más sincero agradecimiento mis revisores, la Dra. María Yasmín Hernández Pérez, a la Dra. Cristina Ramírez Fernández y al Dr. Dante Mújica Vargas por su invaluable guía y consejos durante todo este proceso de enseñanza aprendizaje. Su dedicación y empeño han sido fundamentales para mi crecimiento académico y profesional. Además, valoro enormemente su amistad y el apoyo incondicional que siempre me han brindado. Sus contribuciones han enriquecido profundamente mi trabajo, y estoy verdaderamente agradecido por tener la oportunidad de aprender de su experiencia y conocimiento. ¡Muchas gracias por todo!

Resumen

La evapotranspiración de referencia es un parámetro clave en la gestión eficiente del agua en la agricultura, ya que representa la cantidad de agua que se evapora y transpira de una superficie de referencia. Su estimación precisa es fundamental para la planificación y manejo de los recursos hídricos, especialmente en contextos agrícolas, donde el uso racional del agua es crucial para la sostenibilidad y productividad de los cultivos. La evapotranspiración se mide utilizando diversos métodos, entre los cuales los modelos FAO56-Penman-Monteith (FAO56-PM) y Hargreaves-Samani son ampliamente recomendados por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) debido a su robustez y precisión.

En nuestra investigación, hemos desarrollado una metodología innovadora para estimar la evapotranspiración de referencia tanto en campo abierto como en agricultura protegida, utilizando técnicas avanzadas de programación genética, una rama del cómputo evolutivo. Además, hemos diseñado un modelo predictivo específico para invernaderos, que permite anticipar las necesidades hídricas de los cultivos bajo condiciones controladas.

La validación de los modelos desarrollados demuestra su alta precisión, con un error cuadrático medio (RMSE) de 0.3548 y un coeficiente de determinación (R^2) de 0.9361 en campo abierto en comparación con el modelo de referencia FAO56-PM y un error cuadrático medio (RMSE) de 0.2565 y un coeficiente de determinación (R^2) de 0.9216 en invernaderos. Estos resultados indican que nuestros modelos no solo son altamente confiables, sino que también se presentan como una herramienta valiosa para optimizar la gestión del riego en diversos entornos agrícolas, contribuyendo así a la mejora de la eficiencia en el uso del agua y al aumento de la productividad agrícola.

Palabras clave: Cómputo evolutivo, Programación genética, Evapotranspiración, FAO56-PM, Hargreaves-Samani.

Abstract

Reference evapotranspiration is a key parameter in the efficient management of water in agriculture, as it represents the amount of water evaporated and transpired from a reference surface. Accurate estimation of ETo is crucial for the planning and management of water resources, especially in agricultural contexts where rational water use is essential for crop sustainability and productivity. Evapotranspiration is measured using various methods, among which the FAO56-Penman-Monteith (FAO56-PM) and Hargreaves-Samani models are widely recommended by the Food and Agriculture Organization (FAO) due to their robustness and precision.

In our research, we have developed an innovative methodology to estimate reference evapotranspiration for both open fields and protected agriculture, utilizing advanced genetic programming techniques, a branch of evolutionary computing. Additionally, we have designed a predictive model specifically for greenhouses, enabling the anticipation of crop water needs under controlled conditions.

Validation of the developed models demonstrates their high accuracy, with a root mean square error (RMSE) of 0.3548 and a coefficient of determination (R^2) of 0.9361 for open fields compared to the FAO56-PM reference model, and an RMSE of 0.2565 and an R^2 of 0.9216 for greenhouses. These results indicate that our models are not only highly reliable but also serve as valuable tools for optimizing irrigation management across various agricultural environments, thereby contributing to improved water use efficiency and enhanced agricultural productivity.

Keywords: Evolutionary computing, Genetic programming, Evapotranspiration, FAO56-PM, Hargreaves-Samani

Índice general

1. Introducción	1
1.1. Introducción	1
1.2. Motivación	2
1.3. Planteamiento del problema	4
1.4. Objetivos	5
1.4.1. General	5
1.4.2. Específicos	5
1.5. Metodología de solución	6
1.6. Contribuciones de la Tesis	8
1.7. Organización de la Tesis	9
2. Fundamento teórico	12
2.1. Proceso de evapotranspiración	12
2.2. FAO Penman-Monteith	14
2.3. Hargreaves-Samani	15
2.4. Agricultura protegida	16
2.5. Programación Genética	16
2.5.1. Introducción	16
2.5.2. Pasos preparatorios de la Programación Genética	18
2.5.3. Pasos ejecutivos de la Programación Genética	20
2.6. MATLAB y GPLab	28
2.6.1. Introducción	28
2.7. Conclusiones	29

3. Estado del arte	31
3.1. Aplicación de algoritmos de ensamblaje para modelar la evapotranspiración	31
3.2. Diversas técnicas de computación suave para modelar la evapotranspiración	33
3.3. Implementación de técnicas de cómputo evolutivo para modelar la evapotranspiración	36
3.4. Análisis de los trabajos presentados	40
3.5. Conclusiones	46
4. Desarrollo experimental de un modelo para estimar la evapotranspiración a campo abierto	47
4.1. Recolección y preparación de datos	47
4.1.1. Obtención de datos	47
4.1.2. Preprocesamiento de datos	49
4.1.3. Creación de conjuntos de datos	50
4.2. Proceso evolutivo y generación del modelo	50
4.2.1. Realizar pasos preparatorios	51
4.2.2. Ejecución del algoritmo evolutivo	53
4.2.3. Depurar las mejores soluciones	54
4.2.4. Seleccionar el mejor modelo	55
4.3. Validación del modelo	55
4.4. Conclusiones	59
5. Desarrollo de un modelo para estimar la evapotranspiración en invernaderos	60
5.1. Recolección y preparación de datos	61
5.1.1. Obtención de datos	61
5.1.2. Preprocesamiento de datos	63
5.1.3. Creación de conjuntos de datos	63
5.2. Proceso evolutivo y generación del modelo	64
5.2.1. Realizar pasos preparatorios	65
5.2.2. Ejecución del algoritmo evolutivo	70
5.2.3. Depurar las mejores soluciones	72
5.2.4. Seleccionar el mejor modelo	75
5.3. Validación del modelo	77
5.3.1. Validación del modelo de evapotranspiración	77
5.4. Conclusiones	79

6. Desarrollo de un modelo predictivo a corto plazo para invernaderos	80
6.1. Introducción	80
6.2. Obtención y transformación de datos	81
6.2.1. Obtención de datos	81
6.2.2. Transformación de datos	81
6.2.3. Creación del conjunto de datos	83
6.3. Generación del modelo predictivo	83
6.3.1. Realizar pasos preparatorios	84
6.3.2. Ejecutar el algoritmo evolutivo	87
6.3.3. Mejores soluciones	89
6.3.4. Seleccionar el mejor modelo	91
6.4. Validación del modelo	91
6.4.1. Comparación del modelo predictivo contra el método de medias móviles simple	92
6.4.2. Validación del modelo predictivo usando validación cruzada « <i>Rolling Origin</i> »	95
6.5. Conclusiones	98
7. Caso de estudio	99
7.1. Estimación de la evapotranspiración con ETo_{gp} en un invernadero	99
7.2. Predecir la evapotranspiración con ETo_{t+1}	101
7.3. Conclusiones	102
8. Conclusiones y trabajos futuros	103
8.1. Conclusiones generales	103
8.2. Logros obtenidos	105
8.2.1. Productos	105
8.2.2. Participación en concursos	106
8.2.3. Estancia de investigación	106
8.2.4. Participación en proyectos de investigación	107
8.2.5. Ponencias	107
8.2.6. Trabajos futuros	107
8.2.7. Otras actividades	108
A. Productos	116

B. Retribución Social	123
C. Otras actividades	124

Índice de figuras

1.1. Metodología de solución para desarrollar un modelo predictivo	6
2.1. Ciclo principal de Programación Genética.	17
2.2. Representación básica de un programa en forma de árbol utilizada en Programación Genética.	17
2.3. Representación de un programa multi-árbol.	18
2.4. Diagrama de flujo de Programación Genética	22
2.5. Creación de un árbol de siete puntos utilizando el método de inicialización «Full» (t=tiempo)	23
2.6. Creación de un árbol de cinco puntos utilizando el método de inicialización «Grow» (t=tiempo)	23
2.7. Ejemplo de interpretación de un árbol sintáctico (el terminal x es una variable que tiene un valor de -1)	25
2.8. Ejemplo de cruce de dos hijos entre árboles sintácticos.	27
2.9. Ejemplo de mutación de sub-árbol	28
2.10. Estructura operativa de la caja de herramientas GPLAB	30
4.1. Ubicación geográfica de las estaciones meteorológicas utilizadas para entrenamiento.	49
4.2. Ubicación geográfica de las estaciones meteorológicas utilizadas para validación	56
4.3. Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS04.	57
4.4. Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS05.	58
4.5. Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS06.	58
4.6. Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS07.	58
5.1. Ubicación geográfica del área de invernaderos del Departamento de Agricultura de la Universidad Americana de Beirut (AUB), Líbano.	62
5.2. Mejor aptitud, promedio de la mejor aptitud y desviación estándar a lo largo de 50 generaciones	71
5.3. Frecuencia de aparición del conjunto de terminales	73
5.4. Frecuencia de aparición del conjunto de funciones	73

5.5. Series de tiempo y gráficos de dispersión obtenidos en la etapa de entrenamiento.	76
5.6. Serie temporal y gráficos de dispersión de los modelos ETo_{gp} , el ETo_{AT} y ETo_{HS}	78
6.1. Transformación de la serie temporal en una matriz de características de 5 estados previos y un vector objetivo con el valor de la serie que sigue a cada fila de la matriz.	82
6.2. Mejor aptitud y promedio de la mejor aptitud a lo largo de la evolución	88
6.3. Frecuencia de aparición del conjunto de terminales	90
6.4. Frecuencia de aparición del conjunto de funciones	90
6.5. RMSE (ETo_{t+1} y Med. Móviles)	93
6.6. R^2 ETo_{t+1}	94
6.7. R^2 Med. Móviles	94
6.8. Método de validación cruzada usando <i>Rolling Origin 1</i>	95
6.9. Método de validación cruzada usando <i>Rolling Origin 2</i>	96
6.10. Método de validación cruzada usando <i>Rolling Origin</i>	96
6.11. RMSE del modelo ETo_{t+1} con respecto al modelo de referencia <i>FAO56 – PM</i>	97
6.12. R^2 del modelo ETo_{t+1} y el modelo de referencia <i>FAO56 – PM</i>	97
7.1. R^2 RMSE de ETo_{gp}	100
7.2. R^2 ETo_{gp}	100
7.3. RMSE del modelo ETo_{t+1}	101
7.4. R^2 ETo_{t+1}	102
A.1. Página principal de artículo JCR. DOI:10.1038/s41598-024-56770-3.	116
A.2. Capítulo de libro.	117
A.3. Certificado de presentación por <i>LANGUAJE & KNOWLEDGE ENGINEERING LAB</i>	118
A.4. Artículo publicado en la revista <i>Computación y Sistemas</i> . DOI:10.13053/cys-26-1-4166.	119
A.5. Artículo publicado en la 9a Jornada de Ciencia y Tecnología Aplicada.	120
A.6. Certificado de Asistencia y Presentación en el <i>21st Annual Humies Awards – Melbourne, Australia</i>	121
A.7. Cronograma de actividades de la estancia de investigación en el TecNM/I.T. Ensenada.	122
B.1. Conferencia dictada en el TecNM/IT Cuautla.	123
C.1. Curso «Introducción al análisis de datos meteorológicos con R».	124
C.2. Curso «Introducción al análisis de datos meteorológicos con R».	125
C.3. Constancia de participación en el taller: Macroentrenamiento en Inteligencia Artificial.	126
C.4. Diploma de acreditación del Diplomado en Ciencia de Datos.	127
C.5. Constancia del Colegio Científico de Datos (COCID).	128
C.6. Constancia de asistencia Encuentro Nacional de Computación (ENC).	129

C.7. Constancia de asistencia X Seminario y Escuela Nacional de Aprendizaje e
Inteligencia Computacional (XENAIC). 130

Índice de tablas

3.1. Resumen de los trabajos presentados en el estado del Arte	42
4.1. Estaciones meteorológicas utilizadas para entrenamiento del algoritmo evolutivo y desarrollo del modelo a campo abierto	48
4.2. Conjuntos de datos y sus características	50
4.3. Estructura de los conjuntos de datos	50
4.4. Conjunto de terminales.	51
4.5. Conjunto de funciones	52
4.6. Modelos seleccionados	54
4.7. Estaciones meteorológicas utilizadas para validación	56
4.8. Conjuntos de datos usados para validación	57
4.9. Métricas obtenidas con los modelos ETo_{gp} y Hargreaves-Samani	57
5.1. Estructura del archivo original de datos proporcionado por la AUB	62
5.2. Parámetros meteorológicos de interés	63
5.3. Conjunto de datos utilizado para entrenamiento	64
5.4. Conjunto de datos utilizado para prueba	64
5.5. Conjunto de terminales enriquecido con conocimiento experto* del modelo FAO56-PM	66
5.6. Conjunto de funciones	68
5.7. Parámetros usados en el algoritmo evolutivo	70
5.8. Rendimiento de los modelos ETo_{gpi} , Atmómetro y Hargreaves-Samani	72
5.9. Métricas obtenidas por los modelos ETo_{gp} , Atmómetro (ETo_{AT}) y Hargreaves-Samani (ETo_{HS}).	75
5.10. Métricas obtenidas por los modelos ETo_{gp} , el ETo_{AT} y ETo_{HS}	78
6.1. Estructura de los conjuntos de datos de entrenamiento y prueba	84
6.2. Conjunto de terminales	85
6.3. Conjunto de funciones	86
6.4. Parámetros utilizados durante el entrenamiento del algoritmo evolutivo	87
6.5. Modelos obtenidos	89
6.6. Métricas obtenidas en el proceso de validación	93
8.1. Descripción de actividades realizadas	104

Capítulo 1

Introducción

1.1. Introducción

El objetivo principal de la presente investigación consiste en el desarrollo de un modelo predictivo destinado a calcular la evapotranspiración de referencia en entornos de agricultura protegida. Para alcanzar este objetivo, nos apoyamos en técnicas de cómputo evolutivo, específicamente Programación Genética.

La tesis presenta una metodología de solución que tiene la versatilidad de crear modelos tanto para entornos de campo abierto como para ambientes controlados. La metodología está integrada por dos fases, la primera se enfoca en desarrollar un modelo para estimar la evapotranspiración de referencia. El modelo resultante proporciona una valiosa fuente de información para los productores, ya que ofrece estimaciones precisas de las necesidades hídricas de los cultivos utilizando parámetros meteorológicos fácilmente accesibles, como la temperatura del aire, la humedad relativa y la velocidad del viento. Su aplicación no solo mejora la calidad de los cultivos, sino que también reduce los costos de producción y contribuye significativamente a la conservación del agua.

La segunda fase, se concentra en el desarrollo de un modelo predictivo de evapotranspiración a corto plazo, diseñado para anticipar las necesidades hídricas del próximo día. Este enfoque reviste una importancia crucial para los productores, ya que les proporciona la capacidad de gestionar eficientemente el recurso hídrico, garantizando un riego preciso y oportuno. Este nivel de previsión repercute positivamente en una administración más efectiva del agua, lo que, a su vez, se traduce en un aumento significativo en la producción agrícola.

La metodología propuesta se fundamenta en un enfoque evolutivo que emplea la Programación Genética como técnica principal para la creación de los modelos. Esta metodología, de naturaleza adaptable y progresiva, permite el desarrollo gradual de los modelos a través de generaciones, optimizando su desempeño y precisión. La Programación Genética, como técnica de cómputo evolutivo, ofrece una serie de ventajas clave para el desarrollo de modelos de evapotranspiración. Su flexibilidad y adaptabilidad permiten ajustarse a diversas condiciones ambientales y tipos de cultivos, optimizando automáticamente los

parámetros del modelo para obtener resultados precisos. Además, su capacidad para trabajar con datos limitados facilita la modelización de fenómenos complejos. Al no requerir un conocimiento previo detallado del modelo, puede utilizarse de manera autónoma y automática, lo que amplía su aplicabilidad a diferentes contextos y usuarios. La naturaleza evolutiva de la Programación Genética posibilita la mejora continua de los modelos, permitiendo realizar ajustes y optimizaciones a medida que se obtienen nuevos datos o se identifican áreas de mejora.

La Programación Genética representa una herramienta poderosa y versátil para mejorar el procedimiento de gestión del agua en entornos agrícolas, contribuyendo así a una producción más eficiente y sostenible. Este enfoque proporciona una base sólida para el desarrollo de modelos predictivos de fácil interpretación, confiables y eficientes en entornos de agricultura protegida y campo abierto.

El propósito fundamental de esta tesis radica en desarrollar un modelo predictivo destinado a entornos de agricultura protegida. El cual pueda ser implementado eficazmente incluso en áreas con tecnología limitada, donde únicamente se disponga de mediciones de parámetros comúnmente observados, tales como temperatura, humedad relativa y velocidad del viento. El objetivo es diseñar una herramienta que permita predecir los requerimientos de agua del cultivo, facilitando así la toma de decisiones y mejorando la productividad en contextos con recursos tecnológicos limitados. La Sección 1.2 de este Capítulo discute la motivación de este trabajo de investigación. La Sección 1.3 presenta el planteamiento del problema. La Sección 1.4 presenta los objetivos de la investigación. La Sección 1.5 presenta la metodología de solución, la sección 1.6 se plentean las contribuciones de la tesis y Finalmente, la Sección 1.7 esboza la organización de la tesis.

1.2. Motivación

Actualmente, el uso eficiente del agua en el sector agrícola es fundamental para generar una producción rentable. El 77% del agua aprovechable en México está destinada a la producción de alimentos, pero el uso eficiente de este recurso es aproximadamente del 46%, como afirman (Hernández-Salazar et.al., 2019). La agricultura protegida en México ha crecido y se ha extendido geográficamente. Aproximadamente el 80% de la producción se destina a los mercados de exportación (casi exclusivamente a Estados Unidos) como lo hacen notar (L. Pratt et.al., 2019). La agricultura protegida en México generalmente se divide en tres categorías: invernaderos de tecnología alta, invernaderos de tecnología baja y casas sombra, y una cuarta categoría intermedia, dependiendo de la tecnología utilizada (Hernández-Salazar et.al., 2019).

La programación adecuada del riego, basada en la comprensión de las necesidades hídricas de las plantas, es fundamental para maximizar la calidad y rendimiento de los cultivos. Además, la gestión eficiente del recurso hídrico es esencial para la sostenibilidad y rentabilidad a largo plazo de la agricultura. El enfoque basado en la evapotranspiración (ET) es un método utilizado para programar el riego. Este método se basa en estimar la pérdida de agua del suelo debido a la evaporación y la transpiración de las plantas. La

Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO *Food and Agriculture Organization*), recomienda el método de Penman-Monteith para estimar la evapotranspiración de referencia (ET_o) con parámetros climáticos en grandes extensiones de cultivo (Allen, 1998). Sin embargo, aunque puede proporcionar una estimación precisa de las necesidades hídricas de las plantas, su implementación puede ser compleja y requiere parámetros precisos que en muchas ocasiones no están disponibles.

Se conoce como evapotranspiración de referencia (ET_o) a la tasa de evapotranspiración de una superficie de referencia, que ocurre sin restricciones de agua, y se denomina ET_o. La superficie de referencia corresponde a un cultivo hipotético de pasto con características específicas. La evapotranspiración de referencia (ET_o) como base para programar el riego, evita la erosión del suelo, pérdida de nutrientes, estrés hídrico y en caso extremo la muerte de la planta. La aplicación de un riego limitado o en exceso ocasiona pérdidas en la producción y enfermedades en las plantas. Sin embargo, es importante destacar que la implementación exitosa de controladores basados en la evapotranspiración de referencia (ET_o) en entornos de invernadero requiere una evaluación minuciosa y adaptación a las condiciones específicas del lugar. Las características únicas de un invernadero, como la variabilidad en las condiciones climáticas y la composición del sustrato, pueden influir en la precisión de las estimaciones de la evapotranspiración de referencia (ET_o) y, por lo tanto, en la eficacia del riego.

En la actualidad, técnicas de computación suave son empleadas con frecuencia en respuesta a los problemas presentados en la agricultura, debido a la gran alternativa y ventajas que ofrece la Inteligencia Artificial. Hoy en día, técnicas de cómputo evolutivo como Programación Genética, Redes Neuronales Artificiales, modelos Neuro Difusos entre otras, se han utilizado con éxito para modelar procesos hidrológicos (Hernández-Salazar et.al., 2019; de la Piedra et.al., 2016; Nourani et.al., 2012). Estas técnicas tienen la capacidad de modelar procesos complejos no lineales y pueden estimar la evapotranspiración de referencia con precisión.

A lo largo de la historia, se ha llevado a cabo una extensa investigación con el propósito de estimar la evapotranspiración de referencia utilizando datos meteorológicos estándar. En este contexto, los esfuerzos sostenidos se han centrado en reducir el número de variables y, consecuentemente, los parámetros meteorológicos requeridos para la estimación de ET_o. Sin embargo, no se ha encontrado evidencia sobre algún método estándar para estimar la evapotranspiración de referencia (ET_o) al interior de invernaderos como afirman (Karaca et.al., 2018). Por lo tanto, consideramos que se necesita más investigación y desarrollo de nuevos métodos y asegurarse que sean adecuados para su uso en invernaderos.

En su libro «*Genetic Programming: On the Programming of Computers by Means of Natural Selection*», (Koza, 1992) introdujo la Programación Genética como un método automático, independiente del dominio, con la capacidad de generar programas genéticos para resolver una amplia variedad de problemas en diversos campos. En este trabajo de investigación, se optó por emplear la Programación Genética debido a su habilidad para abordar problemas de alta complejidad y explorar soluciones que podrían no ser inmediatamente evidentes para los seres humanos. La distintiva ventaja del enfoque de Programación Genética radica en su naturaleza de «caja abierta» o «caja blanca». Gracias

al principio de «supervivencia del más apto», este enfoque permite extraer elementos relevantes y útiles para la construcción de modelos a partir del conjunto inicial de datos.

1.3. Planteamiento del problema

El agua es un elemento esencial para la supervivencia, y en el contexto global, más del 70 % del consumo de agua es destinada para fines agrícolas. Se estima que en los países en desarrollo, entre el 70 y el 90 % de los recursos de agua dulce se emplean en la producción de alimentos. Además, el 75 % de las 1,200 millones de personas que viven en condiciones de pobreza en el mundo dependen en gran medida de la agricultura como su principal fuente de ingresos (Ramírez Barraza et.al., 2019).

En términos económicos, la gestión adecuada de la evapotranspiración es crucial para reducir el desperdicio de agua y mejorar la calidad de los cultivos. La pérdida de un milímetro de agua por hectárea equivale a 10,000 litros, lo que destaca la importancia de una administración eficiente del agua en la agricultura. Estos aspectos tienen un impacto significativo en la producción de alimentos y en la economía agrícola.

En México, la superficie agrícola abarca aproximadamente 106,891,000 millones de hectáreas, según el informe de (Gómez et.al., 2024). Considerando que la pérdida de un milímetro de agua por hectárea equivale a 10,000 litros, esto se traduce en una pérdida global de 1,068,910,000 metros cúbicos de agua por milímetro. Económicamente, esta cantidad representa un costo aproximado de \$1,068,910,000 millones de pesos por milímetro de agua perdido, calculado a un costo promedio de \$1.11 pesos por metro cúbico de agua para uso agrícola (Gómez et.al., 2024).

La eficiencia en la gestión del agua no solo reduce costos, sino que también contribuye a la sostenibilidad del recurso hídrico, mitiga los efectos del cambio climático y garantiza la seguridad alimentaria. La implementación de tecnologías y métodos avanzados para controlar la evapotranspiración puede aumentar significativamente la productividad agrícola y la rentabilidad, asegurando que los recursos se utilicen de manera óptima y sostenible.

En la agricultura protegida, un manejo adecuado del agua es fundamental para una producción rentable. La programación precisa del riego, basada en la comprensión de las necesidades hídricas de las plantas, es esencial para maximizar la calidad y el rendimiento de los cultivos. Además, la gestión eficiente del recurso hídrico es crucial para la sostenibilidad y la rentabilidad a largo plazo de la agricultura en invernaderos.

Para lograr una gestión eficiente del agua, se pueden emplear varios métodos. Un enfoque basado en datos para estimar la evapotranspiración (ET) es uno de los métodos más efectivos para programar el riego. Este método se basa en la estimación de la pérdida de agua del suelo debido a la evaporación y la transpiración de las plantas, lo que permite una programación del riego más precisa y eficiente.

No obstante, es esencial subrayar que la implementación efectiva de controladores basados en la evapotranspiración (ET) en entornos de invernadero demanda una evaluación minuciosa y una adaptación precisa a las condiciones específicas del lugar. Las

características únicas de los invernaderos, como la variabilidad en las condiciones climáticas y la composición del sustrato, pueden influir significativamente en la precisión de las estimaciones de la ETo y, por ende, en la eficacia del riego.

Por consiguiente, ante la ausencia de un modelo estándar para estimar la evapotranspiración de referencia dentro de los invernaderos, tal como lo señaló (Karaca et.al., 2018), resulta imperativo el desarrollo de nuevos modelos diseñados específicamente para entornos de agricultura protegida. Esta necesidad se deriva de la singularidad de los invernaderos, donde las condiciones ambientales y los factores de influencia difieren considerablemente de los cultivos al aire libre, lo que demanda enfoques adaptados y más precisos para la estimación de la evapotranspiración en este contexto. Esta falta de un modelo estándar constituye uno de los principales desafíos en el diseño y la gestión eficaz del riego en los invernaderos, resaltando la urgencia de la investigación y el desarrollo en este campo para mejorar la productividad y la sostenibilidad de la agricultura protegida.

Para abordar la problemática planteada, se propone trabajar con técnicas de cómputo evolutivo, en particular con Programación Genética. La Programación Genética se posiciona como una técnica de aprendizaje automático supervisado que utiliza la computación evolutiva para afrontar una amplia gama de desafíos científicos. En el ámbito de la ingeniería de recursos hídricos (WRE, *Water Resources Engineering*), ha despertado un interés considerable debido a su capacidad única para introducir modelos explícitos en un proceso determinado (Mehr et.al., 2018). Esta atención creciente se debe a que la Programación Genética ofrece una herramienta poderosa y versátil para abordar problemas complejos relacionados con la gestión del agua, permitiendo la creación automatizada de modelos que pueden adaptarse y evolucionar para satisfacer las necesidades específicas de cada situación. Además, la Programación Genética proporciona la flexibilidad necesaria para explorar y aprovechar la diversidad de soluciones potenciales en entornos variables y dinámicos, lo que la convierte en una opción prometedora para mejorar la eficiencia y la precisión en la predicción y gestión de los recursos hídricos.

1.4. Objetivos

1.4.1. General

Desarrollar un modelo predictivo para estimar la evapotranspiración de referencia en invernaderos mediante Programación Genética.

1.4.2. Específicos

A continuación, se detallan los objetivos específicos que orientaron y estructuraron el desarrollo de esta investigación:

- Seleccionar y adaptar las variables meteorológicas relevantes para estimar la evapotranspiración de referencia en el contexto específico de los invernaderos, considerando factores como la temperatura, humedad relativa y velocidad del viento.

- Recolectar y preparar una base de datos de alta calidad con datos climáticos históricos y de evapotranspiración de referencia en invernaderos.
- Configurar e implementar un algoritmo evolutivo que pueda generar modelos matemáticos precisos para la estimación de la evapotranspiración de referencia.
- Evaluar la precisión y eficacia del modelo predictivo mediante la comparación de sus predicciones con el modelo de referencia utilizando métricas de validación adecuadas.

Hipótesis

La implementación de un modelo predictivo basado en Programación Genética para estimar la evapotranspiración de referencia en invernaderos ofrecerá predicciones más precisas y adaptativas que los métodos convencionales, al aprovechar la capacidad de la Programación Genética para considerar la complejidad ambiental específica de este entorno.

1.5. Metodología de solución

En esta sección se presenta la metodología propuesta para abordar el desarrollo del modelo predictivo a corto plazo. Esta metodología consta de dos fases principales. La primer fase, titulada «Desarrollo de un Modelo para Evapotranspiración», se divide en dos procesos. El proceso de «recolección y preparación de datos», y el proceso «proceso evolutivo y generación del modelo». Como resultado, se obtiene un modelo explícito capaz de estimar la evapotranspiración de referencia diaria, el cual será utilizado en la siguiente fase. La segunda fase, «Desarrollo de un Modelo Predictivo a Corto Plazo», se compone de dos procesos. El proceso de «obtención y transformación de datos», seguido por el proceso de «generación del modelo predictivo». Como resultado, obtenemos un modelo para predecir la evapotranspiración de referencia a corto plazo. Para una visualización más clara de la metodología de solución, se incluye la Figura 1.1, que representa de manera gráfica el flujo de trabajo y la interacción entre las diferentes fases y procesos.

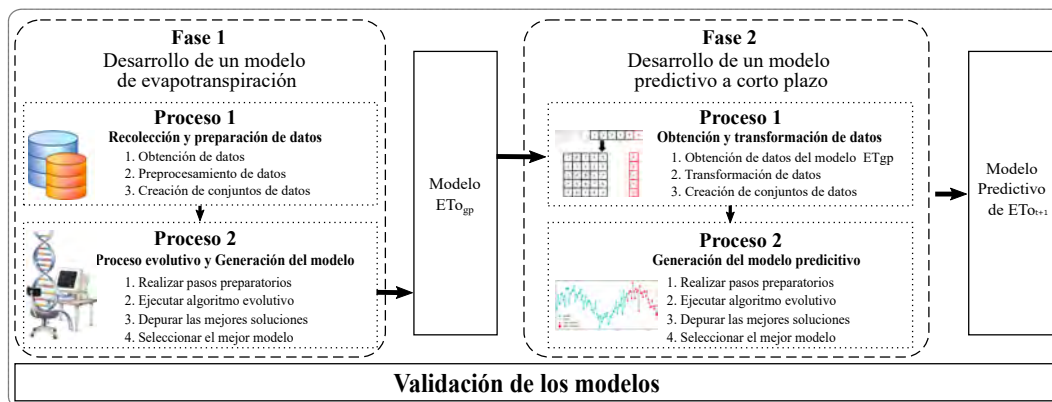


Figura 1.1: Metodología de solución para desarrollar un modelo predictivo

A continuación, proporcionaremos una visión general de los distintos procesos que componen nuestra metodología de solución. Estos procesos están diseñados para abordar de manera integral el desarrollo del modelo predictivo para la evapotranspiración a corto plazo. Posteriormente, en secciones subsiguientes, nos adentraremos en una descripción detallada de cada uno de estos procesos, proporcionando una comprensión más profunda de las etapas específicas que llevaremos a cabo para alcanzar nuestro objetivo.

Fase 1 (Proceso 1): Recolección y preparación de datos

El proceso de recolección y preparación de datos es esencial para asegurar la calidad de los datos, representar adecuadamente el problema en cuestión, preprocesar los datos según sea necesario y mejorar la eficiencia computacional. La calidad de los datos es fundamental para obtener resultados precisos y útiles, mientras que una representación adecuada del problema garantiza que los datos reflejen correctamente las características y relaciones relevantes. El preprocesamiento de datos puede implicar la limpieza y normalización de datos, así como la selección de características, para prepararlos para su análisis. Finalmente, una preparación eficiente de datos puede mejorar la eficiencia computacional al reducir la dimensionalidad de los datos y optimizar su representación, lo que facilita el proceso de búsqueda. En la sección 4.1 del capítulo 4 describe este proceso

Fase 1 (Proceso 2): Proceso evolutivo y generación del modelo

El proceso evolutivo y generación del modelo comienza con la creación inicial de una población de modelos matemáticos, representados como árboles de expresión. Estos modelos se evalúan utilizando una función de aptitud que mide su capacidad para ajustarse a los datos observados de evapotranspiración. A través de operadores genéticos de selección, reproducción y reemplazo, los modelos más aptos se combinan y/o mutan para formar una nueva población con el objetivo de mejorar la precisión de las predicciones. Este proceso continúa durante un número predeterminado de generaciones o hasta que se cumpla un criterio de convergencia, momento en el cual se selecciona el mejor modelo como la solución final. En la sección 4.2 del capítulo 4 ofrece una descripción detallada de este proceso.

Para avanzar hacia la generación del modelo predictivo, nos adentramos en los siguientes dos procesos que conforman la fase 2 de la metodología de solución. Estos pasos son fundamentales para completar el desarrollo del modelo y permitirán obtener una visión más clara y detallada de cómo se llevará a cabo la predicción de la evapotranspiración a corto plazo.

Fase 2 (Proceso 1). Obtención y transformación de datos

La obtención y transformación de datos consiste en transformar un vector de datos en una matriz y un vector con una ventana de tiempo de 5 cinco estados previos, esto implica organizar los datos en una estructura matricial y extraer valores objetivo a partir del desplazamiento de una ventana a lo largo del vector original. Inicialmente, se crea una matriz donde cada fila representa una observación y cada columna, una característica. Luego, se desplaza una ventana de tamaño 5 a lo largo del vector, extrayendo conjuntos de cinco valores consecutivos y seleccionando el siguiente valor como el objetivo de la

predicción. Estos valores se almacenan en un vector separado, mientras que la matriz de datos se actualiza agregando filas adicionales a medida que avanza la ventana. Este proceso es esencial para preparar los datos antes de aplicar técnicas de modelado predictivo, permitiendo que los algoritmos de aprendizaje automático o técnicas de regresión trabajen con la estructura adecuada y los objetivos adecuados para predecir valores futuros con base en datos históricos. En la sección 5.1 del capítulo 5 se describe este proceso.

Fase 2 (Proceso 2). Generación del modelo predictivo

La generación del modelo predictivo implica un proceso iterativo que combina la evolución de modelos matemáticos con la consideración de datos históricos. En primer lugar, se organiza el conjunto de datos en una estructura adecuada, en nuestra investigación utilizamos una ventana de tiempo de 5 estados previos para crear tanto la matriz de características como el vector objetivo (descrito en el proceso anterior). Posteriormente, se utiliza la Programación Genética para evolucionar modelos matemáticos que puedan predecir valores futuros basados en los datos históricos proporcionados. Durante este proceso, se aplican operadores genéticos como la selección, cruce y mutación para generar una población inicial de modelos y mejorar su adaptación a los datos de entrenamiento. Luego, se evalúan los modelos en función de su capacidad para predecir correctamente los valores objetivo, utilizando métricas de rendimiento específicas. Los modelos más aptos son seleccionados y utilizados como base para generar la siguiente generación de modelos, continuando así el ciclo evolutivo hasta que se alcanza un criterio de convergencia o se ha completado un número predeterminado de generaciones.

1.6. Contribuciones de la Tesis

En el ámbito de la investigación, el aporte computacional consistió en una metodología que emplea técnicas de ciencia de datos y optimización para crear modelos computacionales de evapotranspiración. En particular la Programación Genética permitió la creación y optimización de modelos complejos para estimación y predicción. La Programación Genética facilitó la generación automática de algoritmos, aprovechando principios inspirados en la evolución biológica. Su aplicación en la investigación aportó una herramienta poderosa para abordar problemas de optimización y modelado, proporcionando soluciones eficientes y adaptativas que mejoraron significativamente la eficacia de los análisis computacionales. Este enfoque no solo promovió la agilidad en la resolución de problemas, sino que también abrió nuevas posibilidades para la exploración y comprensión de fenómenos complejos a través de la simulación y el modelado de datos meteorológicos. En la investigación, se destacan los siguientes aportes:

- Una metodología para estimar la evapotranspiración de referencia en invernaderos y a campo abierto. esta metodología se destaca por su capacidad para emplear exclusivamente parámetros meteorológicos de humedad relativa y temperatura. Esta singular característica le permite realizar evaluaciones precisas de la evapotranspiración de referencia al interior de invernaderos, superando eficazmente las mediciones obtenidas mediante el modelo de Hargreaves-Samani y el uso de un atmómetro. La

limitación a estos dos parámetros específicos demuestra la eficiencia del algoritmo, al mismo tiempo que subraya su capacidad para ofrecer resultados más exactos que los métodos convencionales; así como, mediciones realizadas con instrumentación especializada. Esta particularidad no solo simplifica el proceso de evaluación, sino que también destaca la mejora significativa en la precisión y utilidad del algoritmo en comparación con técnicas tradicionales como la de Hargreaves-Samani.

- Modelos para predecir la evapotranspiración de referencia de forma diaria en invernaderos y campo abierto. Estos modelos destacan por su habilidad para realizar pronósticos diarios altamente precisos de la evapotranspiración de referencia en el interior de invernaderos y a campo abierto. Su capacidad predictiva se basa únicamente en una ventana de tiempo de cinco días de datos históricos diarios de evapotranspiración. Esta singular característica no solo resalta su eficacia, sino que también subraya su facilidad de implementación. La utilidad de estos modelos se manifiesta de manera significativa en el ámbito de la agricultura protegida, donde la capacidad de predecir con precisión la evapotranspiración de referencia de forma diaria resulta esencial para la gestión efectiva del recurso hídrico y del cultivo en invernaderos. La simplificación de las variables de entrada, sin comprometer la precisión, lo posiciona como una herramienta valiosa para mejorar la eficiencia y la toma de decisiones en la producción agrícola en el contexto de agricultura protegida.
- Conjunto de Datos preprocesados. El conjunto de datos está disponible con el propósito de ser empleado con la metodología propuesta en el marco de la investigación. Este recurso constituye un componente fundamental que amplía la aplicabilidad y validez de la metodología. El conjunto de datos, tras un proceso de preprocesamiento riguroso, se presenta como un recurso listo para ser aprovechado, facilitando así la aplicación efectiva de la metodología y fortaleciendo la robustez de los resultados obtenidos.

La aportación computacional se revela como un activo especialmente significativo en el dominio que se aborda, destacando particularmente en escenarios donde la optimización de parámetros o la selección de una combinación óptima de variables se vuelve imperativa. El enfoque de Programación Genética no solo simplifica tareas complejas, sino que también se presenta como una herramienta fundamental para abordar de manera eficaz problemas que requieren un análisis riguroso y una adaptación constante en busca de soluciones óptimas.

1.7. Organización de la Tesis

El resto de esta tesis está organizado en los siguientes capítulos:

Capítulo 2 Fundamento teórico

Este capítulo proporciona la base conceptual y contextual necesaria para comprender el marco de referencia en el que se desarrolla la investigación. Aquí se abordan los principios,

teorías y conceptos clave que sustentan la propuesta.

Capítulo 3. Estado del arte

Este capítulo se centra en el estado del arte, presentando una revisión exhaustiva de los trabajos de investigación relevantes en el campo de estudio. Este análisis crítico sitúa nuestro trabajo en el contexto más amplio de la investigación existente, identificando brechas de conocimiento y estableciendo conexiones con las contribuciones previas.

Capítulo 4. Desarrollo experimental de un modelo para la estimación de la evapotranspiración a campo abierto

En esta sección se propone experimentar con la primera fase de la metodología de solución, ya que esta es fundamental para el éxito de la segunda fase, la cual se centra en desarrollar el modelo predictivo, que es el objetivo principal de esta investigación. La experimentación en esta etapa busca desarrollar un modelo simple y fácil de interpretar para estimar la evapotranspiración en campo abierto. Para lograrlo, se utilizarán datos meteorológicos obtenidos de estaciones ubicadas en diversas regiones y climas, con fuentes que incluyen la red CIMIS (*California Irrigation Management Information System*) y una estación adicional «El Porvenir» ubicada en el norte-centro de México. Esta diversidad de datos permitirá evaluar la robustez y aplicabilidad del modelo en diferentes condiciones climáticas, asegurando que el modelo no solo sea preciso, sino también versátil y capaz de generalizar bien en distintos entornos.

Capítulo 5. Desarrollo de un modelo para la estimación de la evapotranspiración

En esta sección se ofrece una explicación exhaustiva sobre la recolección y preparación de datos, abordando meticulosamente los métodos, enfoques y técnicas utilizadas para alcanzar los objetivos propuestos. Además, se detalla la generación del modelo de evapotranspiración de referencia. El capítulo cubre los procesos desde la obtención de datos, hasta el proceso evolutivo desde la definición de los parámetros del algoritmo evolutivo hasta la obtención del mejor modelo descubierto. Aquí se analizan los resultados obtenidos, se comparan con el modelo de Hargreaves-Samani y un atmómetro tomando como referencia el modelo FAO Penman-Monteith y se examina la eficacia y precisión del modelo, proporcionando una evaluación crítica de su desempeño.

Capítulo 6. Generación del modelo predictivo

Este capítulo detalla el proceso de obtención y transformación de datos derivados del modelo de evapotranspiración desarrollado en el capítulo anterior. Se explora en detalle el método utilizado para convertir los datos en una matriz de 5 estados previos, y un vector objetivo. El propósito principal de esta transformación es suministrar al algoritmo evolutivo datos precisos y de alta calidad, lo que contribuirá a la obtención de resultados más exactos y confiables. Así mismo, se profundiza en el desarrollo del modelo predictivo mediante la aplicación de técnicas de Programación Genética, un enfoque innovador que permite la creación de algoritmos adaptativos y eficientes para la predicción de la evapotranspiración de referencia a corto plazo. Este enfoque meticuloso busca garantizar que el modelo pueda hacer predicciones acertadas basadas en información histórica relevante, lo que a su vez mejora su capacidad para adaptarse a diferentes condiciones y contextos.

Capítulo 7. Caso de estudio

En este capítulo, se presenta un caso de estudio destinado a la validación de nuestro modelo predictivo utilizando datos previamente no observados. La validación con datos fuera de la muestra representa una evaluación crítica de la efectividad del modelo, ya que permite determinar su capacidad de generalización. Este procedimiento es fundamental para verificar la robustez y precisión del modelo en la predicción de instancias no incluidas en el conjunto de entrenamiento, asegurando así su aplicabilidad y fiabilidad en condiciones reales y diversas configuraciones experimentales.

Capítulo 8. Conclusiones y trabajos futuros

En este capítulo se exponen las conclusiones derivadas de la investigación, enfatizando el papel crucial de la Programación Genética en el desarrollo de los modelos de evapotranspiración. Además, se resaltan los hallazgos más relevantes y se discuten sus implicaciones en el contexto de la agricultura protegida. También se esbozan las direcciones para futuras investigaciones, señalando posibles áreas de expansión, investigación adicional y desarrollo continuo en el tema abordado.

Esta estructura proporciona una disposición lógica y progresiva del contenido, guiando al lector a través de los aspectos teóricos, revisión de la literatura, metodología, validación y conclusiones, brindando así una comprensión completa del trabajo de investigación y sus contribuciones.

Capítulo 2

Fundamento teórico

En este capítulo, se expondrán los conceptos fundamentales necesarios para una comprensión integral del trabajo de investigación. En particular, se abordarán conceptos clave en hidrología e Inteligencia Artificial, como evaporación, transpiración, evapotranspiración, la relación entre evaporación y transpiración, así como, cómputo evolutivo y Programación Genética.

2.1. Proceso de evapotranspiración

Se conoce como evapotranspiración (ET) la combinación de dos procesos separados por los que el agua se pierde a través de la superficie del suelo por evaporación y por otra parte mediante transpiración del cultivo.

La **evaporación** es el proceso por el cual el agua líquida se convierte en vapor de agua (vaporización) y se retira de la superficie evaporante (remoción de vapor). El agua se evapora de una variedad de superficies, tales como lagos, ríos, caminos, suelos y la vegetación mojada. Para cambiar el estado de las moléculas del agua de líquido a vapor se requiere energía. La radiación solar directa y, en menor grado, la temperatura ambiente del aire, proporcionan esta energía. La fuerza impulsora para retirar el vapor de agua de una superficie evaporante es la diferencia entre la presión del vapor de agua en la superficie evaporante y la presión de vapor de agua de la atmósfera circundante. A medida que ocurre la evaporación, el aire circundante se satura gradualmente y el proceso se vuelve cada vez más lento hasta detenerse completamente si el aire mojado circundante no se transfiere a la atmósfera, o en otras palabras no se retira de alrededor de la hoja. El reemplazo del aire saturado por un aire más seco depende de la velocidad del viento. Por lo tanto, la radiación, la temperatura del aire, la humedad atmosférica y la velocidad del viento son parámetros climatológicos a considerar al evaluar el proceso de evaporación (Allen et.al., 2006).

La **transpiración** consiste en la vaporización del agua líquida contenida en los tejidos de la planta y su posterior remoción hacia la atmósfera. Los cultivos pierden agua predominantemente a través de los estomas. Estos son pequeñas aberturas en la hoja de la planta a través de las cuales atraviesan los gases y el vapor de agua de la planta hacia la atmósfera.

El agua, junto con algunos nutrientes, es absorbida por las raíces y transportada a través de la planta. La vaporización ocurre dentro de la hoja, en los espacios intercelulares, y el intercambio del vapor con la atmósfera es controlado por la abertura estomática. Casi toda el agua absorbida del suelo se pierde por transpiración y solamente una pequeña fracción se convierte en parte de los tejidos vegetales.

La transpiración, igual que la evaporación directa, depende del aporte de energía, del gradiente de presión del vapor y de la velocidad del viento. Por lo tanto, la radiación solar, la temperatura del aire, la humedad atmosférica y el viento también deben ser considerados en su determinación (Allen et.al., 2006).

La **evapotranspiración** se expresa normalmente en milímetros [mm] por unidad de tiempo. Esta unidad expresa la cantidad de agua perdida de una superficie cultivada en unidades de altura de agua. La unidad de tiempo puede ser una hora, día, 10 días, mes o incluso un período completo de cultivo o un año. Como una hectárea tiene una superficie de 10,000 [m²] y 1 [mm] es igual a 0.001 [m], la pérdida de 1 [mm] de agua corresponde a la pérdida de 10 [m³] de agua por hectárea. Es decir, la pérdida de agua a razón de una hectarea por día es equivalente 10 [m³].

La altura del agua se puede expresar también en términos de la energía recibida por unidad de área. Esto último se refiere a la energía o al calor requerido para vaporizar el agua. Esta energía, conocida como el calor latente de vaporización λ , es una función de la temperatura del agua. Por ejemplo, a 20[°C], λ tiene un valor cerca de 2.45 [MJ/Kg]. Es decir 2.45 [MJ] son necesarios para vaporizar 1 kilogramo ó 0.001 [m³] de agua. Por lo tanto, un aporte de energía de 2.45 [MJ/m²] puede vaporizar 0.001 [m] o 1 [mm] de agua, y entonces 1 [mm] de agua es equivalente a 2.45 [MJ/m⁻²] (Allen et.al., 2006).

Los principales factores que influyen en la evaporación y la transpiración son: el clima, las características del cultivo, el manejo y medio de desarrollo del cultivo. En esta sección nos enfocaremos en las variables climáticas.

Los principales parámetros climáticos que afectan la evapotranspiración son: la radiación, la temperatura del aire, la humedad atmosférica y la velocidad del viento. Se han desarrollado varios procedimientos para determinar la evaporación a partir de estos parámetros. La fuerza evaporativa de la atmósfera puede ser expresada por la evapotranspiración del cultivo de referencia (ET_o). La evapotranspiración del cultivo de referencia (ET_o) representa la pérdida de agua de una superficie cultivada estándar (Allen et.al., 2006).

La tasa de evapotranspiración de una superficie de referencia, que ocurre sin restricciones de agua, se conoce como evapotranspiración del cultivo de referencia, y se denomina ET_o. La superficie de referencia corresponde a un cultivo hipotético de pasto con características específicas. El concepto de evapotranspiración de referencia se introdujo para estudiar la demanda de evapotranspiración de la atmósfera, independientemente del tipo y desarrollo del cultivo, y de las prácticas de manejo.

Los únicos factores que afectan ET_o son los parámetros climáticos. Por lo tanto, ET_o es también un parámetro climático que puede ser calculado a partir de datos meteorológicos. ET_o expresa el poder evaporante de la atmósfera en una localidad y época del año específicas, y no considera ni las características del cultivo, ni los factores del suelo. Desde este

punto de vista, el método FAO Penman-Monteith se recomienda como el método estándar para determinar la ETo con parámetros climáticos. Este método ha sido seleccionado debido a que aproxima de una manera cercana la ETo de cualquier localidad evaluada, tiene bases físicas sólidas e incorpora explícitamente parámetros fisiológicos y aerodinámicos. Una alternativa recomendada por la FAO para localidades donde no existen datos de radiación solar es el método de la diferencias de temperatura de Hargreaves-Samani (Ecuación 2.5) (Allen et.al., 2006).

2.2. FAO Penman-Monteith

Un panel de expertos e investigadores en riego fue organizado por la FAO en mayo de 1990, en colaboración con la Comisión Internacional para el Riego y Drenaje y con la Organización Meteorológica Mundial, con el fin de revisar las metodologías previamente propuestas por la FAO para el cálculo de los requerimientos de agua de los cultivos y para elaborar recomendaciones sobre la revisión y la actualización de procedimientos a este respecto. El panel de expertos recomendó la adopción del método combinado de Penman-Monteith (Ecuación 2.1) como nuevo método estandarizado para el cálculo de la evapotranspiración de la referencia y aconsejó sobre los procedimientos para el cálculo de los varios parámetros que la fórmula incluye.

Ecuación de Penman-Monteith recomendada por la FAO

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_{mean} + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (2.1)$$

donde ET_o evapotranspiración de referencia [mm/día], R_n radiación neta en la superficie del cultivo [MJ/m²/día], R_a radiación extraterrestre [mm/día/], G flujo del calor de suelo [MJ/m²/ día], T_{mean} temperatura media del aire a 2 m de altura [°C], u_2 velocidad del viento a 2 [m] de altura [m/s], e_s presión de vapor de saturación [kPa], e_a presión real de vapor [kPa], $e_s - e_a$ déficit de presión de vapor [kPa], Δ pendiente de la curva de presión de vapor [kPa/°C], γ constante psicrométrica [kPa/°C]

El método FAO Penman-Monteith fue desarrollado haciendo uso de la definición del cultivo de referencia como un cultivo hipotético con una altura asumida de 0.12 [m], con una resistencia superficial de 70 [s/m] y un albedo de 0.23 y que representa a la evapotranspiración de una superficie extensa de pasto verde de altura uniforme, creciendo activamente y adecuadamente regado. El método reduce las imprecisiones del método anterior de FAO Penman y produce globalmente valores más consistentes con datos reales de uso de agua de diversos cultivos (Allen et.al., 2006). El método de FAO Penman-Monteith para estimar ETo, puede ser derivado de la ecuación original de Penman-Monteith (Ecuación 2.2) y las ecuaciones de la resistencia aerodinámica (Ecuación 2.3) y superficial (Ecuación 2.4)

Ecuación original de Penman-Monteith

$$\lambda ET = \frac{\Delta(R_n - G) + \rho_a C_p \frac{e_s - e_a}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})}, \quad (2.2)$$

donde R_n es la radiación neta, G es el flujo del calor en el suelo, $(e_s - e_a)$ representa el déficit de presión de vapor del aire, ρ_a es la densidad media del aire a presión constante, C_p es el calor específico del aire, Δ representa la pendiente de la curva de presión de vapor de saturación, γ es la constante psicrométrica, y r_s y r_a son las resistencias superficial (total) y aerodinámica.

Ecuación de la resistencia aerodinámica r_a

$$r_a = \frac{\ln\left[\frac{z_m - d}{z_{om}}\right] \ln\left[\frac{z_h - d}{z_{oh}}\right]}{k^2 - U_z}, \quad (2.3)$$

donde r_a resistencia aerodinámica [s/m], z_m altura de medición del viento [m], z_h altura de medición de humedad [m], d plano de altura de desplazamiento cero [m], z_{om} longitud de la rugosidad que gobierna la transferencia del momentum [m], z_{oh} longitud de la rugosidad que gobierna la transferencia de calor y vapor de agua [m], k constante de Von Karman 0.41, u_z velocidad del viento a la altura z [m/s].

Ecuación de resistencia superficial r_s

$$r_s = \frac{r_l}{IAF_{activo}}, \quad (2.4)$$

donde r_s resistencia superficial (total) [s/m], r_l resistencia estomática total de una hoja bien iluminada [s/m], IAF_{activo} índice activo de área foliar [m² (área de la hoja) m² (superficie del suelo)].

2.3. Hargreaves-Samani

Cuando no se tiene disponibilidad de datos meteorológicos de radiación solar, humedad relativa o velocidad del viento, estos deberían ser estimados para utilizar el método de Penman-Monteith. No obstante, como una opción alternativa, la ET_o puede ser estimada usando la (Ecuación 2.5) de Hargreaves-Samani, según recomienda la FAO en (Allen, 1998):

$$ET_o = 0.0023(T_{mean} + 17.8) + (T_{max} - T_{min})^{0.5} R_a, \quad (2.5)$$

donde ET_o es la evapotranspiración de referencia [mm/día], T_{mean} representa la temperatura media [°C], T_{max} temperatura máxima [°C], T_{min} temperatura mínima [°C] y R_a radiación extraterrestre [$MJ/m^2/dia$].

Es importante mencionar que la ecuación de Hargreaves tiende a subestimar los valores de ET_o bajo condiciones de viento fuerte ($u_2 > 3m/s$) y a sobreestimar la ET_o bajo condiciones de elevada humedad relativa (Allen et.al., 2006).

2.4. Agricultura protegida

La Agricultura Protegida (AP) se configura como un sistema de producción meticulosamente diseñado con el objetivo primordial de proporcionar a las plantas las condiciones óptimas para su desarrollo, permitiéndoles expresar su máximo potencial productivo, según lo señalado por (Vargas-Canales et.al., 2018). Este enfoque, al mismo tiempo, facilita el manejo preciso de las condiciones ambientales críticas, como temperatura, humedad relativa, nutrición, agua y luz, así como la gestión de riesgos asociados a plagas y enfermedades, como lo destaca (Allende et.al., 2017).

La Agricultura Protegida (AP), como concepto amplio, engloba diversas técnicas de cultivo que buscan controlar total o parcialmente el microclima que rodea a las plantas, adaptándolo a las necesidades específicas de cada especie durante su periodo de crecimiento. En México, la AP ha evolucionado como una estrategia eficiente para mitigar las emisiones contaminantes contribuyentes al calentamiento global. Esta estrategia se alinea con las recomendaciones de la FAO sobre la intensificación sostenible de la producción agrícola, como indican (P. Pratt y Ortega, 2019). El propósito central de la AP en México se enfoca en incrementar y mantener la productividad en términos de cantidad, calidad y oportunidad comercial, como sugieren (Castañeda-Miranda et.al., 2007) y (Moreno Reséndez et.al., 2011).

Una alternativa a las cubiertas de plástico dentro de la AP es la implementación de mallas sombra, diseñadas para resguardar cultivos, hortalizas y áreas recreativas de las inclemencias solares. Estas mallas, fabricadas con materiales específicos como polietileno, polipropileno y poliéster, poseen propiedades que no solo protegen las plantas, sino que también contribuyen al ahorro de agua y energía. La selección de porcentajes de sombra, que oscilan desde un 35 % hasta un 95 %, se ajusta a las necesidades particulares de cada cultivo, proporcionando un ambiente controlado que evita cambios bruscos de temperatura y mantiene una humedad adecuada entre el suelo y la tela. Además, algunos modelos de mallas incorporan tratamientos aluminizados para optimizar aún más sus propiedades protectoras (Orona-Castillo et.al., 2022).

2.5. Programación Genética

2.5.1. Introducción

El objetivo de lograr que las computadoras resuelvan problemas automáticamente es central para la inteligencia artificial, el aprendizaje automático y el amplio campo abarcado por lo que Turing llamó «inteligencia de máquinas» (Turing, 1948). En su charla de 1983 titulada «IA: Dónde ha estado y hacia dónde se dirige», el pionero del aprendizaje automático Arthur Samuel declaró el objetivo principal de los campos del aprendizaje automático y la inteligencia artificial:

«El objetivo es ... lograr que las máquinas exhiban un comportamiento que, si fuera realizado por humanos, se asumiría que involucra el uso de inteligencia».

La Programación Genética es un método sistemático para hacer que las computadoras resuelvan automáticamente un problema comenzando desde una declaración de alto nivel de lo que necesita ser hecho. La Programación Genética es un método independiente del dominio que crea genéticamente una población de programas de computadora para resolver un problema. Específicamente, la Programación Genética transforma iterativamente una población de programas de computadora en una nueva generación de programas aplicando análogos de operaciones genéticas que ocurren naturalmente como se observa en la Figura 2.1.

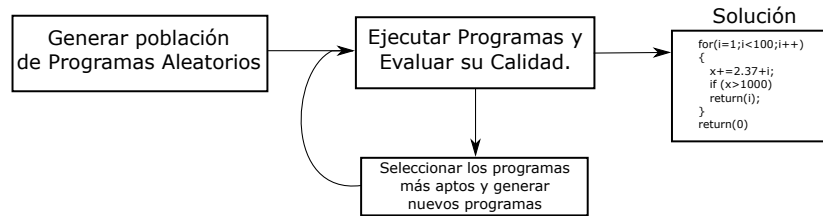


Figura 2.1: Ciclo principal de Programación Genética.

Las operaciones genéticas incluyen el cruce (recombinación sexual), la mutación, la reproducción, la duplicación y eliminación de genes. A veces se utilizan análogos de procesos de desarrollo para transformar un embrión en una estructura completamente desarrollada. La Programación Genética es una extensión del Algoritmo Genético (Holland, 1992) en el que las estructuras en la población no son cadenas de caracteres de longitud fija que codifican soluciones candidatas a un problema, sino programas que, al ejecutarse, son las soluciones candidatas al problema.

Los programas se expresan en Programación Genética en forma de árboles sintácticos en lugar de líneas de código. Por ejemplo, la expresión simple $\frac{x+x}{\sin(9.34*x)}$ se representa como se muestra en la Figura 2.2. El árbol incluye nodos y enlaces. Los nodos indican las instrucciones a ejecutar. Los enlaces indican los argumentos para cada instrucción. En lo siguiente, los nodos internos en un árbol se llamarán funciones, mientras que las hojas del árbol se llamarán terminales.

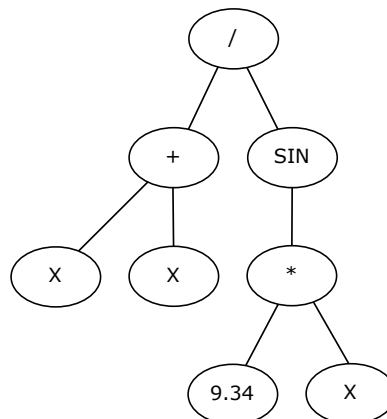


Figura 2.2: Representación básica de un programa en forma de árbol utilizada en Programación Genética.

En formas más avanzadas de Programación Genética, los programas pueden estar compuestos por múltiples componentes (por ejemplo, subrutinas). En este caso, la representación utilizada en la Programación Genética es un conjunto de árboles (uno para cada componente) agrupados bajo un nodo especial llamado raíz, como se ilustra en la Figura 2.3. Llamaremos a estos árboles (sub) ramas. El número y tipo de las ramas en un programa, junto con ciertas otras características de la estructura de las ramas, forman la arquitectura del programa.

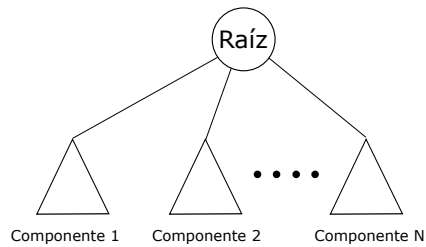


Figura 2.3: Representación de un programa multi-árbol.

2.5.2. Pasos preparatorios de la Programación Genética

La Programación Genética parte de una declaración de alto nivel de los requisitos de un problema e intenta producir un programa de computadora que resuelva el problema. El usuario comunica la declaración de alto nivel del problema al algoritmo de Programación Genética realizando ciertos pasos preparatorios bien definidos.

Los cinco principales pasos preparatorios para la versión básica de la Programación Genética requieren que el usuario especifique:

1. El conjunto de terminales (por ejemplo, las variables independientes del problema, funciones de cero argumentos y constantes aleatorias) para cada rama del programa a evolucionar.
2. El conjunto de funciones primitivas para cada rama del programa a optimizar.
3. La medida de aptitud, para medir explícita o implícitamente la calidad de los programas, es decir la aptitud de los individuos en la población).
4. Ciertos parámetros para controlar la ejecución.
5. El criterio de terminación y el método para designar el resultado de la ejecución.

Los dos primeros pasos preparatorios especifican los elementos disponibles para crear los programas de computadora. Una ejecución de Programación Genética es una búsqueda competitiva entre una población diversa de programas compuestos por las funciones y terminales disponibles.

La identificación del conjunto de funciones y terminales para un problema en particular suele ser un proceso sencillo. Para algunos problemas, el conjunto de funciones puede consistir simplemente de las funciones aritméticas (suma, resta, multiplicación y división). El conjunto de terminales puede consistir de las entradas externas del programa (variables independientes) y constantes numéricas. Para muchos otros problemas, los componentes abarcan tanto funciones como terminales especializados. Por ejemplo, si el objetivo es lograr que la Programación Genética programe automáticamente a un robot para trapear todo el piso de una habitación llena de obstáculos, el usuario humano debe decirle a la Programación Genética qué puede hacer el robot. Por ejemplo, el robot puede ser capaz de ejecutar funciones como moverse, girar y trapear. Si el objetivo es la creación automática de un controlador, el conjunto de funciones puede consistir en integradores, diferenciadores, adelantos, atrasos, ganancias, sumadores, restadores, etc., y el conjunto de terminales puede consistir en señales como la señal de referencia y la salida del sistema. Si el objetivo es la síntesis automática de un circuito eléctrico analógico, el conjunto de funciones puede permitir que la Programación Genética construya circuitos a partir de componentes como transistores, capacitores y resistores. Una vez que el usuario ha identificado los elementos primitivos para un problema de síntesis de circuitos, el mismo conjunto de funciones se puede utilizar para sintetizar automáticamente un amplificador, un circuito computacional, un filtro activo, un circuito de referencia de voltaje o cualquier otro circuito compuesto por estos elementos.

El tercer paso preparatorio se refiere a la medida de aptitud para el problema. La medida de aptitud especifica lo que debe hacerse. Es el mecanismo principal para comunicar los requisitos de alto nivel del problema al sistema de Programación Genética. Por ejemplo, si el objetivo es que la Programación Genética sintetice automáticamente un amplificador, la función de aptitud es el mecanismo para indicar a la Programación Genética que sintetice un circuito que amplifique una señal entrante (en lugar de, por ejemplo, un circuito que suprima las frecuencias bajas de una señal entrante o que calcule la raíz cuadrada de la señal entrante). Los dos primeros pasos preparatorios definen el espacio de búsqueda, mientras que la medida de aptitud específica implícitamente el objetivo deseado de la búsqueda.

El cuarto y quinto pasos preparatorios son administrativos. El cuarto paso preparatorio implica especificar los parámetros de control para la ejecución. El parámetro de control más importante es el tamaño de la población. Otros parámetros de control incluyen las probabilidades de realizar las operaciones genéticas, el tamaño máximo para los programas y otros detalles de la ejecución.

El quinto paso preparatorio consiste en especificar el criterio de terminación y el método para designar el resultado de la ejecución. El criterio de terminación puede incluir un número máximo de generaciones a ejecutar, así como un predicado de éxito específico del problema. El individuo mejor hasta el momento se recoge y se designa como el resultado de la ejecución.

2.5.3. Pasos ejecutivos de la Programación Genética

Después de que el usuario ha realizado los pasos preparatorios para un problema, se puede iniciar la ejecución de la Programación Genética. Una vez que se ha iniciado la ejecución, se llevan a cabo una serie de pasos bien definidos e independientes del problema.

La Programación Genética típicamente comienza con una población de programas de computadora generados aleatoriamente, compuestos por los elementos programáticos disponibles (funciones y terminales).

La Programación Genética transforma iterativamente una población de programas de computadora en una nueva generación de la población aplicando análogos de operaciones genéticas que ocurren naturalmente. Estas operaciones se aplican a individuos seleccionados de la población. Los individuos son seleccionados de forma probabilística para participar en las operaciones genéticas basadas en su aptitud (métrica proporcionada por el usuario en el tercer paso preparatorio). La transformación iterativa de la población se ejecuta dentro del ciclo generacional principal de la ejecución de la Programación Genética.

Los pasos ejecutivos de la Programación Genética son los siguientes:

1. Crear inicialmente una población aleatoria (generación 0) de programas individuales de computadora compuestos por las funciones y terminales disponibles.
2. Realizar iterativamente los siguientes sub-pasos (llamados generación) en la población hasta que se satisfaga el criterio de terminación:
 - a) Ejecutar cada programa en la población y determinar su aptitud (explícita o implícitamente) utilizando la medida de aptitud del problema.
 - b) Seleccionar uno o dos programas individuales de la población con una probabilidad basada en la aptitud (permitiendo la reelección) para participar en las operaciones genéticas en (c).
 - c) Crear nuevos programas individuales para la población aplicando las siguientes operaciones genéticas con probabilidades especificadas:
 - **Reproducción:** Copiar el programa individual seleccionado en la nueva población.
 - **Crossover:** Crear nuevos programas descendientes para la nueva población recombinando partes elegidas al azar de dos programas seleccionados.
 - **Mutación:** Crear un nuevo programa descendiente para la nueva población mutando al azar una parte elegida al azar de un programa seleccionado.
 - **Operaciones de alteración de arquitectura:** Elegir una operación de alteración de arquitectura del repertorio disponible de tales operaciones y crear un nuevo programa descendiente para la nueva población aplicando la operación de alteración de arquitectura elegida a un programa seleccionado.
3. Después de que se satisfaga el criterio de terminación, se recoge el mejor programa único en la población producido durante la ejecución (el mejor individuo hasta el

momento) y se designa como el resultado de la ejecución. Si la ejecución es exitosa, el resultado puede ser una solución (o solución aproximada) al problema.

La Figura 2.4 muestra un diagrama de flujo de la Programación Genética que presenta las operaciones genéticas de crossover, reproducción y mutación, así como las operaciones de alteración de arquitectura. Este diagrama de flujo expone una versión de dos descendientes de la operación de crossover.

A continuación, se presenta una explicación detallada del diagrama de flujo de Programación Genética 2.4. La ejecución del algoritmo se inicia con la fase de inicialización, donde se configuran los parámetros fundamentales: $Run=0$, que indica el número de ejecuciones realizadas, y $Gen=0$, que representa el número de generaciones ejecutadas. En esta etapa, se genera una población inicial de soluciones candidatas de manera aleatoria, estableciendo la base sobre la cual el algoritmo evolutivo comenzará su proceso iterativo de optimización. Luego, se evalúa si se cumple el criterio de terminación, que, si se satisface, finaliza la ejecución; de lo contrario, se procede a aplicar una medida de aptitud a cada individuo de la población. A continuación, se llevan a cabo diversas operaciones genéticas en cada individuo hasta alcanzar el tamaño total de la población ($i=M$), donde i representa el índice del individuo en la población y M el tamaño total de la población. Estas operaciones incluyen la selección de individuos basados en su aptitud, reproducción para copiar individuos a la nueva población, cruce para generar descendencia, mutación para introducir variabilidad y alteración de arquitectura para modificar la estructura de los individuos. Tras completar las operaciones genéticas, se incrementa la generación ($Gen=Gen+1$) y se repite el ciclo hasta cumplir con el criterio de terminación. Al finalizar cada ejecución, se designa el resultado, se incrementa Run , y se verifica si se ha alcanzado el número total de ejecuciones ($Run=N$). Si se cumple, el proceso concluye; de lo contrario, se reinicia el ciclo para una nueva ejecución.

La Programación Genética es independiente del problema en el sentido de que el diagrama de flujo que especifica la secuencia básica de pasos ejecutivos no se modifica para cada nueva ejecución o cada nuevo problema.

Normalmente no hay intervención humana discrecional o interacción durante una ejecución de Programación Genética (aunque un usuario humano puede ejercer juicio sobre si terminar una ejecución).

La Programación Genética comienza con una población inicial de programas de computadora compuestos por funciones y terminales apropiados para el problema. Los programas individuales en la población inicial suelen generarse recursivamente generando un árbol de programa con raíz etiquetada compuesto por selecciones aleatorias de funciones y terminales primitivos (proporcionados por el usuario como parte de los primeros y segundos pasos preparatorios). Los individuos iniciales suelen generarse sujetos a un tamaño máximo preestablecido (especificado por el usuario como un parámetro menor como parte del cuarto paso preparatorio). Por ejemplo, en el método de inicialización «Full», se toman nodos del conjunto de funciones hasta que se alcance una profundidad máxima del árbol. Más allá de esa profundidad, solo se pueden elegir terminales. La Figura 2.5 muestra varias instantáneas de este proceso. Una variante de esto, el método de inicialización «Grow»,

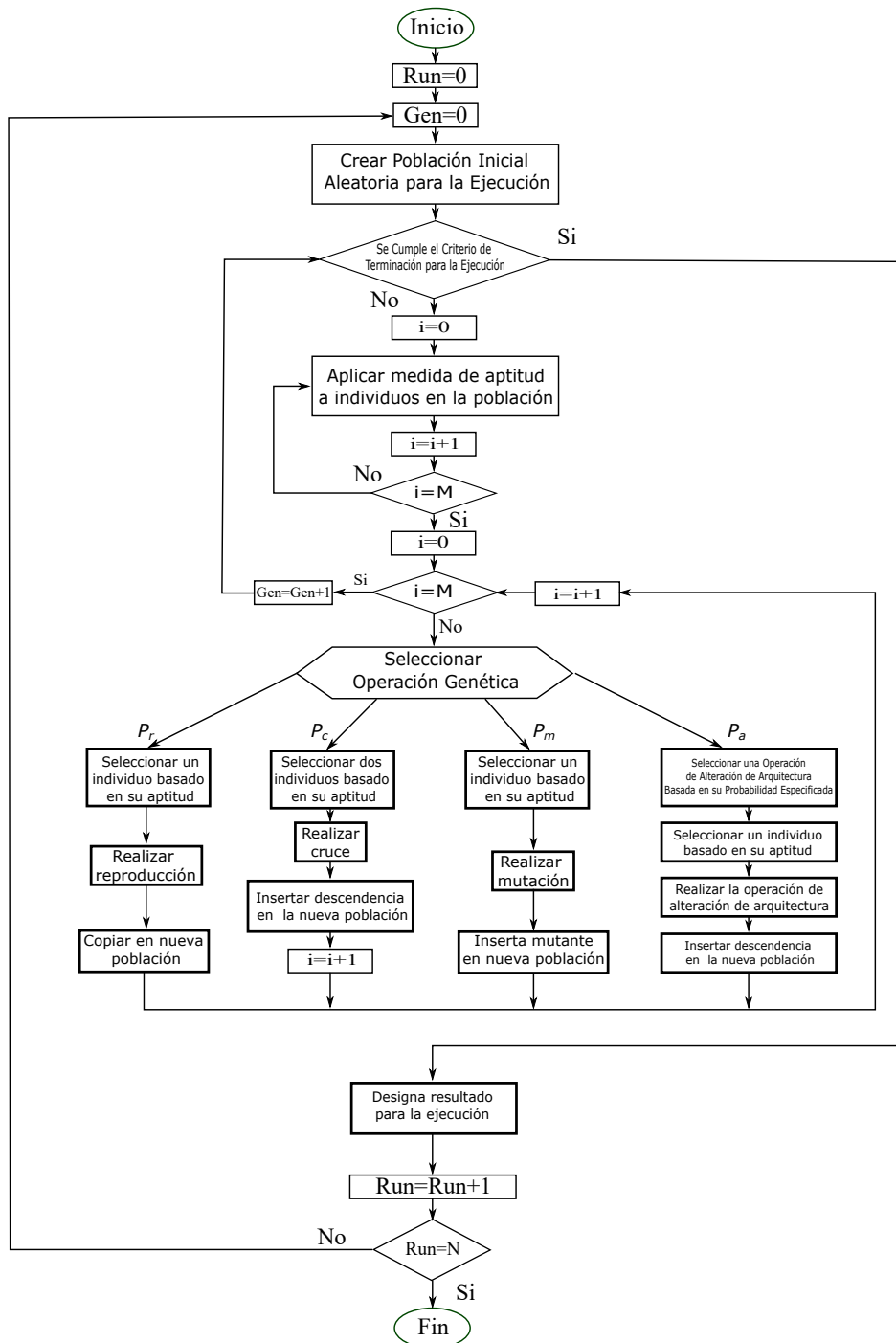


Figura 2.4: Diagrama de flujo de Programación Genética

permite la selección de nodos de todo el conjunto primitivo hasta que se alcance el límite de profundidad. Después de eso, se comporta como el método «Full». La Figura 2.6 ilustra este proceso. Se proporciona un pseudocódigo para una implementación recursiva de ambos métodos, «Full» y «Grow», en el algoritmo 1. El código asume que los programas están representados como expresiones en notación prefija.

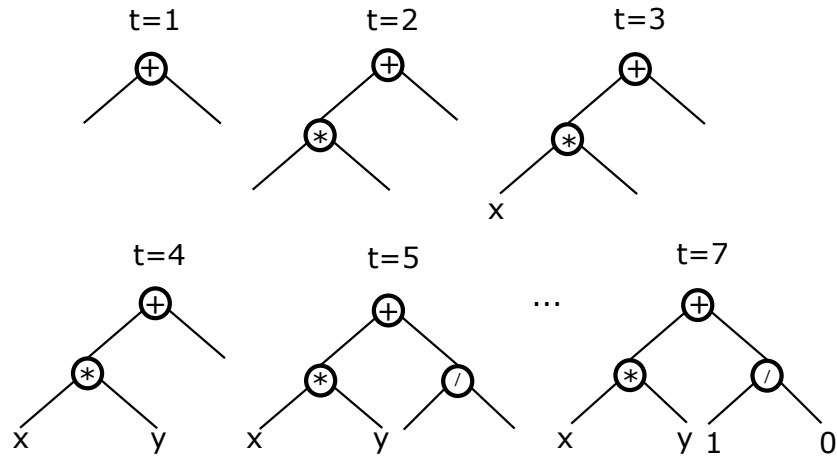


Figura 2.5: Creación de un árbol de siete puntos utilizando el método de inicialización «Full» (t=tiempo)

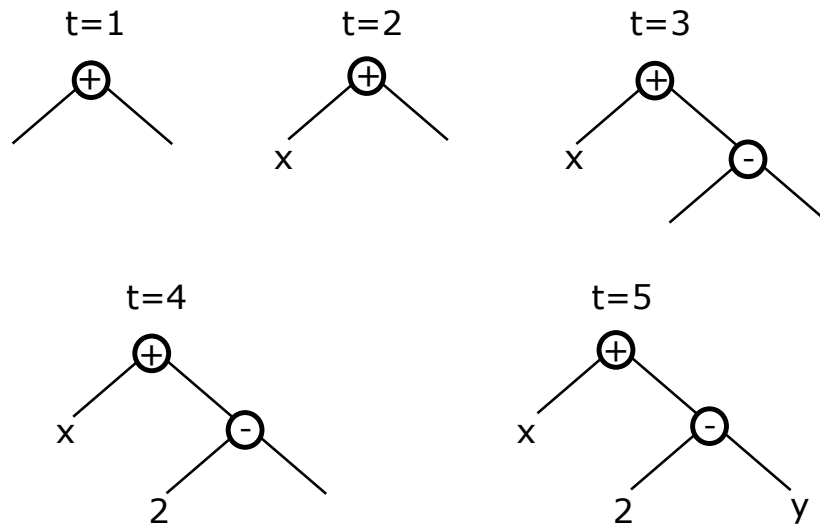


Figura 2.6: Creación de un árbol de cinco puntos utilizando el método de inicialización «Grow» (t=tiempo)

El Algoritmo 1 describe un procedimiento recursivo para generar programas utilizando los métodos «Full» y «Grow» en Programación Genética. Este procedimiento, GEN_RND_EXPR, comienza con la inicialización de parámetros como el conjunto de funciones (*func_set*), el conjunto de terminales (*term_set*), la profundidad máxima de las expresiones (*max_d*), y el método de generación. Si la profundidad máxima es 0 o, bajo el método «Grow», un dígito aleatorio es igual a 1, se selecciona aleatoriamente un terminal del conjunto de terminales. Si no se cumple esta condición, se elige una función del conjunto de funciones y se generan recursivamente los argumentos de esta función, reduciendo la profundidad máxima en uno en cada llamada recursiva. El resultado final es una expresión en notación prefija que integra la función y sus argumentos generados. El método «Full» asegura la generación de árboles completos hasta la profundidad máxima, mientras que el método «Grow» permite una mayor flexibilidad, generando árboles de diversas estructuras antes de alcanzar la profundidad máxima. Este algoritmo es esencial para explorar eficientemente el espacio de soluciones en Programación Genética, proporcionando una variedad de estructuras de árboles de expresión.

Algoritmo 1 Pseudocódigo para la generación recursiva de programas con los métodos «Full» y «Grow»

```

1: procedure GEN_RND_EXPR
2:   arguments:
3:     func_set,                               /* Conjunto de funciones */
4:     term_set,                               /* Conjunto de terminales */
5:     max_d,                                  /* Profundidad máxima para expresiones */
6:     method                                  /* «Full» o «Grow» */
7:   result:
8:     expr                                    /* Expresión en notación prefija */
9:   begin
10:  if max_d = 0 or method = «Grow» and random digit = 1 then
11:    expr = choose_random_element(term_set);
12:  else
13:    func = choose_random_element(func_set);
14:    for i = 1 to arity(func) do
15:      arg_i = GEN_RND_EXPR(func_set, term_set, max_d - 1, method)
16:    end for
17:    expr = (func, arg_1, arg_2, ...)
18:  end if
19:  end
20: end procedure

```

Cada programa individual en la población es medido o comparado en términos de qué tan bien realiza la tarea en cuestión (utilizando la medida de aptitud proporcionada en el tercer paso preparatorio). Para muchos problemas, esta medición produce un único valor numérico explícito, llamado aptitud. Normalmente, la evaluación de la aptitud requiere ejecutar los programas en la población, a menudo varias veces, dentro del sistema de Programación Genética. Existen una variedad de estrategias de ejecución, incluyendo (rela-

tivamente poco común) compilación y enlace fuera de línea o en línea y la (relativamente común) compilación e interpretación de código de máquina virtual.

Interpretar un árbol de programas significa ejecutar los nodos en el árbol en un orden que garantice que los nodos no se ejecuten antes de que se conozca el valor de sus argumentos (si los hay). Esto suele hacerse recorriendo el árbol de manera recursiva comenzando desde el nodo raíz y posponiendo la evaluación de cada nodo hasta que se conozca el valor de sus hijos (argumentos). Este proceso se ilustra en la Figura 2.7, donde los números a la derecha de los nodos internos representan los resultados de evaluar los subárboles con raíz en dichos nodos. En este ejemplo, la variable independiente x se evalúa como -1 . El algoritmo 2 proporciona una implementación del procedimiento de interpretación. El código asume que los programas se representan como expresiones en notación prefija y que dichas expresiones se pueden tratar como listas de componentes (donde se puede usar una construcción como $expr(i)$ para leer o establecer el componente i de la expresión $expr$).

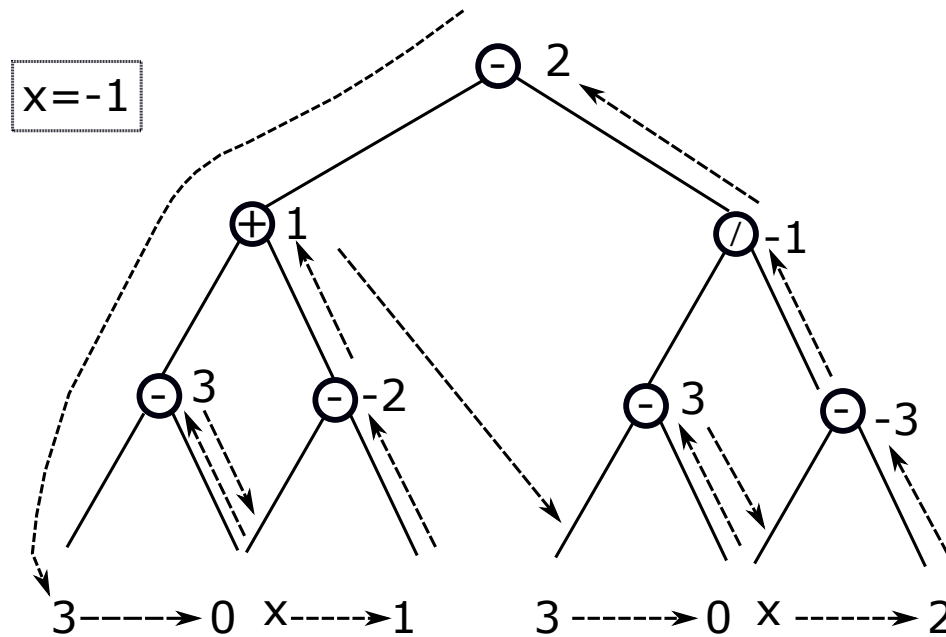


Figura 2.7: Ejemplo de interpretación de un árbol sintáctico (el terminal x es una variable que tiene un valor de -1)

El Algoritmo 2 presenta un intérprete para evaluar expresiones en notación prefija generadas por Programación Genética. Este procedimiento, denominado EVAL, comienza verificando si la expresión es una lista. Si lo es, se interpreta como una expresión no terminal que incluye una función y sus argumentos. Se evalúan recursivamente estos argumentos y se aplica la función correspondiente. Si la expresión no es una lista, se verifica si es una variable, una constante o una función de aridad 0. En este caso, se evalúa directamente o se toma su valor, respectivamente.

El algoritmo utiliza recursión para manejar nodos no terminales, donde cada función se aplica a los resultados de las evaluaciones recursivas de sus argumentos. Los nodos

terminales representan variables, constantes o funciones sin argumentos, evaluándose directamente según corresponda. Esta estructura asegura la evaluación precisa de expresiones complejas, esencial para interpretar programas generados por Programación Genética.

Algoritmo 2 Intérprete típico para Programación Genética

```
1: procedure EVAL
2:   arguments:
3:     expr                                     /* expresión en notación prefija */
4:   result:
5:     value                                     /* número */
6:   begin
7:   if expr is a list then                       /*No Terminal*/
8:     proc = expr(1);
9:     value = proc(EVAL(expr(2)),EVAL(expr(3)),...)
10:  else                                           /* Terminal */
11:    if expr is a variable or a constant then
12:      value = expr
13:    else                                           /* Función de aridad 0 */
14:      value = expr()
15:    end if
16:  end if
17: end
18: end procedure
```

Independientemente de la estrategia de ejecución adoptada, la aptitud de un programa puede medirse de muchas maneras diferentes, incluyendo, por ejemplo, en términos de la cantidad de error entre su salida y la salida deseada, la cantidad de tiempo (combustible, dinero, etc.) requerido para llevar un sistema a un estado objetivo deseado, la precisión del programa en el reconocimiento de patrones o la clasificación de objetos en clases, el rendimiento que produce un programa de juego, o la conformidad de una estructura compleja (como una antena, circuito o controlador) con criterios de diseño especificados por el usuario. La ejecución del programa a veces devuelve uno o más valores explícitos. Alternativamente, la ejecución de un programa puede consistir solo en efectos secundarios en el estado de un mundo (por ejemplo, las acciones de un robot). También es posible que la ejecución de un programa produzca tanto valores de retorno como efectos secundarios. La medida de aptitud es, para muchos problemas prácticos, multiobjetivo en el sentido de que combina dos o más elementos diferentes. En la práctica, los diferentes elementos de la medida de aptitud compiten entre sí hasta cierto grado. Para muchos problemas, cada programa en la población se ejecuta sobre una muestra representativa de diferentes casos de aptitud. Estos casos de aptitud pueden representar diferentes valores de la(s) entrada(s) del programa, diferentes condiciones iniciales de un sistema o diferentes entornos. A veces, los casos de aptitud se construyen de manera probabilística.

La creación de la población inicial aleatoria es, en efecto, una búsqueda aleatoria ciega

del espacio de búsqueda del problema. Proporciona una línea base para juzgar los esfuerzos de búsqueda futuros. Típicamente, los programas individuales en la generación 0 tienen una aptitud extremadamente pobre. No obstante, algunos individuos en la población son (generalmente) más aptos que otros. Las diferencias en la aptitud son luego explotadas por la Programación Genética. La Programación Genética aplica la selección darwiniana y las operaciones genéticas para crear una nueva población de programas descendientes a partir de la población actual. Las operaciones genéticas incluyen el cruce (recombinación sexual), la mutación, la reproducción y las operaciones de alteración de la arquitectura. Dado que se tienen copias de dos árboles parentales, típicamente, el cruce implica seleccionar aleatoriamente un punto de cruce (que puede pensarse equivalente a un nodo o un enlace entre nodos) en cada árbol parental y intercambiar los subárboles con raíces en los puntos de cruce, como se ejemplifica en la Figura 2.8. A menudo, los puntos de cruce no se seleccionan con probabilidad uniforme. Una estrategia frecuente es, por ejemplo, seleccionar nodos internos (funciones) el 90% de las veces, y cualquier nodo para el 10% restante. La mutación tradicional consiste en seleccionar aleatoriamente un punto de mutación en un árbol y sustituir el subárbol con raíz allí con un subárbol generado aleatoriamente, como se ilustra en la Figura 2.9. A veces, la mutación se implementa como un cruce entre un programa y un programa aleatorio recién generado (esto también se conoce como cruce «pollo sin cabeza»). La reproducción implica simplemente copiar ciertos individuos en la nueva población.

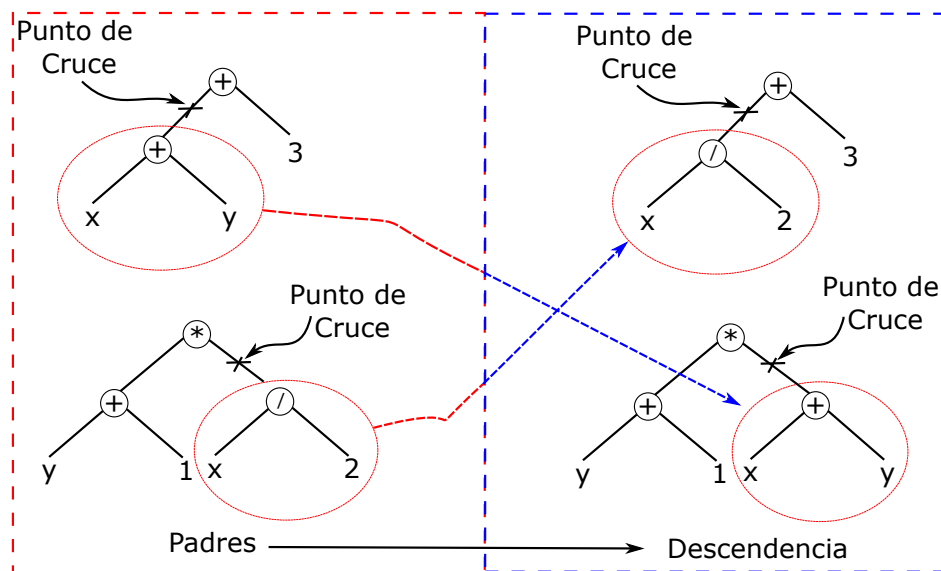


Figura 2.8: Ejemplo de cruce de dos hijos entre árboles sintácticos.

Las operaciones genéticas descritas anteriormente se aplican a individuos que son seleccionados de manera probabilística de la población en función de su aptitud. En este proceso de selección probabilística, se favorece a los individuos mejores sobre los inferiores. Sin embargo, el mejor individuo de la población no necesariamente es seleccionado y el peor individuo de la población no necesariamente es descartado. Después de que las operaciones genéticas se llevan a cabo en la población actual, la población de descendencia (es decir, la nueva generación) reemplaza a la población actual (es decir, la generación

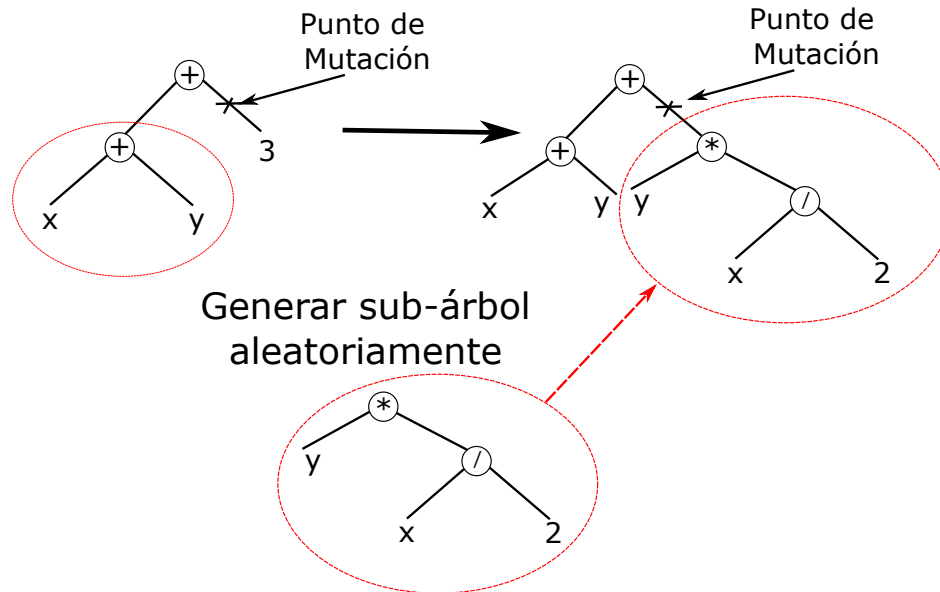


Figura 2.9: Ejemplo de mutación de sub-árbol

ahora vieja). Este proceso iterativo de medir la aptitud y realizar las operaciones genéticas se repite durante muchas generaciones. La ejecución de la Programación Genética termina cuando se satisface el criterio de terminación (según lo proporcionado por el quinto paso preparatorio). El resultado de la ejecución se especifica mediante el método de designación de resultados. El mejor individuo encontrado durante la ejecución (es decir, el mejor individuo hasta el momento) generalmente se designa como el resultado de la ejecución. Todos los programas en la población aleatoria inicial (generación 0) de una ejecución de Programación Genética son programas válidos sintácticamente y ejecutables. Las operaciones genéticas que se realizan durante la ejecución (es decir, cruce, mutación, reproducción y las operaciones de alteración de la arquitectura) están diseñadas para producir descendencia que sean programas válidos sintácticamente y ejecutables. Por lo tanto, cada individuo creado durante una ejecución de Programación Genética (incluido, en particular, el mejor individuo de la ejecución) es un programa válido sintácticamente y ejecutable. Existen numerosas implementaciones alternativas de la Programación Genética que varían de la descripción anterior.

2.6. MATLAB y GPLab

2.6.1. Introducción

MATLAB es un lenguaje de programación de alto nivel que permite realizar cálculos, visualizar resultados y desarrollar algoritmos utilizando notación matemática. El nombre MATLAB proviene de *Matrix Laboratory*. Inicialmente, se creó como una interfaz para las librerías de rutinas de Fortran como *EISPACK* y *LINPACK*, que en ese momento representaban el estado del arte para resolver problemas de álgebra matricial.

En MATLAB, los datos básicos se manejan como arreglos, como vectores y matrices, sin necesidad de declarar variables ni solicitar memoria. Esta característica clave permite manipular vectores o matrices como si fueran variables simples. Actualmente, MATLAB no se limita únicamente a resolver problemas numéricos, sino que ofrece una amplia gama de herramientas que permiten la integración con otros programas, la adquisición de datos, el control en tiempo real, el procesamiento simbólico y más.

El entorno de trabajo de MATLAB, conocido como *Command Window*, es donde se ingresan los comandos y se obtienen los resultados. Aunque tiene pocas palabras clave, la mayoría de las instrucciones se realizan a través de rutinas llamadas funciones, que se agrupan en conjuntos o librerías llamadas toolboxes según las operaciones que realizan. Algunos ejemplos comunes de estas toolboxes son: *Control Systems*, *GPLab*, *Symbolic*, *Signals Processing*, entre otros. Al utilizar MATLAB, es crucial comprender cómo se introducen y manipulan las matrices, cómo se utilizan las funciones en general y cómo obtener información específica de cada función. (Bianchi, 2001).

GPLab es una suite de herramientas de Programación Genética diseñada para MATLAB. Destaca por su versatilidad, adaptabilidad a diferentes usuarios y facilidad de expansión. Ha sido probada en diversas versiones de MATLAB y plataformas informáticas, sin requerir herramientas adicionales. Su arquitectura se basa en una estructura modular y parametrizada (Figura 2.10), permitiendo que usuarios de distintos niveles de experiencia la utilicen de manera efectiva (Silva y Almeida, 2003).

2.7. Conclusiones

En esta sección se enfatizan la importancia de integrar la Programación Genética y la agricultura protegida para mejorar la productividad y sostenibilidad agrícola. Se reconoce la agricultura protegida como una práctica crucial para controlar las condiciones ambientales y optimizar el crecimiento de cultivos. La gestión eficaz de la evapotranspiración se identifica como fundamental para el éxito de esta práctica. Por otro lado, la Programación Genética emerge como una herramienta prometedora para desarrollar modelos y algoritmos que estimen y predigan la evapotranspiración en entornos de agricultura protegida. Se destaca su capacidad para generar soluciones óptimas y su potencial para mejorar la eficiencia y sostenibilidad de la producción agrícola en invernaderos y otros sistemas protegidos. En conjunto, se destaca la importancia de un enfoque integrado y multidisciplinario para abordar los desafíos agrícolas contemporáneos y avanzar hacia sistemas agrícolas más eficientes y sostenibles.

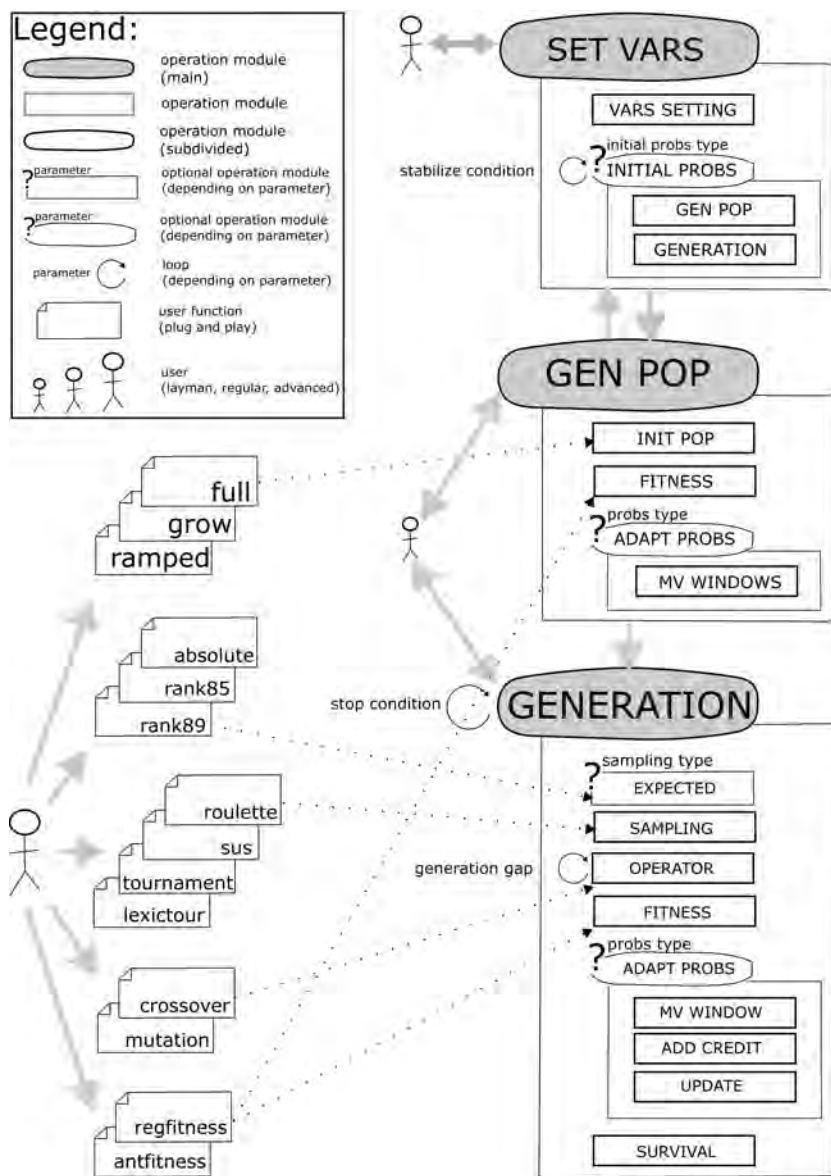


Figura 2.10: Estructura operativa de la caja de herramientas GPLAB

Capítulo 3

Estado del arte

El objetivo principal de este capítulo es analizar exhaustivamente las investigaciones previas que se han centrado en el fenómeno de la evapotranspiración, haciendo especial hincapié en el uso de métodos de cómputo suave. En este contexto, nuestro propósito es contextualizar de manera adecuada nuestro trabajo de investigación, identificando tanto las fortalezas como las debilidades de los métodos previamente analizados. Así mismo, buscamos resaltar de manera clara y precisa nuestra contribución en relación con las propuestas existentes en la literatura. A continuación, se presentan y discuten detalladamente las diferentes técnicas de inteligencia artificial que han sido objeto de nuestro estudio.

Las técnicas mencionadas no solo se mencionan como puntos de comparación, sino que también son algunas de las más utilizadas en el modelado de la evapotranspiración. Estas técnicas han demostrado ser efectivas para manejar la complejidad de los datos climáticos y ambientales asociados con la evapotranspiración, ofreciendo capacidades que van desde la captura de relaciones no lineales hasta la interpretación de cómo diferentes variables influyen en el proceso de evapotranspiración.

3.1. Aplicación de algoritmos de ensamblaje para modelar la evapotranspiración

En el estudio de (Fan et.al., 2018), se evaluaron cuatro algoritmos de aprendizaje automático basados en árboles (RF, M5Tree, GBDT y XGBoost) para estimar la evapotranspiración utilizando datos meteorológicos completos e incompletos de diferentes climas de China, comparándolos con la Máquina de Soporte Vectorial SVM y la Máquina de Aprendizaje Extremo ELM. Los resultados mostraron que la radiación solar es crucial para la estimación de la evapotranspiración en zonas monzónicas tropicales y subtropicales, mientras que la humedad relativa y la velocidad del viento son importantes en otras zonas climáticas. Los modelos XGBoost y GBDT mostraron una precisión y estabilidad comparables a SVM y ELM, pero con menores costos computacionales, destacándose como alternativas prometedoras para predecir la evapotranspiración diaria con datos climáticos limitados. Cada modelo tiene sus propias limitaciones, lo que resalta la importancia de seleccionar cuidadosamente el

modelo adecuado para el problema específico y las características de los datos disponibles.

En la investigación de (Huang et.al., 2019), se evaluó el potencial del algoritmo CatBoost para estimar la evapotranspiración diaria en zonas subtropicales húmedas del sur de China, comparándolo con RF y SVM utilizando datos meteorológicos de cinco estaciones durante 2001-2015. Los tres algoritmos lograron una precisión satisfactoria incluso con datos incompletos, destacando SVM por su precisión y estabilidad en combinaciones incompletas, y CatBoost con combinaciones completas. Ambos modelos mostraron valores de R^2 superiores a 0.80. Concluyeron que CatBoost tiene gran potencial para la estimación precisa de evapotranspiración en regiones húmedas a nivel global. Sin embargo, señalaron la necesidad de controlar el sobreajuste y seleccionar cuidadosamente las características de entrada para evitar capturar ruido en los datos.

Con el objetivo de evaluar diferentes modelos de evapotranspiración en la región subtropical húmeda de China, (Fan et.al., 2019) compararon cuatro modelos empíricos (Hargreaves-Samani, Tabari, Makkink y Trabert) y de computación suave (M5Tree, RF y LightGBM) usando datos meteorológicos de 49 estaciones durante 2001-2015. Los resultados mostraron que el modelo LightGBM superó a los modelos empíricos y a M5Tree y RF en términos de precisión en aplicaciones locales, con un RMSE promedio de 0.08 - 0.58 mm/día, sugiriendo su eficacia para estimar la evapotranspiración en regiones con escasez de datos meteorológicos. Sin embargo, su dependencia de un extenso conjunto de datos históricos podría limitar su aplicabilidad en áreas con información insuficiente.

La investigación de (Han et.al., 2019) presentó el modelo BAXGB, que combina un algoritmo de murciélago con XGBoost para estimar la evapotranspiración de referencia (ET_o) en entornos áridos y semiáridos de China. Compararon los resultados de BAXGB con el modelo FAO56-PM usando datos de tres estaciones meteorológicas: Datong, Yinchuan y Taiyuan. Desarrollaron seis modelos para cada algoritmo con distintas combinaciones de parámetros meteorológicos, incluidos temperatura, velocidad del viento, humedad relativa y radiación solar. Los resultados indicaron que BAXGB superó a los modelos MARS y GPR en la mayoría de los casos. La precisión del modelo dependió de los parámetros meteorológicos, subrayando la necesidad de considerar diversas combinaciones de estos para mejorar la gestión del agua en la agricultura. Una posible limitación del modelo es su dependencia de datos precisos de entrada, como la radiación solar y la humedad relativa.

Un estudio realizado por (Ge et.al., 2022) utilizó datos experimentales de tres años (2019-2021) para modelar los efectos de ocho factores meteorológicos en la evapotranspiración de cultivos de tomate con riego por goteo en invernadero. Los parámetros meteorológicos considerados fueron la radiación solar neta (R_n), la temperatura media (T_a), la temperatura mínima (T_{min}), la temperatura máxima (T_{max}), la humedad relativa (RH), la humedad relativa mínima (RH_{min}), la humedad relativa máxima (RH_{max}) y la velocidad del viento (V). El modelo desarrollado, XGBR-ET, se basó en la regresión XGBoost (XGBR). Este modelo se comparó con otros siete modelos de regresión comunes: regresión lineal (LR), regresión de vectores de soporte (SVR), regresión de K vecinos (KNN), regresión de bosque aleatorio (RFR), regresión AdaBoost (ABR), regresión bagging (BR) y regresión de gradiente boosting

(GBR). Los resultados mostraron que R_n , T_{avg} y T_{max} se correlacionaron positivamente con la evapotranspiración, mientras que T_{min} , RH , RH_{min} , RH_{max} y V se correlacionaron negativamente. R_n tuvo la mayor correlación con ET ($r = 0,89$), y V tuvo la menor correlación ($r = 0,43$).

3.2. Diversas técnicas de computación suave para modelar la evapotranspiración

En el estudio realizado por (Kim y Kim, 2008), se presentaron tres variantes del modelo de algoritmos genéticos (GA) incorporadas en el modelo generalizado de redes neuronales de regresión (GRNNM): COMBINE-GRNNM-GA, EXTREME-GRNNM-GA y Average-GRNNM-GA. Estos modelos fueron utilizados para estimar la evaporación de bandeja y la evapotranspiración de referencia en alfalfa durante el período 1985-1992 en 14 estaciones meteorológicas distribuidas en la República de Corea. Los análisis estadísticos realizados indicaron que COMBINE-GRNNM-GA destacó como el modelo más efectivo entre ellos.

En el estudio realizado por Kisi y colaboradores (Kisi, 2010) utilizaron algoritmos genéticos para calibrar la función de membresía difusa para modelar la evapotranspiración diaria en tres estaciones ubicadas en el centro de California, EE.UU., durante el período 1998-2007. Se empleó un modelo genético difuso (FG) para estimar la evapotranspiración utilizando el método FAO56-PM, y su rendimiento fue comparado con los métodos de Penman, Hargreaves, Ritchie, Turc y Redes Neuronales Artificiales (ANN). Los resultados mostraron que el modelo FG, demostró un mejor rendimiento en la estimación de la evapotranspiración en comparación con otros modelos empíricos.

En otro estudio significativo realizado por (Eslamian et.al., 2012), desarrollaron modelos híbridos de Redes Neuronales Artificiales y algoritmos genéticos (ANN-GA), así como modelos de ANN independientes, para estimar la evapotranspiración. Los resultados indicaron que los modelos ANN-GA lograron una mayor precisión en el cálculo de la evapotranspiración en comparación con otros modelos.

El presente estudio propuso dos modelos de inteligencia artificial, bosques aleatorios (RF) y redes neuronales de regresión generalizada (GRNN), para estimar la evapotranspiración diaria. Los datos meteorológicos utilizados fueron la temperatura del aire máxima/mínima, radiación solar, humedad relativa y velocidad del viento durante 2009-2014 de dos estaciones en el suroeste de China. Los resultados indicaron que los modelos RF y GRNN locales y externos funcionaron bien para estimar la evapotranspiración diaria. Sin embargo, RF en general fue ligeramente mejor que GRNN según lo reportado por (Feng et.al., 2017).

La investigación de (Antonopoulos y Antonopoulos, 2017) estudió la estimación de la evapotranspiración en Grecia usando métodos empíricos y Redes Neuronales Artificiales (ANN) con datos climáticos diarios de cinco años. Compararon varios métodos, incluyendo Hargreaves, Priestley-Taylor, Makkink, y Penman-Monteith. Los métodos empíricos mostraron buena correlación con Penman-Monteith pero tendieron a sobreestimar o subestimar

la ETo. Las Redes Neuronales Artificiales demostraron ser prometedoras para predecir la ETo diaria, especialmente con variables como temperatura y radiación solar. Se concluye que las redes neuronales pueden mejorar la precisión de las estimaciones de la evapotranspiración en comparación con los métodos empíricos, aunque requieren una cuidadosa selección de datos de entrenamiento y arquitectura para evitar problemas de sobreajuste o subajuste, y demandan recursos computacionales significativos.

En este estudio se llevó a cabo una investigación que utilizó los modelos splines de regresión adaptativa multivariante (MARS) y programación de expresión génica (GEP) para determinar la evapotranspiración en seis estaciones climáticas diferentes en Irán: Isfahán y Shiraz (áreas áridas), Urmia y Tabriz (semiáridas), y Yazd y Zahedan (hiperáridas), durante el período 2000-2014. El estudio utilizó datos meteorológicos diarios y sus retardos para la modelar la evapotranspiración. La innovación de este estudio, fue proponer nuevos modelos híbridos basados en datos rezagados de la evapotranspiración a través de la combinación de los modelos MARS y GEP con el modelo de series de tiempo de heterocedasticidad condicional autoregresiva (ARCH). Los investigadores concluyeron que los modelos MARS-ARCH y GEP-ARCH mostraron resultados superiores en comparación con los modelos MARS y GEP de forma independiente (Mehdizadeh, 2018)

Algunos expertos han explorado el uso de algoritmos genéticos para mejorar la precisión de modelos de estimación, como se ilustra en el estudio de Srivastava y colaboradores (Srivastava et al., 2018). En este trabajo, se compararon dos métodos para estimar la evapotranspiración en una región de clima monzónico tropical: el método de Hargreaves-Samani y el modelo VIC-3L, utilizando el estándar de referencia FAO-56-PM. Los parámetros climáticos utilizados incluyeron temperatura máxima y mínima, velocidad del viento, humedad relativa y radiación solar. Ambos métodos utilizaron estos datos meteorológicos para calcular la evapotranspiración. Se observó que las estimaciones de la evapotranspiración basadas en el método de Hargreaves-Samani tendieron a sobreestimarse significativamente, pero mejoraron notablemente después de la estandarización mediante algoritmos genéticos. Por otro lado, el modelo VIC-3L también demostró producir estimaciones precisas de ET.

El estudio de (Ferreira et al., 2019) evaluó la evapotranspiración en Brasil utilizando datos meteorológicos como temperatura y humedad relativa. Compararon varios modelos empíricos (HS, Oudin, Hamon, Valiantzas, Romanenko, Schendel) y modelos de aprendizaje automático como Redes Neuronales Artificiales (ANN) y máquinas de vectores de soporte (SVM). Emplearon enfoques adicionales como el agrupamiento de estaciones meteorológicas similares y el uso de datos de días previos como entrada para los modelos. Los resultados mostraron que incluir la humedad relativa mejoró la generalización y el rendimiento de los modelos. Sin embargo, la opacidad de ANN y SVM como "cajas negras" puede dificultar la comprensión y confianza en sus resultados, así como la identificación de sesgos o errores en los datos de entrenamiento.

El estudio de Balmat (Balmat et al., 2019) tuvo como objetivo desarrollar un nuevo modelo de evapotranspiración para un cultivo en invernadero utilizando un Sistema de Inferencia Difusa Basado en Redes Adaptativas ANFIS. Para el aprendizaje del modelo se utilizaron datos de temperatura y humedad así como, la ETo diaria calculada con el

modelo de referencia FAO56-PM de un invernadero experimental ubicado en la Université de Toulon en el sur de Francia. El desempeño del modelo fue evaluado comparándolo con el modelo de referencia FAO56-PM durante un período de prueba. Los resultados mostraron una eficacia significativa del método propuesto con una $R^2=0.86$, mostrando una buena precisión para programar el riego. Sin embargo, los modelos ANFIS presentan complejidad computacional, lo que puede resultar en un mayor costo computacional durante el entrenamiento y la inferencia, limitando su aplicación en sistemas con recursos limitados. Además, la selección óptima de la estructura y los parámetros del ANFIS puede requerir una exploración y ajuste meticolosos, lo que puede ser una tarea compleja.

La investigación realizada por Sanikhani y colaboradores (Sanikhani et.al., 2019), tuvo como objetivo evaluar la evapotranspiración de referencia ETo utilizando diferentes modelos de inteligencia artificial IA y métodos empíricos, como el Hargreaves-Samani (HS) y su versión calibrada (CHS), utilizando datos climáticos de dos estaciones en Turquía. Los modelos de IA (MLP, GRNN, RBNN, ANFIS-GP, ANFIS-SC y GEP), mostraron un mejor rendimiento en términos de precisión en la estimación de la ETo en comparación con el HS y CHS. La calibración mejoró significativamente la precisión del HS. Sin embargo, la limitación principal del estudio fue el uso de solo dos estaciones de la misma región climática. Por otro lado, los modelos de IA requieren conjuntos de datos completos y precisos para entrenar y producir predicciones confiables. Si los datos disponibles son limitados, incompletos o sesgados, los modelos de IA pueden tener dificultades para generar resultados precisos. Además, la selección adecuada de variables de entrada y la configuración óptima de los parámetros del modelo también son factores críticos que pueden afectar su rendimiento.

En su estudio, (Wu y Fan, 2019) investigaron los efectos de la temperatura y la precipitación en la precisión de la evapotranspiración diaria en diversas zonas climáticas de China. Los investigadores evaluaron ocho modelos de aprendizaje automático en cuatro categorías, es decir modelos basados en neuronas (MLP, GRNN y ANFIS), basados en kernel (SVM y KNEA), basados en árboles (M5Tree y XGBoost) y basados en curvas (MARS). Utilizaron datos de temperatura máxima/máxima y precipitación durante 2001 - 2015 de 14 estaciones meteorológicas. Los resultados evidenciaron que la precisión de la evapotranspiración aumentaba al incorporar precipitación, especialmente en regiones húmedas. Los modelos basados en kernel y curvas fueron los más precisos, destacando KNEA en regiones áridas y SVM en diversos climas. SVM, MLP, MARS y KNEA mostraron estabilidad en sus predicciones. Sin embargo, la dependencia de la calidad de los datos meteorológicos y la complejidad de los modelos son limitaciones significativas.

En este estudio, los investigadores (Yu et.al., 2020) propusieron un nuevo modelo de estimación de la evapotranspiración llamado PSO-XGBoost, que utiliza XGBoost (eXtreme Gradient Boosting) como modelo principal de regresión y emplea el algoritmo de Optimización por Enjambre de Partículas (PSO) para optimizar los parámetros de XGBoost. Utilizaron datos meteorológicos y de humedad del suelo de un invernadero solar en Beijing, China, el modelo se comparó con otros modelos de aprendizaje integrado como CatBoost,

Bagging, XGBoost, AdaBoost y Random Forest, así como con modelos clásicos de aprendizaje automático como Redes Neuronales Artificiales (RNN), Árboles de Decisión (DT) y K-Nearest Neighbor (KNN). Los resultados demostraron que el algoritmo PSO optimizó de manera estable los parámetros del modelo XGBoost, permitiendo al modelo PSO-XGBoost estimar con precisión la evapotranspiración. El modelo PSO-XGBoost mostró la mejor capacidad de ajuste y generalización en comparación con los otros modelos evaluados.

En el estudio realizado por (Jung et.al., 2022) desarrollaron modelos de aprendizaje profundo basados en datos para estimar la evapotranspiración y la humedad en invernaderos de tomates. Utilizando datos de series temporales y modelado de memoria a corto plazo (LSTM), los investigadores diseñaron y validaron el modelo en comparación con el enfoque de Stanghellini. Los parámetros meteorológicos de entrada incluyeron Text, HRext, Rs, u_2 , Tsuelo, Tint, y Thoja. Los resultados indican que el modelo LSTM logra una estimación precisa de la evapotranspiración en relación con el modelo de referencia.

En el estudio de (Yuan et.al., 2023), se empleó una red neuronal artificial (ANN) para estimar la evapotranspiración utilizando parámetros climáticos como temperatura, humedad relativa y radiación solar. Los datos utilizados abarcan desde octubre de 2016 hasta noviembre de 2017, obtenidos de un invernadero experimental. La evapotranspiración de referencia se calculó mediante un sistema de microlisímetros. Los resultados indican que los modelos de ANN logran estimaciones precisas de la evapotranspiración.

3.3. Implementación de técnicas de cómputo evolutivo para modelar la evapotranspiración

Se han realizado numerosos estudios utilizando enfoques de cómputo evolutivo en diversos escenarios climatológicos. Por ejemplo, Parasuraman y colaboradores, investigaron el efecto de diferentes condiciones climáticas y topográficas. En su estudio, estimaron la ETo utilizando Programación Genética (GP) a partir de variables como la temperatura del suelo, la velocidad del viento, el calor latente medido por covarianza de remolinos, la radiación neta, la humedad relativa y la temperatura del aire. Al comparar la predictibilidad del modelo de GP con la formulación empírica clásica de Redes Neuronales Artificiales (ANN) y el modelo de Penman-Monteith (PM), los investigadores encontraron que la Programación Genética ofrecía un potencial significativo como modelo predictivo (Parasuraman et.al., 2007).

El efecto de la evapotranspiración de referencia fue examinada por Guven en 2008, utilizaron datos atmosféricos diarios obtenidos del Sistema de Información de Gestión de Riego de California ((CIMIS)[®], 2022) de múltiples estaciones meteorológicas: Davis, Hasting, Suisun, Dixon y Oakville. En su estudio, los autores evaluaron la eficacia del modelo de Programación Genética (GP) comparándolo con varios modelos empíricos, como el Penman-Monteith de la FAO (FAO56-PM), el Jensen-Haise, la ecuación Hargreaves-Samani, Jones-Ritchie, el método Turc y modelos basados en radiación solar. Los resultados

indicaron que el modelo de GP superó significativamente a los modelos empíricos en términos de precisión, presentando errores relativamente bajos y una alta correlación con los datos observados (Güven et al., 2008).

Por otro lado Kisi y Güven (Kisi y Güven, 2010) exploraron la precisión de la Programación Genética lineal (LGP) y la compararon con la programación de expresión génica (GEP) para modelar la evapotranspiración de referencia (ET_o) en tres estaciones, demostrando que LGP supera en rendimiento a GEP.

La investigación realizada por (El-Baroudy et al., 2010) presenta dos casos de estudios en South West Sand Storage y South Bison Hill, Mildred Lake mine, Canadá, donde utilizaron tres modelos inteligentes para estimar la evapotranspiración: regresión polinómica evolutiva (EPR), Redes Neuronales Artificiales (ANN) y programación de expresión génica (GEP). Los resultados destacaron que el modelo EPR superó a los otros en la precisión de la estimación de evapotranspiración.

En este estudio, los investigadores (Izadifar y Elshorbagy, 2010) desarrollaron y compararon dos modelos basados en datos: Programación Genética (GP) y Redes Neuronales Artificiales (ANNs), junto con modelos de regresión estadística, para estimar la evapotranspiración medida por covarianza de remolinos (AET) en intervalos horarios utilizando variables meteorológicas. Además, los investigadores evaluaron el desempeño de estos modelos frente al modelo físico HYDRUS-1D en Alberta, Canadá. Los resultados señalaron que los modelos de regresión estadística y GP superaron al modelo ANN en la precisión de la estimación de la evapotranspiración.

El modelo de programación de expresión génica (GEP) fue aplicado por (Shiri et al., 2012) para estimar la evapotranspiración de referencia en el País Vasco, ubicado en el norte de España. Los modelos se desarrollaron utilizando parámetros meteorológicos como la humedad relativa, la radiación solar, la temperatura del aire y la velocidad del viento. La capacidad del modelo GEP se validó en comparación con otros modelos de inteligencia artificial y empíricos. En general, los resultados demostraron que el modelo GEP logró una mayor precisión que los modelos Priestley-Taylor (PT), Hargreaves-Samani (HS) y el sistema de inferencia neuro-difuso adaptativo (ANFIS).

En el estudio realizado por (Traore y Güven, 2012) examinaron la capacidad de generalización del modelo de programación de expresión génica (GEP) para modelar la evapotranspiración en Burkina Faso, un país de la región del Sahel. Los autores informaron que la incorporación de múltiples variables meteorológicas mejoró significativamente la precisión del modelo GEP.

El objetivo principal del estudio realizado por (de la Paix Mupenzi et al., 2012) fue evaluar las pérdidas de agua por evapotranspiración, evaporación y filtración en zonas áridas. Para modelar las pérdidas, desarrollaron una combinación de Programación Genética (GP) y el modelo Penman-Monteith. La investigación demostró que la combinación de los modelos de programación de expresión génica (GEP) y FAO56-PM proporciona una estimación más precisa de la evaporación y la filtración en la Cuenca de Datong.

En este estudio los investigadores (Shiri et al., 2013), presentaron procedimientos basados en temperatura y radiación solar para estimar la evapotranspiración mediante el muso

de programación de expresión génica (GEP) y un sistema de inferencia neurodifusa adaptativa (ANFIS). También realizaron una comparación con modelos tradicionales basados en temperatura/radiación solar. Los resultados indicaron que el modelo GEP superó tanto a los modelos tradicionales (Hargreaves-Samani, Makkink, turk) y al modelo ANFIS.

En esta investigación se desarrolló un modelo de Programación de Expresiones Genéticas (GEP) para estimar la evapotranspiración en África Occidental. Los autores reportaron un alto rendimiento de la formulación algebraica obtenida utilizando GEP en la estimación de la evapotranspiración en las regiones de África Subsahariana (Traore y Guven, 2013).

En su investigación (Shiri, Nazemi, et.al., 2014), realizaron un análisis comparativo utilizando varios modelos inteligentes, incluyendo Redes Neuronales Artificiales (ANN), sistemas de inferencia neuro-difusos adaptativos (ANFIS), máquinas de vectores de soporte (SVM) y programación de expresiones genéticas (GEP) para estimar la evapotranspiración de referencia (ET_o). Además, se validó la capacidad de estos modelos de inteligencia artificial frente a diversas formulaciones empíricas, como Priestley-Taylor (PT), Hargreaves-Samani (HS), Turc y Makkink. Los resultados demostraron la superioridad del modelo GEP sobre otras metodologías de inteligencia artificial y empíricas.

En este estudio los investigadores (Shiri, Sadraddini, et.al., 2014) emplearon la programación de la expresión génica (GEP) para estimar la evapotranspiración, utilizaron diversas combinaciones de variables meteorológicas como atributos de entrada para construir un modelo predictivo. El estudio destacó el excelente rendimiento del modelo GEP al incorporar todas las variables meteorológicas disponibles.

En la investigación realizada por (Kisi et.al., 2015) evaluaron el rendimiento de los modelos ANN, ANFIS con agrupación sustractiva y partición en rejilla, y la programación de expresión génica (GEP) para la simulación a largo plazo de la evapotranspiración mensual en 50 estaciones en Irán. Los autores destacaron la eficacia del modelo GEP en comparación con otros modelos de IA desarrollados. Además, observaron una consistente capacidad predictiva en la simulación de la evapotranspiración mediante el uso de GEP.

Este estudio los investigadores (Martí et.al., 2015) llevaron a cabo un estudio para comparar el rendimiento de la programación de expresión génica (GEP) en la estimación de la evapotranspiración con los modelos FAO56-PM y Hargreaves-Samani utilizando datos de lisímetros en dos ubicaciones de España: Las Tiesas (Albacete) y La Orden (Badajoz). El estudio concluyó que los mejores resultados en la estimación de la evapotranspiración se lograron utilizando GEP.

La investigación desarrollada por (Yassin et.al., 2016) tuvo como objetivo desarrollar ecuaciones matemáticas utilizando programación de expresión génica (GEP) para estimar la evapotranspiración a partir de datos meteorológicos de 13 estaciones en el Reino de Arabia Saudita, y reportaron que el modelo GEP proporcionó una estimación precisa de la evapotranspiración. Sin embargo, encontraron que las Redes Neuronales Artificiales (ANN) ofrecían un rendimiento ligeramente superior en comparación con GEP.

En el estudio desarrollado por Kumar y colaboradores (Kumar et.al., 2016) investigaron el rendimiento de la máquina de aprendizaje extremo (ELM), Redes Neuronales Artificiales (ANN), Programación Genética (GP) y máquinas de soporte vectorial (SVM) para estimar

la evapotranspiración diaria en el norte de Bihar, India, durante el período 2001-2005. Concluyeron que el modelo ELM mostró un rendimiento superior respecto a otros enfoques de aprendizaje automático y cómputo suave.

En su investigación (Karimi et.al., 2017) desarrollaron modelos heurísticos que combinan máquinas de vectores de soporte (SVM) y programación de expresión genética (GEP) para estimar la evapotranspiración en la República de Corea. Los resultados indicaron que el modelo GEP mostró un rendimiento superior al del modelo SVM, tanto en escenarios locales como en evaluaciones entre diferentes estaciones meteorológicas.

(Kiafar et.al., 2017) desarrollaron ecuaciones para estimar la evapotranspiración en dos estaciones meteorológicas en una región hiperárida de Irán y dos estaciones en una región húmeda de España. Los investigadores compararon el rendimiento del modelo de programación de la expresión génica (GEP) con tres modelos empíricos para calcular la evapotranspiración: modelos de transferencia de masa, basados en temperatura y basados en radiación. Según el estudio, el modelo GEP con cuatro entradas mostró un rendimiento superior en precisión y robustez comparado con los modelos empíricos tradicionales, destacándose especialmente en la capacidad de adaptarse a diferentes condiciones climáticas y geográficas.

El rendimiento de los modelos Máquina de Soporte Vectorial (SVM) con kernel polinomial (SVM-Pol) y de base radial (SVM-RBF), programación de expresión génica (GEP), regresión spline adaptativa multivariada (MARS) y modelos empíricos en la estimación de la evapotranspiración en 44 estaciones meteorológicas en Irán fue investigado por (Mehdizadeh et.al., 2017). Según los resultados obtenidos se observó que los modelos MARS y SVM-RBF superaron en rendimiento a los modelos SVM-Pol y GEP en la precisión de la estimación de ETo en diversas condiciones climáticas y geográficas de las estaciones evaluadas.

El algoritmo de programación de expresión génica (GEP) fue utilizado para predecir la escala diaria de la evapotranspiración en la estación de Gaoyou, China, según el estudio de (Traore et.al., 2017). Los resultados sugieren que el modelo GEP no solo es efectivo para estimar ETo, sino que también puede servir como una herramienta útil en la programación de riego a corto plazo para la toma de decisiones agrícolas.

En el estudio realizado por (Shiri, 2017), se evaluaron varios enfoques empíricos y semi-empíricos, incluidos los modelos de Kimberly-Penman, basados en temperatura y en radiación, junto con el modelo de programación de expresión génica (GEP) para estimar la evapotranspiración diaria en cinco estaciones meteorológicas en Irán. La estrategia utilizada fue dividir el conjunto de datos en tres segmentos para la calibración, prueba y validación del modelo. La comparación de los resultados reveló que el modelo empírico basado en temperatura desarrollado con GEP mostró la mayor precisión en la estimación de la evapotranspiración.

Este estudio representa la primera investigación en Ghana sobre la capacidad de modelos basados en inteligencia artificial, como Redes Neuronales Artificiales (ANN) y programación de expresión génica (GEP), junto con enfoques auxiliares, para modelar la evapotranspiración utilizando datos meteorológicos limitados. Utilizaron datos recopilados durante 7 años (2006-2012) de los cuales 4 años (2009-2012) fueron utilizados para

prueba. Según los resultados de este estudio, los modelos GEP han demostrado ser una alternativa muy prometedora para la estimación precisa de la evapotranspiración en todas las ubicaciones evaluadas en Ghana, incluso bajo diferentes escenarios de disponibilidad de datos meteorológicos (Landeras et.al., 2018).

El estudio de (Jovic et.al., 2018) investigó el potencial de la Programación Genética (GP) para desarrollar modelos de estimación de la evapotranspiración. Desarrollaron varios modelos que incorporaron una combinación de diversos parámetros meteorológicos de entrada, como temperatura mínima y máxima, presión de vapor, velocidad del viento, horas de sol y humedad relativa. Los investigadores concluyeron que el mejor rendimiento se obtuvo al utilizar la totalidad de estos parámetros en la implementación del modelo.

En esta investigación, la evapotranspiración se modela como un componente fundamental en aplicaciones hidrológicas mediante diferentes combinaciones de variables climáticas utilizando dos técnicas distintas: programación de expresión génica (GEP) y regresión lineal múltiple (MLR). Los datos utilizados fueron recopilados de 27 estaciones meteorológicas en Egipto a partir de la base de datos CLIMWAT. Se utilizó la ecuación Penman-Monteith FAO56-PM como modelo de referencia, considerando todas las variables climáticas disponibles. Según el estudio, la temperatura del aire, la velocidad del viento y la humedad relativa fueron identificadas como variables críticas para una modelización precisa de la evapotranspiración. Además, los resultados resaltaron que la combinación de MLR y GEP, utilizando la humedad media y la velocidad del viento a 2 metros de altura como variables de entrada, proporcionó la mejor estimación (Mattar y Alazba, 2019).

En esta investigación se evaluó la aplicación de modelos de estimación de la evapotranspiración en entornos insulares de Irán utilizando el modelo FAO-Penman-Monteith (FAO56-PM) como referencia. Se compararon ecuaciones basadas en temperatura, radiación y transferencia de masa con modelos heurísticos de programación de expresión génica (GEP), evaluados con pruebas k-fold. Los resultados evidenciaron un mejor rendimiento del modelo basado en transferencia de masa, mientras que los modelos basados en GEP, temperatura y radiación mostraron un rendimiento relativamente inferior (Shiri, 2019).

En este estudio se presenta una innovación al integrar entradas auxiliares externas en la estimación de evapotranspiración mediante la programación de expresión génica (GEP), utilizando el modelo de Hargreaves-Samani basado en temperatura y el modelo de Presley-Taylor basado en radiación solar con datos meteorológicos locales. La evaluación comparativa de la precisión de los modelos mostró que el enfoque GEP logró la mayor precisión entre todas las metodologías aplicadas (Shiri et.al., 2019).

3.4. Análisis de los trabajos presentados

En esta sección, se ha llevado a cabo un análisis detallado de los trabajos presentados en el estado del arte, destacando los artículos más relevantes relacionados con la investigación propuesta en este documento. Se han examinado particularmente estudios relacionados con el cómputo evolutivo, fundamentales para comprender su aplicación en diferentes entornos y su impacto en la estimación de la evapotranspiración de referencia.

Los modelos utilizados para estimar la evapotranspiración de referencia presentan diversas desventajas. Los modelos de caja negra, como RF, GBDT, XGBoost, SVM, ELM, CatBoost, LightGBM, BAXGB, ANN, GRNN, MLP, RBNN y KNEA, aunque pueden proporcionar estimaciones precisas, carecen de interpretabilidad y requieren grandes cantidades de datos para ser eficaces. Además, son propensos al sobreajuste y pueden ser computacionalmente demandantes, lo que limita su implementación en escenarios donde la transparencia y la eficiencia son cruciales.

Además, los trabajos revisados a menudo dependen de variables que no siempre están disponibles debido a los altos costos asociados con los dispositivos de medición, como los sensores para medir la radiación solar. En entornos con baja o nula tecnología, esta dependencia se vuelve aún más problemática, limitando la recopilación de datos a parámetros más accesibles como la temperatura, la humedad relativa y la velocidad del viento. Este contexto subraya la necesidad de nuevas investigaciones orientadas a desarrollar modelos simples y fáciles de interpretar que utilicen únicamente variables básicas.

Dichos modelos no solo serían más accesibles y económicos, sino que también podrían ser implementados de manera más amplia, facilitando la estimación de la evapotranspiración de referencia en una variedad de entornos, especialmente en aquellos con limitaciones tecnológicas y presupuestarias. La búsqueda de soluciones eficientes y prácticas que empleen variables fácilmente disponibles es crucial para avanzar en la precisión y aplicabilidad de las estimaciones de evapotranspiración. Esta búsqueda no solo investiga como mejorar la exactitud de los modelos actuales, sino también hacerlos más accesibles y utilizables en una variedad de entornos, especialmente aquellos con limitaciones tecnológicas y recursos financieros restringidos. Al enfocarnos en variables básicas como la temperatura, la humedad relativa y la velocidad del viento, aspiramos a desarrollar modelos que puedan ser implementados de manera amplia y sin necesidad de costosos equipos de medición. De esta manera, nuestra investigación no solo contribuirá a la mejora de las prácticas de gestión hídrica, sino que también apoyará la sostenibilidad agrícola y la adaptación a los desafíos locales de acceso a tecnología y datos, promoviendo una gestión de recursos más equitativa y efectiva.

En resumen, los artículos mencionados anteriormente son una referencia importante para el desarrollo de esta tesis. A continuación, en la Tabla 3.1, se presenta un resumen del estado del arte, destacando el autor, la tecnología utilizada, los parámetros de entrada y las métricas empleadas. Este resumen servirá como una guía para entender las metodologías actuales y cómo nuestra propuesta innovará en este campo.

Tabla 3.1: Resumen de los trabajos presentados en el estado del Arte

Algoritmos de ensamblaje para estimar la evapotranspiración				
No.	Autor	Técnica	Parámetros Entrada	Métricas
1.	(Fan et.al., 2018)	RF, M5Tree, GBDT, XGBoost, SVM y ELM	Tmax, Tmin, HRmin, HRmax, u_2 , Rs y Ra	MAE, RMSE, R^2
2.	(Huang et.al., 2019)	CatBoost, RF, SVM	Rs, Tmax, Tmin, HRmin, HRmax y u_2	R^2 , RMSE, MBE, MAPE
3.	(Fan et.al., 2019)	M5Tree, RF, LightGBM y Modelos empíricos	Tmax, Tmin, HRmin, HRmax, u_2 , Ra y Rs	RMSE, R^2 , NMRSE
4.	(Han et.al., 2019)	BAXGB, XGBoost, MARS, GPR	Tmin, Tmax, u_2 , HRmin, HRmax, Rs	R^2 , RMSE, MAE
5.	(Ge et.al., 2022)	XGBR, RL, SVR, KNR, RFR, ABR, BR, GBR	Tavg, Tmax, Tmin, HRavg, HRmin, HRmax, u_2 , Rn	RMSE, R^2 , MSE, MAE, MAPE
Diversas técnicas de computación suave para modelar la evapotranspiración				
No.	Autor	Técnica	Parámetros Entrada	Métricas
6.	(Kim y Kim, 2008)	COMBINE-GRNNM-GA, EXTREME-GRNNM-GA and Average-GRNNMGA based on GRNN and GA	Tmax, Tavg, Tmin, HRmin, HRmax, Foroperiodo, u_{2avg} , u_{2max} , $PRocio_{avg}$	RMSE, MAE, R^2
7.	(Kisi, 2010)	LGP, SVR, ANN	Tmax, HRmin, HRmax, Rs, u_2	RMSE, R^2 , MAE
8.	(Eslamian et.al., 2012)	ANN-GA, ANN	Tmax, Tavg, Tmin, HRmin, HRmax, Rs, u_2	MSE, NMSE, MAE, R^2
9.	(Feng et.al., 2017)	RF, GRNN	Tmax, Tmin, HRmin, HRmax, Rs, Fotoperiodo, u_2	RRMSE, MAE, NS
10.	(Antonopoulos y Antonopoulos, 2017)	RNA y Modelos empíricos	Tavg, HRmin, HRmax, u_2 , Rs	RMSE, r, Media, DesvEst
11.	(Mehdizadeh, 2018)	MARS, GEP	Tmax, Tmin, HRmin, HRmax, Rs, u_2	RMSE, R^2 , MAE, MAPE
12.	(Srivastava et.al., 2018)	VIC-3L, HS, GA	Tmax, Tmin, HRmax, HRmin, u_2 , Rs	RMSE, R^2 , NSE

13.	(Ferreira et.al., 2019)	ANN,SVM, Modelos empíricos	Tmin, Tmax, HRmin, HRmax	RMSE, MBE, R^2
14.	(Sanikhani et.al., 2019)	MLP, GRNN, RBFNN	Tmax, Tavg, Tmin, HRmin, HRmax, u_2	RMSE, R^2 , MAE, CRM, NS
15.	(Wu y Fan, 2019)	MLP, GRNN, ANFIS, SVM, KNEA, M5Tree, XGBoost, y MARS	Precipitación, Temperatura	RMSE,NRMSE, R^2
16.	(Balmat et.al., 2019)		Rn, u_2 ,Tavg, Thoja, Tsuperficie, HRmin, HRmax, IAF, VPD, Ra	R^2
17.	(Yu et.al., 2020)	XGBoost, PSO, CatBoost, Bagging, AdaBoost, RF, RNN, DT, KNN	Tavg, HRavg, Rs, VP, HSuelo	MAE, MSE, RMSE, R^2
18.	(Jung et.al., 2022)	RNN, CNN, LSTM, CNN-LSTM	Text, HRext, Rs, u_2 , IAF	RMSE, SEP, R^2 , MAE
19.	(Yuan et.al., 2023)	ANN	T,HR, Rs	RMSE, R^2

Cómputo evolutivo para estimar la evapotranspiración

No.	Autor	Técnica	Parámetros Entrada	Métricas
20.	(Parasuraman et.al., 2007)	GP, ANN	Rn, Tavg, u_2 , HRmin, HRmax	RMSE, MARE, R
21.	(Güven et.al., 2008)	GEP	Rs, Tavg, TSuelo, HRmin, HRmax, VPD, u_2 , DirViento, Precip	MSE, R^2
22.	(Kisi y Güven, 2010)	LGP, SVR, ANN	Tmax, HRmin, HRmax, Rs, u_2	RMSE, R^2 , MAE
23.	(El-Baroudy et.al., 2010)	GEP, EPR, ANN	Rn, Tavg, TSuelo, u_2 , HRmin, HRmax	RMSE, MARE, R
24.	(Izadifar y Elshorbagy, 2010)	GP, ANN, Modelos estadísticos, Penman	Rn, TSuelo, Tavg, HRmin, HRmax, u_2	RMSE, MARE, R
25.	(Shiri et.al., 2012)	GEP, ANFIS, Modelos empíricos	Tmax, Tavg, Tmin, HRmin, HRmax, Rs, u_2	RMSE, R^2 , SI
26.	(Traore y Güven, 2012)	GEP	NR, Tavg, HRmin, HRmax, u_2	RMSE, R^2

27.	(de la Paix Mupenzi et.al., 2012)	GEP, PM	Tavg, Rs, Flujo de calor, Rn, u_2	R^2
28.	(Shiri et.al., 2013)	GEP, ANFIS, Modelos empíricos	Tmax, Tavg, Tmin, HRmin, HRmax, Rs, u_2	MAE, SI, NS, R^2
29.	(Traore y Guven, 2013)	GEP	Tmax, Tmin, Precipitación, HRmin, HRmax, luminosidad, u_2	RMSE, R^2
30.	(Shiri, Nazemi, et.al., 2014)	ANN, ANFIS, SVR, GEP, modelos empíricos	Tmax, Tavg, Tmin, HRmin, HRmax, Rs, u_2	RMSE, R^2 , SI, MAE
31.	(Shiri, Sadraddini, et.al., 2014)	GEP_{1-3}	Tmax, Tavg, Tmin, HRmin, HRmax, Rs, u_2	RMSE, R^2 , AARE, MAE
32.	(Kisi et.al., 2015)	ANN, ANFIS-GP, ANFIS-SC, GEP	Tmax, Tmin, HRmin, HRmax, u_2 , Rs, latitude, longitude y altitud	RMSE, R^2
33.	(Martí et.al., 2015)	GEP, Modelos empíricos	evapotranspiración	RMSE, R^2 , MAE
34.	(Yassin et.al., 2016)	ANN, GEP, Penman	Tmax, Tavg, Tmin, HRmax, HRavg, HRmin, Rs, u_2	RMSE, R^2 , OI, MAE
35.	(Kumar et.al., 2016)	ELM, ANN, SVM, GP	Tmax, Tmin, HRmax, HRmin, Rs, Precipitación, u_2	RMSE, R^2 , Tiempo
36.	(Karimi et.al., 2017)	SVM, GEP	Tmax, Tmin, HRmin, HRmax, Rs, u_2	MAE, CRM
37.	(Kiafar et.al., 2017)	GEP, Modelos empíricos	Tmax, Tavg, Tmin, HRmin, HRmax, Rs, u_2	RMSE, R^2 , CRM
38.	(Mehdizadeh et.al., 2017)	MARS, SVR, GEP	Tmax, Tavg, Tmin, HRmin, HRmax, Rs, u_2	RMSE, R^2 , MAE
39.	(Traore et.al., 2017)	GEP, Penman	Tmax, Tavg, Tmin, HRmin, HRmax, luminosidad, u_2	RMSE, R, MAE, RRSE
40.	(Shiri, 2017)	KP modelos empíricos y GEP_{1-8}	Tmax, Tmin, Tavg, HRavg, Rs, u_2	SI, MAE, NS
41.	(Landeras et.al., 2018)	ANN, GEP	Tmax, Tmin, Tavg, HRmin, HRmax, HRavg, Fotoperiodo, u_2	RMSE, MAE, SI

42.	(Jovic et.al., 2018)	GP	Tmax, Tavg, Tmin, HRmin, HRmax, Fotoperiodo, u_2	RMSE, R^2
43.	(Mattar y Alazba, 2019)	GEP, MLR, Penman	Tmax, Tmin, HRavg, Rs, u_2	RMSE, R, MAE
44.	(Shiri, 2019)	GEP	Tmax, Tavg, Tmin, HRmin, HRmax, Fotoperiodo, u_2	IOA, RMSE, NS
45.	(Shiri et.al., 2019)	GEP, Modelos empíricos	Tmax, Tavg, Tmin, Ra, Rs	SI, NS

Tmax = Temperatura máxima; Tmin = Temperatura mínima; Tavg = Temperatura promedio; Text = Temperatura exterior, Tint = Temperatura interior; TSuelo = Temperatura del suelo; HRExt = Humedad relativa; exterior HRmax = Humedad relativa máxima; HRmin = Humedad relativa mínima; HSuelo = Humedad del suelo; u_2 = Velocidad del viento; Rs = Radiación solar; Rn = Radiación neta; Ra = Radiación extraterrestre; THoja = Temperatura de la hoja; IAF = Índice de área foliar; $PRoci_{avg}$ = Punto de rocío; TSuelo = Temperatura del suelo; VP = Presión de vapor.

3.5. Conclusiones

En este capítulo, se han revisado diversas investigaciones que tienen relación con el trabajo desarrollado en esta tesis.

La investigación sobre modelos para estimar la evapotranspiración ha experimentado avances notables en los últimos años, sin embargo, persisten limitaciones significativas que merecen ser abordadas. Si bien numerosos modelos, incluidos aquellos basados en redes neuronales o algoritmos de aprendizaje automático como Random Forest, Support Vector Machine (SVM) y Light Gradient Boosting Machine (LightGBM), han mostrado un rendimiento predictivo prometedor, su complejidad los convierte en herramientas menos interpretables y a menudo requieren grandes conjuntos de datos y recursos computacionales considerables. Además, modelos como CatBoost, LightGBM o ANFIS pueden verse afectados por el sobreajuste, capturando relaciones espurias en los datos de entrenamiento, mientras que otros, como las redes neuronales profundas, a menudo se perciben como «cajas negras», dificultando la comprensión de su funcionamiento interno.

Una limitación común que se destaca es la dependencia de datos precisos y completos, lo que puede restringir la aplicabilidad de estos modelos en contextos donde la disponibilidad de datos es limitada o incompleta. Además, cabe señalar que la mayoría de las investigaciones se centran en la estimación de la evapotranspiración en grandes extensiones de superficie, dejando de lado las estructuras utilizadas en agricultura protegida, a pesar del crecimiento significativo que ha experimentado esta técnica de producción agrícola (L. Pratt et.al., 2019). Por otro lado, no se ha encontrado evidencia sobre algún método estándar para estimar la evapotranspiración de referencia al interior de invernaderos (Karaca et.al., 2018).

En resumen, si bien estos modelos pueden ser efectivos en ciertos contextos, la necesidad de modelos más simples y fácilmente interpretables resalta la importancia de una mayor investigación en este campo. Es fundamental desarrollar modelos que sean tanto precisos como comprensibles, especialmente en áreas donde los recursos son limitados y se requieren soluciones prácticas y accesibles para la estimación de la evapotranspiración.

Capítulo 4

Desarrollo experimental de un modelo para estimar la evapotranspiración a campo abierto

En esta sección, se describe la primera fase de la metodología propuesta presentada en la sección 1.5. El objetivo principal de esta sección es experimentar con la generación de modelos y validar estimaciones precisas de evapotranspiración en entornos de campo abierto. Esta etapa es crucial, ya que sentará las bases para el desarrollo de modelos adaptados a entornos de agricultura protegida, como es el caso de los invernaderos. Al ajustar el modelo a las condiciones específicas del entorno, se busca mejorar la precisión y la aplicabilidad de las estimaciones, lo que contribuirá significativamente a una gestión más eficiente del riego y de los recursos hídricos en diversas modalidades agrícolas.

A continuación se presentan las actividades desarrolladas para obtener el modelo de evapotranspiración a campo abierto.

4.1. Recolección y preparación de datos

4.1.1. Obtención de datos

área de estudio

El área de estudio fue un conjunto de estaciones meteorológicas ubicadas en localidades distantes geográficamente incluso en países diferentes como se muestra en la Tabla 4.1. Tres estaciones meteorológicas pertenecen al «*The California Irrigation Management Information System (CIMIS)*» ubicado en el estado de California, Estados Unidos ((CIMIS)[®], 2022) y una estación ubicada en el estado de Coahuila al centro-norte de México. A continuación, se describen las características principales de las estaciones meteorológicas utilizadas. El tipo de clima fue definido en función de la clasificación climática de Köppen and Geiger (Köppen y Geiger, 1930).

- a. Estación «Davis» su ubicación geográfica es latitud 38°32'8N, longitud 121°46'35W con una altitud de 18.288 metros sobre el nivel del mar, con clima cálido templado (Csa). La temperatura promedio anual es de 16.8°C y una precipitación anual de 613 mm.
- b. Estación «Calipatria/Mulberry» su ubicación geográfica es latitud 33°2'35N, longitud 115°24'57W con una altura de -33.528 metros sobre el nivel del mar, con clima árido cálido (BWh). La temperatura promedio anual oscila entre los 6°C y 42°C y una precipitación promedio anual de 76 mm.
- c. Estación «McArthur» su ubicación geográfica es latitud 41°3'50N, longitud 121°27'22W con una altura de 1008.888 metros sobre el nivel del mar, con clima cálido templado (Csb). La temperatura promedio anual es 9.4°C y una precipitación promedio anual de 1441 mm
- d. Estación «El Porvenir» ubicada geográficamente latitud 25°46'52N, longitud 103°18'46"W, con una elevación de 1112 metros sobre el nivel del mar, ubicada al norte de México. Con clima árido cálido (BWh). La temperatura promedio anual oscila entre los 18°C y 22°C. La precipitación promedio anual es de 250 mm.

Tabla 4.1: Estaciones meteorológicas utilizadas para entrenamiento del algoritmo evolutivo y desarrollo del modelo a campo abierto

No.	Estación Meteorológica	Latitud	Longitud	Clima	País
1	El Porvenir	25.7811	-103.3130	Árido	México
2	Davis	38.5357	-121.7763	Templado	USA
3	Calipatria/Mulberry	33.0431	-115.4158	Árido	USA
4	McArthur	41.0637	-121.4560	Templado	USA

La Figura 4.1 muestra la ubicación geográfica de las estaciones meteorológicas utilizadas para entrenamiento.

Base de datos

Las bases de datos de las estaciones: Davis, Calipatria/Mulberry y McArthur, utilizadas para la construcción del modelo, comprenden el período de 2011 a 2015. La estación El Porvenir contiene datos diarios desde el 1 de febrero hasta el 31 de diciembre de 2019. Los atributos utilizados de estas bases de datos son: temperatura máxima, temperatura mínima, humedad relativa máxima, humedad relativa mínima, velocidad del viento y radiación solar.

Las bases de datos utilizadas para validar el modelo son de los siguientes periodos: estación Modesto año 2019, estación Oakville año 2015, estación Meloland año 2019 y estación Ferndale año 2019.



Figura 4.1: Ubicación geográfica de las estaciones meteorológicas utilizadas para entrenamiento.

4.1.2. Preprocesamiento de datos

De acuerdo con la metodología de solución las bases de datos fueron sometidas a un análisis exploratorio para detectar datos atípicos o datos faltantes. Los datos atípicos fueron analizados y se tomó la decisión de eliminarlos ya que provenían de mediciones imprecisas, por ejemplo, temperaturas de 100°C o humedad relativa mayor a 100% y los datos faltantes fueron generados a través de interpolación simple cuando se trataba de un día. Como resultado, se obtuvieron 334 registros de la estación El Porvenir, de la estación Davis se obtuvieron 1826 registros, la estación Calipatria/Mulberry presento mayor número de datos faltantes, y se obtuvieron 1792 registros y de la estación McArthur se obtuvieron 1826 registros. Posteriormente se realizó un preprocesamiento para estandarizar las unidades en las que estaban registrados los datos originales. Por ejemplo, la temperatura estaba registrada en $^{\circ}\text{F}$ se realizó la conversión a $^{\circ}\text{C}$, las unidades de la velocidad del viento estaban en $[\text{mph}]$ se convirtieron a $[\text{m/s}]$, la radiación solar viene reportada en $[\text{Ly/day}]$ se convirtió a $[\text{W/m}^2]$.

Las bases de datos utilizadas para la validación recibieron el mismo tratamiento que las bases de datos utilizadas para el entrenamiento. Los resultados muestran que se obtuvieron los siguientes registros: 265 registros de la estación Modesto, 342 registros de la estación Oakville, 336 registros de la estación Meloland y 149 registros de la estación Ferndale.

4.1.3. Creación de conjuntos de datos

Con los datos preprocesados y estandarizados en unidades se formaron tres conjuntos de datos para entrenamiento el primero DS01 con 334 registros de la estación El Porvenir, el segundo DS02 con 1826 registros de la estación Davis y el tercer conjunto de datos DS03 fue un clúster con las estaciones El Porvenir, Davis, Calipatria/Mulberry y McArthur agrupando un total de 5778 registros. en la Tabla 4.2 se muestran los tres conjuntos de datos y sus características.

Tabla 4.2: Conjuntos de datos y sus características

No.	Conjuntos de datos	Descripción	Registros
1	DS01	El Porvenir	334
2	DS02	Davis	1826
3	DS03	El Porvenir, Davis, Calipatria y McArthur	5778

Los conjuntos de datos fueron separados en 80% para entrenamiento y el resto para prueba. Se realizó un muestreo uniforme para obtener los datos de prueba, es decir, se tomó uno de cada cinco datos para el conjunto de prueba; y el resto para el conjunto de entrenamiento. La estructura de los conjuntos de datos de entrenamiento y prueba se muestra en la Table 4.3.

Tabla 4.3: Estructura de los conjuntos de datos

Tmax[°C]	Tmin[°C]	HRmax[%]	HRmin[%]	$u_2[m/s]$	DJ
25.9	3.9	69	22	0.75	1
25.4	6.4	76	21	0.70	2
...
26.8	9.0	73	26	0.74	365

4.2. Proceso evolutivo y generación del modelo

De acuerdo a la metodología de solución, se deben realizar cinco pasos preparatorios. Estos pasos consisten en: determinar el conjunto de terminales, el conjunto de funciones, la función de aptitud, los parámetros para controlar la ejecución, y el método para designar un resultado. El proceso de ejecución del algoritmo evolutivo utilizado para generar el modelo llevó cuatro días.

4.2.1. Realizar pasos preparatorios

De acuerdo con Jonh Koza, 1992 (Koza, 1992), antes de comenzar el proceso evolutivo es esencial llevar a cabo cinco pasos preparatorios fundamentales. Estos pasos son cruciales para establecer las bases del algoritmo de programación genética y asegurar que el proceso evolutivo se desarrolle de manera efectiva. Koza subraya que una preparación adecuada no solo optimiza el rendimiento del modelo, sino que también facilita la adaptación del algoritmo a diferentes problemas, aumentando así la probabilidad de encontrar soluciones óptimas en un espacio de búsqueda complejo. A continuación se describen los pasos preparatorios

Determinar el conjunto de terminales

El conjunto de terminales representa las variables independientes del modelo aún no descubierto. En nuestra investigación primero se identificaron las variables de entrada independientes (humedad relativa máxima y mínima, temperatura máxima y mínima, velocidad del viento), posteriormente se buscó aprovechar el conocimiento experto que proporciona el modelo FAO56-PM haciendo un análisis de aquellos parámetros que pueden ser utilizados como terminales. En la Tabla 4.4 se muestra el conjunto de terminales seleccionadas.

Tabla 4.4: Conjunto de terminales.

No.	Terminal	Descripción	Expresión matemática
1	*DJ	Día del año	
2	Tmax	Temperatura máxima	
3	Tmin	Temperatura mínima	
4	HRmax	Humedad relativa máxima	
5	HRmin	Humedad relativa mínima	
6	U_2	Velocidad del viento	
7	*G	Temperatura del suelo	
8	Tprom	Temperatura promedio	
9	*gt	Grados totales	
10	*lt	Latitud	
11	*N	Fotoperíodo	$N = \frac{24}{\pi} * ws$
12	*P	Presión atmosférica	$P = 101.3 \left(\frac{293 - 0.0065z}{293} \right)^{5.26}$
13	* γ	Constante psicrométrica	$\gamma = \frac{c_p P}{\epsilon \lambda}$
14	* d_r	Distancia relativa de la Tierra al Sol	$d_r = 1 + 0.033 * \cos\left(\frac{2\pi}{365} DJ\right)$
15	* d_s	Declinación solar	$\delta = 0.409 * \sin\left(\frac{2\pi}{365} DJ - 1.39\right)$
16	* e_a	Presión de vapor actual	$e_a = \frac{e^o T_{min} * \frac{HR_{max}}{100} + e^o T_{max} * \frac{HR_{min}}{100}}{2}$
17	* e_s	Presión de vapor a saturación	$e_s = \frac{e^o T_{min} + e^o T_{max}}{2}$
18	* R_a	Radiación Extraterrestre	$R_a = \frac{24 * 60}{\pi} G_{sc} d_r [w_s \sin(lt) \sin(\delta) + \cos(lt) \cos(\delta) \sin(w_s)]$
19	* w_s	Ángulo de radiación	$w_s = \arcsin[-\tan(\theta) \tan(\delta)]$

No.	Terminal	Descripción	Expresión matemática
20	* Δ	Pendiente de la curva de presión de vapor	$\Delta = \frac{4098[0.618 * \exp(\frac{17.27 * T}{T+237.3})]}{(T+237.2)^2}$
21	* $e^o T_{max}$	Presión de vapor a temperatura máxima	$e^o T_{max} = 0.618 * \exp(\frac{17.27 * T_{max}}{T_{max}+237.3})$
22	* $e^o T_{min}$	Presión de vapor a temperatura mínima	$e^o T_{min} = 0.618 * \exp(\frac{17.27 * T_{min}}{T_{min}+237.3})$
23	*reflexion	Constante de reflexión solar	
24-28	*Constantes	900, 273, 1, 0.34, 0.408	

* Conocimiento experto obtenido del modelo FAO56-PM

Determinar el conjunto de funciones

El conjunto de funciones se detalla en la Tabla 4.5, y abarca una variedad de operadores básicos, así como funciones trigonométricas, exponenciales y logarítmicas que son comúnmente empleadas en estudios hidrológicos, según lo descrito por (Mehdizadeh et.al., 2017; Mattar, 2018). Además, se incluyen las funciones hiperbólicas, las cuales son ampliamente utilizadas para el análisis de fenómenos físicos, como se señala en (Pandir y Ulusoy, 2013).

Tabla 4.5: Conjunto de funciones

No.	Función	Expresión	No.	Función	Expresión
1	Seno	$\sin(\cdot)$	10	Tangente hiperbólica	$\tanh(\cdot)$
2	Coseno	$\cos(\cdot)$	11	Función exponencial de base e	$e^{(\cdot)}$
3	Tangente	$\tan(\cdot)$	12	Raíz cuadrada	$\sqrt{(\cdot)}$
4	Arco seno	$\text{asin}(\cdot)$	13	Exponente	$x^{(\cdot)}$
5	Arco coseno	$\text{acos}(\cdot)$	14	Suma	+
6	Arco tangente	$\text{atan}(\cdot)$	15	Resta	-
7	Seno hiperbólico	$\sinh(\cdot)$	16	Multiplicación	*
8	Coseno hiperbólico	$\cosh(\cdot)$	17	División	\div
9	x al cuadrado	x^2	18	x al cubo	x^3

Definir la función de aptitud «fitness»

La función de aptitud, también conocida como fitness, es fundamental en el proceso evolutivo. Cada miembro de la población se ejecuta y luego se evalúa utilizando esta función para determinar su rendimiento. Las funciones y terminales definen el espacio de búsqueda, mientras que la medida de aptitud específica implícitamente el objetivo deseado de la búsqueda y permite valorar la calidad de los individuos. En este caso, la función de aptitud debe producir valores reales positivos.

Según la metodología de solución empleada, se ha seleccionado la raíz del error cuadrático medio (RMSE) como medida de aptitud (Ecuación. 5.2).

Cuanto mejor sea la aptitud de un individuo, mayor será su probabilidad de pasar a la siguiente generación.

Determinar los parámetros del algoritmo evolutivo

Los principales parámetros para controlar la ejecución del algoritmo evolutivo son el tamaño de la población, el número máximo de generaciones, método de inicialización, método de selección, probabilidad de cruce y probabilidad de mutación. En nuestra investigación los valores asignados a los parámetros son los siguientes:

- a. El tamaño de la población se estableció en 200 individuos,
- b. El número máximo de generaciones fue de, 200,
- c. El método de inicialización establecido fue, «Mitad-Mitad»,
- d. La selección de los individuos se llevó a cabo a través del método, «lexicotour»,
- e. La probabilidad de cruce se estableció en, 80% y
- f. La probabilidad de mutación fue del, 20%.
- g. Elitismo Mantener el mejor

Determinar el método para designar un resultado

El proceso evolutivo fue implementado en MatLab utilizando la caja de herramientas GPLab (Silva y Almeida, 2003). Se estableció como criterio de terminación el número máximo de generaciones, fijado en 200 en esta instancia. De este modo, cada ejecución del proceso evolutivo busca alcanzar un valor óptimo dentro de ese límite de generaciones. Este enfoque asegura una exploración exhaustiva del espacio de búsqueda en cada iteración, permitiendo así obtener resultados más precisos y consistentes.

4.2.2. Ejecución del algoritmo evolutivo

El algoritmo implementado con GPLab inicia generando una población inicial de soluciones, combinando aleatoriamente los elementos de los conjuntos de terminales y funciones especificados en las tablas 4.4 y 4.5, respectivamente. Cada individuo en esta población inicial, representado gráficamente en forma de árbol (como se observa en la figura 2.2), se evalúa utilizando la función de aptitud (fitness) definida por la (Ecuación 5.2). A continuación, se seleccionan las soluciones más prometedoras para clasificar a todos los individuos, eliminando aquellos con baja aptitud. Luego, se realiza la recombinación genética entre los individuos seleccionados, empleando operadores de cruce y mutación, tal como se muestra en las figuras 2.8 y 2.9, respectivamente. La siguiente generación se forma a partir de la selección de las mejores soluciones entre los padres y los hijos. Este proceso se repite hasta alcanzar el número máximo de generaciones especificado en los parámetros de ejecución.

4.2.3. Depurar las mejores soluciones

Las soluciones encontradas con Programación Genética fueron 3754 de las cuales fueron seleccionadas las mejores dando prioridad a las de mejor aptitud, menor complejidad estructural y menor número de parámetros de entrada. En la Tabla 4.6 se muestran los modelos seleccionados y el fitness para cada conjunto de datos. Los modelos fueron etiquetados como EToi (donde i representa un numero consecutivo).

Los modelos ETo1 y ETo4 están estructurados en función de la temperatura, mientras que los modelos ETo2, ETo6, ETo7, ETo9, ETo10, ETo12, ETo13 y ETo14 tienen en cuenta la temperatura y la velocidad del viento. Por otro lado, los modelos ETo3, ETo5, ETo8 y ETo11 están basados en la humedad relativa, la temperatura y la velocidad del viento. Este enfoque permite capturar diferentes aspectos climáticos que influyen en el cálculo de la evapotranspiración, proporcionando así una mayor precisión en la estimación de este parámetro clave en la hidrología y la agricultura.

Tabla 4.6: Modelos seleccionados

No.	Modelo	Fitness		
		DS01	DS02	DS03
1	$ETo1 = \frac{e_s}{0.4718}$	1.131	1.720	1.839
2	$ETo2 = \operatorname{atan}\left(\frac{u_2}{d_r + e_s + \frac{e_s}{d_r}}\right) + \frac{e_s}{d_s} + \frac{e_s}{\frac{d_r}{d_r}}$	0.973	1.663	1.819
3	$ETo3 = \sqrt{e_a + u_2} + 2d_s * \operatorname{acos}(u_2 \sqrt{\sqrt{e_a + u_2} * u_2 + \sqrt{d_s}}) + e^o Tmax$	0.895	1.639	1.657
4	$ETo4 = d_s + 2e_s + \frac{e_s - e_a}{\sinh(e_s)}$	1.094	1.581	1.655
5	$ETo5 = e^o Tmax - \frac{\tanh(Tmin)}{u_2} + e^o Tmin + d_s + \tanh(HRmax)$	0.894	1.861	1.945
6	$ETo6 = \sqrt{e_s + u_2 + \tanh(d_s + R_a)} + 2d_s + e_s + \tanh(\tanh(\sqrt{\tanh(e_s) + 2d_s + \Delta + e_s + e_s * \tanh(d_s)}) + e^o Tmin)$	0.878	1.664	1.590
7	$ETo7 = w_s(\operatorname{acos}(lt^{u_2}) + d_s + \Delta + e_s - e_a)$	1.723	0.852	0.834
8	$ETo8 = \frac{\tanh\left(\frac{d_s + \tanh(\tanh(lt))}{1 + \tanh(\operatorname{atan}(lt + \tanh(e_s - e_a)) + \tanh(2e^o Tmax + \tanh(HRmin) + i)}\right)}{d_s + \frac{u_s + \tanh(d_s) + ds}{\operatorname{atan}(\operatorname{atan}(\Delta + e^o Tmax) + \operatorname{atan}(\sin(e^o Tmax))) + \tanh(\Delta) + \tanh(lt + e^o Tma. 2e_s - e_a)}$	1.783	0.688	0.779
9	$ETo9 = \sqrt{u_2 + d_s} + 2(e_s - e_a) + 2d_s + \sqrt{d_s}$	1.726	0.806	0.898
10	$ETo10 = 3d_s + 2(e_s - e_a) + \operatorname{atan}(u_2) + \operatorname{atan}(\operatorname{atan}(\operatorname{atan}(\operatorname{atan}(\operatorname{acos}(900) + \sin(\operatorname{atan}(u_2) + \operatorname{atan}(900) + \sin(u_2)) + u_2)))) + u_2$	1.793	0.864	0.922

No.	Modelo	Fitness		
		DS01	DS02	DS03
11	$ETo11 = atan(u_2 - (e_s - e_a)) + 3(e_s - e_a) + d_s + atan(d_s) + atan(d_s + atan(e_s) + atan(d_s) - (e_s - e_a))$	1.837	0.753	0.839
12	$ETo12 = e^o Tmax + d_s + 2tanh(d_s) + \frac{d_s + \frac{d_s + u_2}{\pi} + u_2}{\pi}$	1.579	0.904	1.021
13	$ETo13 = d_s + e^o Tmax * u_2^{reflexion}$	1.804	0.915	1.083
14	$ETo14 = e^o Tmax + u_2^4 * 0.1325 * 3d_s$	2.092	0.923	1.004
15	$HS^* = 0.0023R_a \sqrt{Tmax - Tmin} * (Tprom + 17.8)$	2.501	1.157	1.016

$HS^* = \text{Hargreaves-Samani}$

4.2.4. Seleccionar el mejor modelo

Para seleccionar el modelo óptimo, se tomó la recomendación de la metodología de solución, que destaca la importancia de realizar un análisis exhaustivo considerando tanto el nivel de aptitud (fitness) como la complejidad estructural de las soluciones. En la investigación, aunque la mejor solución identificada fue la ETo8, como se muestra en la Tabla 4.6, su estructura demostró ser altamente compleja. Por lo tanto, optamos por elegir un modelo que equilibre de manera adecuada el rendimiento y la complejidad estructural. En consecuencia, seleccionamos el modelo ETo11 de la Tabla 4.6, que a partir de este punto denominaremos ETo_{gp} . (Ecuación 4.1).

$$ETo_{gp} = atan(u_2 - (e_s - e_a)) + 3(e_s - e_a) + d_s + atan(d_s) + atan(d_s + atan(e_s) + atan(d_s) - (e_s - e_a)) \quad (4.1)$$

Donde:

u_2	Velocidad del viento [ms^{-1}]
$e_s = \frac{e^o Tmin + e^o Tmax}{2}$	Presión de vapor a saturación [kPa]
DJ	Día Juliano
$d_s = 0.409 * \sin(\frac{2\pi * DJ}{365} - 1.39)$	Declinación Solar [Rad]
$e_a = \frac{\frac{e^o Tmin * HRmax}{100} + \frac{e^o Tmax * HRmin}{100}}{2}$	Déficit de presión de vapor [kPa]
$e^o Tmin = 0.6108 * \exp(\frac{17.27 * Tmin}{Tmin * 237.3})$	Presión de vapor a temperatura mínima [kPa]
$e^o Tmax = 0.6108 * \exp(\frac{17.27 * Tmax}{Tmax * 237.3})$	Presión de vapor a temperatura máxima [kPa]

4.3. Validación del modelo

Para validar el modelo se utilizaron cuatro estaciones meteorológicas obtenidas de «*The California Irrigation Management Information System (CIMIS)*» ubicadas en el estado de California, Estados Unidos ((CIMIS)[®], 2022). A continuación, se describen en la tabla 4.7 sus principales características. El tipo de clima fue definido en función de la clasificación climática de Köppen and Geiger (Köppen y Geiger, 1930)

- Estación «Modesto» latitud 37°38'43N, longitud 121°1'16W con una altura de 10.668 metros sobre el nivel del mar, con clima cálido templado (Csa). La temperatura promedio anual es de 17.6 °C con una precipitación promedio anual de 474 mm.
- Estación «Oakville» latitud 38°25'43N, longitud 122°24'37W con una altura de 57.912 sobre el nivel del mar, con clima cálido templado (Csc). La temperatura promedio anual es de 15.0°C con una precipitación promedio anual de 594 mm.
- Estación «Ferndale» latitud 40°36'16N, longitud 124°14'35W con una altura de 6.4008 metros sobre el nivel del mar, con clima cálido templado (Csb). La temperatura promedio anual es de 9.9°C con una precipitación promedio anual de 1617 mm.
- Estación «Meloland» latitud 32°48'22N, longitud 115°26'47W con una altura de -15.24 metros sobre el nivel del mar. con clima árido cálido (BWh). la temperatura promedio anual oscila entre los 6°C a 42°C con una precipitación promedio anual de 76 mm.

Tabla 4.7: Estaciones meteorológicas utilizadas para validación

No.	Estación meteorológica	Latitud	Longitud	Clima	Período
1.	Modesto	37.6452	-121.1877	templado-cálido	2019
2.	Oakville	38.4284	-122.4102	templado-cálido	2015
3.	Meloland	32.8061	-115.4462	árido-cálido	2019
4.	Ferndale	40.6044	-124.2431	templado-cálido	2019



Figura 4.2: Ubicación geográfica de las estaciones meteorológicas utilizadas para validación

Con las bases de datos obtenidas se formaron cuatro conjuntos de datos: DS04 con 365 registros, DS05 con 342 registros, DS06 con 336 registros y DS07 con 149 registros, como se observa en la Tabla 4.8.

Tabla 4.8: Conjuntos de datos usados para validación

No.	Conjunto de datos	Descripción	Registros
1.	DS04	Modesto	365
2.	DS05	Oakville	342
3.	DS06	Meloland	336
4.	DS07	Ferndale	149

Para validar el modelo ET_{ogp} (Ecuación 4.1) se utilizaron los conjuntos de datos DS04, DS05, DS06 y DS07 con climas árido-cálido y cálido-templado obtenidos de estaciones diferentes y distantes geográficamente de las utilizadas para su entrenamiento.

La Tabla 4.9 muestra los valores de RMSE y R^2 obtenidos con los modelos ET_{ogp} y Hargreaves-Samani. Los gráficos de dispersión y series de tiempo de la Figura 4.3, Figura 4.4, Figura 4.5 y Figura 4.6 muestran que el modelo ET_{ogp} presenta mayor precisión que el modelo Hargreaves-Samani con respecto al modelo de referencia FAO-PM.

Tabla 4.9: Métricas obtenidas con los modelos ET_{ogp} y Hargreaves-Samani

Conjunto de datos	ET_{ogp}		Hargreaves-Samani	
	RMSE	R^2	RMSE	R^2
DS04	0.31488	0.9783	0.63578	0.9166
DS05	0.68269	0.9188	0.71055	0.9187
DS06	1.4834	0.8184	1.6166	0.7678
DS07	0.29217	0.937	0.40086	0.8672

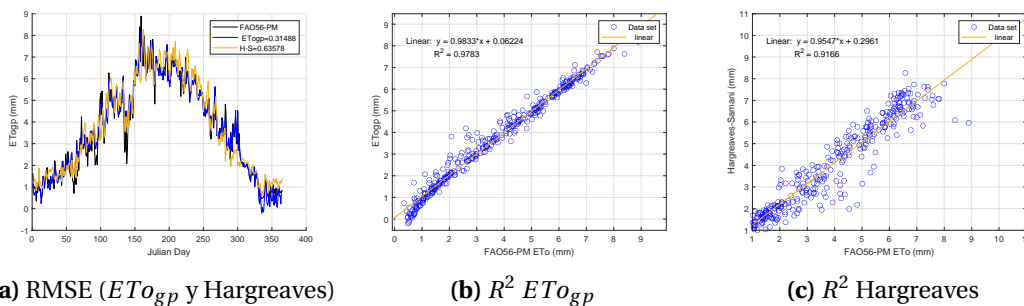
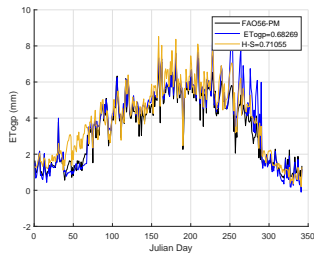
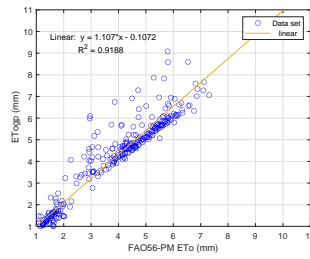


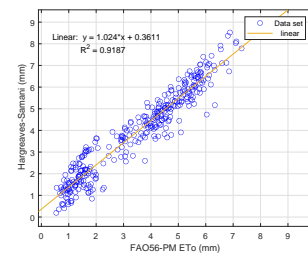
Figura 4.3: Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS04.



(a) RMSE (ET_{top} y Hargreaves)

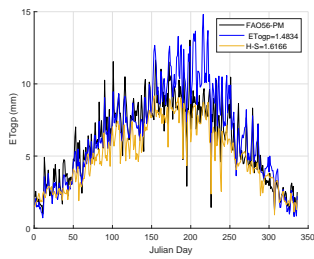


(b) R^2 ET_{top}

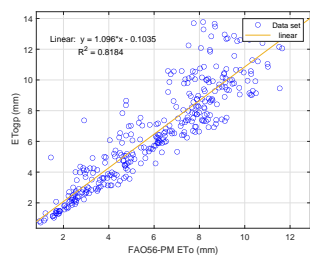


(c) R^2 Hargreaves

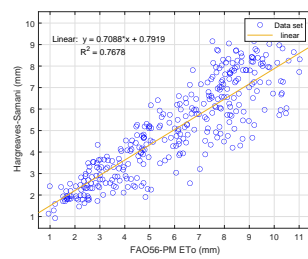
Figura 4.4: Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS05.



(a) RMSE (ET_{top} y Hargreaves)

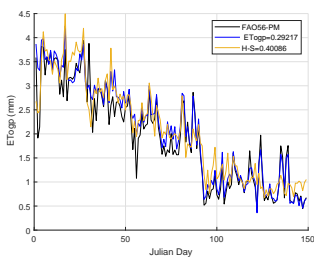


(b) R^2 ET_{top}

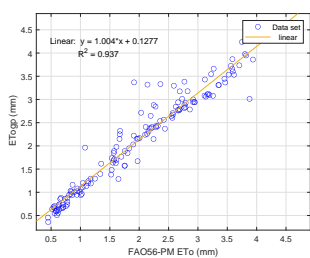


(c) R^2 Hargreaves

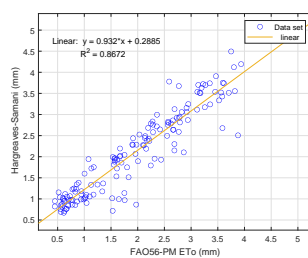
Figura 4.5: Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS06.



(a) RMSE (ET_{top} y Hargreaves)



(b) R^2 ET_{top}



(c) R^2 Hargreaves

Figura 4.6: Series de tiempo y gráficos de dispersión obtenidos con el conjunto de datos DS07.

De acuerdo a los resultados obtenidos en el proceso de validación, se observa que el modelo ET_{ogp} muestra una mayor precisión en la estimación de la evapotranspiración en comparación con el modelo Hargreaves-Samani en la mayoría de los conjuntos de datos evaluados. Para los conjunto de datos DS04, DS05, DS06 y DS07, el modelo ET_{ogp} exhibe un menor RMSE y un mayor R^2 en comparación con el modelo Hargreaves-Samani, lo que indica una mejor capacidad de predicción. Estos hallazgos sugieren que el modelo ET_{ogp} , desarrollado mediante Programación Genética, ofrece una precisión notable en la estimación de la evapotranspiración de referencia utilizando parámetros meteorológicos como la velocidad del viento, la temperatura y la humedad relativa.

4.4. Conclusiones

En este capítulo, se presentó el desarrollo de un modelo para estimar la evapotranspiración a campo abierto utilizando la Fase 1 de la metodología planteada en la sección 1.5. La metodología, demostró ser altamente eficaz para generar modelos precisos. La identificación de fuentes fiables y el preprocesamiento meticuloso de los datos resultaron en un conjunto de datos robusto, facilitando la generación de un modelo preciso. El modelo desarrollado mostró un rendimiento superior en comparación con modelos de referencia tradicionales como el FAO-PM y Hargreaves-Samani, evidenciado por la alta precisión. Las estaciones para validación provinieron de ubicaciones geográficas y regiones diferentes, lo que destacó la capacidad del modelo para generalizar y adaptarse a diversas condiciones climáticas. Los resultados subrayan la relevancia y aplicabilidad del modelo en la gestión agrícola y de recursos hídricos, destacando su capacidad para contribuir a una mejor planificación del riego y optimización del uso del agua, aspectos críticos en la agricultura sostenible y en la mitigación del cambio climático. Además, la utilización de Programación Genética representa una innovación significativa en el campo de la hidrología, aportando una nueva perspectiva metodológica y abriendo la puerta a futuras investigaciones y aplicaciones en diversas áreas.

Capítulo 5

Desarrollo de un modelo para estimar la evapotranspiración en invernaderos

El desarrollo de un modelo para estimar la evapotranspiración implica dos procesos fundamentales, cada uno compuesto por actividades esenciales. El primero de estos procesos, denominado «Recolección y Preparación de Datos», se subdivide en tres actividades principales. En primer lugar, se lleva a cabo la adquisición de datos relevantes que pueden venir de diversas fuentes, como estaciones meteorológicas, sensores remotos, archivos históricos de parámetros meteorológicos u otras fuentes de información meteorológica. Posteriormente, se realiza el preprocesamiento de estos datos, lo que implica una serie de pasos para garantizar su calidad, integridad y coherencia, como la limpieza de datos, la detección y corrección de errores, y la interpolación de valores faltantes. Finalmente, se procede a la creación de conjuntos de datos específicos destinados al entrenamiento y prueba del modelo, dividiendo los datos disponibles en conjuntos de entrenamiento y prueba, con el fin de garantizar la generalización y robustez del modelo ante diferentes condiciones. Este proceso de recolección y preparación de datos sienta las bases sólidas para el desarrollo y validación del modelo de evapotranspiración.

El segundo proceso, denominado «Proceso Evolutivo y Generación del Modelo», se compone de una serie de cinco actividades fundamentales. En primer lugar, se llevan a cabo los pasos preparatorios necesarios para la ejecución del algoritmo evolutivo, los cuales incluyen la configuración de parámetros y la preparación de los datos de entrada. Posteriormente, se procede a la ejecución del algoritmo evolutivo en sí mismo, con el fin de explorar y optimizar soluciones potenciales para el modelo de evapotranspiración. Durante este proceso, se generan y evalúan múltiples soluciones candidatas, utilizando métodos de búsqueda basados en principios evolutivos. Una vez completada la ejecución del algoritmo, se lleva a cabo la depuración de las soluciones más prometedoras. Luego, se procede a la selección del mejor modelo resultante, basándose en criterios de desempeño predeterminados, complejidad estructural y evaluación de su capacidad para representar de manera precisa el fenómeno de la evapotranspiración.

5.1. Recolección y preparación de datos

El proceso de recolección y preparación de datos desempeña un papel fundamental en la construcción de nuestro modelo predictivo. Su objetivo principal radica en organizar y depurar los datos disponibles para crear conjuntos de datos coherentes y de alta calidad que servirán como base para el análisis y modelado subsiguiente. Este proceso implica una serie de actividades esenciales que se detallan a continuación, cada una de las cuales contribuye a garantizar la integridad y la idoneidad de los datos para su uso en el desarrollo del modelo predictivo.

5.1.1. Obtención de datos

En esta actividad es crucial recopilar una amplia gama de datos relevantes para el análisis y modelado posterior. Se obtienen datos clave como la temperatura máxima y mínima, la humedad relativa máxima y mínima, la velocidad del viento y la radiación solar. Estos datos se pueden recolectar de diversas fuentes, como archivos históricos de parámetros meteorológicos, estaciones meteorológicas ubicadas en entornos de campo abierto, así como también en estructuras específicas utilizadas en la agricultura protegida, tales como invernaderos, microtúneles, casas sombra, entre otros. La diversidad de las fuentes de datos permite capturar las variaciones climáticas tanto en ambientes al aire libre como en ambientes controlados, lo cual es fundamental para comprender y predecir con precisión los patrones de evapotranspiración en diferentes contextos agrícolas. Las estaciones meteorológicas situadas en campo abierto proporcionan datos representativos de las condiciones climáticas generales de la región, mientras que las estaciones dentro de estructuras agrícolas protegidas ofrecen información específica sobre las condiciones microclimáticas presentes en esos entornos particulares. Para el desarrollo de nuestro modelo de ETo se obtuvieron datos del sistema de invernaderos del Departamento de Agricultura de la Universidad Americana de Beirut (AUB), Líbano.

Base de datos

Los datos obtenidos son de un invernadero con una superficie total de 72 m^2 ($5.25 \times 13.7\text{ m}$) y una altura de 4.55 m con orientación de este a oeste. El recubrimiento del invernadero fue con una película de policarbonato transparente. La ventilación fue inducida mediante un ventilador eléctrico de succión ubicado en el muro sur. La velocidad media del viento horizontal diaria generada por el ventilador de succión fue de 1.0 m/s (medida con un anemómetro fuera del invernadero) (Jaafar y Ahmad, 2018). La Figura 5.1 muestra la ubicación geográfica del área de invernaderos del Departamento de Agricultura de la Universidad Americana de Beirut (AUB), Líbano.

La base de datos obtenida cuenta con 476 registros que corresponden a un invernadero ventilado y un invernadero no ventilado y van desde el 1 de enero de 2015 al 5 de enero de 2017 periodo que contempló el crecimiento de un cultivo de orégano «*Origanum*» (Jaafar y Ahmad, 2018). La estructura del archivo original de datos se observa en la Tabla 5.1



Figura 5.1: Ubicación geográfica del área de invernaderos del Departamento de Agricultura de la Universidad Americana de Beirut (AUB), Líbano.

Tabla 5.1: Estructura del archivo original de datos proporcionado por la AUB

No.	Campo	No.	Campo
1.	Fecha	8.	Temperatura promedio [°C]
2.	Día juliano	9.	Humedad relativa máxima [%]
3.	Radiación solar externa [MJm^2]	10.	Humedad relativa mínima [%]
4.	Radiación extraterrestre [MJm^2d^{-1}]	11.	Humedad relativa promedio [%]
5.	Transmisibilidad [%]	12.	ETo Hargreaves-Samani [mm]
6.	Temperatura máxima [°C]	13.	ETo Atmómetro [mm]
7.	Temperatura mínima [°C]		

5.1.2. Preprocesamiento de datos

El preprocesamiento de datos es una etapa crítica en el desarrollo de modelos predictivos, donde se abordan diversos aspectos para garantizar la calidad y la coherencia de los datos. Esto implica la detección y tratamiento de datos atípicos, que pueden ser eliminados, generados o reducidos en su influencia, así como la gestión de datos faltantes. La base de datos original fue sometida a un proceso de separación para aislar los parámetros meteorológicos de interés y registros del invernadero ventilado, lo que arrojó un total de 306 registros. Posteriormente, se llevó a cabo un análisis exploratorio para detectar posibles datos faltantes o anomalías. Tras este análisis, se llegó a la conclusión de que los datos carecían de anomalías significativas, aunque se identificaron algunas lagunas en la información. Cuando se detectaron datos faltantes correspondientes a un día, se optó por completarlos utilizando interpolación simple, como se describe en la ecuación (Ecuación 5.1). Sin embargo, en los casos en los que la ausencia de datos se extendía a períodos más largos, se tomó la decisión de eliminar dichos registros para preservar la integridad y precisión del conjunto de datos utilizado en el análisis. Los parámetros meteorológicos de interés junto con sus unidades se observan en la Tabla 5.2.

$$y_i = y_a + \frac{(t_i - t_a)(y_p - y_a)}{(t_p - t_a)} \quad (5.1)$$

Donde:

- y_i = dato a interpolar
- y_a = dato anterior
- y_p = dato posterior
- t_i = tiempo actual
- t_a = tiempo anterior
- t_p = tiempo posterior

Tabla 5.2: Parámetros meteorológicos de interés

Temperatura Máxima [°C]	Temperatura Mínima [°C]	Humedad Relativa Máxima [%]	Humedad Relativa Mínima [%]	Radiación solar [MJm ² dia]
28.50	15.70	72.50	27.70	17.87
31.10	18.00	68.60	26.80	16.60
29.30	17.00	78.50	28.20	15.43
28.20	17.70	83.50	37.70	14.77
...
27.80	15.40	83.20	35.40	2.80

5.1.3. Creación de conjuntos de datos

Después del preprocesamiento de la base de datos original se procedió a la creación de conjuntos de datos para entrenamiento y prueba. Esto se llevó a cabo mediante un proceso

de muestreo uniforme de la base de datos. Para obtener los datos de prueba, se implementó una estrategia de muestreo que consistía en seleccionar uno de cada cinco registros de la base de datos preprocesada, reservando el resto para el conjunto de entrenamiento. Esta metodología de selección aseguró la representatividad de los datos de prueba sin comprometer la integridad del conjunto de entrenamiento.

Como resultado de este enfoque, se generaron dos conjuntos de datos: uno diseñado para el entrenamiento, con 245 registros, y otro utilizado para evaluar el rendimiento del modelo, con 61 registros. Los conjuntos de datos para entrenamiento y prueba se observan en la Tabla 5.3 y la Tabla 5.4. Esta división estratégica permitió realizar una evaluación precisa del modelo mediante la comparación de su desempeño con datos independientes. De esta manera, se garantizó una evaluación robusta y fiable del modelo desarrollado.

Tabla 5.3: Conjunto de datos utilizado para entrenamiento

Día Juliano	Temperatura Máxima [°C]	Temperatura Mínima [°C]	Temperatura Promedio [°C]	Humedad Relativa Máxima [%]	Humedad Relativa Mínima [%]	Radiación solar [$MJm^2 dia$]	Velocidad Viento [ms^{-1}]
66	28.50	15.70	21.52	72.50	27.70	17.86	1
67	31.10	18.00	22.83	68.60	26.80	16.59	1
68	29.30	17.00	21.69	78.50	28.20	15.43	1
69	28.20	17.70	22.35	83.50	37.70	14.77	1
71	25.40	16.00	19.39	81.90	39.00	9.31	1
72	28.50	15.30	20.49	83.20	37.50	18.42	1
...
337	27.80	15.40	18.03	83.20	35.40	2.79	1

Tabla 5.4: Conjunto de datos utilizado para prueba

Día Juliano	Temperatura Máxima [°C]	Temperatura Mínima [°C]	Temperatura Promedio [°C]	Humedad Relativa Máxima [%]	Humedad Relativa Mínima [%]	Radiación solar [$MJm^2 dia$]	Velocidad Viento [ms^{-1}]
70	25.50	19.20	21.73	83.60	52.60	9.21	1
75	28.30	15.30	20.46	69.80	27.60	19.87	1
80	26.50	15.10	18.43	78.36	40.10	5.28	1
85	30.40	16.50	22.03	76.70	28.60	22.09	1
90	27.20	17.30	21.80	79.60	40.00	19.34	1
...
334	26.50	16.10	19.99	54.30	33.00	5.57	1

5.2. Proceso evolutivo y generación del modelo

El proceso evolutivo y generación del modelo de evapotranspiración, además de abordar los factores ambientales convencionales como humedad relativa, radiación solar, temperatura del aire y velocidad del viento, también aprovechó la técnica avanzada de cómputo evolutivo mediante Programación Genética. Esta técnica, inspirada en la evolución biológica, busca optimizar modelos matemáticos mediante la selección, combinación y modificación de componentes de manera similar a la selección natural. En este contexto, la Programación Genética fue utilizada para evolucionar modelos de evapotranspiración explorando una amplia gama de posibles estructuras y parámetros, adaptándose dinámicamente a los cambios del entorno y datos observados. Este proceso consta de cuatro actividades

fundamentales que son críticas para la obtención del modelo de evapotranspiración. Estas actividades deben ser llevadas a cabo de manera cuidadosa y precisa para garantizar la precisión y la fiabilidad del modelo. A continuación, se detallan y se amplían estas actividades:

5.2.1. Realizar pasos preparatorios

Antes de iniciar con la Programación Genética de acuerdo con (Koza, 1992), es esencial realizar cinco pasos preparatorios básicos para garantizar un proceso eficaz y exitoso. Estos pasos incluyen definir el conjunto de funciones y terminales, seleccionar una función de aptitud adecuada, configurar los parámetros de ejecución del algoritmo evolutivo y designar resultados. Estos pasos son cruciales para establecer una base sólida y maximizar las posibilidades de obtener resultados óptimos al aplicar la Programación Genética.

Definir el conjunto de terminales

Definir el conjunto de terminales es un paso crítico en la preparación para la Programación Genética debido a su impacto directo en la calidad y eficacia del algoritmo evolutivo. Los terminales representan los componentes básicos o variables que se utilizarán para construir los modelos candidatos. En el contexto de la evapotranspiración las terminales son los parámetros meteorológicos que influyen en el proceso. Es esencial seleccionar cuidadosamente estas terminales para garantizar que abarquen todos los aspectos relevantes del problema y proporcionen una representación completa de los datos disponibles. Esto implica considerar tanto los parámetros directamente relacionados con el fenómeno de evapotranspiración, como también variables derivadas o calculadas que pueden mejorar la precisión del modelo. La selección de las terminales debe basarse en un conocimiento profundo del problema y en la revisión de la literatura especializada para identificar los factores más influyentes.

En nuestra investigación, se llevó a cabo una cuidadosa selección del conjunto de terminales, abarcando los principales parámetros meteorológicos cruciales para la estimación precisa de la evapotranspiración de referencia. Estos incluyeron temperaturas máximas y mínimas, humedad relativa máxima y mínima, velocidad del viento, que son factores fundamentales en el proceso de evapotranspiración. Además de estos parámetros básicos, se enriqueció el conjunto de terminales al incorporar variables derivadas del modelo de referencia FAO56-PM (Ecuación 2.1), una fuente altamente reconocida y validada en hidrología. Esta adición de conocimiento experto permitió capturar una mayor variedad de matices en el proceso de evapotranspiración, mejorando así la capacidad predictiva del algoritmo evolutivo. El conjunto final de terminales utilizado en el algoritmo se presenta detalladamente en la Tabla 5.5, resaltando tanto la diversidad como la relevancia de las variables consideradas. Esta meticulosa selección y diversificación de terminales proporciona una base sólida para la estimación precisa de la evapotranspiración de referencia, vital para la gestión efectiva de recursos hídricos en diversos contextos agrícolas y medioambientales.

Tabla 5.5: Conjunto de terminales enriquecido con conocimiento experto* del modelo FAO56-PM

No.	Terminal	Descripción	Expresión matemática
1.	* Δ	Pendiente de la curva de presión de vapor	$\Delta = \frac{4098[0.618 * \exp(\frac{17.27 * T}{T+237.3})]}{(T+237.2)^2}$
2.	Tmax	Temperatura máxima	
3.	Tmin	Temperatura mínima	
4.	Tprom	Temperatura promedio	
5.	HRmax	Humedad relativa máxima	
6.	HRmin	Humedad relativa mínima	
7.	U_2	Velocidad del viento	
8.	*DJ	Día del año	
9.	* π	Número Pi	
10.	*reflexion	Constante de reflexión solar	
11.	*gt	Grados totales	
12.	*lt	Latitud	
13.	*P	Presión atmosférica	$P = 101.3(\frac{293-0.0065z}{293})^{5.26}$
14.	* γ	Constante psicrométrica	$\gamma = \frac{c_p P}{\epsilon \lambda}$
15.	* $e^o T_{min}$	Presión de vapor a temperatura mínima	$e^o T_{min} = 0.618 * \exp(\frac{17.27 * T_{min}}{T_{min}+237.3})$
16.	* $e^o T_{max}$	Presión de vapor a temperatura máxima	$e^o T_{max} = 0.618 * \exp(\frac{17.27 * T_{max}}{T_{max}+237.3})$
17.	* e_s	Presión de vapor a saturación	$e_s = \frac{e^o T_{min} + e^o T_{max}}{2}$
18.	* e_a	Presión de vapor actual	$e_a = \frac{e^o T_{min} * \frac{HR_{max}}{100} + e^o T_{max} * \frac{HR_{min}}{100}}{2}$
19.	* $e_s - e_a$	Déficit de presión de vapor	
20.	* d_r	Distancia relativa de la Tierra al Sol	$d_r = 1 + 0.033 * \cos(\frac{2\pi}{365} DJ)$
21.	* d_s	Declinación solar	$\delta = 0.409 * \sin(\frac{2\pi}{365} DJ - 1.39)$
22.	* w_s	Ángulo de radiación	$w_s = \arcsin[-\tan(\theta) \tan(\delta)]$
23.	*N	Fotoperíodo	$N = \frac{24}{\pi} * w_s$
24.	* R_a	Radiación Extraterrestre	$R_a = \frac{24 * 60}{\pi} G_{sc} d_r [w_s \sin(lt) \sin(\delta) + \cos(lt) \cos(\delta) \sin(w_s)]$
25.	*Constante	0.48	
26.	*Constante	900	
27.	*Constante	273	
28.	*Constante	1	
29.	*Constante	0.34	
30.	*Constante	0.4718	

Definir el conjunto de funciones

El conjunto de funciones es fundamental por varias razones. En primer lugar, estas funciones constituyen el «lenguaje» con el cual se construyen los modelos candidatos durante la evolución del algoritmo evolutivo. Al seleccionar cuidadosamente estas funciones, se garantiza que el espacio de búsqueda sea lo suficientemente amplio como para abarcar una amplia gama de posibles soluciones. Además, el conjunto de funciones tiene un impacto directo en la capacidad del algoritmo evolutivo para explorar eficientemente el espacio de búsqueda. Una selección diversa y equilibrada de funciones permite una exploración más completa y efectiva de las posibles soluciones, lo que puede conducir a una convergencia más rápida hacia soluciones óptimas o cercanas a óptimas. Otro aspecto importante es que las funciones deben ser adecuadas para el problema específico que se está abordando. Esto significa que deben ser capaces de manipular y transformar adecuadamente los datos relevantes para el problema, ya sea realizando operaciones matemáticas, lógicas o de otro tipo. La elección incorrecta de funciones puede limitar la capacidad del algoritmo para resolver el problema de manera efectiva.

En nuestra investigación, el conjunto de funciones se ha seleccionado cuidadosamente para abarcar una amplia gama de operadores y operaciones matemáticas fundamentales utilizadas en estudios hidrológicos, basándonos en referencias previas (Mehdizadeh et.al., 2017; Mattar, 2018). Este conjunto incluye operadores básicos junto con funciones trigonométricas, exponenciales y logarítmicas, que son comúnmente empleadas en el análisis de datos hidrológicos. Además, hemos incorporado funciones hiperbólicas, las cuales han demostrado ser útiles para explorar fenómenos físicos más complejos, como se detalla en estudios relevantes (Pandir y Ulusoy, 2013). La Tabla 5.6 proporciona una descripción detallada del conjunto completo de funciones utilizadas en el algoritmo evolutivo.

Tabla 5.6: Conjunto de funciones

No.	Función	Expresión	No.	Función	Expresión
1.	Seno	$\sin(\cdot)$	13	Secante hiperbólica	$\operatorname{sech}(\cdot)$
2.	Coseno	$\cos(\cdot)$	14	Cosecante hiperbólica	$\operatorname{csch}(\cdot)$
3.	Arcoseno	$\operatorname{asin}(\cdot)$	15	Cosecante	$\operatorname{csc}(\cdot)$
4.	Arcotangente	$\operatorname{atan}(\cdot)$	16	Secante	$\operatorname{sec}(\cdot)$
5.	Tangente hiperbólica	$\operatorname{tanh}(\cdot)$	17	Cotangente	$\operatorname{cot}(\cdot)$
6.	Raíz cuadrada	$\sqrt{\cdot}$	18	Suma	+
7.	Exponente	$x^{(\cdot)}$	19	Resta	-
8.	Funcion exponencial base e	$e^{(\cdot)}$	20	Multiplicación	*
9.	Tangente	$\operatorname{tan}(\cdot)$	21	División	÷
10.	Coseno hiperbólico	$\operatorname{cosh}(\cdot)$	22	Logaritmo natural	$\log(\cdot)$
11.	Seno hiperbólico	$\operatorname{sinh}(\cdot)$	23	Logaritmo base 10	$\log_{10}(\cdot)$
12.	Cotangente hiperbólica	$\operatorname{coth}(\cdot)$			

Definición de la función de aptitud «Fitness»

La función de aptitud desempeña un papel fundamental en la Programación Genética, ya que actúa como el criterio para evaluar qué tan bien se desempeñan los modelos candidatos en la resolución del problema dado. Es esencial, porque guía el proceso evolutivo al proporcionar una medida objetiva del rendimiento de cada individuo en la población. Una función de aptitud bien diseñada y adecuada al problema permite identificar y seleccionar las soluciones más prometedoras para la reproducción y la creación de la próxima generación. Además, influye en la dirección y velocidad de la evolución al proporcionar retroalimentación sobre la calidad de las soluciones, lo que ayuda al algoritmo evolutivo a converger hacia soluciones óptimas o cercanas a óptimas en un tiempo razonable. La elección de una función de aptitud adecuada es crucial, ya que una función inapropiada puede conducir a resultados subóptimos o incluso a una convergencia prematura hacia soluciones no óptimas. Por lo tanto, la importancia de la función de aptitud radica en su capacidad para dirigir y guiar el proceso de búsqueda hacia soluciones efectivas y útiles para el problema que se está abordando en la Programación Genética.

En nuestra investigación, la función de aptitud se fundamentó en el cálculo de la raíz del error cuadrático medio (RMSE) (Ecuación 5.2). Esta métrica se utiliza para evaluar la discrepancia promedio entre los valores predichos por los programas candidatos y los valores objetivos. Una menor discrepancia indica un mejor ajuste del modelo a los datos. En consecuencia, los individuos que presentan una mayor aptitud, es decir, aquellos cuyas predicciones se acercan más a los valores objetivos, tienen una mayor probabilidad de ser seleccionados para pasar a la siguiente generación durante el proceso evolutivo. La elección

del RMSE como función de aptitud proporciona una medida objetiva y cuantitativa del desempeño de los programas candidatos. Esto ayuda a dirigir la búsqueda hacia soluciones precisas y efectivas para el problema abordado en nuestra investigación.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (5.2)$$

donde n es el número total de datos, P_i y O_i son los valores de evapotranspiración estimados y observados, respectivamente.

Parámetros del algoritmo evolutivo y designación de resultados

La importancia de definir adecuadamente los parámetros del algoritmo evolutivo y la designación de resultados radica en su influencia directa en el rendimiento y la eficacia de la Programación Genética aplicada al problema específico que se está abordando. Estos parámetros incluyen aspectos como el tamaño de la población, la tasa de mutación, la tasa de recombinación, el número de generaciones, entre otros. La elección adecuada de estos parámetros puede afectar significativamente la capacidad del algoritmo para explorar eficazmente el espacio de búsqueda de soluciones. Por ejemplo, un tamaño de población demasiado pequeño puede limitar la diversidad genética y reducir la capacidad del algoritmo para encontrar soluciones óptimas, mientras que un tamaño de población demasiado grande puede aumentar los requisitos computacionales sin proporcionar beneficios significativos en la calidad de las soluciones encontradas. Además, los parámetros del algoritmo también pueden influir en la convergencia del algoritmo hacia soluciones óptimas o subóptimas. Por ejemplo, una tasa de mutación alta puede introducir una mayor diversidad genética en la población, lo que puede ser beneficioso para evitar la convergencia prematura hacia soluciones subóptimas, pero también puede aumentar el riesgo de perder soluciones potencialmente prometedoras. Por otro lado, una tasa de mutación baja puede conducir a una convergencia más rápida, pero también puede aumentar el riesgo de estancamiento en mínimos locales. Además de definir los parámetros del algoritmo, la designación adecuada de resultados también es crucial para evaluar el desempeño del algoritmo y comparar diferentes enfoques o configuraciones. Esto implica definir métricas de evaluación adecuadas que reflejen la calidad y la relevancia de las soluciones encontradas en relación con el problema específico que se está abordando. Estas métricas pueden incluir la precisión, la eficiencia computacional, la robustez, entre otras.

En nuestro estudio, detallamos los parámetros de ejecución para el entrenamiento del algoritmo evolutivo en la Tabla 5.7. Para asegurar una configuración óptima, llevamos a cabo múltiples iteraciones de prueba y error. Algunos parámetros fueron ajustados a través de este proceso, mientras que otros se tomaron de configuraciones previamente establecidas, reconocidas por su eficacia en diversas aplicaciones. Este enfoque meticuloso garantizó que los valores de los parámetros estuvieran óptimamente adaptados al contexto específico de nuestra investigación. Además, establecimos criterios de terminación mediante la fijación de un número máximo de generaciones, en este caso, 50. Este límite asegura que el proceso evolutivo converja hacia una solución óptima en cada ejecución, evitando ejecuciones

prolongadas y permitiendo un control efectivo del tiempo y los recursos computacionales utilizados en el entrenamiento del algoritmo. Estos parámetros y criterios de terminación son fundamentales para garantizar la eficacia y la consistencia del proceso de optimización en nuestra investigación.

Tabla 5.7: Parámetros usados en el algoritmo evolutivo

Parámetro	Valor
Tamaño de la población	400
Número de generaciones	50
Probabilidad de cruce	80%
Probabilidad de mutación	20%
Profundidad máxima de árbol	14
Método de selección	lexicotour
Método de iniciación	Mitad-Mitad
Elitismo	Mantener el mejor

5.2.2. Ejecución del algoritmo evolutivo

Una ventaja del uso de computación evolutiva para modelar la evapotranspiración de referencia es su capacidad para explorar y buscar soluciones en un espacio de búsqueda complejo y multidimensional de manera eficiente. La evapotranspiración de referencia es influenciada por una variedad de factores meteorológicos y ambientales, lo que hace que su modelado sea un problema desafiante. Los algoritmos evolutivos pueden adaptarse y aprender de manera autónoma a partir de iteraciones sucesivas, lo que les permite encontrar soluciones óptimas o cercanas a óptimas en entornos complejos y dinámicos. Además, estos algoritmos pueden manejar datos ruidosos o incompletos y no requieren suposiciones específicas sobre la forma funcional de la relación entre las variables, lo que los hace flexibles y adaptables a una variedad de situaciones y condiciones climáticas.

En el marco de nuestra investigación, se empleó la biblioteca GPLab dentro del entorno de MatLab para llevar a cabo la ejecución del algoritmo evolutivo y desarrollar nuestro modelo. Esta biblioteca se reveló como una herramienta fundamental que no solo nos permitió diseñar el modelo de manera efectiva, sino también analizarlo con gran precisión. Sus funcionalidades avanzadas fueron clave para la implementación de los algoritmos específicos requeridos por nuestros objetivos de investigación. La elección de GPLab se basó en su versatilidad, efectividad y capacidad para abordar los desafíos particulares de nuestro proyecto, lo que resultó en la obtención de resultados precisos y significativos dentro del ámbito de nuestro estudio.

El algoritmo implementado con GPLab genera la población inicial de soluciones combinando de manera aleatoria elementos de los conjuntos de terminales y funciones proporcionados en la tabla 5.5 y la tabla 5.6 respectivamente. Luego, cada individuo de esta población inicial (representado gráficamente en forma de árbol como se observa en la figura 2.2), se evalúa utilizando la función de aptitud (fitness) (Ecuación 5.2). Posteriormente, se seleccionan soluciones candidatas para clasificar a todos los individuos y eliminar aquellas con baja aptitud. A continuación, se lleva a cabo la recombinación genética entre los individuos seleccionados mediante operadores de cruce y mutación como se observa en la figura 2.8 y la figura 2.9 respectivamente. Por último, la siguiente generación se forma a partir de la selección de las mejores soluciones entre padres e hijos. Estos pasos se repiten hasta que se alcance el número máximo de generaciones establecido en los parámetros de ejecución. Cabe mencionar que el algoritmo evolutivo se ejecutó 30 iteraciones independientes con 50 generaciones cada una, y el proceso completo tuvo una duración de 3 horas y 3 minutos. El diagrama de flujo del proceso evolutivo se muestra en la Figura 2.4.

Estadística evolutiva

Una vez terminado la ejecución del algoritmo evolutivo realizamos un análisis estadístico con el objetivo de mostrar la tasa de convergencia del algoritmo evolutivo. La Figura 5.2 presenta el mejor ajuste, el promedio de los mejores ajustes y la desviación estándar de las 30 ejecuciones.

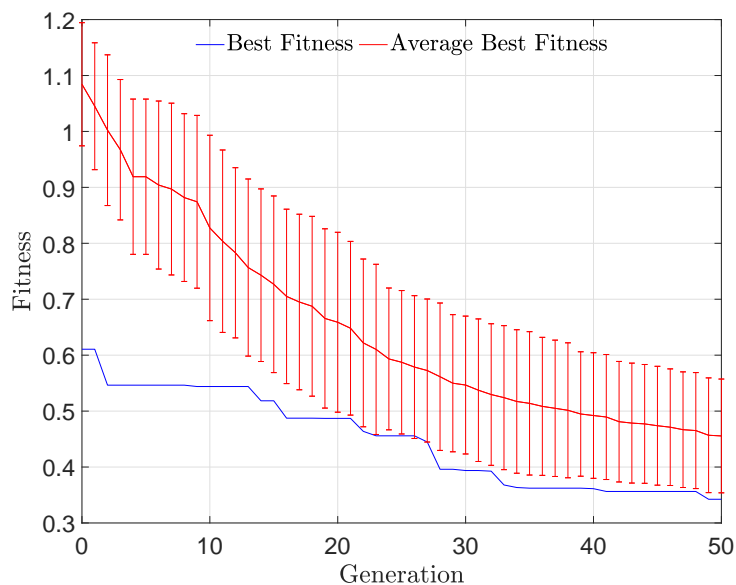


Figura 5.2: Mejor aptitud, promedio de la mejor aptitud y desviación estándar a lo largo de 50 generaciones

Es importante destacar que, en promedio, se encontraron soluciones con una medida de fitness menor a 0.7 a partir de la generación 17. El fitness mínimo se alcanzó alrededor de la generación 49, y el mejor fitness convergió a 0.61 a partir de la primera generación. Estos resultados indican una rápida convergencia del algoritmo hacia soluciones óptimas, lo que demuestra su eficacia en la optimización del problema planteado.

5.2.3. Depurar las mejores soluciones

Depurar las mejores soluciones implica identificar y seleccionar aquellas que ofrecen un mejor desempeño o una mayor aptitud dentro de la población de individuos generados por el algoritmo evolutivo. Sin embargo, a pesar de que la solución que se encuentra en la última generación se considera la mejor en términos de precisión, en ciertas ocasiones, es necesario sacrificar algo de esta precisión en favor de encontrar una estructura más simple o menos compleja en generaciones anteriores. Esta estrategia se fundamenta en la idea de que una estructura más simple puede ser más fácil de comprender, interpretar y aplicar en situaciones del mundo real. Además, una solución menos compleja podría ser más eficiente computacionalmente y requerir menos recursos para su implementación y mantenimiento. Por lo tanto, en la búsqueda de un equilibrio entre precisión y simplicidad, es importante evaluar cuidadosamente las opciones disponibles en diferentes generaciones para encontrar la solución óptima para un problema dado.

Después del proceso evolutivo, se obtuvieron 402 soluciones diferentes con fitness menor a 0.7, de las cuales fueron seleccionadas nueve de ellas. El criterio de selección se basó primero en la aptitud y luego en la complejidad estructural. La Tabla 5.8 presenta estas soluciones, así como los resultados del atmómetro y el modelo Hargreaves-Samani, junto con su rendimiento, evaluado mediante la raíz del error cuadrático medio (RMSE) (Ecuación 5.2) en comparación con el modelo de referencia FAO56-PM (Ecuación 2.1). Los resultados muestran que el enfoque de Programación Genética es altamente preciso en la estimación de la evapotranspiración de referencia. El nuevo modelo de evapotranspiración se denomina ETo_{gpi} , donde «i» representa un número consecutivo. Utilizaremos el acrónimo ETo_{gp} (Evapotranspiración de referencia de Programación Genética) para referirnos a nuestros modelos obtenidos.

Tabla 5.8: Rendimiento de los modelos ETo_{gpi} , Atmómetro y Hargreaves-Samani

No.	Modelo	RMSE
1.	$ETo_{gp1} = e_s + 4d_s + \cos(e_a) + \cosh(e_s) + 0.408$	0.4078
2.	$ETo_{gp2} = \tan(d_s) + 2d_s + \operatorname{sech}(e_s) + \operatorname{atan}(d_s) + \log(\sinh(\operatorname{sech}(e_a))) + \operatorname{sech}(e^o T_{min}) + e_s$	0.4136
3.	$ETo_{gp3} = \exp(w_s) * \log_{10}(w_s + 3d_s + e_s) + \frac{\log(e_s)}{e_s}$	0.4491
4.	$ETo_{gp4} = w_s^{w_s} + (e_s - e_a)^{\tanh(\tanh(\tanh(\tanh(\tanh(e_s - e_a) * \tanh(e_s - e_a))) * w_s)) * w_s}$	0.4394
5.	$ETo_{gp5} = \exp(w_s) - \operatorname{csch}(w_s - 0.34 - \operatorname{delta}) - \operatorname{csch}(e_s - e_a)$	0.4561
6.	$ETo_{gp6} = 3d_s + e_s + \sin(e^o T_{min})$	0.4744
7.	$ETo_{gp7} = \pi * d_s + e^o T_{max} * \tanh(\operatorname{coth}((e_s - e_a)_a^e))$	0.4865
8.	$ETo_{gp8} = \tanh(\tanh(\tanh(2d_s + d_r + 0.408))) + 4d_s + \sin(e_s) + e_s$	0.4989
9.	$ETo_{gp9} = \operatorname{csch}(\operatorname{atan}(e_s)) + e_s - \sin(0.7854 - 2d_s) - 2d_s - 0.6937$	0.532
10.	$ETo_{AT} = \text{Atmómetro}$	1.056
11.	$ETo_{HS} = 0.0023(T_{med} + 17.78)R_a * \sqrt{T_{max} - T_{min}}$	2.360

Frecuencia de uso de las funciones y terminales

El análisis estructural de los modelos ETo_{gp} , como se muestra en la Tabla 5.8, ofrece una visión interesante de las estrategias identificadas por el proceso evolutivo al combinar el conjunto primitivo de terminales y funciones. En este estudio, utilizamos una unidad de frecuencia de uso para medir el número de veces con la que se emplea cada elemento en la generación de nuevos modelos. Las Figuras 5.3 y 5.4 ilustran la frecuencia de ocurrencia de los conjuntos de terminales y funciones, respectivamente. Los números asignados a los terminales y funciones se corresponden con la descripción detallada en las Tablas 5.5 y 5.6.

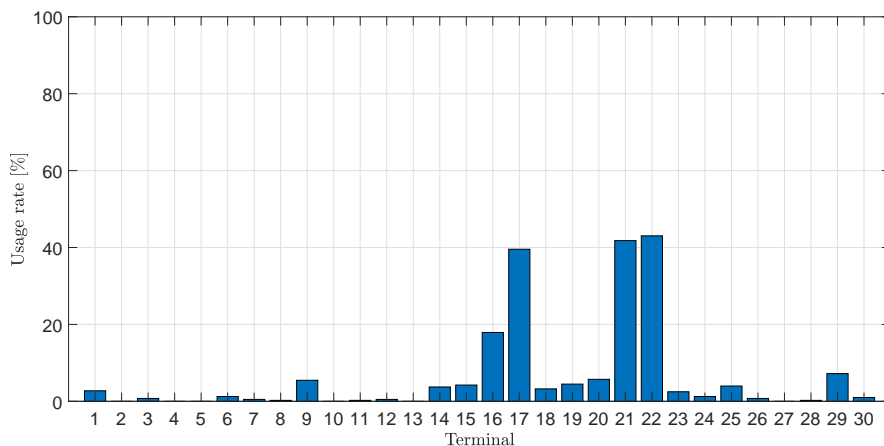


Figura 5.3: Frecuencia de aparición del conjunto de terminales

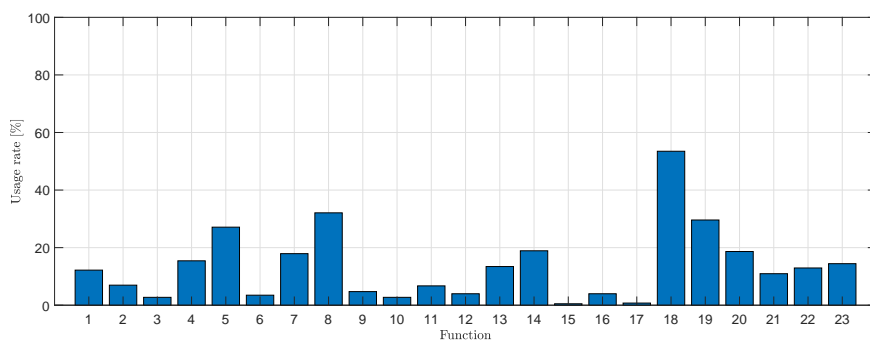


Figura 5.4: Frecuencia de aparición del conjunto de funciones

En cuanto a los elementos pertenecientes al conjunto terminal, se puede observar que el ángulo de radiación solar (ws) tiene la frecuencia de ocurrencia más alta, siendo utilizado en el 42% de los modelos generados. Le sigue la declinación solar (ds) con un 41%, posteriormente la presión de vapor de saturación (es) con un 40% y por último la presión de vapor a temperatura máxima (e^oTmax) con un 18%. Se nota que el algoritmo evolutivo no utilizó la mayoría de las terminales proporcionadas, lo que indica que el algoritmo de Programación Genética los identificó como no estrictamente necesarios para estimar la ETo . Además, es importante destacar que los modelos ETo_{gp} presentados no emplean directamente los parámetros climáticos básicos. Sin embargo, son utilizados implícitamente en variables que representan conocimiento experto, como la presión de vapor de

saturación (e_s), la declinación solar (ds), el déficit de presión de vapor (e_a) y la presión de vapor a temperatura mínima y máxima ($e^o T_{max}$, $e^o T_{min}$). Esto es importante ya que solo se utilizan parámetros climáticos de temperatura y humedad relativa, lo que sugiere que la Programación Genética puede identificar parámetros climáticos fundamentales para estimar la ETo.

Estos parámetros climáticos son cruciales para el cálculo de la evapotranspiración (ETo) debido a su influencia directa en la transferencia de agua desde el suelo y las plantas hacia la atmósfera. El ángulo de radiación solar (ws) y la declinación solar (ds) determinan la cantidad y la intensidad de la radiación solar, la principal fuente de energía para la evapotranspiración. La presión de vapor de saturación (e_s) y la presión de vapor a temperatura máxima ($e^o T_{max}$) y mínima ($e^o T_{min}$) reflejan la capacidad del aire para retener vapor de agua, mientras que el déficit de presión de vapor (e_a) indica la demanda atmosférica de agua. Estos parámetros integran información sobre temperatura y humedad relativa, proporcionando una medida completa de las condiciones climáticas que controlan la tasa de evapotranspiración, y son esenciales para modelos precisos de ETo (Allen, 1998).

En el caso del conjunto de funciones, la importancia de cada operador estaba distribuida de manera uniforme. El operador de adición se utilizó en el 55% de los casos, mientras que el operador de seno fue el menos utilizado con el 12%. Esto es importante ya que la Programación Genética puede construir modelos estructuralmente simples para estimar la ETo. Este hallazgo sugiere que el algoritmo evolutivo reconoció la relación entre parámetros climáticos fundamentales pero efectivos en la estimación del fenómeno ETo.

Las funciones matemáticas como la arcotangente, el seno, la tangente hiperbólica, el exponente, el exponente base e , la secante hiperbólica, la cosecante hiperbólica, el logaritmo natural y el logaritmo base 10 resultan interesantes en el cálculo de la evapotranspiración (ETo) debido a su capacidad para capturar relaciones no lineales y transformaciones complejas en los procesos climáticos. La arcotangente y la tangente hiperbólica suavizan cambios abruptos y normalizan datos, mientras que el seno modela variaciones cíclicas y periódicas, esenciales para las variaciones diurnas y estacionales de radiación solar y temperatura. Las funciones exponenciales y el exponente base e son fundamentales para representar influencias exponenciales de variables como la temperatura en la tasa de evaporación. La secante y la cosecante hiperbólicas son útiles para modelar efectos atenuantes y comportamientos inversos, como la disminución de radiación solar con la altitud y la relación inversa entre humedad relativa y déficit de presión de vapor. Finalmente, los logaritmos natural y base 10 transforman datos exponenciales en escalas más manejables, linealizando relaciones multiplicativas y proporcionando una forma intuitiva de interpretar datos que varían en órdenes de magnitud (Allen, 1998; Hargreaves y Samani, 1985; Monteith, 1965; McMahon et al., 2013; Jensen et al., 1990). Estas funciones permiten modelar con precisión la complejidad y no linealidad de los procesos climáticos que afectan la ETo, mejorando así la robustez de los modelos.

5.2.4. Seleccionar el mejor modelo

En esta actividad para seleccionar el mejor modelo se tomó como referencia la tabla 5.8 ya que presenta una comparación detallada de varios modelos de evapotranspiración desarrollados mediante Programación Genética, así como los resultados de un atmómetro y el modelo Hargreaves-Samani. Cada modelo de evapotranspiración (ETo_{gp1} a ETo_{gp9}) se evalúa en términos de su rendimiento, mediante la raíz del error cuadrático medio (RMSE) en relación con el modelo de referencia FAO56-PM. El análisis revela que el modelo ETo_{gp1} se destaca como el más efectivo, con un RMSE de 0.4078, indicando una alta capacidad de ajuste, además cuenta con una complejidad estructural simple y fácil de interpretar. En contraste, los otros modelos y las referencias del atmómetro y el modelo Hargreaves-Samani muestran valores de RMSE más altos, lo que sugiere un rendimiento inferior en términos de precisión. La figura 5.5 muestra los gráficos de dispersión y series de tiempo con los resultados obtenidos durante el proceso de entrenamiento. Por lo tanto, el modelo ETo_{gp1} emerge como la opción más confiable y adecuada para la estimación precisa de la evapotranspiración de referencia en invernaderos. A partir de este punto, nos referiremos al modelo seleccionado como ETo_{gp} (Ecuación 5.3).

$$ETo_{gp} = e_s + 4d_s + \cos(e_a) + \cosh(e_s) + 0.408 \quad [mm] \quad (5.3)$$

Donde:

$$e_s = \frac{e^o T_{min} + e^o T_{max}}{2} \quad \text{Presión de vapor a saturación [kPa]}$$

$$d_s = 0.409 * \sin\left(\frac{2\pi * DJ}{365} - 1.39\right) \quad \text{Declinación Solar [Rad]}$$

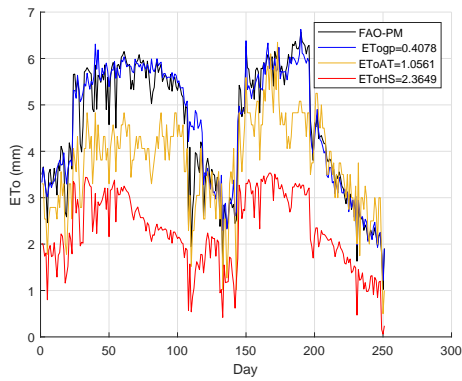
$$e_a = \frac{\frac{e^o T_{min} * HR_{max}}{100} + \frac{e^o T_{max} * HR_{min}}{100}}{2} \quad \text{Déficit de presión de vapor [kPa]}$$

$$e^o T_{min} = 0.6108 * \exp\left(\frac{17.27 * T_{min}}{T_{min} + 237.3}\right) \quad \text{Presión de vapor a temperatura mínima [kPa]}$$

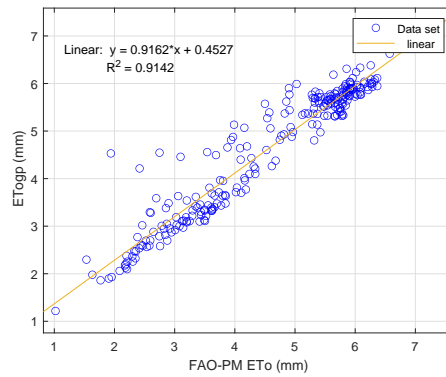
$$e^o T_{max} = 0.6108 * \exp\left(\frac{17.27 * T_{max}}{T_{max} + 237.3}\right) \quad \text{Presión de vapor a temperatura máxima [kPa]}$$

Tabla 5.9: Métricas obtenidas por los modelos ETo_{gp} , Atmómetro (ETo_{AT}) y Hargreaves-Samani (ETo_{HS}).

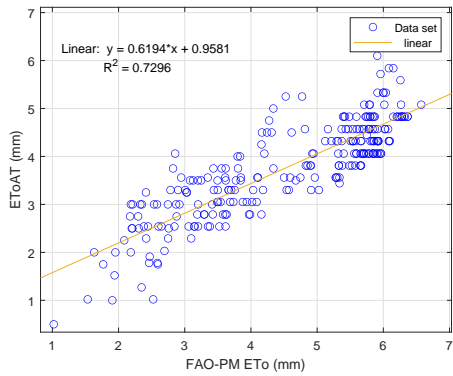
Modelo	RMSE	R^2
ETo_{gp}	0.4078	0.9142
ETo_{AT}	1.056	0.729
ETo_{HS}	2.364	0.841



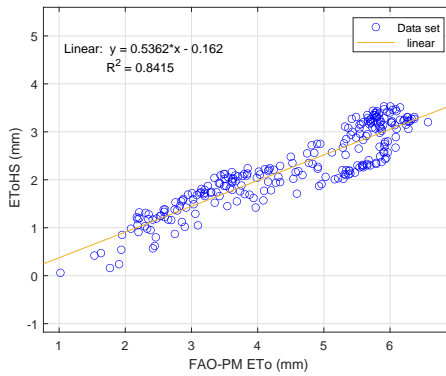
(a) RMSE ET_{0gp} , ET_{0AT} , ET_{0HS}



(b) $R^2 ET_{0gp}$



(c) $R^2 ET_{0AT}$



(d) $R^2 ET_{0HS}$

Figura 5.5: Series de tiempo y gráficos de dispersión obtenidos en la etapa de entrenamiento.

Los resultados obtenidos con el modelo ET_{ogp} muestran una alta precisión en la estimación de la evapotranspiración de referencia dentro de invernaderos, utilizando datos de temperatura y humedad relativa. Este hallazgo es de gran relevancia científica, ya que indica la capacidad del modelo para capturar con precisión los procesos de evaporación y transpiración en un entorno controlado como es el caso de los invernaderos. La precisión de estas estimaciones es fundamental para la gestión eficiente del riego y el control ambiental en la producción de cultivos bajo condiciones de invernadero. Los resultados obtenidos respaldan la utilidad y fiabilidad del modelo ET_{ogp} como una herramienta viable para la estimación precisa de la evapotranspiración de referencia en este contexto específico.

5.3. Validación del modelo

La validación con datos no conocidos es fundamental para evaluar la capacidad de un modelo de extrapolar más allá de los datos empleados en su entrenamiento, permitiéndole operar de manera efectiva en situaciones reales. Este paso es crucial para identificar y reducir el riesgo de sobreajuste, un problema en el cual el modelo se adapta en exceso a las particularidades del conjunto de entrenamiento, limitando su rendimiento cuando se enfrenta a nuevos datos.

La validación del modelo ET_{ogp} en campo abierto con datos no conocidos es esencial para asegurar que el modelo sea robusto y aplicable en diversas condiciones climáticas y ambientales. Este proceso es crucial para garantizar la confiabilidad de las estimaciones, proporcionando una base sólida para tomar decisiones informadas en la gestión de recursos hídricos y en la planificación agrícola a gran escala.

5.3.1. Validación del modelo de evapotranspiración

Para validar el modelo de evapotranspiración ET_{ogp} (Ecuación 5.3), se utilizó el conjunto de datos de prueba que contiene 61 registros y cuya estructura se muestra en la tabla 5.4. Realizamos una comparación de la precisión entre el modelo de Hargreaves-Samani, un atmómetro y el modelo obtenido ET_{ogp} , tomando como referencia el modelo FAO56-PM. Las métricas utilizadas para la evaluación incluyeron el error cuadrático medio (*Root Mean Square Error*) $RMSE$ (Ecuación 5.2) y el coeficiente de determinación (*coefficient of determination*) R^2 (Ecuación 5.4).

El $RMSE$ es una medida de la diferencia promedio entre los valores observados y los valores predichos por el modelo; un valor menor indica un mejor ajuste. El R^2 , por otro lado, indica la proporción de la variabilidad en la variable dependiente que es explicada por el modelo; un valor más cercano a 1 indica un mejor ajuste.

El coeficiente de determinación R^2 se calcula utilizando la siguiente fórmula:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5.4)$$

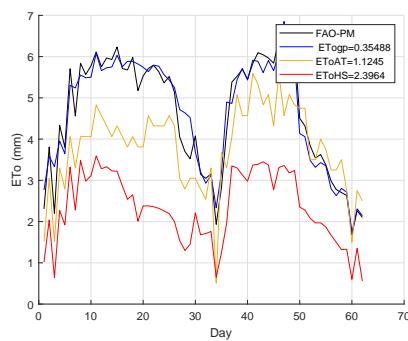
donde:

- y_i son los valores observados.
- \hat{y}_i son los valores predichos por el modelo.
- \bar{y} es la media de los valores observados.
- n es el número de observaciones.

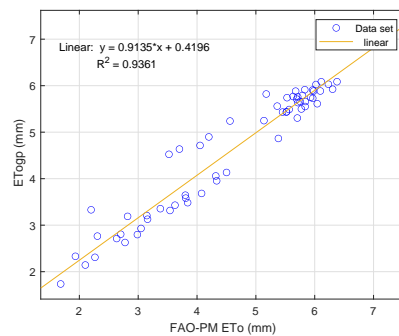
Los resultados mostraron que el modelo ETo_{gp} alcanzó un $RMSE$ de 0.3548 y un R^2 de 0.9361, el atmómetro obtuvo un $RMSE$ de 1.1245 y un R^2 de 0.7158, mientras que el modelo ETo_{HS} logró un $RMSE$ de 2.3964 y un R^2 de 0.8531, como se muestra en la Tabla 5.10 y la Figura 5.6.

Tabla 5.10: Métricas obtenidas por los modelos ETo_{gp} , el ETo_{AT} y ETo_{HS}

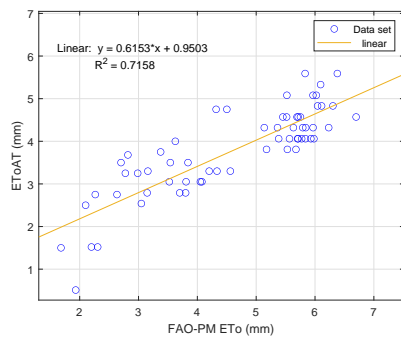
Modelo	RMSE	R^2
ETo_{gp}	0.3548	0.9361
ETo_{AT}	1.1245	0.7158
ETo_{HS}	2.3964	0.8531



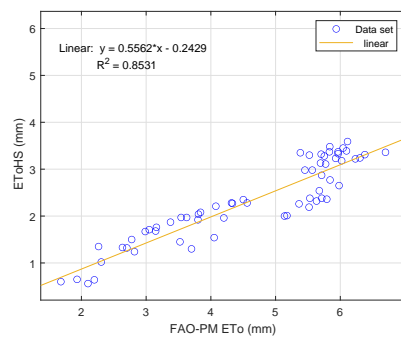
(a) $RMSE ETo_{gp}, ETo_{AT}, ETo_{HS}$



(b) $R^2 ETo_{gp}$



(c) $R^2 ETo_{AT}$



(d) $R^2 ETo_{HS}$

Figura 5.6: Serie temporal y gráficos de dispersión de los modelos ETo_{gp} , el ETo_{AT} y ETo_{HS}

Los resultados obtenidos con el modelo ET_{ogp} muestran una alta precisión en la estimación de la evapotranspiración de referencia dentro de invernaderos utilizando datos de temperatura y humedad relativa. Este hallazgo es significativo, ya que indica la capacidad del modelo para capturar con precisión los procesos de evaporación y transpiración en un entorno controlado como un invernadero. La precisión de estas estimaciones es crucial para la gestión eficiente del riego y el control ambiental en la producción de cultivos bajo condiciones de invernadero. Los resultados respaldan la utilidad y la fiabilidad del modelo ET_{ogp} como una herramienta viable para la estimación precisa de la evapotranspiración de referencia en este contexto específico.

5.4. Conclusiones

En este capítulo, se presentó el desarrollo de un modelo para estimar la evapotranspiración en invernaderos utilizando la metodología presentada en la sección 1.5 basada en Programación Genética. Los datos para esta investigación fueron proporcionados por el Departamento de Agricultura de la Universidad Americana de Beirut, Líbano. La limpieza de datos, la imputación de valores faltantes y la normalización de las variables aseguraron la compatibilidad y calidad de los datos, facilitando la generación de modelos precisos. El uso de datos provenientes de una fuente académica de confianza y su correcta preparación permitió que los modelos generados fueran capaces de generalizar y adaptarse a las condiciones específicas de los invernaderos. El modelo desarrollado fue validado y comparado con el método de Hargreaves-Samani y un atmómetro, superando ambos en términos de precisión y fiabilidad. El método de referencia utilizado para esta comparación fue el FAO56-PM, lo que proporciona una sólida base para la evaluación del desempeño del modelo.

La utilización de Programación Genética representa una innovación significativa en el campo de la hidrología, proporcionando una nueva perspectiva metodológica y abriendo la puerta a futuras investigaciones y aplicaciones en diversas áreas. En conclusión, el desarrollo de modelos de estimación de evapotranspiración para invernaderos utilizando Programación Genética ha demostrado ser una herramienta poderosa y precisa, con un gran potencial para mejorar la gestión de recursos hídricos y la sostenibilidad agrícola, estableciendo una base sólida para futuras investigaciones y aplicaciones prácticas en el campo de la hidrología y la agricultura.

Capítulo 6

Desarrollo de un modelo predictivo a corto plazo para invernaderos

6.1. Introducción

Los modelos predictivos son fundamentales por su capacidad de anticipar resultados futuros mediante el análisis de datos históricos y patrones. Su importancia radica en su capacidad para informar la toma de decisiones, mejorar la eficiencia operativa al prever demandas y fallos, reducir riesgos al anticipar problemas potenciales, personalizar servicios en función del comportamiento del cliente, mejorar la precisión y rapidez en las predicciones, y fomentar la innovación y competitividad al detectar oportunidades emergentes. En conjunto, los modelos predictivos permiten transformar datos en información valiosa para tomar decisiones más inteligentes y obtener beneficios significativos en diversos campos.

En esta sección, se aborda la Fase 2 de la metodología de solución, centrada en el desarrollo de un modelo predictivo a corto plazo. El objetivo principal es predecir con precisión la evapotranspiración de referencia diaria dentro de un invernadero, utilizando datos históricos diarios de evapotranspiración. En este contexto, el uso de Programación Genética para desarrollar el modelo predictivo es de suma importancia. La Programación Genética permite la creación de modelos complejos y adaptativos capaces de capturar de manera precisa y dinámica las interacciones entre múltiples variables ambientales y agronómicas que influyen en la evapotranspiración en entornos controlados como los invernaderos. Estos métodos genéticos tienen la capacidad de evolucionar y ajustarse automáticamente a medida que se recopilan más datos, mejorando así su precisión predictiva con el tiempo. Esto es crucial para los agricultores, ya que les proporciona herramientas sólidas para tomar decisiones informadas sobre el riego y la gestión de recursos hídricos, lo que a su vez contribuye a maximizar la producción de cultivos y optimizar el uso eficiente del agua en condiciones de agricultura protegida.

6.2. Obtención y transformación de datos

El proceso de obtención y transformación de datos de una serie temporal para su uso en Programación Genética implica varios pasos cruciales para estructurar los datos adecuadamente. Primero, se identifican y extraen los datos históricos relevantes de diversas fuentes. Luego, los datos se limpian para eliminar valores atípicos y faltantes, y se normalizan o estandarizan para asegurar consistencia. Posteriormente, la serie temporal se divide en ventanas temporales de tamaño fijo, y cada ventana se convierte en un vector de características, formando una matriz donde cada fila representa un conjunto de observaciones consecutivas en el tiempo. Esta matriz de vectores resultante es adecuada para alimentar el algoritmo evolutivo de Programación Genética facilitando el desarrollo de modelos predictivos.

6.2.1. Obtención de datos

Esta actividad implica la recopilación de datos históricos diarios de evapotranspiración de referencia, los cuales pueden provenir de diversas fuentes, tales como archivos históricos de parámetros meteorológicos, estaciones meteorológicas ubicadas en entornos de campo abierto y estructuras específicas utilizadas en agricultura protegida (invernaderos, microtúneles y casas sombra). Para llevar a cabo esta tarea de manera efectiva, en nuestra investigación hemos decidido utilizar una serie temporal con datos históricos diarios de evapotranspiración de referencia previamente estimados mediante el modelo $ET_{o_{gp}}$ desarrollado en la Fase 1 (Ecuación 5.3). Este modelo proporciona estimaciones precisas de evapotranspiración.

6.2.2. Transformación de datos

La transformación de una serie temporal en una matriz de características y un vector objetivo es un paso fundamental en el preprocesamiento de datos para el modelado predictivo. Este proceso es trascendental para aplicar métodos de aprendizaje automático que requieren entradas estructuradas como vectores de características.

Concepto de Serie Temporal

Una serie temporal es una secuencia de datos observados en momentos sucesivos, generalmente espaciados uniformemente en el tiempo. Ejemplos comunes incluyen datos meteorológicos, precios de acciones, y medidas fisiológicas. El análisis de series temporales busca modelar y predecir comportamientos futuros basados en patrones históricos.

Transformación a Matriz de Características y Vector Objetivo

La técnica más común para la transformación de una serie temporal en una matriz de características es el uso de ventanas deslizantes. Esta técnica consiste en definir un tamaño de ventana w , que determina cuántos valores consecutivos de la serie temporal se tomarán como un vector de características. Cada ventana desplazada crea una fila en

la matriz de características. Si se tiene una serie temporal y_t , el vector de características correspondiente a la observación en el tiempo t será $[y_{t-w}, y_{t-w+1}, \dots, y_{t-1}]$

El vector objetivo (valor a predecir), que corresponde al valor de la serie temporal en el instante actual y_t , se almacena en un vector objetivo (target vector). Así, para cada vector de características generado, se asocia el valor de la serie en el tiempo t , formando así pares de datos de entrenamiento.

Ejemplo matemático

Dada una serie temporal $y = [y_1, y_2, \dots, y_n]$ y un tamaño de ventana w , la matriz de características X y el vector objetivo y_{target} se definen como:

$$X = \begin{bmatrix} y_1 & y_2 & \dots & y_w \\ y_2 & y_3 & \dots & y_{w+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n-w} & y_{n-w+1} & \dots & y_{n-1} \end{bmatrix}, \quad y_{target} = \begin{bmatrix} y_{w+1} \\ y_{w+2} \\ \vdots \\ y_n \end{bmatrix}$$

Cada fila de X representa un conjunto de características que describen el estado de la serie temporal en un instante específico, mientras que y_{target} contiene los valores correspondientes que se desean predecir.

En nuestra investigación, la transformación de los datos implicó convertir la serie temporal que contiene la evapotranspiración de referencia. Esta transformación es importante para capturar la relación entre los datos históricos y la evapotranspiración de referencia actual, siendo fundamental para el desarrollo del modelo predictivo a corto plazo. La Figura 6.1 ilustra este proceso al mostrar cómo la serie temporal de ETo es convertida en una matriz con 5 estados previos, donde cada fila representa una ventana de tiempo y cada columna corresponde a una característica. Además, se incluye un vector objetivo que contiene el valor de la serie que sigue a cada fila de la matriz, completando así la representación de los datos transformados. Este enfoque permitió analizar y utilizar eficazmente la información histórica para predecir la evapotranspiración futura en el contexto de agricultura protegida.

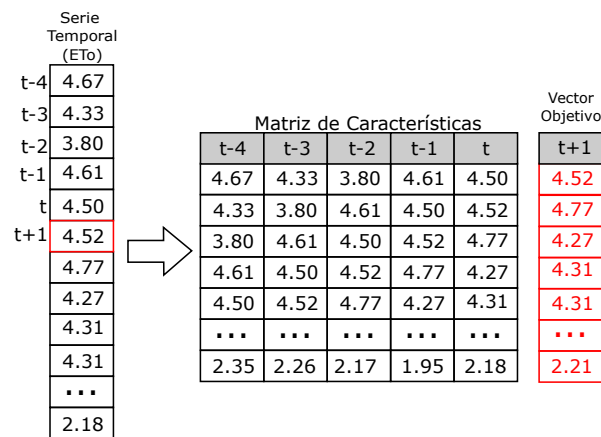


Figura 6.1: Transformación de la serie temporal en una matriz de características de 5 estados previos y un vector objetivo con el valor de la serie que sigue a cada fila de la matriz.

Después de finalizar esta actividad, se avanza hacia la formación de dos conjuntos de datos: uno diseñado para el entrenamiento del modelo y otro destinado a su evaluación en la fase de prueba. La actividad subsiguiente ofrece una explicación detallada del proceso de creación de estos conjuntos y su importancia en el desarrollo y validación del modelo predictivo.

6.2.3. Creación del conjunto de datos

El proceso de creación de conjuntos de datos implica una serie de pasos clave que son fundamentales para el desarrollo y validación de modelos predictivos. En primer lugar, se realiza la transformación de la serie temporal en una matriz de características y un vector objetivo, lo que facilita el análisis y procesamiento de los datos como se muestra en la figura 6.1. Luego, se procede a la formación de conjuntos de datos distintos para entrenamiento y prueba del modelo.

Usualmente, se sigue una práctica común donde se asigna un porcentaje entre el 70% y el 80% de los datos para el conjunto de entrenamiento, mientras que el 20% al 30% restante se destina al conjunto de prueba. Esta división se basa en la experiencia empírica de la comunidad de aprendizaje automático. Muchos estudios han demostrado que una división 80/20 proporciona una buena medida de equilibrio para muchos tipos de problemas y tamaños de datos, aunque no es una regla rígida y puede ajustarse según las necesidades específicas del problema (Géron, 2022)

En nuestra investigación, tomamos la decisión de asignar el 20% de los datos para el conjunto de pruebas, reservando el resto, para el entrenamiento del modelo. Esta división busca asegurar que el modelo pueda generalizar bien a datos no vistos previamente, evitando el sobreajuste y garantizando que el modelo no se ajuste demasiado a los datos específicos del entrenamiento. Al utilizar un conjunto de prueba separado, se puede evaluar de manera precisa el desempeño del modelo mediante métricas calculadas en este conjunto. Esto permite comprobar su capacidad para funcionar adecuadamente con datos nuevos y proporcionar una evaluación imparcial de su rendimiento. La estructura de los conjuntos de datos para entrenamiento y prueba se observa en la Tabla 6.1.

6.3. Generación del modelo predictivo

La generación del modelo predictivo implica un proceso iterativo que combina la evolución de modelos matemáticos con la consideración de datos históricos. En primer lugar, se organiza el conjunto de datos en una estructura adecuada, por lo tanto, en nuestra investigación utilizamos una ventana de tiempo de 5 estados previos para crear una matriz de características y un vector objetivo como se observa en la Figura 6.1. Posteriormente, se utiliza la Programación Genética para desarrollar modelos matemáticos capaces de predecir valores futuros a partir de los datos históricos diarios disponibles. Durante este proceso, se aplican operadores genéticos como la selección, cruce y mutación para generar una población inicial de modelos y mejorar su adaptación a los datos de entrenamiento.

Tabla 6.1: Estructura de los conjuntos de datos de entrenamiento y prueba

Matriz de características					Vector objetivo
t-4	t-3	t-2	t-1	t	t+1
4.67	4.33	3.80	4.61	4.50	4.52
4.33	3.80	4.61	4.50	4.52	4.77
3.80	4.61	4.50	4.52	4.77	4.27
4.61	4.50	4.52	4.77	4.27	4.31
4.50	4.52	4.77	4.27	4.31	4.31
...
2.35	2.26	2.17	1.95	2.18	2.21

Luego, los modelos se evalúan mediante la función objetivo (aptitud) para determinar su capacidad de predecir con precisión los valores objetivo, utilizando métricas de rendimiento específicas. Los modelos más aptos son seleccionados y utilizados como base para generar la siguiente generación, continuando el ciclo evolutivo hasta alcanzar un criterio de convergencia o completar un número predeterminado de generaciones.

6.3.1. Realizar pasos preparatorios

Como se mencionó en el capítulo anterior, antes de iniciar con la Programación Genética, es esencial seguir los cinco pasos preparatorios descritos por John Koza en su libro (Koza, 1992). Estos pasos son fundamentales para configurar correctamente el algoritmo y asegurar su efectividad en la resolución del problema planteado. A continuación, se detalla cómo abordamos los principales pasos preparatorios en nuestra investigación.

Definir el conjunto de terminales

Para definir el conjunto de terminales, se tomó cada columna de la matriz de características creada en el proceso previo (Figura 6.1). Además de ello, utilizamos la serie de Taylor de primer 6.1, segundo 6.2 y tercer orden 6.3 con aproximaciones finitas hacia atrás con una ventana temporal de 5 puntos. Esta técnica nos permite calcular derivadas precisas para cada punto en la serie temporal. Las ecuaciones para estas derivadas se presentan a continuación:

$$f'(x_t) = \frac{3f(x_t) - 4f(x_{t-1}) + f(x_{t-2})}{2h} \quad (6.1)$$

$$f''(x_t) = \frac{2f(x_t) - 5f(x_{t-1}) + 4f(x_{t-2}) - f(x_{t-3})}{h^2} \quad (6.2)$$

$$f'''(x_t) = \frac{5f(x_t) - 18f(x_{t-1}) + 24f(x_{t-2}) - 14f(x_{t-3}) + 3f(x_{t-4})}{2h^3} \quad (6.3)$$

Este enfoque nos permite obtener una representación detallada y precisa de la variación de los datos en la serie temporal, lo que a su vez contribuye a una definición más robusta del conjunto de terminales para nuestros modelos de predicción. El conjunto de terminales proporcionado al algoritmo evolutivo se observa en la Tabla 6.2

Tabla 6.2: Conjunto de terminales

No.	Terminal	Descripción	No.	Terminal	Descripción
1.	$t - 4$	Cuatro días atrás	7.	$pd t1$	Primera derivada de $t - 1$
2.	$t - 3$	Tres días atrás	8.	$pd t2$	Primera derivada de $t - 2$
3.	$t - 2$	Dos días atrás	9.	$sd t$	Segunda derivada de t
4.	$t - 1$	Un día atrás	10.	$sd t1$	Segunda derivada de $t - 1$
5.	t	Día actual	11.	$td t$	Tercera derivada de t
6.	$pd t$	Primera derivada de t			

El conjunto de terminales representa las variables independientes que se proporcionan al algoritmo evolutivo. Estas variables son fundamentales para el proceso de evolución, ya que actúan como entradas que el algoritmo utiliza para generar, evaluar y optimizar las soluciones candidatas. Al definir un conjunto adecuado de terminales, se puede asegurar que el algoritmo tenga acceso a la información necesaria para modelar el problema de manera efectiva y encontrar soluciones de alta calidad.

Definir el conjunto de funciones

La definición del conjunto de funciones es un aspecto crucial, ya que estas funciones constituyen una parte fundamental del espacio de búsqueda; una definición inadecuada podría restringir significativamente dicho espacio. Para definir nuestro conjunto de funciones, nos basamos en la frecuencia de aparición de las funciones utilizadas en experimentos previos, eliminando aquellas con una frecuencia de aparición baja o nula. El conjunto resultante incluye operadores básicos, así como funciones trigonométricas, exponenciales y logarítmicas, comúnmente empleadas en estudios hidrológicos (Mehdizadeh et.al., 2017; Mattar, 2018). Además, hemos incorporado funciones hiperbólicas, que son esenciales para el análisis de fenómenos físicos complejos (Pandir y Ulusoy, 2013). La Tabla 6.3 presenta el conjunto de funciones empleadas en el algoritmo evolutivo, garantizando una representación robusta y diversificada del espacio de búsqueda.

Una definición adecuada permite explorar soluciones más efectivas y variadas, mejorando así el rendimiento y facilitando la evolución hacia soluciones óptimas. Además, un conjunto de funciones bien definido asegura que el algoritmo sea flexible y adaptable a diferentes problemas y circunstancias. Por otro lado, una definición inadecuada puede restringir significativamente el espacio de búsqueda, limitando la capacidad del algoritmo para encontrar soluciones óptimas.

Definir la función de aptitud

Tabla 6.3: Conjunto de funciones

No.	Función	Descripción	No.	Función	Descripción
1.	$e^{(\cdot)}$	Función exponencial base	8.	$\log(\cdot)$	Logaritmo natural
2.	$\tanh(\cdot)$	Tangente hiperbólica	9.	$\log_{10}(\cdot)$	Logaritmo base 10
3.	$x^{(\cdot)}$	Exponente	10.	+	Suma
4.	$\operatorname{atan}(\cdot)$	Arcotangente	11.	-	Resta
5.	$\sin(\cdot)$	Seno	12.	*	Multiplicación
6.	$\operatorname{sech}(\cdot)$	Secante hiperbólica	13.	\div	División
7.	$\operatorname{csch}(\cdot)$	Cosecante hiperbólica			

La función de aptitud es fundamental para evaluar y seleccionar las soluciones candidatas generadas por el algoritmo evolutivo. Asigna un valor numérico a cada individuo en la población, reflejando su rendimiento en relación con el problema específico. Este valor guía la selección de padres para la siguiente generación, favoreciendo a las soluciones con mayor aptitud y dirigiendo la evolución hacia regiones del espacio de búsqueda más prometedoras. Al mismo tiempo, la función de aptitud ayuda a mantener la diversidad en la población, evitando la convergencia prematura y permitiendo una exploración más amplia del espacio de búsqueda. En problemas dinámicos, puede adaptarse a cambios en las condiciones, asegurando que el algoritmo evolutivo continúe encontrando soluciones óptimas.

En nuestra investigación tomamos la decisión de utilizar como medida de aptitud la raíz del error cuadrático medio (RMSE) (Ecuación 5.2), una métrica que evalúa la discrepancia promedio entre las predicciones y los valores de referencia. Esta medida proporciona una evaluación precisa de la calidad de las predicciones realizadas por un individuo dentro del modelo. Es importante destacar que a medida que la aptitud de un individuo mejora, su probabilidad de ser seleccionado para pasar a la siguiente generación aumenta significativamente. Esta relación directa entre la aptitud y la probabilidad de selección asegura que los individuos más aptos sean favorecidos en el proceso evolutivo, contribuyendo así a la mejora continua del rendimiento del algoritmo.

Definir los parámetros del algoritmo evolutivo y la designación de resultados

Con el fin de optimizar el desempeño del algoritmo evolutivo, se llevó a cabo una exploración sistemática (probando con diferentes combinaciones de parámetros) para identificar los parámetros que generaron resultados óptimos en el proceso evolutivo. Este enfoque garantizó una configuración óptima aceptable tanto la calidad de las soluciones obtenidas como el uso eficiente de los recursos computacionales disponibles. La Tabla 6.4 proporciona un detalle de los parámetros utilizados durante el entrenamiento del algoritmo.

Tabla 6.4: Parámetros utilizados durante el entrenamiento del algoritmo evolutivo

Parámetro	Valor
Tamaño de la población	2000
Número de generaciones	500
Probabilidad de cruce	80%
Probabilidad de mutación	20%
Profundidad máxima de árbol	14
Método de selección	lexicotout
Método de iniciación	Mitad-Mitad
Elitismo	Mantener el mejor

6.3.2. Ejecutar el algoritmo evolutivo

El algoritmo implementado con GPLab genera la población inicial de soluciones combinando de manera aleatoria elementos de los conjuntos de terminales y funciones proporcionados en la tabla 6.2 y la tabla 6.3 respectivamente. Luego, cada individuo de esta población inicial (representado gráficamente en forma de árbol como se observa en la figura 2.2), se evalúa utilizando la función de aptitud (Ecuación 5.2). Posteriormente, se seleccionan soluciones candidatas para clasificar a todos los individuos y eliminar aquellas con baja aptitud. A continuación, se lleva a cabo la recombinación genética entre los individuos seleccionados mediante operadores de cruce y mutación como se observa en la figura 2.8 y la figura 2.9 respectivamente. Por último, la siguiente generación se forma a partir de la selección de las mejores soluciones entre padres e hijos. Estos pasos se repiten hasta que se alcance el número máximo de generaciones establecido en los parámetros de ejecución. El algoritmo evolutivo se ejecutó de manera independiente 10 veces, cada una con 500 generaciones, y el proceso completo tuvo una duración de 2 horas y 35 minutos. El diagrama de flujo del proceso evolutivo se presenta en la Figura. 2.4.

Análisis estadístico sobre el proceso de búsqueda

El análisis estadístico busca demostrar la tasa de convergencia del algoritmo evolutivo. En la Figura 6.2, se presenta una comparación detallada del desempeño del algoritmo. Esta figura incluye el mejor ajuste, el promedio de los mejores ajustes y la desviación estándar a lo largo de 10 ejecuciones. El mejor ajuste representa el resultado más óptimo logrado en cualquiera de las ejecuciones, proporcionando una medida de la capacidad máxima del algoritmo. El promedio de los mejores ajustes ofrece una visión general del desempeño típico del algoritmo, mostrando cuán consistentemente puede alcanzar soluciones cercanas al óptimo. Finalmente, la desviación estándar refleja la variabilidad en los resultados entre las diferentes ejecuciones, indicando la estabilidad y la robustez del algoritmo. Este análisis estadístico es fundamental para evaluar la eficiencia y la fiabilidad del algoritmo evolutivo, permitiendo identificar tanto su potencial máximo como su comportamiento esperado en

aplicaciones repetidas. Además, ayuda a comprender la dispersión de los resultados.

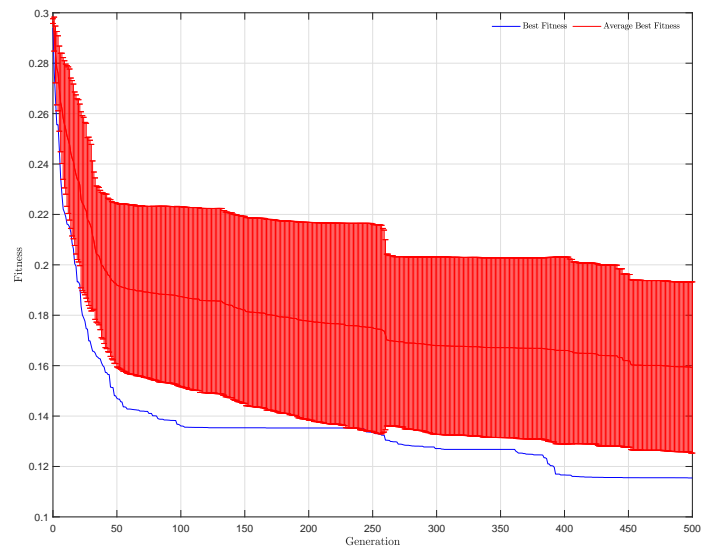


Figura 6.2: Mejor aptitud y promedio de la mejor aptitud a lo largo de la evolución

Es importante resaltar que, en promedio, el algoritmo evolutivo encontró soluciones con una medida de fitness inferior a 0.3 desde la primera generación. Este hallazgo indica que el algoritmo fue capaz de identificar soluciones bastante eficientes en una etapa temprana del proceso evolutivo. El valor de fitness mínimo alcanzado se observó aproximadamente en la generación 450, lo que sugiere que el algoritmo continuó mejorando las soluciones hasta llegar a un punto de optimización significativa. Además, el fitness promedio mostró una rápida convergencia desde la primera generación con un fitness de 0.2967.

Estos resultados destacan la capacidad del algoritmo para aproximarse rápidamente a soluciones óptimas. La rápida convergencia no solo mejora la eficiencia computacional, sino que también reduce el tiempo necesario para obtener soluciones viables, lo que convierte al algoritmo en una herramienta altamente efectiva para aplicaciones prácticas.

La convergencia temprana a un fitness bajo y la posterior mejora hasta el fitness mínimo indican una exploración efectiva del espacio de soluciones y una explotación óptima de las mejores soluciones encontradas. La estabilidad del fitness promedio desde el inicio sugiere que el algoritmo tiene una robusta capacidad para mantener un rendimiento consistente a lo largo de múltiples ejecuciones, garantizando que las soluciones no solo sean óptimas, sino también reproducibles y confiables. Esto refuerza la validez del algoritmo como una herramienta poderosa en la optimización y modelación de procesos complejos.

6.3.3. Mejores soluciones

Elegir las mejores soluciones implica seleccionar las que ofrecen mejor desempeño dentro de la población. Aunque la solución de la última generación suele ser la más precisa, a veces es necesario sacrificar precisión por una estructura más simple de generaciones anteriores. Esto se debe a que es más fácil de comprender, interpretar y aplicar en el mundo real, además de ser más eficiente computacionalmente. Por lo tanto, es crucial equilibrar precisión y simplicidad al evaluar las opciones disponibles para un problema dado.

Después de ejecutar el algoritmo evolutivo, se obtuvieron 6232 soluciones diferentes con un fitness menor a 0.3, de las cuales se seleccionaron cuatro. El criterio de selección se basó primero en la aptitud y luego en la complejidad estructural. La Tabla 6.5 presenta estas soluciones, junto con el fitness alcanzado.

Tabla 6.5: Modelos obtenidos

No.	Modelo	Fitness
1	$ETo_{t+1} = \left(\frac{1}{(exp^{tan(exp^{f'(ETo_t)})ETo_t + ETo_{t-1}})} \right) + tanh(exp^{tanh(exp^{tanh(tanh(f'(ETo_t)))})}) * ETo_{t-1}$	0.222
2	$ETo_{t+1} = \left(\frac{f'(ETo_t)}{ETo_{t-4}} \right) ETo_{t-1} + tanh(exp^{tanh(exp^{ETo_t})}) * ETo_{t-1}$	0.248
3	$ETo_{t+1} = ETo_{t-2} + \log_{10}(exp^{f'(ETo_{t-1})}) - \log_{10}(atan(ETo_{t-1}))$	0.248
4	$ETo_{t+1} = ETo_{t-1} + sech(\log_{10}(ETo_{t-4} - \frac{ETo_{t-1}}{f'(ETo_t)}))$	0.259

El enfoque de Programación Genética ha demostrado su capacidad para generar modelos predictivos precisos a corto plazo en diversos contextos y aplicaciones. Este método se destaca por su capacidad para adaptarse y evolucionar, utilizando algoritmos evolutivos inspirados en procesos biológicos para encontrar soluciones óptimas a problemas complejos. En el contexto específico de la predicción a corto plazo, la Programación Genética ha mostrado su eficacia al explorar y combinar diversas funciones y terminales para generar modelos que se ajusten de manera precisa y dinámica a los datos disponibles. Al aprovechar la diversidad y la flexibilidad inherentes al proceso evolutivo, la Programación Genética es capaz de capturar patrones y tendencias relevantes en los datos, lo que resulta en modelos predictivos que ofrecen resultados confiables y certeros en el corto plazo.

Frecuencia de uso de las funciones y terminales

El análisis estructural de los modelos presentados en la Tabla 6.5, ofrece una perspectiva reveladora sobre las estrategias desarrolladas por el proceso evolutivo para combinar el conjunto primitivo de terminales y funciones. En este estudio, empleamos una métrica de frecuencia de uso para evaluar con qué frecuencia se utiliza cada elemento en la generación de nuevos modelos. Las Figuras 6.3 y 6.4 ilustran la frecuencia de ocurrencia de los conjuntos de terminales y funciones, respectivamente. Los números asignados a los terminales y funciones se corresponden con la descripción detallada proporcionada en las Tablas 6.2 y 6.3.

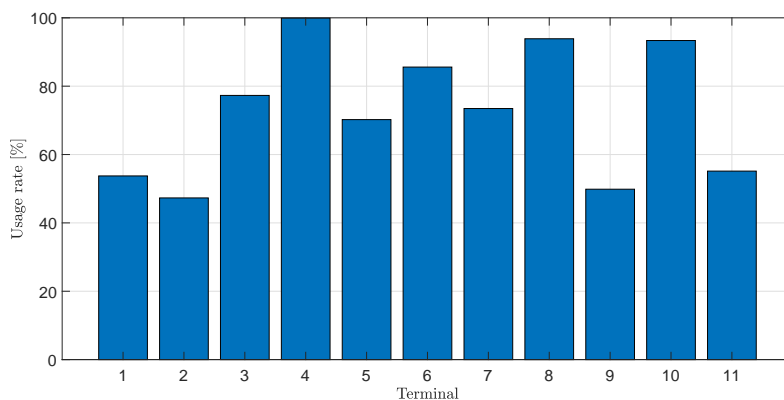


Figura 6.3: Frecuencia de aparición del conjunto de terminales

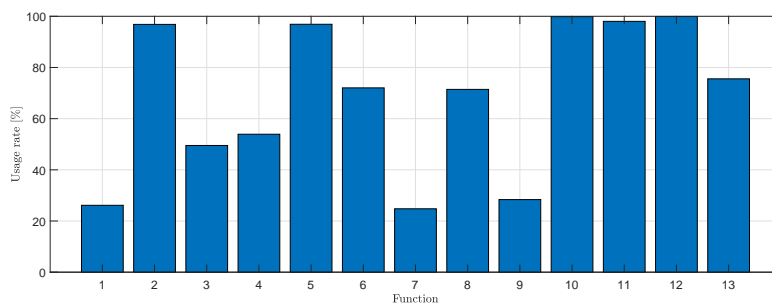


Figura 6.4: Frecuencia de aparición del conjunto de funciones

Este enfoque nos permite identificar las funciones y terminales más utilizadas por el algoritmo evolutivo, lo que a su vez revela las preferencias del modelo en términos de componentes esenciales para predecir la evapotranspiración. Al observar la frecuencia de uso de estos elementos, se pueden inferir patrones y tendencias que explican la eficacia del modelo y su capacidad para generalizar en diferentes escenarios. Este análisis es fundamental para entender la dinámica del proceso evolutivo y para optimizar futuras versiones del modelo, asegurando una mayor precisión y eficiencia en la predicción de la evapotranspiración.

6.3.4. Seleccionar el mejor modelo

En el proceso de selección del modelo, llevamos a cabo un análisis minucioso que consideró tanto el nivel de aptitud (fitness) como la complejidad estructural de las soluciones obtenidas presentadas en la Tabla 6.5. Este enfoque nos permitió evaluar de manera integral cada una de las opciones generadas por el algoritmo evolutivo, identificando aquellas que ofrecían un equilibrio óptimo entre precisión y simplicidad. Tras un detallado análisis, determinamos que la mejor solución correspondía al modelo descrito en la ecuación (Ecuación 6.4) y catalogado con el índice 4 en la Tabla 6.5. Este modelo sobresalió no solo por su alto nivel de aptitud, indicativo de su capacidad para ajustarse y predecir con precisión los datos de entrada, sino también por su estructura simplificada que facilita su comprensión e implementación práctica. Este hallazgo resalta la importancia de considerar tanto la eficacia predictiva como la simplicidad en la selección de modelos óptimos, ya que estos atributos son fundamentales para garantizar su utilidad y aplicabilidad en entornos reales.

$$ETo_{t+1} = ETo_{t-1} + sech(\log_{10}(ETo_{t-4} - \frac{ETo_{t-1}}{f'(ETo_t)})) [mm] \quad (6.4)$$

6.4. Validación del modelo

La validación con datos no conocidos es crucial para determinar la capacidad del modelo de generalizar más allá de los datos utilizados durante el entrenamiento, permitiéndole funcionar de manera efectiva en escenarios del mundo real. Al validar con datos no vistos, se puede detectar y mitigar el sobreajuste, un fenómeno en el cual el modelo se ajusta demasiado a las particularidades del conjunto de entrenamiento, capturando patrones específicos que no se trasladan adecuadamente a nuevos datos.

Además, la validación proporciona una evaluación precisa y objetiva del rendimiento del modelo, revelando su verdadera precisión y robustez. Este análisis es indispensable para identificar y corregir sesgos presentes en los datos de entrenamiento, lo que contribuye a la equidad y representatividad del modelo.

En el ámbito de la predicción de la evapotranspiración de referencia, la validación con datos no conocidos garantiza que los modelos desarrollados sean aplicables y precisos en diversos entornos, como invernaderos, y bajo distintas condiciones climáticas. Esta

validación asegura que las predicciones generadas sean fiables, lo que es esencial para la toma de decisiones informadas en la gestión de recursos hídricos y la planificación agrícola.

6.4.1. Comparación del modelo predictivo contra el método de medias móviles simple

Cuando se trabaja con un conjunto de datos pequeño, el uso del método de medias móviles simples para la validación conlleva ciertas consideraciones clave a pesar de su vulnerabilidad a datos atípicos. Las medias móviles simples son elegidas por su simplicidad y facilidad de interpretación: consisten en calcular el promedio de un número fijo de observaciones pasadas, lo cual proporciona una visión intuitiva de la tendencia de los datos a lo largo del tiempo (Wheelwright et.al., 1998). Aunque estas medias móviles son sensibles a datos extremadamente atípicos, pueden ayudar a suavizar las fluctuaciones en los datos y facilitar la identificación de patrones generales de comportamiento, especialmente cuando los datos son ruidosos o irregulares. Además, aunque no son tan sensibles a cambios abruptos en los datos como otros métodos más sofisticados, las medias móviles simples pueden capturar tendencias a largo plazo, lo que las hace útiles para análisis preliminares o validaciones iniciales en conjuntos de datos limitados (Makridakis et.al., 2020), (Beyaztas y Shang, 2019). La cantidad de datos limitados puede ser restrictiva para la aplicación de métodos más complejos, por lo que las medias móviles simples ofrecen una opción robusta y de bajo costo computacional para explorar patrones generales y validar tendencias (Hyndman y Athanasopoulos, 2018).

En nuestro estudio, validamos el modelo predictivo utilizando el método de medias móviles simples con una serie temporal que cuenta con 61 registros no vistos previamente durante el entrenamiento. Aunque el método de medias móviles simples puede ser vulnerable a datos atípicos, decidimos utilizarlo debido a que, en nuestra investigación, los datos fueron sometidos a un procesamiento previo para eliminar estos valores atípicos. Esta validación nos permitió evaluar su capacidad para hacer predicciones precisas en situaciones desconocidas, destacando su utilidad y confiabilidad en diversas aplicaciones. Los resultados de la validación se presentan en la Figura 6.6, donde se comparan visualmente las predicciones del modelo predictivo con el método de medias móviles simples y valores de referencia. Además, la Tabla 6.6 detalla las métricas (RMSE, MAE y R^2) obtenidas durante la validación, proporcionando una evaluación cuantitativa de la precisión y desempeño de los modelos en escenarios reales.

La métrica MAE (Mean Absolute Error, o Error Absoluto Medio) (Ecuación 6.5) es una medida de precisión utilizada para evaluar modelos de predicción. Se utiliza comúnmente en el contexto de modelos de regresión y series temporales para cuantificar la diferencia promedio entre los valores predichos por el modelo y los valores reales observados.

La fórmula del Error Absoluto Medio (MAE) es la siguiente:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6.5)$$

donde:

- n es el número total de observaciones,
- y_i es el valor real observado,
- \hat{y}_i es el valor predicho por el modelo,
- $|y_i - \hat{y}_i|$ es el valor absoluto de la diferencia entre la predicción y la observación.

Tabla 6.6: Métricas obtenidas en el proceso de validación

Modelo	RMSE	MAE	R^2
ETo_{t+1}	0.286	0.082	0.8976
Med Mov	0.286	0.082	0.8936

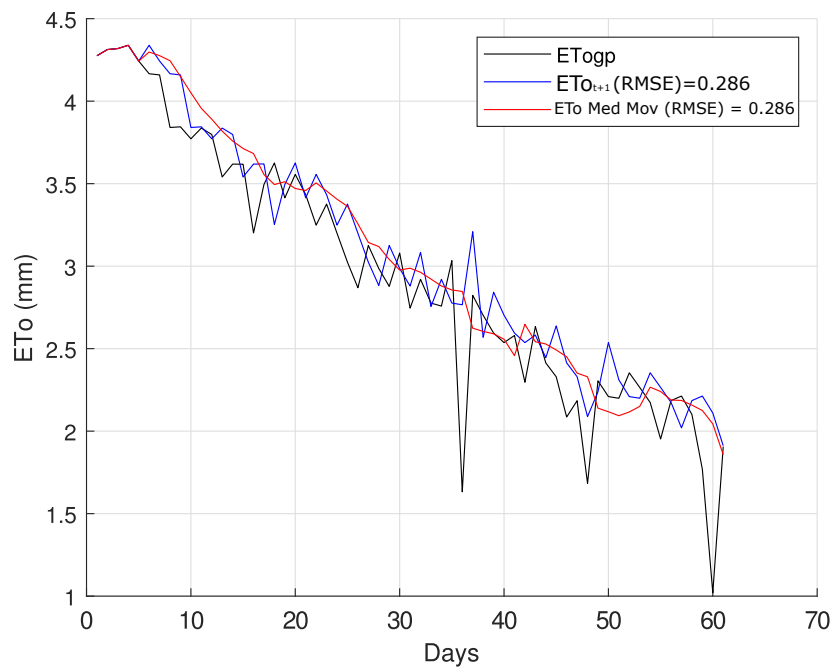


Figura 6.5: RMSE (ETo_{t+1} y Med. Móviles)

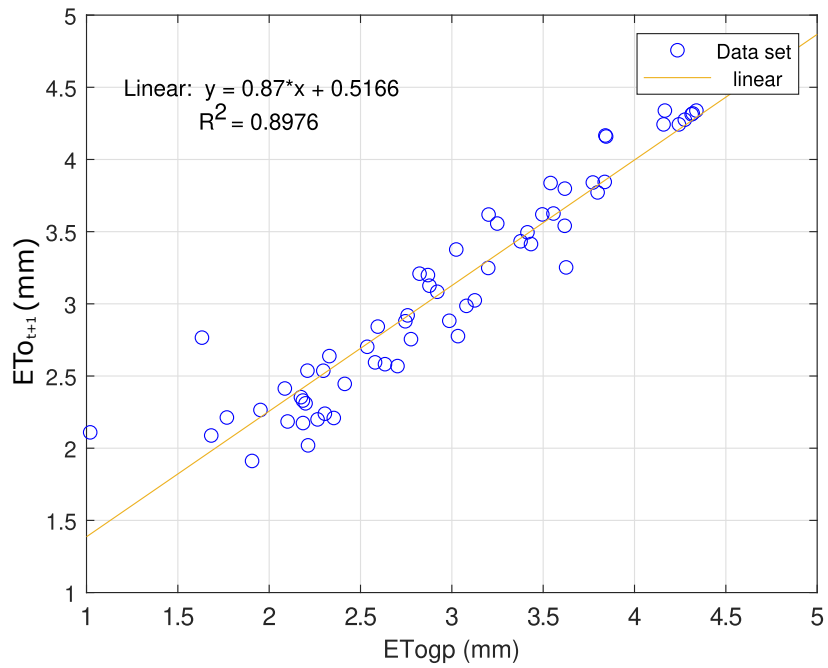


Figura 6.6: R^2 ET_{t+1}

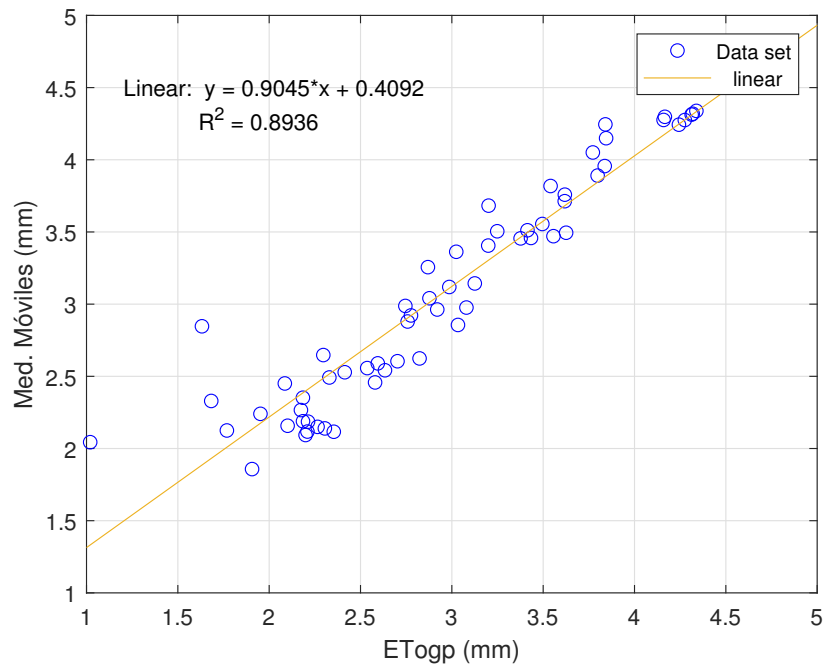


Figura 6.7: R^2 Med. Móviles

El análisis comparativo entre el modelo predictivo y el modelo de medias móviles revela diferencias sutiles pero importantes en su desempeño. Aunque el modelo predictivo exhibe una precisión igual al modelo de medias móviles, no se alcanza significancia estadística, como se refleja en el RMSE para ambos modelos (0.286). Sin embargo, sugiere una tendencia hacia una menor discrepancia promedio entre las predicciones del modelo predictivo y los valores reales, lo cual es relevante para la precisión de las predicciones.

Además, al analizar el Error Absoluto Medio (MAE), el modelo predictivo muestra un desempeño comparable al del modelo de medias móviles, con valores idénticos de 0.082 en ambos casos. Esta igualdad promedio en las predicciones del modelo predictivo indica una exactitud similar en las estimaciones realizadas.

En cuanto al coeficiente de determinación (R^2), se destaca una diferencia sutil a favor del modelo predictivo, con un valor de 0.8976 frente a 0.8936 para el modelo de medias móviles. Este indicador señala la capacidad del modelo predictivo para explicar una mayor proporción de la variabilidad presente en los datos, lo que se traduce en una capacidad más efectiva para modelar y predecir la evapotranspiración de referencia.

6.4.2. Validación del modelo predictivo usando validación cruzada «*Rolling Origin*»

El método «*Rolling Origin*» es una técnica de evaluación según la cual el origen del pronóstico se actualiza sucesivamente y los pronósticos se producen a partir de cada origen (Tashman, 2000). Esta técnica permite obtener varios errores de pronóstico para series de tiempo, lo que brinda una mejor comprensión del desempeño de los modelos. Existen diferentes opciones sobre esta técnica.

La figura 6.8 muestra la idea básica del origen rodante (Svetunkov y Petropoulos, 2018). Las celdas blancas corresponden a los datos de la muestra, mientras que las celdas de color gris corresponden a los pronósticos de tres pasos por delante. La serie de tiempo tiene 25 observaciones en esa figura y los pronósticos se producen a partir de 8 orígenes, comenzando desde el origen 15. El modelo se reestima en cada iteración y se producen los pronósticos. Después de eso se agrega una nueva observación al final de la serie y el procedimiento continúa. El proceso se detiene cuando no hay más datos para agregar. Esto podría considerarse como un origen móvil con un tamaño de muestra constante. Como resultado de este procedimiento se elaboran previsiones de uno a tres pasos por delante. En base a ellos podemos calcular las medidas de error preferidas y elegir el modelo con mejor rendimiento.

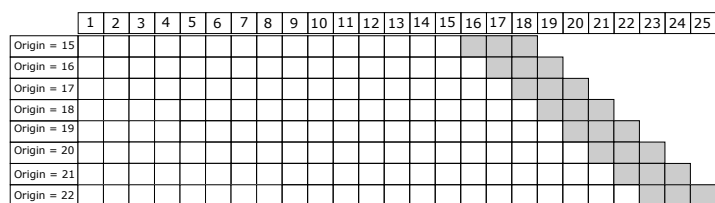


Figura 6.8: Método de validación cruzada usando *Rolling Origin* 1

Otra opción de producir pronósticos a partir de 8 orígenes sería comenzar desde el origen 17 en lugar de 15 (ver Figura 6.9). En este caso, el procedimiento continúa hasta el origen 22, cuando se produce el último pronóstico de tres pasos hacia adelante, y luego continúa con el horizonte de pronóstico decreciente. Así, el pronóstico de dos pasos por delante se produce desde el origen 23 y sólo el pronóstico de un paso por delante se produce desde el origen 24. Como resultado obtenemos 8 pronósticos de un paso por delante, 7 pronósticos de dos pasos por delante y 6 previsiones de tres pasos por delante. Esto puede considerarse como un origen móvil con un tamaño de muestra no constante . Esto puede resultar útil en casos de muestras pequeñas, cuando no disponemos de observaciones de sobra.

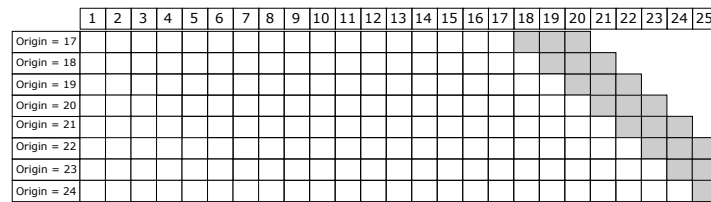


Figura 6.9: Método de validación cruzada usando *Rolling Origin 2*

Finalmente, en los dos casos anteriores tuvimos un tamaño creciente en la muestra . Sin embargo, para algunos fines de investigación, es posible que necesitemos una muestra constante . La figura 6.10 demuestra tal situación. En este caso, en cada iteración agregamos una observación al final de la serie y eliminamos una del comienzo de la serie (celdas de color gris oscuro) .

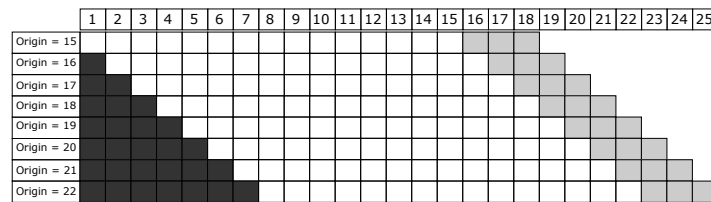


Figura 6.10: Método de validación cruzada usando *Rolling Origin*

Para validar el modelo predictivo, se implementó la técnica de validación cruzada *Rolling Origin*. Se utilizó esta técnica, debido a la estructura temporal de los datos y la necesidad de maximizar el uso de los 306 registros disponibles.

Los resultados de la validación de nuestro modelo predictivo, a través del metodo de validacion cruzada *Rolling Origin* con una ventana de tiempo de 5 registros diarios con saltos de 3 días de un total de 306 registros de datos, muestran un excelente desempeño. Esto se evidencia por un RMSE de 0.2565, un MAE de 0.1499 y un coeficiente de determinación (R^2) de 0.9216. El bajo valor de RMSE indica que, en promedio, las predicciones del modelo están a una distancia de 0.2565 unidades de los valores reales, sugiriendo una alta precisión. Además, el MAE, al ser también muy bajo, refuerza esta precisión promedio, destacando que los errores no están fuertemente influenciados por valores atípicos. El alto valor de R^2 sugiere que el modelo explica el 92.16% de la variabilidad en los datos, demostrando una

fuerte relación entre las variables predictoras y la variable de respuesta. Estos resultados indican que el modelo tiene un alto poder explicativo y una capacidad predictiva robusta como se observa en las Figuras 6.11 y 6.12. Esto lo convierte en una herramienta confiable para la estimación de la evapotranspiración, con significativas aplicaciones prácticas en el manejo y la planificación agrícola.

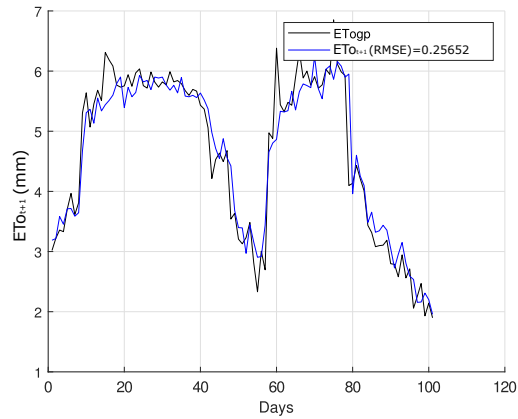


Figura 6.11: RMSE del modelo ET_{t+1} con respecto al modelo de referencia $FAO56 - PM$

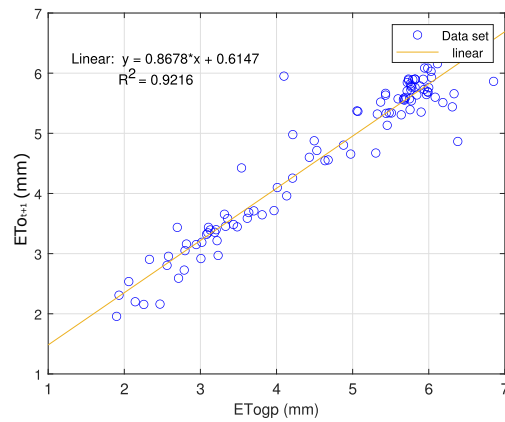


Figura 6.12: R^2 del modelo ET_{t+1} y el modelo de referencia $FAO56 - PM$

6.5. Conclusiones

En conclusión, el modelo predictivo desarrollado mediante Programación Genética demostró ser altamente efectivo y robusto en la estimación de la evapotranspiración. La validación del modelo se llevó a cabo utilizando dos enfoques distintos: primero, una comparación con el modelo de medias móviles simples, donde nuestro modelo superó sutilmente al método de medias móviles en términos de precisión. Este resultado preliminar sugirió una mejora en la capacidad predictiva, aunque marginal.

Sin embargo, fue en la segunda validación, empleando la técnica de validación cruzada *Rolling Origin*, donde nuestro modelo mostró su verdadero potencial. Este método permitió evaluar el desempeño del modelo en diferentes segmentos de la serie temporal, respetando la estructura temporal de los datos y asegurando una utilización eficiente de los 306 registros diarios disponibles. Los resultados de esta validación destacaron el alto poder explicativo y la robusta capacidad predictiva de nuestro modelo, evidenciado por métricas sólidas como un RMSE de 0.2565, un MAE de 0.1499 y un coeficiente de determinación (R^2) de 0.9216. Estos valores indican que el modelo no solo captura la variabilidad inherente de los datos de manera efectiva, sino que también proporciona predicciones precisas y confiables.

En resumen, el modelo predictivo basado en Programación Genética no solo supera ligeramente a los modelos tradicionales como el de medias móviles, sino que también ofrece una solución robusta y precisa para la estimación de la evapotranspiración, con potenciales aplicaciones prácticas significativas en el manejo y la planificación agrícola. La validación mediante *Rolling Origin* confirma su capacidad de generalización y su adecuación para aplicaciones en entornos dinámicos y exigentes.

Capítulo 7

Caso de estudio

Este capítulo proporciona una descripción detallada del caso de estudio llevado a cabo en este trabajo de investigación, con el objetivo de validar el modelo predictivo desarrollado. En particular, se analizó el desempeño del modelo con un mes de datos de un invernadero de la Universidad Americana de Beirut, Líbano (AUB), evaluando su precisión, robustez y capacidad de generalización. Se esperaba que este análisis no solo demostrara la eficacia del modelo predictivo basado en Programación Genética, sino que también aportara valiosos conocimientos sobre su aplicabilidad en diversos contextos y condiciones climáticas, reforzando su potencial como herramienta confiable para la estimación de la evapotranspiración.

7.1. Estimación de la evapotranspiración con $ET_{o_{gp}}$ en un invernadero

En primera instancia se estimó la evapotranspiración de referencia con el modelo $ET_{o_{gp}}$ (Ecuación 5.3) obtenido en la primera fase de nuestra metodología de investigación 1.1.

El modelo de evapotranspiración $ET_{o_{gp}}$ fue sometido a un proceso de evaluación en un entorno real. El conjunto de datos utilizado en la evaluación abarca del 1 al 31 de octubre de 2016. Tras la evaluación del modelo $ET_{o_{gp}}$, las métricas obtenidas fueron un error cuadrático medio (RMSE) de 0.189 y un coeficiente de determinación (R^2) de 0.931. Estos resultados demuestran un rendimiento excepcional del modelo, con estimaciones extremadamente cercanas a los valores observados, reflejando una alta precisión en la minimización de los errores. El alto R^2 indica que el modelo explica aproximadamente el 93.1% de la variabilidad en los datos de evapotranspiración, señalando un excelente ajuste a los datos observados.

Aunque estos resultados están basados en un periodo corto, la combinación de un RMSE muy bajo y un R^2 alto sugiere que el modelo es extremadamente preciso y confiable. Este desempeño validado destaca la capacidad del modelo para proporcionar estimaciones precisas de la evapotranspiración, siendo una herramienta altamente eficaz para la gestión

de recursos hídricos y la agricultura de precisión, incluso en periodos cortos y contextos específicos.

En las figuras 7.1 y 7.2 podemos observar gráficamente el rendimiento del modelo con pocos datos.

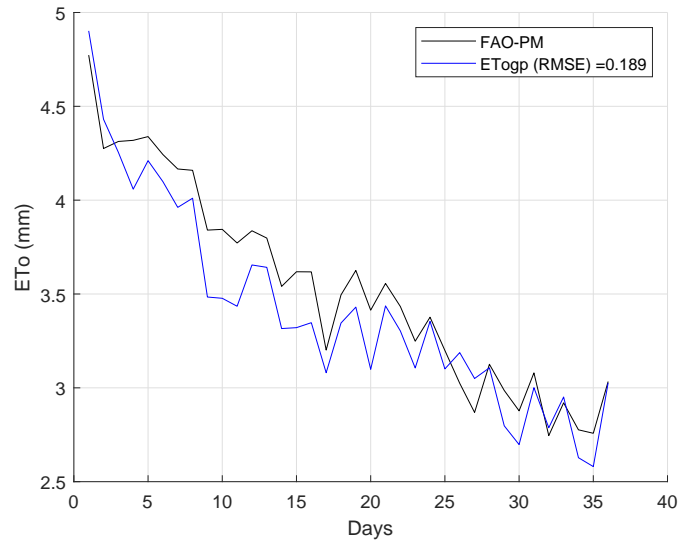


Figura 7.1: R^2 RMSE de ET_{ogp}

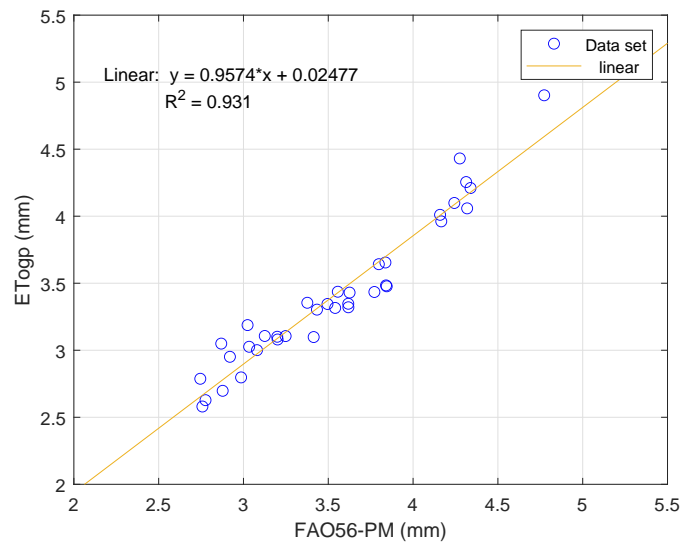


Figura 7.2: R^2 ET_{ogp}

7.2. Predecir la evapotranspiración con ETo_{t+1}

Para predecir la evapotranspiración de referencia en invernaderos se utilizó el modelo ETo_{t+1} (Ecuación 6.4) desarrollado en la segunda fase de la metodología de solución (figura 1.1). Como datos de entrada al modelo ETo_{t+1} se utilizó la evapotranspiración de referencia histórica diaria obtenida con el modelo ETo_{gp} (Ecuación 5.3) la cual cuenta con 31 datos. Los resultados obtenidos fueron un RMSE de 0.373 y el coeficiente de determinación, R^2 con un valor de 0.7577.

Interpretación y Significado

La evaluación del modelo ETo_{t+1} (Ecuación 6.4) con solo 31 datos proporciona una perspectiva específica sobre su rendimiento en un corto plazo. La combinación de un bajo $RMSE$ y un alto R^2 sugiere que el modelo predictivo es preciso y confiable para este conjunto de datos limitado. Sin embargo, es importante considerar que estos resultados están basados en un periodo corto y podrían variar con datos de periodos más largos o diferentes condiciones climáticas.

En resumen, el desempeño del modelo predictivo en términos de $RMSE$ y R^2 valida su potencial como una herramienta eficaz para la estimación de la evapotranspiración en periodos cortos. Estos resultados contribuyen a la comprensión de la aplicabilidad del modelo en situaciones específicas y destacan su precisión y confiabilidad en la estimación de la evapotranspiración.

En las figuras 7.3 y 7.4 podemos observar gráficamente las predicciones del modelo ETo_{t+1} con pocos datos.

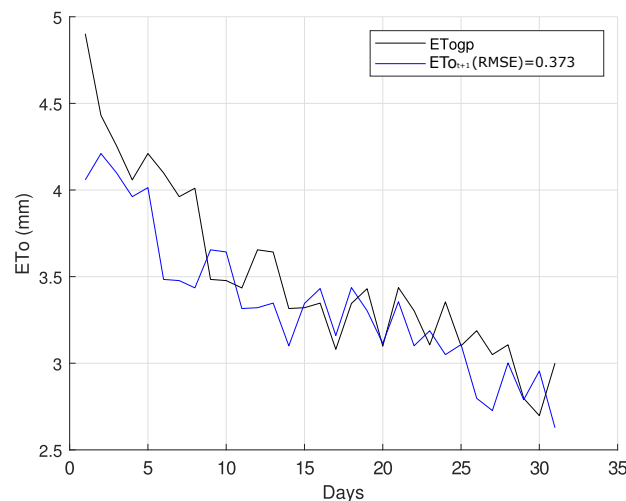


Figura 7.3: RMSE del modelo ETo_{t+1}

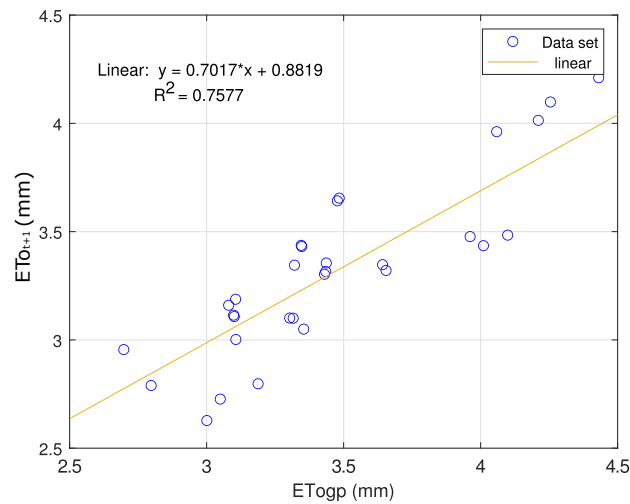


Figura 7.4: R^2 ETo_{t+1}

7.3. Conclusiones

La aplicación del caso de estudio fue sumamente útil, ya que demostró tanto la precisión como la robustez del modelo predictivo en condiciones de invernadero. A través de este estudio, se evidenció la capacidad del modelo para proporcionar estimaciones concisas de la evapotranspiración, ajustándose de manera precisa a las variaciones específicas de los datos dentro del entorno controlado de un invernadero.

Además, el análisis mostró que el modelo es altamente robusto, capaz de mantener su desempeño incluso frente a fluctuaciones en las condiciones ambientales, lo que es esencial para la gestión eficiente del riego y la optimización del uso del agua en la agricultura de precisión. Estos hallazgos no solo resaltan la eficacia del modelo en escenarios controlados, sino que también subrayan su potencial aplicabilidad en diversas situaciones agrícolas, promoviendo prácticas sostenibles y mejorando la producción agrícola.

Capítulo 8

Conclusiones y trabajos futuros

En este capítulo, tras realizar las evaluaciones pertinentes, se realiza un análisis detallado de las conclusiones de la investigación, basándose en los resultados obtenidos y alineados con los objetivos planteados inicialmente. Se detallan los productos generados y se destacan sus principales aportaciones. Este análisis proporciona una visión integral del impacto y la relevancia del trabajo realizado, así como de las oportunidades para futuras investigaciones que puedan continuar avanzando en este ámbito.

8.1. Conclusiones generales

En este trabajo de investigación, se planteó como objetivo principal desarrollar un modelo predictivo para estimar la evapotranspiración de referencia en invernaderos mediante Programación Genética. Las actividades llevadas a cabo en concordancia con los objetivos establecidos se detallan en la Tabla **8.1**.

Tabla 8.1: Descripción de actividades realizadas

Objetivo	Actividad
Seleccionar y adaptar las variables climáticas y ambientales relevantes para estimar la evapotranspiración de referencia en el contexto específico de los invernaderos, considerando factores como la radiación solar, temperatura, humedad relativa y velocidad del viento.	Para alcanzar este objetivo, se realizó una exhaustiva revisión de la literatura, donde se identificó que el Boletín 56 de la FAO detalla los principales parámetros climáticos que influyen en la evapotranspiración: radiación solar, temperatura del aire, humedad atmosférica y velocidad del viento (Allen, 1998), descrito en la sección 2. Este marco teórico se empleó como fundamento para seleccionar las variables pertinentes en la estimación precisa de la evapotranspiración en el contexto específico de los invernaderos.
Recolectar y preparar una base de datos de alta calidad con datos climáticos históricos y de evapotranspiración de referencia en invernaderos.	Las actividades desarrolladas para alcanzar este objetivo incluyeron la obtención de datos detallados del área de invernaderos proporcionados por la Universidad Americana de Beirut en Líbano. Estos datos fueron cuidadosamente procesados y estructurados en conjuntos de datos que sirvieron como entrada para el algoritmo evolutivo. Se realizaron preparativos meticulosos para asegurar la calidad y coherencia de los datos antes de ser utilizados en la modelización. Posteriormente, se implementaron técnicas de optimización mediante Programación Genética, con el fin de desarrollar un modelo preciso para estimar la evapotranspiración de referencia en condiciones específicas de invernaderos. La actividad se describe en la sección 5.1

Objetivo	Actividad
Configurar e implementar un algoritmo evolutivo que pueda generar modelos matemáticos precisos para la estimación de la evapotranspiración de referencia.	En esta actividad se definió la estructura del algoritmo, incluyendo la representación de los individuos (modelos matemáticos), las funciones de aptitud (criterios de evaluación) y los operadores evolutivos (selección, cruce, mutación). Se establecieron los parámetros del algoritmo, como el tamaño de la población, el número de generaciones, la tasa de mutación y la tasa de cruce, a través de experimentación y ajuste fino. La implementación del algoritmo, se desarrolló en el entorno de programación Matlab y la librería GPLab. La actividad se describe en la sección 5.2
Evaluar la precisión y eficacia del modelo predictivo mediante la comparación de sus predicciones con el modelo de referencia FAO Penman-Monteith en invernaderos, utilizando métricas de validación adecuadas.	Esta actividad consistió en una evaluación comparativa para determinar la precisión de nuestro modelo predictivo en relación con otras técnicas de estimación de la evapotranspiración. Se llevó a cabo una comparación detallada contra el modelo de Hargreaves-Samani y un atmómetro, a través del modelo de referencia FAO56-Penman-Monteith (FAO56-PM). La evaluación del modelo se detalla en la sección ??

8.2. Logros obtenidos

La difusión de los resultados obtenidos fue una parte crucial de este trabajo de investigación. A continuación, se detallan los logros alcanzados a lo largo del desarrollo de esta investigación.

8.2.1. Productos

- Reporte del estado del arte. El reporte identifica las fortalezas y debilidades de los métodos analizados. Asimismo, buscamos resaltar de manera clara y precisa nuestra contribución en relación con las propuestas existentes en la literatura. Sección 3
- Desarrollo de una metodología integral para la creación de modelos destinados a estimar la evapotranspiración y predecirla mediante un enfoque evolutivo. Descrita en la sección 1.5

-
- Artículo científico publicado en la revista *scientific reports* con factor de impacto de 3.8 (F. J. Ruiz-Ortega et.al., 2024). Figura A.1
 - Capítulo de libro presentado y aceptado en la convocatoria «Experiencias del Tecnológico Nacional de México y sus aliados estratégicos en el sector agroalimentario» de la Red de Atención al Sector Agroalimentario RASA del TecNM y el Colegio del Estado de Hidalgo (CEH). Figura A.2
 - Artículo presentado *In the 8th International Symposium on Language & Knowledge Engineering*. Figura A.3
 - Artículo publicado en la revista *Computación y Sistemas*, con el ISSN 2007-9737. Esta revista se encuentra indexada en diversas bases de datos y plataformas reconocidas, incluyendo el Índice de Excelencia de Revistas Mexicanas de CONACYT, Scopus, Web of Science, E-Journal, REDIB, Latindex, Biblat, Periodica, DBLP y SciELO. (J. Ruiz-Ortega et.al., 2022). Figura A.4
 - Artículo presentado como colaborador en la 9a Jornada de Ciencia y Tecnología Aplicada. Figura A.5

8.2.2. Participación en concursos

Certificado de asistencia y participación en el concurso «*21st Annual Humies Awards – Melbourne, Australia*». Nuestro trabajo fue seleccionado como uno de los ocho finalistas. Figura A.6

8.2.3. Estancia de investigación

Realicé estancia de investigación en el TecNM/I.T. Ensenada, durante el período del 15 de agosto del 2021 al 31 de enero de 2022. Figura A.7

Profesor responsable: Dr. Eddie H. Clemente Torres. Profesor investigador de tiempo completo.

Pertenece al Sistema Nacional de Investigadores Nivel I.

Perfil deseable PRODEP.

Es miembro del cuerpo académico: Sistemas cognitivos.

Líder de línea de investigación: Sistemas cognitivos.

- Visión por Computadora.
- Cómputo Evolutivo.
- Aprendizaje de máquinas.

8.2.4. Participación en proyectos de investigación

Título del proyecto: Invernadero con tecnología de IoT para el monitoreo de variables climáticas

Clave: 14813.22-P

Vigencia: Del 1 de enero al 31 de diciembre de 2022

Director(a) responsable: Martínez-Rebollar, Alicia

Estatus: Liberado

8.2.5. Ponencias

«PROTOTIPO DE INTERNET DE LAS COSAS PARA EL MONITOREO DE VARIABLES CLIMÁTICAS EN AGRICULTURA PROTEGIDA». Figura B.1

Realizada el 13 de septiembre de 2021 en conmemoración al “DÍA DEL PROGRAMADOR”, en el TecNM/I.T.Cuautla.

8.2.6. Trabajos futuros

Algunos trabajos futuros de investigación emanados de nuestra investigación pueden incluir:

- **Mejora del modelo:** Optimizar el modelo actual utilizando técnicas avanzadas de Programación Genética, como la incorporación de operadores evolutivos más sofisticados o el uso de enfoques híbridos que combinen Programación Genética con otros métodos de inteligencia artificial.
- **Ampliación de la base de datos:** Recopilar y analizar datos adicionales de diferentes invernaderos y condiciones climáticas para aumentar la robustez y la capacidad de generalización del modelo, permitiendo su aplicación en una variedad más amplia de entornos.
- **Aplicaciones prácticas:** Implementar el modelo en sistemas de gestión del riego en invernaderos para evaluar su desempeño en tiempo real y su impacto en la eficiencia del uso del agua y la productividad de los cultivos.
- **Desarrollo de interfaces usuario-amigables:** Crear herramientas de software con interfaces gráficas intuitivas que permitan a los agricultores y gestores de invernaderos utilizar el modelo fácilmente para la toma de decisiones.
- **Comparación con otras técnicas de IA:** Realizar estudios comparativos entre la Programación Genética y otras técnicas de inteligencia artificial, como redes neuronales profundas y máquinas de soporte vectorial, para identificar las ventajas y desventajas de cada enfoque en la estimación de la evapotranspiración.

-
- **Evaluación del impacto del cambio climático:** Investigar cómo el cambio climático puede afectar los patrones de evapotranspiración y adaptar el modelo para predecir estos cambios de manera precisa, contribuyendo a la resiliencia agrícola.
 - **Automatización y control en tiempo real:** Desarrollar sistemas automatizados que integren el modelo de Programación Genética con sensores en tiempo real para ajustar dinámicamente las prácticas de riego en función de las condiciones ambientales actuales.

Estos trabajos futuros no solo profundizarán en el conocimiento actual, sino que también mejorarán la aplicabilidad y la eficacia del modelo, contribuyendo a la sostenibilidad y eficiencia en la gestión del agua en la agricultura protegida.

8.2.7. Otras actividades

Durante la investigación, se llevaron a cabo diversas actividades con el propósito de reforzar y ampliar los conocimientos adquiridos dentro del marco del plan de trabajo establecido. Estas actividades no solo consolidaron la comprensión de los conceptos fundamentales, sino que también enriquecieron la formación académica y profesional durante mis estudios de doctorado. A través de estas experiencias, se logró una integración más profunda de los conocimientos teóricos y prácticos, facilitando el desarrollo de habilidades críticas y técnicas esenciales para el éxito de la investigación. A continuación, enlisto las diversas actividades realizadas.

Participación en:

- Curso: «Introducción al Cómputo Evolutivo». Figura C.1
- Curso: «Introducción al análisis de datos meteorológicos con R». Figura C.2
- «Macroentrenamiento en Inteligencia Artificial MeIA». Figura C.3
- Diplomado «Ciencia de Datos ofrecido por el TecNM». Figura C.4
- Curso: «Aprendizaje profundo con redes neuronales». Figura C.5
- Asistente: «Encuentro Nacional de Computación 2020». Figura C.6
- Asistente: «X Seminario y Escuela Nacional de Aprendizaje e Inteligencia Computacional». Figura C.7

Referencias

- Allen, R. G. (1998). Crop evapotranspiration-guideline for computing crop water requirements. *Irrigation and drain*, 56, 300.
- Allen, R. G., Pereira, L. S., Raes, D., y Smith, M. (2006). Evapotranspiración del cultivo: guías para la determinación de los requerimientos de agua de los cultivos. *Roma: FAO*, 298(0).
- Allende, M., Salinas, L., Torres, A., y cols. (2017). *Manual de cultivo del tomate bajo invernadero*. Instituto de Investigaciones Agropecuarias (Chile).
- Antonopoulos, V. Z., y Antonopoulos, A. V. (2017). Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Computers and Electronics in Agriculture*, 132, 86–96.
- Balmat, J.-F., Lafont, E., Ali, A. M., Pessel, N., y Fernández, J. C. R. (2019). Evaluation of the reference evapotranspiration for a greenhouse crop using an adaptive-network-based fuzzy inference system (anfis). En *Proceedings of the 3rd international conference on machine learning and soft computing* (pp. 203–205).
- Beyaztas, U., y Shang, H. L. (2019). Robust forecasting procedure based on weighted likelihood estimators for functional time series. *Journal of Forecasting*, 38(8), 688–702.
- Bianchi, F. D. (2001). Introducción a matlab. *pp. Universidad Nacional de la Plata, Argentina*.
- Castañeda-Miranda, R., Ventura-Ramos Jr, E., Peniche-Vera, R. d. R., y Herrera-Ruiz, G. (2007). Análisis y simulación del modelo físico de un invernadero bajo condiciones climáticas de la región central de México. *Agrociencia*, 41(3), 317–335.
- (CIMIS)[®]. (2022, mes-nommarzo). *The california irrigation management information system*. <https://cimis.water.ca.gov/Default.aspx>. (Accedido en marzo de 2022 (ejemplo))
- de la Paix Mupenzi, J., Li, L., Ge, J., Ngamije, J., Achal, V., Habiyaremye, G., y de Dieu Habumugisha, J. (2012). Water losses in arid and semi-arid zone: Evaporation, evapotranspiration and seepage. *Journal of Mountain Science*, 9, 256–261.
- de la Piedra, M. C., Martínez, A. G. C., de Tuxtla Gutiérrez, I., Molina, J. A. L., Navarro, N. A. M., y Sánchez, J. O. G. (2016). Modelo neurodifuso para el control de humedad del

-
- suelo en cultivo hidropónico para la planta de tomate. *Revista Tecnología Digital Vol, 6(1)*, 43–56.
- El-Baroudy, I., Elshorbagy, A., Carey, S., Giustolisi, O., y Savic, D. (2010). Comparison of three data-driven techniques in modelling the evapotranspiration process. *Journal of hydroinformatics, 12(4)*, 365–379.
- Eslamian, S. S., Gohari, S. A., Zareian, M. J., y Firoozfar, A. (2012). Estimating penman–monteith reference evapotranspiration using artificial neural networks and genetic algorithm: a case study. *Arabian Journal for Science and Engineering, 37*, 935–944.
- Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., y Zeng, W. (2019). Light gradient boosting machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural water management, 225*, 105758.
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., . . . Xiang, Y. (2018). Evaluation of svm, elm and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of china. *Agricultural and forest meteorology, 263*, 225–241.
- Feng, Y., Cui, N., Gong, D., Zhang, Q., y Zhao, L. (2017). Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. *Agricultural Water Management, 193*, 163–173.
- Ferreira, L. B., da Cunha, F. F., de Oliveira, R. A., y Fernandes Filho, E. I. (2019). Estimation of reference evapotranspiration in brazil with limited meteorological data using ann and svm—a new approach. *Journal of Hydrology, 572*, 556–570.
- García, S., Luengo, J., y Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer.
- Ge, J., Zhao, L., Yu, Z., Liu, H., Zhang, L., Gong, X., y Sun, H. (2022). Prediction of greenhouse tomato crop evapotranspiration using xgboost machine learning model. *Plants, 11(15)*, 1923.
- Géron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and tensorflow*. O'Reilly Media, Inc.
- Gómez, C. S., Cauich, I. C., Batista, B. d. P., y José, C. C. L. (2024). Análisis de la situación agrícola de la república mexicana. *Revista de la Universidad del Zulia, 15(42)*, 8–36.
- Güven, A., AYTEK, A., YUCE, M. I., y AKSOY, H. (2008). Genetic programming-based empirical model for daily reference evapotranspiration estimation. *Clean-Soil, Air, Water, 36(10-11)*, 905–912.
- Han, Y., Wu, J., Zhai, B., Pan, Y., Huang, G., Wu, L., y Zeng, W. (2019). Coupling a bat algorithm with xgboost to estimate reference evapotranspiration in the arid and semiarid regions of china. *Advances in Meteorology, 2019*, 1–16.

-
- Hargreaves, G. H., y Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, 1(2), 96–99.
- Hernández-Salazar, J. A., Hernández-Rodríguez, D., Hernández-Cruz, R. A., Ramos-Fernández, J. C., Márquez-Vera, M. A., y Trejo-Macotela, F. R. (2019). Estimation of the evapotranspiration using anfis algorithm for agricultural production in greenhouse. En *2019 IEEE International Conference on Applied Science and Advanced Technology (ICASAT)* (pp. 1–5).
- Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., ... Zhou, H. (2019). Evaluation of catboost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, 574, 1029–1041.
- Hyndman, R. J., y Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). Melbourne, Australia: OTexts. (<https://OTexts.com/fpp2> [Accessed on June 2024])
- Izadifar, Z., y Elshorbagy, A. (2010). Prediction of hourly actual evapotranspiration using neural networks, genetic programming, and statistical models. *Hydrological processes*, 24(23), 3413–3425.
- Jaafar, H. H., y Ahmad, F. (2018). Evaluating atmometer performance for estimating reference evapotranspiration in ventilated and unventilated greenhouses. *Journal of Irrigation and Drainage Engineering*, 144(7), 04018014.
- Jensen, M. E., Burman, R. D., y Allen, R. G. (1990). Evapotranspiration and irrigation water requirements: a manual. *ASCE manuals and reports on engineering practice (USA)*. no. 70..
- Jovic, S., Nedeljkovic, B., Golubovic, Z., y Kostic, N. (2018). Evolutionary algorithm for reference evapotranspiration analysis. *Computers and electronics in agriculture*, 150, 1–4.
- Jung, D.-H., Lee, T. S., Kim, K., y Park, S. H. (2022). A deep learning model to predict evapotranspiration and relative humidity for moisture control in tomato greenhouses. *Agronomy*, 12(9), 2169.
- Karaca, C., Tezcan, A., Büyüктаş, K., Büyüктаş, D., y Baştuğ, R. (2018). Equations developed to estimate evapotranspiration in greenhouses. *Yuzuncu Yıl University Journal of Agricultural Sciences*, 28(4), 482–489.
- Karimi, S., Kisi, O., Kim, S., Nazemi, A. H., y Shiri, J. (2017). Modelling daily reference evapotranspiration in humid locations of south korea using local and cross-station data management scenarios. *International Journal of Climatology*, 37(7), 3238–3246.
- Kiafar, H., Babazadeh, H., Marti, P., Kisi, O., Landeras, G., Karimi, S., y Shiri, J. (2017). Evaluating the generalizability of gep models for estimating reference evapotranspiration in distant humid and arid locations. *Theoretical and Applied Climatology*, 130, 377–389.

-
- Kim, S., y Kim, H. S. (2008). Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. *Journal of hydrology*, 351(3-4), 299–317.
- Kisi, O. (2010). Fuzzy genetic approach for modeling reference evapotranspiration. *Journal of irrigation and drainage engineering*, 136(3), 175–183.
- Kisi, O., y Guven, A. (2010). Evapotranspiration modeling using linear genetic programming technique. *Journal of Irrigation and Drainage Engineering*, 136(10), 715–723.
- Kisi, O., Sanikhani, H., Zounemat-Kermani, M., y Niazi, F. (2015). Long-term monthly evapotranspiration modeling by several data-driven methods without climatic data. *Computers and Electronics in Agriculture*, 115, 66–77.
- Köppen, W., y Geiger, R. (1930). *Handbuch der klimatologie* (Vol. 1). Gebrüder Borntraeger Berlin.
- Koza, J. R. (1992). Genetic programming, on the programming of computers by means of natural selection. a bradford book. *MIT Press*.
- Kumar, D., Adamowski, J., Suresh, R., y Ozga-Zielinski, B. (2016). Estimating evapotranspiration using an extreme learning machine model: case study in north bihar, india. *Journal of Irrigation and Drainage Engineering*, 142(9), 04016032.
- Landeras, G., Bekoe, E., Ampofo, J., Logah, F., Diop, M., Cisse, M., y Shiri, J. (2018). New alternatives for reference evapotranspiration estimation in west africa using limited weather data and ancillary data supply strategies. *Theoretical and Applied Climatology*, 132, 701–716.
- Makridakis, S., Wheelwright, S. C., y Hyndman, R. J. (2020). *Forecasting: Methods and applications* (5th ed.). Wiley.
- Martí, P., González-Altozano, P., López-Urrea, R., Mancha, L. A., y Shiri, J. (2015). Modeling reference evapotranspiration with calculated targets. assessment and implications. *Agricultural Water Management*, 149, 81–90.
- Mattar, M. A. (2018). Using gene expression programming in monthly reference evapotranspiration modeling: a case study in egypt. *Agricultural Water Management*, 198, 28–38.
- Mattar, M. A., y Alazba, A. (2019). Gep and mlr approaches for the prediction of reference evapotranspiration. *Neural Computing and Applications*, 31, 5843–5855.
- McMahon, T., Peel, M., Lowe, L., Srikanthan, R., y McVicar, T. (2013). Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis. *Hydrology and Earth System Sciences*, 17(4), 1331–1363.
- Mehdizadeh, S. (2018). Estimation of daily reference evapotranspiration (eto) using artificial intelligence methods: Offering a new approach for lagged eto data-based modeling. *Journal of hydrology*, 559, 794–812.

-
- Mehdizadeh, S., Behmanesh, J., y Khalili, K. (2017). Using mars, svm, gep and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and electronics in agriculture*, 139, 103–114.
- Mehr, A. D., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A. M., y Yaseen, Z. M. (2018). Genetic programming in water resources engineering: A state-of-the-art review. *Journal of hydrology*, 566, 643–667.
- Monteith, J. L. (1965). Evaporation and environment. En *Symposia of the society for experimental biology* (Vol. 19, pp. 205–234).
- Moreno Reséndez, A., Aguilar Durón, J., y Luévano González, A. (2011). Características de la agricultura protegida y su entorno en México. *Revista Mexicana de agronegocios*, 29(1345-2016-104296), 763–774.
- Nourani, V., Komasi, M., y Alami, M. T. (2012). Hybrid wavelet–genetic programming approach to optimize ann modeling of rainfall–runoff process. *Journal of Hydrologic Engineering*, 17(6), 724–741.
- Orona-Castillo, I., Del-Toro-Sánchez, C. L., Fortis-Hernández, M., Preciado-Rangel, P., Espinoza-Arellano, J., Rueda-Puente, E., ... Cano-Ríos, P. (2022). Indicadores técnico-económicos de la producción del cultivo de tomate bajo agricultura protegida en la comarca lagunera, México. *Biotecnia*, 24(3), 70–76.
- Pandir, Y., y Ulusoy, H. (2013). New generalized hyperbolic functions to find new exact solutions of the nonlinear partial differential equations. *Journal of Mathematics*, 2013.
- Parasuraman, K., Elshorbagy, A., y Carey, S. K. (2007). Modelling the dynamics of the evapotranspiration process using genetic programming. *Hydrological Sciences Journal*, 52(3), 563–578.
- Pratt, L., Ortega, J. M., Nieto, E., y Braly-Cartillier, I. (2019). Agricultura protegida en México. *Elaboración de la metodología para el primer bono verde agrícola certificado*, 1–57.
- Pratt, P., y Ortega, J. (2019). Elaboración de la metodología para el primer bono verde certificado. *Agricultura Protegida en México*, Nieto, E. y Braly CI Editores. Banco Interamericano de Desarrollo Nota Técnica No. IDB-TM-1668.
- Ramírez Barraza, B. A., González Estrada, A., Valdivia Alcalá, R., Salas González, J. M., y García Salazar, J. A. (2019). Tarifas eficientes para el agua de uso agrícola en la comarca lagunera. *Revista mexicana de ciencias agrícolas*, 10(3), 539–550.
- Ruiz-Ortega, F. J., Clemente, E., Martínez-Rebollar, A., y Flores-Prieto, J. J. (2024). An evolutionary parsimonious approach to estimate daily reference evapotranspiration. *Scientific Reports*, 14(1), 6736.
- Ruiz-Ortega, J., Martínez-Rebollar, A., Flores-Prieto, J., y Estrada-Esquivel, H. (2022). Design on a low cost iot architecture for greenhouses monitoring. *Computación y Sistemas*, 26(1), 221–232.

-
- Sanikhani, H., Kisi, O., Maroufpoor, E., y Yaseen, Z. M. (2019). Temperature-based modeling of reference evapotranspiration using several artificial intelligence models: application of different modeling scenarios. *Theoretical and applied climatology*, 135, 449–462.
- Shiri, J. (2017). Evaluation of fao56-pm, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of iran. *Agricultural water management*, 188, 101–114.
- Shiri, J. (2019). Modeling reference evapotranspiration in island environments: Assessing the practical implications. *Journal of Hydrology*, 570, 265–280.
- Shiri, J., Kişi, Ö., Landeras, G., López, J. J., Nazemi, A. H., y Stuyt, L. C. (2012). Daily reference evapotranspiration modeling by using genetic programming approach in the basque country (northern spain). *Journal of Hydrology*, 414, 302–316.
- Shiri, J., Marti, P., Karimi, S., y Landeras, G. (2019). Data splitting strategies for improving data driven models for reference evapotranspiration estimation among similar stations. *Computers and Electronics in Agriculture*, 162, 70–81.
- Shiri, J., Nazemi, A. H., Sadraddini, A. A., Landeras, G., Kisi, O., Fard, A. F., y Marti, P. (2014). Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in iran. *Computers and Electronics in Agriculture*, 108, 230–241.
- Shiri, J., Sadraddini, A. A., Nazemi, A. H., Kisi, O., Landeras, G., Fard, A. F., y Marti, P. (2014). Generalizability of gene expression programming-based approaches for estimating daily reference evapotranspiration in coastal stations of iran. *Journal of hydrology*, 508, 1–11.
- Shiri, J., Sadraddini, A. A., Nazemi, A. H., Kisi, O., Marti, P., Fard, A. F., y Landeras, G. (2013). Evaluation of different data management scenarios for estimating daily reference evapotranspiration. *Hydrology Research*, 44(6), 1058–1070.
- Silva, S., y Almeida, J. (2003). Gplab-a genetic programming toolbox for matlab. En *Proceedings of the nordic matlab conference* (pp. 273–278).
- Srivastava, A., Sahoo, B., Raghuwanshi, N. S., y Chatterjee, C. (2018). Modelling the dynamics of evapotranspiration using variable infiltration capacity model and regionally calibrated hargreaves approach. *Irrigation Science*, 36, 289–300.
- Svetunkov, I., y Petropoulos, F. (2018). Old dog, new tricks: a modelling view of simple moving averages. *International Journal of Production Research*, 56(18), 6034–6047.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 16(4), 437–450.
- Traore, S., y Guven, A. (2012). Regional-specific numerical models of evapotranspiration using gene-expression programming interface in sahel. *Water resources management*, 26, 4367–4380.

-
- Traore, S., y Guven, A. (2013). New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of west africa. *Irrigation Science*, 31, 1–10.
- Traore, S., Luo, Y., y Fipps, G. (2017). Gene-expression programming for short-term forecasting of daily reference evapotranspiration using public weather forecast information. *Water resources management*, 31, 4891–4908.
- Turing, A. (1948). Intelligent machinery [technical report]. *Teddington: National Physical Laboratory (see also Copeland BJ (ed) 2004 The Essential Turing: seminal writings in Computing Logic, Philosophy, artificial Intelligence, and Artificial Life plus The Secrets of Enigma. Oxford University Press, Oxford).*
- Vargas-Canales, J. M., Palacios-Rangel, M. I., Aguilar-Ávila, J., Camacho-Vera, J. H., Ocampo-Ledesma, J. G., y Medina-Cuellar, S. E. (2018). Efficiency of small enterprises of protected agriculture in the adoption of innovations in mexico. *Estudios Gerenciales*, 34(146), 52–62.
- Wheelwright, S., Makridakis, S., y Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.
- Wu, L., y Fan, J. (2019). Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. *PLoS one*, 14(5), e0217520.
- Yassin, M. A., Alazba, A. A., y Mattar, M. A. (2016). Comparison between gene expression programming and traditional models for estimating evapotranspiration under hyper arid conditions. *Water Resources*, 43, 412–427.
- Yu, J., Zheng, W., Xu, L., Zhangzhong, L., Zhang, G., y Shan, F. (2020). A pso-xgboost model for estimating daily reference evapotranspiration in the solar greenhouse. *Intelligent Automation & Soft Computing*, 26(5).
- Yuan, H., Feng, C., Li, J., Zhang, J., Wang, X., y Cheng, M. (2023). A method to estimate evapotranspiration in greenhouse conditions by artificial neural networks using limited climate parameters. En *Journal of physics: Conference series* (Vol. 2650, p. 012029).

Apéndice A

Productos

scientific reports

 Check for updates

OPEN

An evolutionary parsimonious approach to estimate daily reference evapotranspiration

F. Javier Ruiz-Ortega^{1,2}, Eddie Clemente³, Alicia Martínez-Rebollar¹ & J. Jassón Flores-Prieto¹

The reference evapotranspiration (ET₀) is an essential component in hydrological and ecological processes. The objective of this research is to develop an explicit model to estimate ET₀ only using commonly measurable meteorological parameters such as relative humidity, air temperature, and wind speed, where the measurements corresponding to solar radiation are omitted. The model was generated using Genetic Programming (GP), evaluated, and validated with reference data ET₀ using FAO56-PM. This reference data was obtained from different climates (warm-temperate and arid-warm) and latitudes, acquired from CIMIS stations in the state of California, United States, and the El Porvenir station in the state of Coahuila, located in north-central Mexico. After applying the proposed methodology, a total of 3754 results were generated, demonstrating a significant improvement in the estimation of ET₀ compared to the Hargreaves–Samani model. A particularly noteworthy result revealed that our approach outperformed the Hargreaves–Samani model in the training phase by 27%, and in the testing phase by 16%, on average. In order to achieve a generalized model, a dataset encompassing meteorological stations in two different climates (warm-temperate and arid-warm) and various latitudes was utilized. The obtained outcome unveiled a highly effective model for estimating ET₀ in diverse climatic contexts, eliminating the need for local adjustments. This model significantly surpassed the Hargreaves–Samani model, exhibiting superior performance by 17% during the training phase and 18% during the testing phase. These results conclusively underscore the capability of our approach to provide more accurate and reliable ET₀ estimates. These results conclusively underscore the capability of our approach to provide more accurate and reliable ET₀ estimates. Finally, to validate the model, four different datasets with climates similar to those used for model creation (warm-temperate, warm-arid) and different latitudes were employed. The validation stage results clearly indicate the superiority of our reference evapotranspiration ET₀₁₁ model over the Hargreaves–Samani model by 51% in warm-temperate climates. For the dataset with arid-warm climate, our model continued to show satisfactory results, surpassing the Hargreaves–Samani model by 8%. GP emerges as an innovative and effective alternative for simplified model development. This approach introduces a novel paradigm that facilitates the efficient development of models, standing out for its simplicity and effectiveness in generating solutions.

Figura A.1: Página principal de artículo JCR. DOI:10.1038/s41598-024-56770-3.

Un enfoque evolutivo para modelar la evapotranspiración de referencia diaria con variables comúnmente medibles en una estación climatológica

FJ Ruiz Ortega, Eddie Clemente, Alicia Martínez Rebollar, Dante Mújica Vargas, Manuel Fortis Hernández

Resumen

El propósito de este estudio fue utilizar un enfoque de cómputo evolutivo para modelar la evapotranspiración de referencia diaria (ET_o). Donde este modelo utilice variables comúnmente medibles en una estación meteorológica, lo anterior como alternativa al modelo de Hargreaves-Samani. Particularmente se empleó Programación Genética (GP) para crear el modelo y se comparó con el modelo de referencia FAO-PM utilizando datos del Sistema de Información para la Gestión de Riego de California (CIMIS). Las variables incluidas en el modelo evolucionado fueron la temperatura del aire, la humedad relativa y la velocidad del viento. Los resultados revelan que los modelos evolucionados logran mejores estimaciones en términos de precisión de ET_o con respecto al modelo de Hargreaves-Samani bajo un clima árido, y un clima templado. Al ser un enfoque basado en programación genética es posible obtener una gran variedad de modelos explícitos y cada uno de ellos puede considerarse una alternativa eficaz para la estimación de la ET_o. Por tanto, el presente trabajo puede tomarse como una metodología para la creación de modelos de ET_o.

Figura A.2: Capítulo de libro.

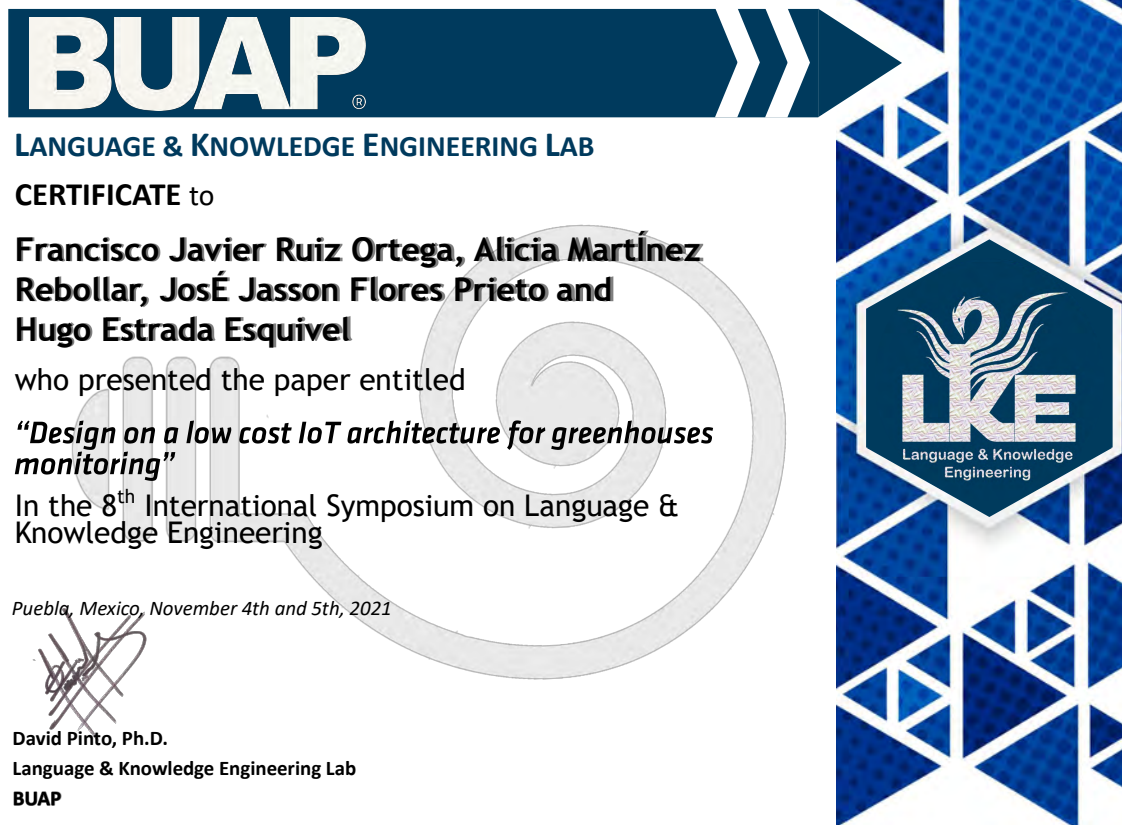


Figura A.3: Certificado de presentación por *LANGUAJE & KNOWLEDGE ENGINEERING LAB*.

Design on a low Cost IoT Architecture for Greenhouses Monitoring

Javier Ruiz-Ortega, Alicia Martínez-Rebollar, Jassón Flores-Prieto,
Hugo Estrada-Esquivel

Tecnológico Nacional de México,
Centro Nacional de Investigación y Desarrollo Tecnológico,
Mexico

{d20ce082, alicia.mr, jose.fp, hugo.ee}@cenidet.tecnm.mx

Abstract. Currently, conventional agriculture has been modified by new quality and quantity requirements and also by changes in climatic factors. In this context, it is necessary to implement emerging technologies that help to optimize natural and economic resources. One of the technological alternatives for the agricultural sector is the implementation of automated greenhouses using the Internet of things concept. One of the difficulties in implementing automated greenhouses is the technical difficulty and cost associated with building these new approaches. This work presents an internet of things architecture to monitor the variables involved within a greenhouse, using wireless connection and internet communication to publish the collected data. The system evaluation was conducted using a mesh covered greenhouse and demonstrated that it is possible to develop low-cost architectures that can connect to the cloud via *WiFi* or *GSM / GPRS* technology to upload data to a web server. This information is the input for new approaches that enable the automatic management of greenhouses.

that the surface decreases mainly due to the degradation of the water table, pollution and the change in the use of the land to build houses [1, 2]. Therefore, the optimization of the yield of production units has been an option for the sustainability of agricultural production, rather than enabling new farmland.

In this change of strategy, producers are focused on improving the yield per unit of production, rather than finding new cultivable farms. The producers are adopting cutting-edge technologies in order to better manage the control of the elements involved in the production process (climate, pests, diseases, nutrients, and water supply). This adoption refers to technological migration and new trends such as applications that integrate internet of things technologies, cloud computing, fog computing, embedded systems, wireless sensors, artificial intelligence, and

Figura A.4: Artículo publicado en la revista *Computación y Sistemas*. DOI:10.13053/cys-26-1-4166.

Utilizando el IoT Para calcular la evapotranspiración de referencia en agricultura protegida

Irvin Velasquez-Mora*, Alicia Martínez-Rebollar*, Yesica I. Saavedra-Benitez*,
Javier Ruiz-Ortega**

*Tecnológico Nacional de México/CI-NDDET, Int. Internado Palmira S/N,
C.P. 62490, Col. Palmira, Cuernavaca, Morelos, México

(e-mail: { m22ev078, alicia.mr, yesica.sb, d20ev082 } @ cenidet.tecnm.mx)

** Tecnológico Nacional de México/Torreón, Carr. Torreón-San Pedro Km 7.5,
Ejido Anna, Torreón, Coahuila, México

Resumen: El agua utilizada para fines agrícolas es un tema central en cualquier debate sobre los recursos hídricos y la seguridad alimentaria. En promedio, en la agricultura se ocupa el 70 % del agua que se extrae en el mundo. De esta manera, en la actividad agrícola es necesario la adopción de tecnologías emergentes para atender las necesidades alimentarias actuales haciendo un uso responsable de los recursos naturales, insumos y económicos. En este contexto, el Internet de las Cosas se está utilizando como tecnología emergente para el apoyo en la producción agrícola. Este trabajo de investigación propone el desarrollo de una arquitectura de Internet de las Cosas para recoger datos climáticos (humedad relativa, temperatura del aire, radiación solar y velocidad del viento) al interior de un invernadero a cada segundo. Los datos obtenidos serán procesados y transformados en un modelo de datos FIWARE para ser enviados al *Orien Context Broker* (OCB) a intervalos de un minuto usando tecnología GSM. Las mediciones servirán para estimar la evapotranspiración de referencia al interior de un invernadero.

Palabras claves: Evapotranspiración, Internet de las cosas, Invernadero

Figura A.5: Artículo publicado en la 9a Jornada de Ciencia y Tecnología Aplicada.

**GECCO 2024 Certificate
of Attendance and
Presentation**



Confirmation: 3FNWNL6D27K
Registration Type: ACM/SIGEVO Professional Member
Attendee: FRANCISCO JAVIER RUIZ ORTEGA
Company/Organization: TecNM
Country: Mexico
Address: C JAZMIN 320, COL LOS ARRAYANES GOMEZ PALACIO,
MX-DUR Mexico

To Whom It May Concern:

This certification is confirmation of FRANCISCO JAVIER RUIZ ORTEGA's attendance and participation in the GECCO 2024 Conference organized by the Association for Computing Machinery's (ACM) Special Interest Group on Genetic and Evolutionary Computation (SIGEVO) held in-person and virtually from Melbourne, Australia from July 14th through July 18th, 2024.

Sincerely,

A handwritten signature in black ink that reads "Stephanie Matal".

Stephanie Matal
GECCO 2024 Registration Manager
Executivevents c/o ACM
137 2nd Avenue
Niwot, CO 80544



ACM 1601 Broadway, 10th Floor New York, NY 10019-7434. Tel: 212-869-7440. Fax: 212-302-5826

Figura A.6: Certificado de Asistencia y Presentación en el *21st Annual Humies Awards – Melbourne, Australia*.



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Ensenada
Departamento de Sistemas y Computación

Ciudad de Ensenada, Baja California **15/agosto/2021**
Asunto: Estancia de investigación

FRANCISCO JAVIER RUIZ ORTEGA
ESTUDIANTE DE DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN
TECNM/CENIDET; CUERNAVACA, MORELOS
PRESENTE

Por medio de la presente tengo el agrado de confirmar su invitación para realizar una estancia de investigación en el TecNM/Ensenada. Las actividades a realizar durante la estancia se muestran en el siguiente cronograma.

No.	Actividad	Inicio	Fin	2021																		
				Sep	Oct	Nov	Dic															
1	Aprender las técnicas del cómputo evolutivo (Algoritmos Genéticos y Programación Genética) como herramientas de optimización	31/08/2021	17/09/2021	█	█																	
2	Aprender el uso de librerías en Matlab (GPLab) y en Python (DEAP)							█	█	█												
3	Analizar y adecuar las bases de datos para llevar un proceso de optimización utilizando cómputo evolutivo.																					
4	Plantear una representación adecuada para aplicar un proceso de optimización.																					
5	Implementar de forma preliminar un proceso de optimización y comparar resultados con los existentes.																					

La estancia de investigación será del 31 de agosto del 2021 al 19 de Dic del 2021. Y para que así conste donde proceda, firmo la presente en Ensenada, Baja California, México a 15 de agosto del 2021.

ATENTAMENTE
Excelencia en Educación Tecnológica
Por la Tecnología de Hoy y del Futuro

EDDIE HELBERT CLEMENTE TORRES
DOCENTE INVESTIGADOR DEL TECN-ENSENADA



Bld. Tecnológico #150, Ex Ejido Chapultepec, C.P. 22780,
Ensenada, Baja California, Tels. (646) 177 5680 y 177 5682
e-mail: eddiect@tecnm.mx
tecnm.mx/ensenada.tecnm.mx



Figura A.7: Cronograma de actividades de la estancia de investigación en el TecNM/I.T. Ensenada.

Apéndice B

Retribución Social



Figura B.1: Conferencia dictada en el TecNM/IT Cuautla.

Apéndice C

Otras actividades



Figura C.1: Curso «Introducción al análisis de datos meteorológicos con R».



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

**EL TECNOLÓGICO NACIONAL DE MÉXICO
A TRAVÉS DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO**

OTORGA LA PRESENTE

CONSTANCIA

A

FRANCISCO JAVIER RUIZ ORTEGA

TECNM/CENIDET

POR HABER ASISTIDO Y ACREDITADO EL CURSO: INTRODUCCIÓN AL ANÁLISIS DE DATOS
METEOROLÓGICOS CON R, DEL 17 AL 21 DE ENERO DE 2022, CON UNA DURACIÓN DE 30
HORAS.

CUERNAVACA, MORELOS., 14 DE FEBRERO DE 2022



DRA. YESICA IMELDA SAAVEDRA BENÍTEZ
**DIRECTORA DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO**

Sello Digital:

Amb7xrG8OPKNRg/DALVLOOMCxo1iUZxjHxDyB5S5qRndhNPb13Mw1Cy+G1sNF8VnMkTzE1s1MftwpEgkdMBp74
qSmeMEKq0BEWob5+T2Hx4V+sIPKI1F1nkHEgdwrWp8x8oUZc857PBi7b3chc1fdLQCOTdVhJxSLh616vmC/bFT
9sDbNALQdaGY+lwOcltE8dFGbb7DiY+KFR9J88/ktA9p1z1b1hvbowvY4JKBMLjFfkUc4uSGkgFku0UaT1Zy6y
m4TX004ua5TEDiL1+0Cay2I+vKnStE25agXvDq6YgVNwKqImn78RL3om6K1bKeuCB02ZdBDai aMXv3uUVbjA==

ROF009422

<http://constancias.cenidet.tecnm.mx>

cenidet[®]
Centro Nacional de Investigación
y Desarrollo Tecnológico



Figura C.2: Curso «Introducción al análisis de datos meteorológicos con R».



Figura C.3: Constancia de participación en el taller: Macroentrenamiento en Inteligencia Artificial.



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

EL TECNOLÓGICO NACIONAL DE MÉXICO

OTORGA EL PRESENTE

DIPLOMA

A

FRANCISCO JAVIER RUIZ ORTEGA

POR HABER ACREDITADO

EL DIPLOMADO EN CIENCIA DE DATOS, LLEVADO A CABO DEL 25 DE SEPTIEMBRE DEL 2023 AL 28 DE ENERO DEL 2024, CON UNA DURACIÓN DE 120 HORAS.

CIUDAD DE MÉXICO, A 08 DE FEBRERO DE 2024.

RAMÓN JIMÉNEZ LÓPEZ
DIRECTOR GENERAL



Folio: 3y0B2sMPq

Figura C.4: Diploma de acreditación del Diplomado en Ciencia de Datos.



**COLEGIO CIENTÍFICO DE DATOS
OTORGA LA PRESENTE**

CONSTANCIA

A: FRANCISCO JAVIER RUIZ ORTEGA

Por haber concluido satisfactoriamente el curso:
“Aprendizaje Profundo con Redes Neuronales”
Con una duración de 10 horas,
durante los días 16 y 23 de abril de 2023.

Cuatla Mor. 05 de mayo de 2023




MTR. TOMÁS ALBERTO SALMERÓN ENCISO
DIRECTOR DE PLANTEL COLEGIO CIENTÍFICO DE DATOS

Figura C.5: Constancia del Colegio Científico de Datos (COCID).



**La Sociedad Mexicana de Ciencias de la Computación
y la Universidad Autónoma de Coahuila a través del
Centro de Investigación en Matemáticas Aplicadas y la Facultad de Sistemas**

otorgan la presente

CONSTANCIA

a

Francisco Javier Ruíz Ortega

Por su destacada participación como ASISTENTE a los talleres y conferencias
en el Encuentro Nacional en Computación 2020
efectuado del 24 al 26 de agosto de 2020

Dra. María Lucía Barrón Estrada
Presidenta de la SMCC

Dra. Valeria Soto Mendoza
Organizador General Local



Valida el reconocimiento en nuestro sitio escaneando el QR.
O ingresa en <https://enc-2020.web.app/validacion/ENC-0241>

Figura C.6: Constancia de asistencia Encuentro Nacional de Computación (ENC).



El Instituto Nacional de Astrofísica, Óptica y Electrónica,
y
la Red Temática CONACyT en Inteligencia Computacional

Otorgan la presente constancia de asistencia a:

Francisco Javier Ruíz Ortega

Por haber asistido a las actividades

del X Seminario y Escuela Nacional de Aprendizaje e
Inteligencia Computacional, realizados en formato virtual del
26 al 30 de septiembre de 2022

Dr. Carlos A. Reyes García
Presidente Comité Nacional INAOE
Responsable RedICA

Dra. Alicia Morales Reyes
Presidente Comité Nacional INAOE



GOBIERNO DE
MÉXICO



CONACyT
Comisión Nacional de Investigación Científica y Tecnológica



inaoe.mx

Figura C.7: Constancia de asistencia X Seminario y Escuela Nacional de Aprendizaje e Inteligencia Computacional (XENAIC).