

**TECNOLÓGICO NACIONAL DE MEXICO**

**Instituto Tecnológico de Culiacán**



**TITULACIÓN INTEGRAL**

***TESIS PROFESIONAL***

**Análisis temporal de datos meteorológicos y su relación con el  
cambio climático y la sequía en la zona centro de Sinaloa**

**ALUMNO**

**Fernando Alonso Moreno Millán**

**CARRERA**

**Ingeniería de Sistemas Computacionales**

**Num. De Control.**

**20170759**

**Culiacán, Sinaloa (2025)**

# AUTORIZACIÓN DE IMPRESIÓN



**Educación**  
Secretaría de Educación Pública



TECNOLÓGICO  
NACIONAL DE MÉXICO



Instituto Tecnológico de Culiacán

DIVISION DE ESTUDIOS PROFESIONALES  
ASUNTO: **Autorización de Impresión**

Culiacán, Sinaloa, 23/09/2025 19:23

**C. FERNANDO ALONSO MORENO MILLÁN**  
**PASANTE(S) DE LA CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**  
**PRESENTE.**

De acuerdo con fallo emitido por la Comisión Revisora Integrada por:

Cédula Profesional

|             |                                                                                                               |          |
|-------------|---------------------------------------------------------------------------------------------------------------|----------|
| Presidente: | Maestría en Ciencias de la Computación<br><b>Zuriel Dathan Mora Felix</b>                                     | 7085761  |
| Secretario: | Doctor en Ciencia y Tecnología con Especialidad en Ingeniería Ambiental<br><b>Jesús Gabriel Rangel Peraza</b> | 8327423  |
| Vocal:      | Maestría en Ciencias en Ciencias Computacionales<br><b>Arcelia Judith Bustillos Martínez</b>                  | 13118410 |

Y considerando que reúne los requisitos establecidos en la Opción: **TITULACIÓN INTEGRAL: TESIS PROFESIONAL "ANÁLISIS TEMPORAL DE DATOS METEOROLÓGICOS Y SU RELACIÓN CON EL CAMBIO CLIMÁTICO Y LA SEQUÍA EN LA ZONA CENTRO DE SINALOA"**, damos a usted nuestra autorización para que proceda a imprimir su trabajo profesional.

Sin otro particular, quedo de usted.

**ATENTAMENTE**

*Excelencia en Educación Tecnológica®*

*CONCEPCION MENDOZA ROSALES*

**LIC. CONCEPCION MENDOZA ROSALES**

**JEFA DE LA DIVISION DE ESTUDIOS PROFESIONALES**



**Educación**  
Secretaría de Educación Pública



TECNOLÓGICO  
NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE CULIACÁN

DEPARTAMENTO DE  
DIVISIÓN DE ESTUDIOS PROFESIONALES



**2025**  
Año de  
**La Mujer**  
**Indígena**

Juan de Dios Bátiz 310 Pte. Col. Guadalupe C.P. 80220  
Culiacán, Sinaloa, Tel. 667-454-0100  
tecnm.mx www.culiacan.tecnm.mx



# Dedicatoria

Este trabajo de investigación está especialmente dedicado a mis abuelos Ramón y Norma, a mi madre María Luisa y mi hermana mayor Luisa Marié, quienes han sacrificado parte de su vida para apoyarme en todos los aspectos, durante más de 23 años. Espero que este sea el primero de los frutos de todos sus esfuerzos.

Y a Romi, mi pareja actual, quien ha soportado las explicaciones innecesarias y lluvias de ideas para esta tesis y mis traspasadas constantes durante el tiempo de desarrollo.

No soy un hombre que exprese demasiado sus emociones actualmente, pero quiero decirles que los amo con todo mi ser.

# Agradecimientos

Primeramente, quisiera agradecer al Instituto Tecnológico de Culiacán, que se encargó de capacitarme por casi 5 años como ingeniero y futuro profesionista.

También quiero agradecer a mis asesores, el Doctor Zuriel Dathan Mora Félix y Doctor Jesús Gabriel Rangel Peraza, el, quienes me apoyaron y no me soltaron por más que pareciera que esta tesis no tendría fin, de igual forma agradezco al Doctor Antonio Jesús Sanhouse García, quien también estuvo apoyándome a corroborar datos a lo largo de esta investigación.

Veo necesario agradecer de igual forma al Servicio Meteorológico Nacional y a la Comisión Nacional del Agua, por proporcionar el acceso público a los datos con los que fue posible llevar a cabo este trabajo.

Finalmente me gustaría agradecer también a mis compañeros de carrera y de la universidad, quienes egresaron antes que yo, así como a quienes están a punto de egresar, en especial a Sebastián, Bryan, Maximiliano, Luis Eduardo, Ryan y Ricardo, con quienes he tenido el gusto de convivir desde bachillerato.

# **Declaración de autenticidad**

Por la presente hago de conocimiento al lector que, a menos que se haga referencia o mención de un tercero, toda la información y datos expuestos dentro de este documento de investigación son de mi autoría intelectual y de naturaleza inédita para el Tecnológico Nacional de México y cualquier otra Universidad.

Fernando Alonso Moreno Millán

Culiacán, Sinaloa, México, 2025

# Resumen

El presente trabajo habla de la importancia de los fenómenos meteorológicos para la nación mexicana en un aspecto cultural, social, legislativo y económico, se proporcionan diferentes conceptos y definiciones que ayudan a medir cuantitativa y cualitativamente estos fenómenos y se trabaja en la creación de una metodología que, mediante la extracción, transformación y cargar de datos meteorológicos clave extraídos de registros de sensores hidrometeorológicos de la Comisión Nacional del Agua (CONAGUA) en conjunto con el Servicio Meteorológico Nacional (SMN). Esto con el fin de llevar un análisis de los datos enfocado a la región comprendida por el estado de Sinaloa.

# Palabras clave

|                 |                                                                                               |
|-----------------|-----------------------------------------------------------------------------------------------|
| <b>SMN</b>      | Servicio Meteorológico Nacional                                                               |
| <b>CONAGUA</b>  | Comisión Nacional del Agua                                                                    |
| <b>ETL</b>      | Extraer, Transformar, Cargar                                                                  |
| <b>WWF</b>      | World Wildlife Fund                                                                           |
| <b>WMO</b>      | World Meteorological Organization                                                             |
| <b>RAE</b>      | Real Academia Española                                                                        |
| <b>UNC</b>      | Universidad Nacional de Córdoba                                                               |
| <b>EMA</b>      | Estación Meteorológica Automática                                                             |
| <b>DW</b>       | Almacén de datos                                                                              |
| <b>RNA</b>      | Redes Neuronales Artificiales                                                                 |
| <b>GIS</b>      | Geographical Information Systems                                                              |
| <b>QGIS</b>     | Quantum Geographical Information Systems                                                      |
| <b>MDP</b>      | Millones de Pesos                                                                             |
| <b>ML</b>       | Machine Learning                                                                              |
| <b>PNUD</b>     | Programa de las Naciones Unidas para el Desarrollo                                            |
| <b>PDSI</b>     | Palmer's Drought Severity Index                                                               |
| <b>SPI</b>      | Standard Precipitation Index                                                                  |
| <b>BI</b>       | Business Intelligence                                                                         |
| <b>ESRI</b>     | Environmental Systems Research Institute                                                      |
| <b>PERSIANN</b> | Precipitation Estimation From Remotely Sensed Information Using<br>Artificial Neural Networks |
| <b>CMORPH</b>   | Climate Prediction Center MORPHing technique                                                  |
| <b>CEQUEAU</b>  | CEntre QUEbécois de l'EAU                                                                     |
| <b>GNA</b>      | Generative Adversarial Networks                                                               |

# Índice general

|       |                                                                       |    |
|-------|-----------------------------------------------------------------------|----|
| 1     | Introducción .....                                                    | 1  |
| 1.1   | Planteamiento del problema .....                                      | 3  |
| 1.2   | Objetivos.....                                                        | 5  |
| 1.2.1 | Objetivo General.....                                                 | 5  |
| 1.2.2 | Objetivos específicos.....                                            | 6  |
| 1.3   | Justificación.....                                                    | 7  |
| 2     | Marco teórico .....                                                   | 9  |
| 2.1   | Análisis de series de tiempo .....                                    | 9  |
| 2.1.1 | Tendencia en el contexto de series de tiempo .....                    | 10 |
| 2.1.2 | Estacionalidad en el análisis de variables hidrometeorológicas .....  | 11 |
| 2.1.3 | Tipos de validaciones .....                                           | 13 |
| 2.2   | Conceptos hidrometeorológicos .....                                   | 14 |
| 2.3   | Variables bioclimáticas y medio ambiente.....                         | 15 |
| 2.3.1 | Causas y consecuencias de la contaminación ambiental .....            | 19 |
| 2.3.2 | Legislaciones en materia ambiental.....                               | 25 |
| 2.4   | Procesamiento y análisis de datos .....                               | 26 |
| 2.4.1 | Extracción, Transformación y Carga.....                               | 27 |
| 2.4.2 | Técnica de raspado de datos (data scraping) .....                     | 27 |
| 2.4.3 | Técnica de web scraping.....                                          | 27 |
| 2.5   | Elementos tecnológicos relevantes para el procesamiento de datos..... | 28 |
| 2.5.1 | Lenguaje de programación Python.....                                  | 28 |
| 2.5.2 | Formato de archivo Valores Separados por Coma (CSV).....              | 28 |
| 2.5.3 | Herramienta Pandas para Python.....                                   | 28 |
| 2.6   | Herramientas disponibles para análisis de datos .....                 | 29 |
| 2.6.1 | Software propietario para análisis de datos Power BI .....            | 29 |
| 2.6.2 | Software de Sistemas de Información Geográfica (GIS) .....            | 29 |
| 2.7   | Generación de datos sintéticos .....                                  | 30 |
| 2.7.1 | Deep Learning para la generación de datos sintéticos.....             | 30 |
| 2.7.2 | Mixture Density Network (MDN).....                                    | 31 |
| 2.7.3 | Generación de datos sintéticos mediante modelos de Deep Learning..... | 31 |
| 3     | Estado del arte .....                                                 | 32 |

|       |                                                                                 |    |
|-------|---------------------------------------------------------------------------------|----|
| 3.1   | Fuentes de datos disponibles .....                                              | 33 |
| 3.1.1 | Sensores hidrometeorológicos.....                                               | 33 |
| 3.1.2 | Variables de análisis meteorológicos .....                                      | 34 |
| 3.2   | Variables Bioclimáticas.....                                                    | 35 |
| 3.2.1 | Generación de datos sintéticos para sensores hidrometeorológicos.....           | 36 |
| 4     | Desarrollo del proyecto .....                                                   | 37 |
| 4.1   | Área de Estudio .....                                                           | 37 |
| 4.2   | Metodología de desarrollo.....                                                  | 39 |
| 4.2.1 | Etapa 1: Extracción (datascraper.py).....                                       | 41 |
| 4.2.2 | Etapa 2: Transformación .....                                                   | 44 |
| 4.2.3 | Etapa 3 Transferencia de archivos.....                                          | 53 |
| 4.2.4 | Etapa 4: Análisis de datos.....                                                 | 61 |
| 5     | Pruebas y Resultados.....                                                       | 62 |
| 5.1   | Etapa de extracción de los archivos de datos de CONAGUA.....                    | 62 |
| 5.1.1 | Configurar porcentaje de datos faltantes aceptable .....                        | 63 |
| 5.1.2 | Extraer datos .....                                                             | 63 |
| 5.2   | Transformación de archivos .....                                                | 66 |
| 5.2.1 | Limpiar archivos.....                                                           | 66 |
| 5.2.2 | Generar registros de datos mensuales.....                                       | 68 |
| 5.2.3 | Generar concentrado de registros diarios .....                                  | 69 |
| 5.2.4 | Generar concentrado de registros mensuales.....                                 | 69 |
| 5.3   | Informes gráficos de información hidrometeorológica .....                       | 70 |
| 5.4   | Análisis de los datos obtenidos.....                                            | 71 |
| 5.4.1 | Caso de estudio general: comparativa entre periodos 1974-1999 y 2000-2025<br>72 |    |
| 5.4.2 | Análisis hidrometereológico específico I (Caso de estudio: Huracán Manuel)      | 78 |
| 5.4.3 | Análisis hidrometeorológico II (Caso de estudio: Sequia extrema 2024) .....     | 81 |
| 6     | Conclusiones y trabajo futuro .....                                             | 84 |
| 6.1   | Más trabajo a futuro.....                                                       | 85 |
| 6.1.1 | Ampliación del alcance del proyecto.....                                        | 85 |
| 6.1.2 | Aplicaciones estadísticas .....                                                 | 85 |
| 6.1.3 | Aplicaciones en la IA .....                                                     | 85 |
| 7     | Bibliografía.....                                                               | 86 |

|                                                  |     |
|--------------------------------------------------|-----|
| A. Código fuente.....                            | 93  |
| A.1 datascraper.py.....                          | 93  |
| A.2 limpiar_datos.py .....                       | 95  |
| A.3 GUI.py .....                                 | 100 |
| B. Manual de Usuario para informes .....         | 102 |
| B.1 Tablero inicial.....                         | 102 |
| B.2 Filtro superior para datos diarios .....     | 103 |
| B.3 Tablero de datos diarios.....                | 104 |
| B.4 Cuerpo del informe de datos diarios.....     | 105 |
| B.5 Informe de temperaturas diarias .....        | 105 |
| B.6 Informe de actividad hidrológica diaria..... | 107 |
| B.7 Informes de datos mensuales .....            | 108 |
| B.8 Filtro superior para datos mensuales .....   | 109 |
| C. Lista de estaciones meteorológicas .....      | 110 |
| D. Carta de no inconveniente de titulación.....  | 113 |

# Índice de figuras

|                                                                                                  |    |
|--------------------------------------------------------------------------------------------------|----|
| Figura 2.1. Precipitación acumulada mensual de 2000 a 2001 (Fuente: Elaboración propia)          | 12 |
| Figura 3.1. Mapa de estaciones meteorológicas operativas de Sinaloa (Fuente: Elaboración propia) | 35 |
| Figura 4.1. Mapa de estaciones meteorológicas operativas de Sinaloa (Fuente: Elaboración propia) | 37 |
| Figura 4.2. Diagrama de metodología ETL-A (Fuente: Elaboración propia)                           | 40 |
| Figura 4.3. Estructura de página de CONAGUA (Fuente: Elaboración propia)                         | 41 |
| Figura 4.4. Código de búsqueda por fuerza bruta. (Fuente: Elaboración propia)                    | 42 |
| Figura 4.5. Descarga por búsqueda en lote. (Fuente: Elaboración propia)                          | 42 |
| Figura 4.6. Función que se referencia en ambos métodos de descarga (Fuente: Elaboración propia)  | 43 |
| Figura 4.7. Inicio de paso 1 dentro del código de Python (Fuente: Elaboración propia)            | 45 |
| Figura 4.8. Código para primeras 4 instrucciones de paso 2 (Fuente: elaboración propia)          | 47 |
| Figura 4.9. Código para últimos 4 pasos del paso 2 (Fuente: elaboración propia)                  | 48 |
| Figura 4.10. Código para paso 3 (Fuente: elaboración propia)                                     | 49 |
| Figura 4.11. Código para creación de datos mensuales (Fuente: elaboración propia)                | 50 |
| Figura 4.12. Resto del código para la crear archivo mensual (Fuente: elaboración propia)         | 51 |
| Figura 4.13. Código para concatenar concentrados mensuales (Fuente: Elaboración propia)          | 52 |
| Figura 4.14. Pantalla de inicio PowerBI (Fuente: Elaboración propia)                             | 53 |
| Figura 4.15. Pantalla de selección de fuentes de datos PowerBI (Fuente: Elaboración propia)      | 54 |
| Figura 4.16. Seleccionando archivo a cargar (Fuente: elaboración propia)                         | 55 |
| Figura 4.17. Seleccionando archivo a cargar (Fuente: Elaboración propia)                         | 56 |
| Figura 4.18. Vista de informe en blanco (Fuente: Elaboración propia)                             | 57 |
| Figura 4.19. Cargar archivo de datos históricos mensuales (Fuente: elaboración propia)           | 58 |
| Figura 4.20. Vista de modelo PowerBI (Fuente: elaboración propia)                                | 58 |

|                                                                                                          |     |
|----------------------------------------------------------------------------------------------------------|-----|
| Figura 4.21. Ventana emergente administrar relaciones (Fuente: elaboración propia) .....                 | 59  |
| Figura 4.22. Creando relación (Fuente: elaboración propia) .....                                         | 60  |
| Figura 5.1. Interfaz de Usuario ETL. (Fuente: elaboración propia) .....                                  | 62  |
| Figura 5.2. Ventana emergente: Seleccionar carpeta (Fuente: elaboración propia) .....                    | 63  |
| Figura 5.3. Log de extracción de datos (Fuente: Elaboración propia) .....                                | 64  |
| Figura 5.4. Archivos extraídos (Fuente: Elaboración propia) .....                                        | 64  |
| Figura 5.5. Ejemplo de la cabecera de un archivo de texto (Fuente: Elaboración propia) ...               | 65  |
| Figura 5.6. Registro de consola mostrando archivo siendo procesado (Fuente: Elaboración propia).....     | 66  |
| Figura 5.7. Archivos limpios en el directorio de salida (Fuente: Elaboración propia).....                | 66  |
| Figura 5.8. Primeros 20 registros de NR_25001.csv (Fuente: Elaboración propia).....                      | 67  |
| Figura 5.9. Carpeta que contiene los registros mensuales (Fuente: Elaboración propia).....               | 68  |
| Figura 5.10. Primeros 20 registros de CM_25001.csv (Fuente: Elaboración propia) .....                    | 68  |
| Figura 5.11. Archivo de Concentrados Diarios (CD) (Fuente: Elaboración propia).....                      | 69  |
| Figura 5.12. Archivo de concentrado mensual (FINAL.csv) (Fuente: Elaboración propia) 70                  |     |
| Figura 8.1. Informe de tablero inicial en PowerBI (Fuente: Elaboración propia) .....                     | 102 |
| Figura 8.2. Filtro superior para informes diarios (Fuente: Elaboración propia).....                      | 103 |
| Figura 8.3. Tablero central para informes hidrometeorológicos diarios (Fuente: Elaboración propia).....  | 104 |
| Figura 8.4. Cuerpo del informe del informe general para datos diarios. (Fuente: Elaboración propia)..... | 105 |
| Figura 8.5 Cuerpo del informe de temperaturas diarias (Fuente: Elaboración propia) .....                 | 106 |
| Figura 8.6. Cuerpo del informe de actividad hidrológica diaria (Fuente: Elaboración propia) .....        | 107 |
| Figura 8.7. Informe concentrado mensual (Fuente: Elaboración propia) .....                               | 108 |
| Figura 8.8. Filtro superior de datos mensuales .....                                                     | 109 |
| Figura 8.9. Carta de no inconveniente de titulación .....                                                | 113 |

# Índice de tablas

|                                                    |    |
|----------------------------------------------------|----|
| Tabla 1. Análisis hidrológico.....                 | 76 |
| Tabla 2. Análisis térmico.....                     | 77 |
| Tabla 3. Datos hidrometeorológicos 2013, 2020..... | 79 |
| Tabla 4. Datos hidrometeorológicos 2020-2024 ..... | 82 |

# Capítulo 1

## 1 Introducción

El estudio de la naturaleza y los fenómenos naturales, han sido un punto de interés clave para México desde sus orígenes; partiendo de la existencia de observatorios en zonas designadas tales como templos, plataformas, pirámides, maquetas y pocitas en peñas (Moreno, 2017) que datan de la era precolonial, estos son vestigios que nos permiten observar claramente la importancia que tenían este tipo de eventos a un nivel cosmológico.

En tiempos más actuales podemos apreciar el proyecto del Observatorio Meteorológico y Astronómico de México el 6 de marzo de 1887, la posterior creación del Servicio Meteorológico Nacional (SMN) y la adhesión de esta a la Organización Meteorológica Mundial (WMO) en 1947.

Desde el año 2002, México, Estados Unidos y Canadá implementaron el Monitor de Sequías de América del Norte (NADM, por sus siglas en inglés), un monitor de sequías clasificable por sus diversos índices o indicadores de sequía (precipitación, evaporación, temperaturas, humedad, entre otros) basados en el Monitor de Sequía de los Estados Unidos (USDM) del año 1999. Utiliza siete índices diferentes, así como análisis de diferentes expertos para validar la información. (Lobato-Sánchez, 2016). Actualmente se ha logrado la implementación de estaciones meteorológicas, radares y sistemas satelitales que proporcionan una cobertura nacional completa (Servicio Meteorológico Nacional, 2021).

Aunado a los esfuerzos internacionales para la monitorización de fenómenos climatológicos y los siniestros que éstos puedan ocasionar en México, el SMN cuenta con más de 200 EMA, llegando a contar actualmente más de 5,400 estaciones climatológicas y aproximadamente 2,800 estaciones funcionales que tienen la labor de dar informes diarios acerca de factores como la temperatura máxima, temperatura mínima, precipitación y evaporación de la zona en la que están ubicadas (CONAGUA, 2025).

En el año 2021, Sinaloa figuró en el quinto lugar de estados con ‘cinco estrellas’ en producción agrícola siendo uno de los mayores productores a nivel nacional, en donde sus dos exportaciones principales fueron el maíz grano y el jitomate, avaluándose de acuerdo en

756,667 Millones de pesos (MDP) a nivel nacional, lo cual significaría un 9.5% del Producto Interno Bruto (PIB) del país (Secretaría de Agricultura y Desarrollo Rural, 2022). Otras de las actividades más redituables de Sinaloa, son la ganadería, acuicultura e incluso minería, las cuales son actividades que, de no ser apropiadamente realizadas, provocan daños ambientales que resultarían extremadamente difíciles de revertir.

Todo lo mencionado anteriormente son algunos de los precedentes modernos de la intención de llevar a cabo una recolección de datos meteorológicos precisa, esto con el fin de llegar a desarrollar modelos predictivos más eficientes, que nos permitan predecir eventos climatológicos y evitar o minimizar los rezagos que estos traigan consigo.

Para poder medir correctamente los parámetros medioambientales de Sinaloa, es necesario llevar a cabo un proceso de análisis de la información captada por los diversos sensores hidrometeorológicos localizados en puntos estratégicos dentro del estado de Sinaloa.

## **1.1 Planteamiento del problema**

A pesar de las tecnologías e infraestructura con la que cuenta el estado, de acuerdo con la información proporcionada anteriormente, un reporte actual (Reuters, 2024) indica que Sinaloa observaría una caída del 31% de su producción de maíz para 2024 dentro del año agrícola 2023 – 2025, sumando pérdidas del 10% de la producción con respecto al año agrícola anterior.

Las afectaciones a la producción agrícola pueden tener diversos orígenes y factores, ya sean directamente provocados por la actividad humana y la transformación del suelo, cambios no previstos de las zonas geográficas, fallos logísticos como problemas en los sistemas de distribución de recursos, cambio climático, hasta problemas en la captación y análisis de la información.

Con respecto a los diferentes problemas derivados de la obtención y análisis de la información, es posible visualizar la información captada por los diferentes sensores climatológicos dentro del territorio mexicano, sin embargo, dentro de los mismos registros se presentan irregularidades en cuanto a los formatos: no son visualizables más que como formato de texto plano, es imposible llevar a cabo una descarga de datos en lote, en los que se añaden cabeceras que los hacen incompatibles con los formatos aceptados por herramientas de análisis de datos. Otro problema es que la información misma contiene lagunas temporales en los registros (que en algunos casos dichos datos faltantes corresponden a periodos anuales completos) y también hay casos particulares relacionados con la falta de mantenimiento, clausura o mantenimiento programado de la estación, los sensores no llegan a captar datos de algunos o incluso de todos los campos.

Otra problemática presentada dentro de los sensores climatológicos son la precisión de estos, como es posible observar dentro del estudio sobre la calidad de los datos de las EMA hecho por Cervantes-Osornio (2020), en el que se indica que existen fallas de medición de hasta un máximo detectado de 23.25% en algunos parámetros.

Las irregularidades mencionadas anteriormente hacen del trabajo de procesamiento y análisis una labor extremadamente compleja, ya que, comenzando de la obtención, se tienen que descargar manualmente los datos de las estaciones meteorológicas (tan solo en Sinaloa son

aproximadamente 50 estaciones funcionales actualmente). Tras descargar los datos y con el fin de realizar un análisis por series de tiempo, es necesario depurar texto de encabezado que contienen los archivos para poder cambiar su formato de archivo de texto a CSV. Una vez transformado, se tienen que rellenar manualmente todas las fechas faltantes en los registros de todas las estaciones. Una vez se regenera la serie de tiempo, y a causa de los datos faltantes por alguna falla en los sensores, es posible llevar a cabo procedimientos de simulación de los datos utilizando estrategias estadísticas, otra vez, de forma manual, o semiautomática utilizando herramientas como Microsoft Excel. Para finalmente poder realizar un análisis estadístico y tener la posibilidad de llevar a cabo análisis de tendencias.

## **1.2 Objetivos**

### **1.2.1 Objetivo General**

Realizar un análisis de tendencia y estacionalidad mediante series de tiempo obtenidos de las estaciones hidrometeorológicas establecidas por el Sistema Meteorológico Nacional (SMN) y de la CONAGUA en el estado de Sinaloa. Para la posterior construcción de modelos meteorológicos proyectivo-predictivos más precisos.

### **1.2.2 Objetivos específicos**

1. Realizar el proceso de limpieza y transformación de datos mediante algoritmos desarrollados en el lenguaje de programación Python.
2. Generar archivos de datos en formato de valores separados por comas (CSV) con una resolución anual, mensual y diaria.
3. Llevar a cabo procesos de generación de datos sintéticos utilizando estrategias estadísticas por medio de algoritmos desarrollados en Python.
4. Estimar la precisión y la calidad de los archivos de datos generados.
5. Generar un análisis en series de tiempo y evaluaciones temporales y espaciales mediante la generación de pronóstico de datos.

### **1.3 Justificación**

A causa de la problemática mencionadas anteriormente, que se puede resumir como la baja calidad de los datos recolectados por los sensores del organismo gubernamental de CONAGUA, las estaciones hidrometereológicas generan archivos de datos con faltantes en periodos extensos, además, el formato en el que se generan los archivos de datos hidrometeorológicos son archivos de texto plano, en este sentido para el estudio de las consecuencias climatológicas que trae consigo el cambio de uso de suelo se requieren datos de gran calidad y precisión para el desarrollo del análisis de las variables ambientales, en este sentido se vuelve viable este proyecto, debido a que el tiempo que consume realizar todo el procesamiento de forma manual actualmente, es decir, descargar los archivos, limpiarlos, transformarlos a un formato válido para herramientas de descarga y analizarlos, es una tarea compleja, cuya dificultad aumenta de forma exponencial conforme más estaciones se quieran considerar.

Es importante mencionar que cada uno de los pasos mencionados anteriormente son propensos a error humano, y, que, dependiendo del alcance del análisis, el tiempo que toma realizar este proceso se aumenta de forma compleja mediante diversos factores, como la antigüedad de las estaciones y el número de estas en la región de interés.

De acuerdo con diferentes investigaciones realizadas sobre el estado de Sinaloa en diversas áreas, como repercusiones del cambio climático en la investigación de Flores, et al. (2012), o en los estudios sobre la preservación del agua y los índices de calidad de esta, de Amabilis, et al. (2023), es posible observar que, aunque existen sistemas de recolección de datos, estos no son realmente utilizados para la toma de decisiones.

Por todo lo mencionado anteriormente, y con el fin de acelerar los tiempos de procesamiento, y aumentar el interés sobre la realización de análisis climáticos realizados sobre el estado de Sinaloa (y México en general), es necesario automatizar los procesos de extracción, limpieza y transformación de los datos, que facilite la conversión de los archivos extraídos desde el sitio de la CONAGUA y facilite su posterior análisis mediante herramientas de explotación de datos.

La realización de este proyecto permitirá la realización de nuevos estudios utilizando los datos proporcionados por éste organismo público, como la elaboración de nuevos modelos predictivos basados en datos más completos que terminarían beneficiando a la comunidad científica en áreas como meteorología, obteniendo predicciones climáticas más precisas, prediciendo eventos climatológicos tales como los huracanes, las sequías, el inicio y término de las temporadas de lluvia de manera más específica, lo cual evitaría pérdidas a los diferentes sectores de la industria tales como: agricultura y ganadería, mediante la oportuna gestión integral y la distribución del agua de forma más precisa.

# Capítulo 2

## 2 Marco teórico

### 2.1 Análisis de series de tiempo

Dentro de la investigación de Antúnez (2011), una serie de tiempo se define como “aquel conjunto de observaciones sobre una variable, que generalmente es espaciada en el tiempo”.

Es decir que, un análisis por series de tiempo es un método de análisis estadístico que permite obtener datos importantes sobre el comportamiento de un sistema de una, -o más- variables a través del tiempo, todo esto por medio del comportamiento dicho conjunto de parámetros y las variaciones que tenga a lo largo dado por el periodo de tiempo analizado.

Los análisis de series de tiempo tienen amplio uso comprobado y validado en diversas áreas de aplicación como la economía, finanzas, climatología, ingeniería de control, astronomía, entre otras, como lo es demostrado en los análisis de Botero & Cano (2008), las aplicaciones de Vallvé & Vide (1994) o en el libro de Ogata (2003).

Es posible llevar a cabo la descomposición de una serie de tiempo utilizando los siguientes tres elementos: Estacionalidad, Tendencia y Ciclo, considerando Tendencia y Ciclo como un solo elemento Tendencia-Ciclo.

### 2.1.1 Tendencia en el contexto de series de tiempo

Dentro del contexto de las series de tiempo, la tendencia refiere al comportamiento que tienen la o las variables consideradas para el modelo dentro de un rango considerado.

Mientras más datos contenga la serie de tiempo, habrá mayor facilidad de identificar una tendencia, pero a su vez, mientras más variables considere, podría ser más difícil identificarla de forma precisa.

La tendencia es uno de los elementos que conforma una serie de tiempo, y refleja el comportamiento a medio y largo plazo de la o las variables de la serie (Antúnez, 2011). A su vez, la tendencia se puede clasificar dentro de una de dos categorías:

- **Tendencia estocástica/aleatoria:** Se define de esta forma cuando la pendiente de la curva varía
- **Tendencia determinística:** Es cuando la pendiente de una serie de tiempo no varía a lo largo de esta.

De acuerdo con Hyndman & Athanasopoulos (2021), la tendencia también puede referir al comportamiento de la dirección de la gráfica, es decir, que la gráfica vaya en tendencia creciente o decreciente a lo largo del tiempo.

### **2.1.2 Estacionalidad en el análisis de variables hidrometereológicas**

La estacionalidad refiere a patrones en los datos que están relacionados con algún espacio de tiempo específico, por ejemplo, las temporadas de lluvia dentro de una región son ejemplo de la estacionalidad, o el aumento de ventas en temporadas como navidad, día del niño, día de las madres, entre otras fechas importantes. Dentro del ámbito climatológico, según Burbano (2023), la estacionalidad es fundamental por la influencia e importancia que tiene sobre la tierra, ya que influye en la persistencia de todos los organismos vivos, y cambios en sus variables, como la temperatura o la precipitación afectan a diferentes componentes de su sistema, llevando a cambios en vegetación y en los requisitos ecológicos de las diversas plantas y animales, que a su vez influyen en la disponibilidad, tipo y cantidad de alimento disponible.

La estacionalidad es importante debido a que ofrece un punto de referencia más acerca del comportamiento de una serie de tiempo, además, relaciona un cambio en los datos con un periodo de tiempo que se repite a lo largo de todas las entradas de la serie en un punto o en un rango específico.

La estacionalidad, al tratarse de un comportamiento repetitivo, se puede identificar fácilmente de forma visual, algunos de los gráficos que se pueden utilizar serían los gráficos de barras o de puntos.

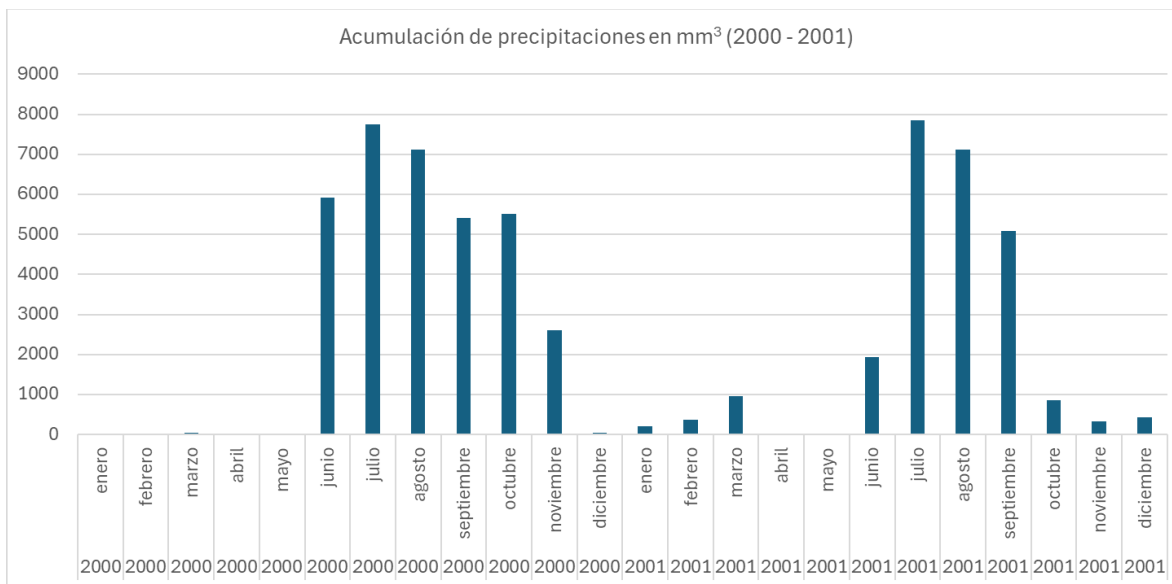


Figura 2.1. Precipitación acumulada mensual de 2000 a 2001 (Fuente: Elaboración propia)

En la figura 2.1 es posible detectar un patrón de estacionalidad: se comprende de lluvias recurrentes a mediados del año (entre junio y septiembre) y una clara falta de estas en los meses de enero a mayo, con presencia de precipitaciones atípicas a finales de año.

Algunas de las técnicas que se utilizan para validar la estacionalidad se mencionan a continuación.

### 2.1.3 Tipos de validaciones

El informe de resultados presentado por Ramírez (2024) hace referencia a los siguientes criterios de validación de series de tiempo:

- **Calidad de los datos**
  - **Compleitud:** No hay valores faltantes.
  - **Consistencia:** Coherencia en los registros.
  - **Precisión:** Exactitud de los datos registrados.
- **Estabilidad y Estacionariedad**
  - **Estacionariedad:** Las propiedades estadísticas (media, varianza, desviación) son constantes a lo largo del tiempo.
  - **Estabilidad:** La serie no presenta cambios estructurales significativos.
- **Componentes de la serie**
  - **Tendencia:** El comportamiento general va en aumento o disminución a lo largo del tiempo.
  - **Estacionalidad:** Hay patrones repetitivos en intervalos regulares
  - **Ciclos:** Hay fluctuaciones irregulares a largo plazo.
- **Ruido y anomalías**
  - **Ruido blanco:** La componente de ruido no muestra autocorrelación.
  - **Anomalías:** Existen valores atípicos.
- **Modelado y pronóstico**
  - **Adecuación del modelo:** El modelo elegido se ajusta a los datos históricos.
  - **Capacidad predictiva:** El modelo puede predecir valores futuros de forma precisa.
- **Ajuste a la realidad operacional:**
  - **Relevancia:** La serie de tiempo refleja correctamente las condiciones reales.
  - **Acción:** Los resultados del análisis permiten la toma de decisiones prácticas.

## **2.2 Conceptos hidrometeorológicos**

La hidrometeorología se define como una disciplina que proporciona una visión completa del estudio del agua dentro de la naturaleza a partir de un enfoque meteorológico, convergiendo la investigación sobre fenómenos atmosféricos con el análisis de los eventos hidrológicos que ocurren en el suelo y el subsuelo (Universidad Nacional de Córdoba, 2023). El objetivo principal de la hidrometeorología es proporcionar información sobre las condiciones del tiempo, clima y el comportamiento de los ríos por medio de la captación, integración y difusión de información meteorológica, climatológica e hidrométrica (Servicio Meteorológico Nacional, 2025).

### 2.3 Variables bioclimáticas y medio ambiente

Las variables bioclimáticas nos permiten llevar a cabo diversos análisis dentro de una región geográfica, tales como la relación entre estos diversos parámetros como la precipitación, evaporación, humedad, temperatura, y la fauna de una zona (Ruiz-Jiménez, C., et al, 2012). Así como la existencia y adaptabilidad de fauna específica en regiones determinadas según sus condiciones bioclimáticas (Hernández, 2023)

Es posible definir al medio ambiente como el conjunto de regiones afectadas por variables tales como temperatura, precipitación, evaporación, humedad, composición química del aire, altitud, ubicación geográfica, entre otros, que alteran (de forma positiva o negativa) al desarrollo de seres vivos en regiones con diferentes combinaciones en las entradas mencionadas anteriormente.

Sin embargo, una definición más precisa dice que, el medio ambiente es conformado por el entorno natural, el cual incluye aire, agua, tierra, flora, fauna, además de los recursos no renovables, tales como los combustibles fósiles y minerales (Cuesta, F., & Cabezas, A. M., 2002).

El agua es un elemento fundamental para todos los seres vivos dentro de la tierra, constituye la mayor parte del planeta y tanto flora como fauna en general dependen de este elemento para poder subsistir y proliferar dentro del entorno. De igual forma, su disponibilidad refleja el estado general del comportamiento hidrológico en relación con las actividades humanas, proporcionando métricas para calificar el estado promedio de esta relación (Sánchez-Ortiz, 2013).

La CONAGUA (2025) menciona que mediante radares meteorológicos de banda C es posible detectar las siguientes variables o indicadores:

- **Indicador de Posición en Planta (PPI):** Generalmente se usa para vigilancia, configurándolo por medio del momento de reflectividad una elevación y rango definidos, y se puede utilizar en diferentes momentos de operación del radar meteorológico. Usa una escala de reflectividad. Por medio de la intensidad de la reflectividad y diversos algoritmos matemáticos es factible llevara a cabo la relación entre intensidad reflectiva y la precipitación consecuente.

- **Indicador de Posición en Planta a Altitud Constante (CAPPI):** Es un producto generado al realizar un corte horizontal en la superficie de la tierra, utilizando el volumen de datos obtenidos a una altitud predefinida. Este indicador se usa en alturas someras para la obtención de productos de acumulación de precipitación y en alturas más elevadas para realizar predicciones. También se usa para observar el desarrollo de una tormenta. De la misma forma que el indicador anterior, es posible utilizarlo con diferentes momentos del radar.
- **VEL – Velocidad radial:** Es un producto que se genera por medio de los momentos de velocidad, y representa la velocidad media a la que el viento es detectado por el radar y su dirección. Para plasmar de forma gráfica los datos obtenidos por este indicador, se utilizan colores cálidos para velocidades positivas (es decir, que se alejan del radio del radar) y colores fríos para velocidades negativas (que se acercan al cono del radar).
- **CMAX / MAX – Columna máxima:** Los píxeles individuales representan el máximo de reflectividad en los datos capturados. Facilita la identificación de tormentas en desarrollo. Usando los datos de reflectividad se genera una imagen 2D por medio de los datos tridimensionales. Es factible para identificar granizo junto a la Reflectividad Diferencial (ZDR) de doble polaridad debido a que las zonas que cuentan con una mayor intensidad de CMAX tienen que coincidir con los sectores sin datos del producto ZDR.
- **HMAX – Altura de la intensidad máxima:** Proporciona un estimado de la altitud en donde fue detectada la intensidad máxima de cada cuadrante del área de detección en un momento dado. Se utiliza el producto de este indicador para llevar a cabo la evaluación de desarrollos de tormentas.
- **EBASE/BASE – Base de los ecos:** Proporciona el nivel base de los ecos de señal detectados, usualmente por medio del reflejo de la parte inferior del área de precipitaciones o de la base de la nube observada.
- **ETOPS/TOPS – Tope de los ecos:** Conocido también como tope de altura de los ecos, permite identificar el límite superior de determinada zona de precipitación o el rango de la capa de nubes de una tormenta, es decir, es la sección superior de una

región abarcada por ecos de radar cuyo rango supera o iguala cierto umbral predefinido de reflectividad.

- **ETHICKNESS/THICK – Espesor de los ecos:** Permite observar de forma gráfica la densidad de la tormenta, nubes, o de las zonas de precipitación. Es calculable restando los valores de tope de los ecos con los de la base. Proporciona datos de la naturaleza de una tormenta, así como de su potencial de generar efectos severos.
- **VIL – Líquido integrado verticalmente:** Proporciona un estimado del potencial de precipitación contenido en una capa de la atmósfera. Delimita eficientemente las regiones donde se producen precipitaciones de gran magnitud, por lo cual es una herramienta de vigilancia para fenómenos severos. También permite visualizar áreas con potencial de lluvias por medio del contenido de líquido.
- **Reflectividad diferencial (ZDR):** Relaciona la reflectividad en ejes X y Y. Sus valores positivos indican que la precipitación detectada es más larga en el eje horizontal, mientras que valores negativos dicen lo mismo, pero para el eje vertical. Si el valor se acerca a cero se concluye que los hidrometeoros tienen tamaños parecidos.
- **RHOHV – Coeficiente de correlación:** Brinda una escala de la consistencia en los tamaños y las formas de la precipitación detectadas. Valores superiores indican una mayor consistencia en cada uno de los objetos detectados, por lo contrario, valores inferiores indican una variabilidad mayor en la forma y dimensiones de los objetivos. Su uso principal es para diferenciar elementos meteorológicos de elementos ajenos a este campo, diferenciar granizo de gran tamaño de lluvia convencional y llevar a cabo un ajuste de calidad de otros indicadores y productos del radar.
- **PHIDP – Fase diferencial:** Es una medida para normalizar los pulsos del radar con respecto a los datos volumétricos que obtiene. Ya que los pulsos de radar se atenúan al penetrar un objetivo, la fase de cada pulso cambia, por lo cual es posible medir la diferencia en las variaciones de fase en cada pulso de radar, obteniendo datos sobre la forma y concentración de los blancos del radar. Su uso principal es definir el área de mayor precipitación.
- **KDP – Fase diferencial específica:** Es una derivación de otros productos que muestra el cambio de la PHIDP contrastando la información con la distancia

explorada. Se usa con el fin de determinar el porcentaje de precipitación dentro del volumen de datos captados mediante el radar meteorológico, debido a que un aumento en el KDP significa un aumento en tamaño y concentración meteoros, aumentando la proporción de precipitación. De igual manera es auxiliar para la identificación de zonas en donde hay precipitaciones de mayor gravedad. Se utiliza junto a la PHIDP para marcar sectores de importancia.

- **HCLASS/HMC – Clasificación de hidrometeoros:** Representa a los diferentes factores hidrometeorológicos que se encuentren presentes en la atmósfera, ya sean en la forma de lluvias, trozos de granizo, caída de nieve, formación de escarcha. También es capaz de detectar objetivos no meteorológicos tales como reflectores antirradar, bichos y ruido captado desde el océano. Para funcionar utiliza un algoritmo que requiere parámetros tales como los momentos de intensidad corregida (Z), ZDR y factores de doble polaridad.

### **2.3.1 Causas y consecuencias de la contaminación ambiental**

Anzules & Castro (2022) definen a la contaminación ambiental como la existencia de algún factor (ya sea este de origen químico, biológico e incluso físico), o la sobreposición de estos diferentes elementos en lugares, cosas o porcentajes suficientes para presentar un riesgo a la salud, seguridad y bienestar de una población o de seres vivos en general.

Domínguez (2015) identifica que, en su mayoría, los factores causantes de contaminación ambiental tienen origen antropogénico, y son provocados por el desarrollo humano, derivado de procesos industriales, agrícolas, agropecuarios y clínicos, etc. También menciona que estas actividades tienen efecto sobre todos los ambientes: aire, agua y suelo, afectando a la flora y fauna en general, ya sean estos emisores de elementos contaminantes o simplemente receptores.

Dentro de su artículo de divulgación, Campos (2022) clasifica la contaminación ambiental en diferentes tipos:

- Causada por residuos solidos
- Dentro de la composición del agua
- Provocada por la presencia de metales pesados
- Derivada de ruido excesivo

A su vez, también identifica diferentes fuentes de las que estas originan, entre las cuales están: la agricultura, presencia de minería ilegal, fuentes de origen natural, rápido crecimiento urbano e industrial, la eliminación de zonas boscosas y humedales, la propagación de elementos contaminantes hacia esteros y ríos explotación intensiva del suelo para actividades petroleras, entre otras.

En los subtemas presentados a continuación se explorarán algunas de los factores detonantes de contaminación ambiental para el estado de Sinaloa.

#### **2.3.1.1 Deforestación**

Se puede definir a la deforestación como la transformación de forma consciente o inconsciente de la extensión de los bosques a otra clase de bioma en un espacio de tiempo dado, usualmente provocado con el fin de paso a actividades humanas, tales como actividades

agrícolas, ganaderas, urbanísticas o mineras. El énfasis sobre las actividades es importante, ya que nos muestra que la transformación de los ecosistemas es un proceso usualmente permanente o irreversible. Dicho proceso tiene como consecuencias el aumento de la temperatura global (tomando en cuenta la escala a la que se realizan estas actividades), mayores incidencias de incendios forestales, degradación del suelo, desde la disminución hasta la desaparición de especies de regiones, entre muchas otras (World Wildlife Fund, 2024).

En el trabajo realizado por Monjardín-Armenta, et al. (2017) se llegó a la conclusión de que existe una deforestación de  $126.50\text{km}^2/\text{año}$  y una tasa promedio al año de un  $0.41\%$ , siendo uno de sus factores detonantes la expansión de la actividad agrícola, así como el aumento de cobertura para infraestructura, generando impactos de  $49.40\%$  y  $18.8\%$  respectivamente.

Lo cual quiere decir que, en un lapso de 18 años se deforestaron alrededor de  $2,227\text{ km}^2$  de árboles en diferentes ecosistemas de Sinaloa y, tomando en cuenta que el estado abarca un área de  $58,092\text{km}^2$ , aproximadamente un  $3.8336\%$  del territorio estatal sufrió de afectaciones producto de esta actividad.

A partir de la información consultada anteriormente, es posible llevar a cabo una correlación entre las actividades comunes que provocan deforestación mencionadas dentro del artículo con las conclusiones destacadas dentro de los diferentes estudios realizados sobre el estado de Sinaloa.

### 2.3.1.2 Sequías

*Sequía* refiere a la ausencia o insuficiencia de lluvias durante un periodo prolongado de tiempo (Real Academia Española, 2024).

Las sequías pueden variar tanto en la intensidad y en su duración dependiendo de las zonas geográficas y sus parámetros, como altitud, humedad, fenómenos meteorológicos presentes en la atmósfera, la actividad humana, entre otros factores.

Por lo tanto, de acuerdo con Coronel (2013), la sequía se puede categorizar en cuatro tipos diferentes, dependiendo el ámbito científico desde el cuál es estudiada:

- **Sequía meteorológica:** Sustentada por medio de datos climatológicos, se trata de una desviación de precipitaciones con respecto a la media de cierto periodo.
- **Sequía agrícola:** Debido a que es el primer nivel afectado por la ausencia de precipitaciones.
- **Sequía hidrológica:** Refiriendo a una deficiencia en el caudal el tránsito del agua por la superficie terrestre o de forma subterránea.
- **Sequía socioeconómica:** Cuando se ve afectada la disponibilidad del agua para la población en una zona, afectándola personal o económicamente.

La duración y la intensidad de una sequía provocan estrés hídrico, el cuál es definido como la relación entre oferta y demanda de agua en una región, combinada con la disponibilidad de fuentes renovables. El estrés hídrico, a su vez, generan problemas en la población y la economía, causando cortes en el servicio de distribución de agua a nivel civil, así como reduciendo el suministro a industrias agrónomas y agropecuarias, afectando la producción anual y la economía de la zona (PNUD, 2024).

Las temporadas de sequía están estrechamente relacionadas con el cambio climático, tal y como dice Magaña, et al. (2012), al realizar un recuento de sequías recientes la zona norte del territorio mexicano es posible ver que los costes provocados por la variabilidad del clima son cada vez más altos.

#### **2.3.1.2.1 Indicadores de sequía**

El trabajo de investigación de Coronel (2013) menciona que gran parte de los artículos de investigación acerca de sequías solo toman en cuenta como índices el Índice de Severidad de Sequía de Palmer (PDSI), el cuál es una ecuación que permite obtener el balance de agua en el suelo, y el Índice de Precipitación Estandarizado (SPI), que lleva a cabo un cálculo de precipitación por medio de un enfoque probabilístico.

Dentro del trabajo de investigación presentado por Alley (1984), reconoce que el PDSI es probablemente uno de los indicadores de sequía regional más utilizados, a pesar de utilizar reglas muy arbitrarias para medir sus propiedades. Dentro de este documento también se menciona que, a pesar de que el PDSI es un indicador de sequía meteorológica, éste toma a consideración factores como la precipitación, evapotranspiración y humedad del suelo, pero no toma a consideración la disponibilidad de agua en arroyos, ríos y reservas.

En México, el monitor de la sequía (CONAGUA, 2025) menciona que utiliza el SPI por su sencillez, enfatizando que la sencillez de dicho modelo está en que solamente usa cantidad de lluvia obtenida para la obtención del índice, lo cual es efectivo para llevar a cabo análisis por periodos y ciclos de humedad.

### 2.3.1.3 Inundaciones

La Coordinación Nacional de Protección Civil de México (2018) dice que las inundaciones son eventos causados por lluvia, variaciones en el oleaje, marea provocada por tormentas o fallas dentro de la infraestructura hidráulica federal que aumente la cantidad de agua presente en cuerpos como ríos, lagos, lagunas o el mar, provocando de esa forma el crecimiento de estos, llegando a invadir sitios donde usualmente no llega el agua.

De acuerdo con Arreguín-Cortés, et al. (2016), las inundaciones pueden ser clasificadas de la siguiente manera:

- **Fluviales:** Provocadas por desbordes de ríos.
- **Súbitas:** También conocidas como *flash floods*, precipitaciones de un alto volumen de agua en un espacio de tiempo reducido.
- **Urbanas:** Provocadas por falla dentro del sistema de drenaje pluvial y alcantarillado
- **Pluviales:** Causadas por el volumen de precipitación
- **Costeras:** Provocada por fenómenos naturales súbitos como tsunamis, ciclones o huracanes o derivado del cambio climático por el aumento del nivel del mar.
- **Por deslizamiento de glaciares:** Aumento en el nivel del mar derivado al desprendimiento de cuerpos glaciares.

#### **2.3.1.4 Cambio de uso de suelo**

Las alteraciones del suelo dentro una región va de la mano con el crecimiento poblacional de esta; ya sea para satisfacer necesidades de vivienda, de infraestructura, de industria o de servicios, eventualmente el suelo sufrirá alguna transformación en mayor o menor medida. Estos cambios dentro del uso del suelo tienen como línea de base procesos como la deforestación, y pueden conllevar a la desaparición de regiones naturales y la degeneración de estos, ocasionando problemas derivados del cambio climático, como el aumento de temperaturas, la intensificación de fenómenos naturales y la concentración de gases de efecto invernadero, además de contaminación del suelo y agua en zonas colindantes. Esto es lo visto dentro del estudio realizado por Bocco, et al. (2001), en el cual se analiza el cambio que sufre Michoacán debido a la deforestación causada por la tala intensiva, al igual que en otros estudios, como lo son el de Palacio et al. (2002), que analiza patrones y cambios provocados por deforestación, degradación, desertificación y otras alteraciones dentro del uso de suelo dentro en territorio mexicano.

De acuerdo con el estudio de Velázquez, et al. (2002), se recomienda la siguiente categoría jerárquica utilizada para medir los factores de cambio de uso de suelo:

1. Cultivos
2. Bosques
3. Selvas
4. Matorrales
5. Pastizales
6. Vegetación hidrófila
7. Otros tipos de vegetación
8. Otras coberturas (como las zonas urbanas)

El estudio anterior destaca que los datos históricos del cambio de uso de suelo son de baja fiabilidad y que, debido a esto, precisa que el mejorar los datos sería algo benéfico para el estudio de esta área, pero que, aun con los datos faltantes, las tasas de cambio estimadas para México fueron notoriamente menores a las alteraciones que realmente ocurrieron, lo cual indica una transformación de suelo más acelerada de lo predicho.

### 2.3.2 Legislaciones en materia ambiental

Con el fin de preservar el medio ambiente evitando el cambio climático, existen diversas medidas de protección y cuidado, impartidas por el gobierno de México a diferentes niveles legislativos.

A nivel federal existen:

- **Ley General del Equilibrio Ecológico y la Protección al Ambiente (DOF 01-04-2024):** Esta ley da derecho a todas las personas a vivir en un medio ambiente sano, define principios de políticas de preservación ambiental y proporciona instrumentos para aplicar dichas políticas. También busca garantizar la participación de todo individuo con el objetivo de preservar y restaurar el equilibrio ecológico dentro de territorio nacional (L.G.E.E.P.A., 2024).
- **Ley Federal de Responsabilidad Ambiental (DOF 20-05-2021):** Esta ley sirve como marco de referencia para la determinación de responsabilidad, imposición de sanciones y compensación de daños y reparaciones para todo individuo o entidad que cause daños al medio ambiente (L.F.R.A., 2021).
- **Ley General del Desarrollo Forestal Sustentable (DOF 01-04-2024):** Esta ley busca la regulación y fomentar el manejo integral y sustentable territorios forestales federales, con el fin de conservarlos, repararlos y aprovecharlos de tal forma que el territorio forestal no vaya en declive (L.G.D.F.S., 2024).

También existen estrategias gubernamentales a nivel estatal, como por ejemplo implementadas por la Ciudad de México (Secretaría del Medio Ambiente, 2025):

- Saneamiento del arbolado
- Hoy no circula
- Cambio climático
- Basura cero
- Infraestructura verde
- Cosecha de lluvia

## **2.4 Procesamiento y análisis de datos**

Toda investigación requiere de procesar datos, los cuales son obtenidos mediante diferentes técnicas y deben de ser procesados con el fin de construir estos datos crudos en información utilizable (Castañeda, 2010), a este proceso de construcción de información nueva a partir de diferentes técnicas o métodos es a lo que se le denomina procesamiento de datos.

Procesar datos de forma manual es una tarea compleja y repetitiva, la realización incorrecta de esta labor puede llevar a errores al momento de realizar un análisis, y localizar el punto donde se cometió el error puede llegar a ser una tarea extremadamente exhaustiva.

El software de procesamiento de datos busca acelerar la limpieza de los datos, por medio de correcciones de formato, de tipo de dato, rellenado o supresión de nulos, entre otros, con el objetivo de aumentar la calidad de la información analizable. Ejemplos de uso de software para procesar datos hay varios, como lo es el procesamiento de datos de sensor LiDAR con la herramienta ArcGIS Desktop 10 (San José, 2011), en donde hacen tratamiento de los datos obtenidos directamente del sensor para permitir su uso posterior en aplicaciones forestales, o el uso de Quimiometrix para procesar datos químicos y bioquímicos, como el caso de Nuñez-Cuadra, et al. (2011), en el que lograron obtener mejores resultados en menor tiempo, analizando un mayor volumen de datos.

El análisis de datos implica realizar operaciones estadísticas sobre los datos procesados para detectar elementos de interés a analizar, como la mediana, el promedio, el valor de moda, desviación estándar, varianza y, dependiendo del formato de los datos, también es posible llevar a cabo análisis de tendencias, permitiendo encontrar así patrones más complejos de comportamiento.

Existen diversas herramientas de libre acceso y de paga que permiten acelerar o incluso automatizar la labor de procesamiento y facilitar el análisis de los datos. pero como es de esperar, no existe una sola herramienta que se pueda considerar como infalible, única o definitiva, sino que la herramienta a utilizar debe de ser seleccionada bajo el contexto específico de los datos y la labor que se quiere realizar y, en caso de ser una herramienta de naturaleza genérica, se deben de realizar ajustes y adaptaciones pertinentes al formato o estructura de lo que se quiere analizar.

### **2.4.1 Extracción, Transformación y Carga**

Vassiliadis, et al. (2009) afirman que las actividades que conforman el proceso Extract-Transform-Load (ETL) se definen como módulos de software que en conjunto cargan con la responsabilidad de alimentar un Almacén de Datos (AD). El artículo menciona que, a pesar de que existen diversas aproximaciones para diferentes flujos de procesos de ETL, existen cualidades genéricas que éstos siguen, dando una taxonomía que los caracteriza en los términos de la relación de los datos originales con los datos procesados, y provee una forma normal basada en la semántica para las actividades dentro del proceso.

### **2.4.2 Técnica de raspado de datos (data scraping)**

El data scraping o raspado de datos, a partir de la definición de Mustapha, et al. (2024), es un concepto que involucra la extracción de información relevante de una base de datos contenida dentro de una computadora. Uno de los principales lenguajes de programación para elaborar scripts de raspado de datos es Python.

### **2.4.3 Técnica de web scraping**

Comúnmente se generaliza el concepto de data scraping con el de web scraping por el hecho de que la web es una base de datos de gran tamaño, sin embargo, el web scraping involucra la búsqueda de palabras, conceptos e información clave dentro de una página web, automatizando la búsqueda dentro de todos los directorios accesibles mediante un navegador (Khder, 2021), estén éstos claramente visibles o no, incluso analizando los metadatos contenidos en el script.

## **2.5 Elementos tecnológicos relevantes para el procesamiento de datos**

Existen diferentes elementos tecnológicos disponibles en el mercado para procesar datos, pueden ser gratuitas o de pago por licencia, pero su uso está mayormente atado a las necesidades de los datos a procesar.

### **2.5.1 Lenguaje de programación Python**

Python, de acuerdo con su documentación, es un lenguaje de programación de alto nivel caracterizado por ser amigable y fácil de aprender e interpretar, que tiene aplicaciones en desarrollo web, conexión con bases de datos, interfaces gráficas, manejo de datos científicos y numéricos, programación de redes, programación de software y juegos en general y educación en sus módulos base, y permite el uso de librerías externas para aumentar su funcionalidad (Python Software Foundation, 2013).

### **2.5.2 Formato de archivo Valores Separados por Coma (CSV)**

Los archivos CSV (Comma Separated Values), de acuerdo con la definición de Shafranovich (2005), es una estructura de archivos utilizada para intercambiar y convertir datos entre varios programas de hoja de cálculo. Como en estos programas, utiliza un formato de filas y columnas, cada columna siendo separada por comas, en el cual la primera fila es la cabecera, es decir, contiene el nombre o identificador de la columna, y las demás filas o registros son los datos asociados a cada columna.

### **2.5.3 Herramienta Pandas para Python**

Pandas es una herramienta integrable a Python desarrollada por Wes McKinney. Permite llevar a cabo diversas transformaciones y análisis de datos en diferentes formatos (CSV, XLSX, HDF5, JSON, entre otros), realizar operaciones matemáticas con estos, obtener datos estadísticos, generar gráficas, entre otras muchas funcionalidades.

## **2.6 Herramientas disponibles para análisis de datos**

Al igual que con las tecnologías para procesar datos, existen varias herramientas para su análisis y explotación. De igual manera existen herramientas gratuitas y de licencia, que son capaces de manejar un gran volumen de información para analizar.

### **2.6.1 Software propietario para análisis de datos Power BI**

Microsoft define a Power BI como una plataforma diseñada especialmente para el manejo de Inteligencia Empresarial (BI). Es bastante potente, puesto que permite crear almacenes de datos a partir de diferentes orígenes desde archivos de Excel, CSV, hasta bases de datos de diferentes compañías (Microsoft Corporation, 2015). Además, permite visualizar de forma conveniente y flexible los datos, a partir de diferentes ejes, realizando diferentes operaciones estadísticas, algoritmos y visualizaciones gráficas para la creación de informes.

### **2.6.2 Software de Sistemas de Información Geográfica (GIS)**

Es posible llevar a cabo la validación del análisis de datos por medio del uso de Sistemas de Información Geográfica (GIS), los cuales permiten utilizar información geoespacial para visualizar la información por medio de geolocalización.

Existen diversos programas informáticos para realizar análisis por medio de GIS, como Quantum Geographic Information System (QGIS), el cual es un software gratuito que hace posible leer y procesar volúmenes enormes de imágenes geográficas. Se encuentra disponible para gran parte de los sistemas operativos actuales, proporcionando actualizaciones constantes (Moyroud, N. & Portet, F., 2018).

Otro programa para la realización de análisis por medio de GIS es ArcGIS, el cual es un software propietario del Environmental Systems Research Institute (ESRI) diseñado para trabajar información a partir de mapas e información geográfica. A diferencia de su contraparte gratuita, este software está diseñado principalmente para las plataformas de Microsoft Windows (Khan, S. & Mohiuddin, K., 2018).

## **2.7 Generación de datos sintéticos**

No existe como tal una convención formal para el término “datos sintéticos”, aunque existen diversas fuentes que buscan definirla de forma similar. Citando una de estas, los datos sintéticos son información, es decir, datos, generados utilizando un modelo matemático o algoritmo hecho a medida, con el fin de resolver uno o varios objetivos de ciencia de datos (Raghunathan, 2021). Otra fuente afirma que los datos sintéticos son una colección de datos para consumo público, generados mediante el procesamiento de datos reales por medio de un proceso estadístico, los cuales son re-expresados con el fin de limitar la exposición de información sensible (Jordon, et al, 2022).

Ambas definiciones aportadas denominan a los datos sintéticos como una colección de datos creadas por medio de algoritmos o modelos matemáticos y estadísticos, con el fin de resolver un problema asociado con el análisis de datos (séase protección de estos, falta de estos, entre otras razones).

La generación de datos sintéticos es una técnica ampliamente aceptada debido a que es una forma bastante general de mejorar los resultados de los análisis de datos.

Existen diversos métodos para generar datos sintéticos, unos más sencillos que otros, y abarcan desde colocar valores duplicados en campos con registros faltantes, utilizar estrategias matemáticas como la interpolación, hasta utilizar IA o ML para rellenar los datos faltantes e incluso predecir datos futuros.

### **2.7.1 Deep Learning para la generación de datos sintéticos**

El Deep Learning (DL) es un tipo de red neuronal más compleja, diferenciándose del Machine Learning (ML) porque el sistema tiene más de una capa oculta de aprendizaje (Kufel, et al. 2023). Estas capas ocultas pueden realizar una operación en particular, o seguir algún proceso, gracias a estas capas ocultas existen diferentes modelos de DL.

Algunos modelos de generación de datos sintéticos utilizan este tipo de redes neuronales, como lo es el caso de Alzanzot, et al, (2017) o de Lu, et al. (2023). Utilizan diferentes capas dentro de un modelo de DL para mejorar los resultados.

### **2.7.1.1 Long-Short Term Memory (LSTM)**

Los creadores de esta red neuronal recurrente (Hochreiter, et al., 1997) definen al LSTM como un algoritmo de aprendizaje de red capaz de saltarse retrasos temporales, reforzando el flujo de errores en secciones especiales del código. Es un algoritmo local y su carga computacional es de  $O(1)$ . Sirve para resolver problemas complejos que generan muchos tiempos de espera a los otros algoritmos de redes neuronales recurrentes.

### **2.7.2 Mixture Density Network (MDN)**

De acuerdo con la investigación de Davis, et al. (2020), una MDN es un conjunto de componentes: una red neuronal y un mixture model. Un mixture model es un modelo distribución de probabilidades construido a partir de la suma pesada de distribuciones más simples. El principal fin de las MDN son emular modelos estocásticos, es decir, simular situaciones en donde los datos son inciertos o cuentan con el factor de aleatoriedad.

### **2.7.3 Generación de datos sintéticos mediante modelos de Deep Learning**

Teniendo un registro decentemente lleno de datos, es posible llevar a cabo la generación de datos sintéticos utilizando un modelo entrenado de DL, evidentemente, mientras más porcentaje de los datos estén disponibles y completos, mejor serán las capacidades predictivas de este, aunque también es posible tener una cantidad considerable de datos faltantes dentro de un conjunto de datos extremadamente extenso.

Existen diversos ejemplos de la utilización de modelos de ML para la generación de data sintética, por ejemplo el trabajo de investigación presentado por Alzanzot, et al, (2017), en el cual implementan SenseGen que es una arquitectura nueva para sintetizar datos de sensores por medio de un modelo de DL que utiliza varias capas de redes de LSTM y una MDN para distinguir entre datos reales y datos sintéticos, y la posterior generación de datos nuevos, más semejantes a los datos captados por los sensores.

# Capítulo 3

## 3 Estado del arte

El procesamiento y análisis de datos hidrometeorológicos otorgan un nivel más de comprensión de estos fenómenos, y mediante el uso de esta información, que es captada por diferentes métodos, es posible elaborar modelos predictivos cada vez más precisos de sistemas más complejos. Integrando a estos modelos la capacidad de aprendizaje y entrenamiento continuo de un sistema de Redes Neuronales Artificiales RNA se vuelve viable la generación de datos sintéticos por este medio, con el fin de reconstruir de la forma más precisa posible y mejorar el conjunto de datos obtenidos mediante sensores terrestres. Todo esto es posible gracias a que actualmente contamos con sistemas para validar esta información como lo es el mostrado por Madueño, et al. (2016) para corroborar datos hidrológicos usando RNA, que dan pasos a sistemas para llevar a cabo proyecciones, como el de Rojas et al. (2024), en el que se logró llevar a cabo la estimación de la precipitación sobre la cuenca del río Pastaza en Ecuador.

A pesar de que la generación de datos sintéticos por medio de Inteligencia Artificial es un proceso con bases sólidas, aplicados en diversas áreas de aplicación, como lo puede ser el predecir brotes de COVID-19 en el caso de Khayyat, et al. (2021), optimizar las predicciones hidrológicas de series de tiempo tal como realizaron Dodig, et al (2024), hasta crear predicciones hidrometeorológicas nuevas por medio de diferentes variables, como fue el caso de Brodeur, et al. (2021). En México, no es una práctica utilizada comúnmente, lo cual es notorio por las carencias que padecen sus sistemas de información. Sin embargo, actualmente, es posible desarrollar herramientas útiles que se pueden implementar en las instituciones gubernamentales estatales y federales en términos de estudios hidrometeorológicos, el primer paso es llevar a cabo la transformación de los datos crudos proporcionados por los sensores con los que se cuentan.

### **3.1 Fuentes de datos disponibles**

Actualmente, existen diversas fuentes de datos que se utilizan que sirven como índices para medir los cambios en el medio ambiente, CONAGUA (2025) mide parámetros hidrometeorológicos como lo son temperaturas máximas y mínimas, precipitación, evaporación, entre otros.

Medir apropiadamente variables bioclimáticas, además de proporcionar un registro histórico de los eventos meteorológicos que suceden en el país, ayudan a la generación de modelos predictivos más fiables y robustos.

#### **3.1.1 Sensores hidrometeorológicos**

Los sensores hidrometeorológicos son herramientas que permiten medir diferentes parámetros concernientes a estudios hidrológicos y meteorológicos, tales como precipitación, evaporación, humedad, temperaturas, entre otros dentro del rango de alcance de estos.

De acuerdo con el estudio desarrollado por Vilchis-Mata, et al. (2015) , algunos de los sensores meteorológicos disponibles en México son las Estaciones Meteorológicas Automáticas (EMA), el radar meteorológico de banda C de Cerro Catedral y sistemas satelitales como el PERSIANN y el CMORPH, todos dentro del modelo hidrológico distribuido CEQUEAU. Aunado a esto, México también cuenta con radares Doppler manejados por la CONAGUA.

Las Estaciones Meteorológicas Automáticas, de acuerdo con la documentación proporcionada (CONAGUA, 2025), son sistemas autónomos automatizados formados por varios sensores encargados de medir diferentes índices, dispositivos eléctricos, electrónicos y mecánicos, ensamblados en una estructura y conectados a un Sistema de Adquisición. Procesamiento y Almacenamiento de Datos (SAPAD) local, que tienen el fin de llevar a cabo mediciones y registros de índices meteorológicos de la zona en la que es desplegada, transmitiendo sus datos obtenidos hacia una oficina central, para su posterior uso y almacenamiento.

### **3.1.2 Variables de análisis meteorológicos**

Las variables analizadas dependen del tipo de sensor, sin embargo, es de notar que, aunque dos fuentes de datos midan el mismo parámetro, es posible que su método de obtención de datos difiera, haciendo posible la interrelación y validación de datos por medio de comparación entre dos sensores diferentes.

Por lo tanto, una EMA cuenta con los siguientes sensores básicos para medir los siguientes parámetros:

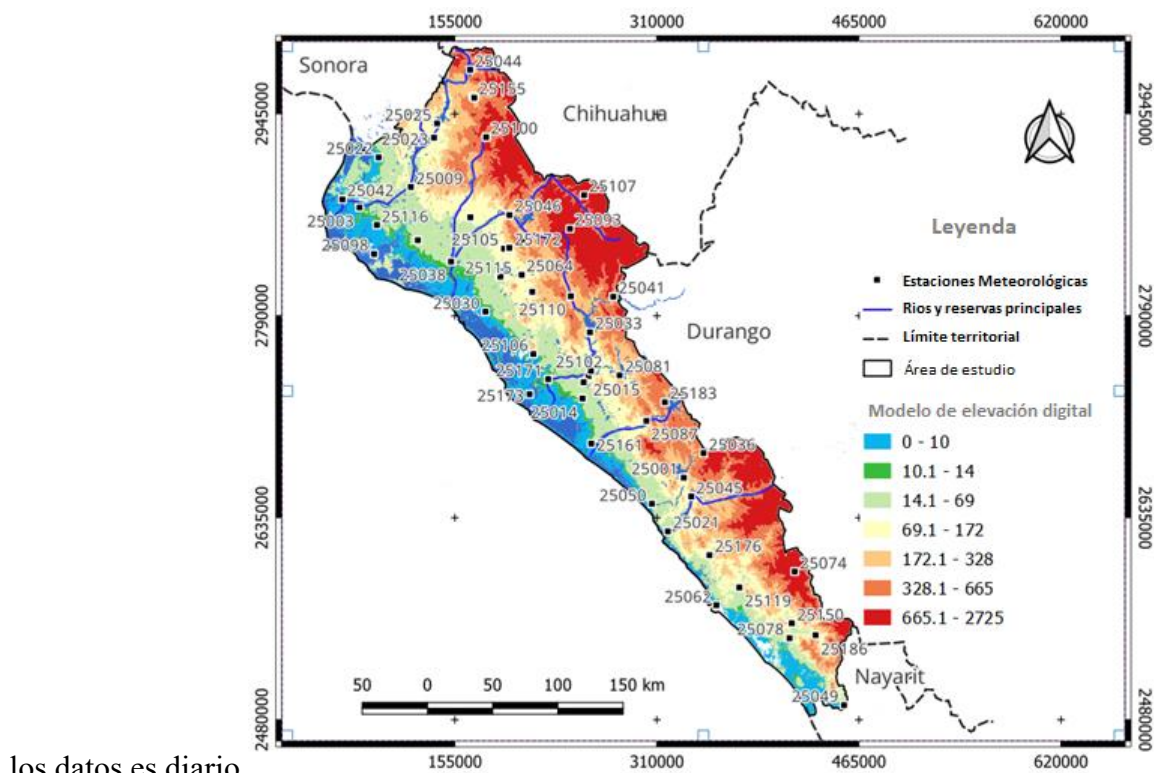
- Anemómetros (para detección de velocidad y dirección del viento)
- Barómetros (para medición de presión atmosférica)
- Termómetros
- Higrómetros (para medir humedad relativa)
- Piranómetros (para captar radiación solar)
- Pluviómetros (para medir cantidad de precipitación)

### 3.2 Variables Bioclimáticas

Dentro del portal oficial de Información Estadística Climatológica (CONAGUA, 2025) es posible obtener datos de sensores acerca de los siguientes parámetros:

- Precipitación (medida en milímetros)
- Evaporación (medida en milímetros)
- Temperatura máxima (medida en °C)
- Temperatura mínima (medida en °C)

Estos datos son obtenidos a través de sensores colocados en lugares estratégicos (cauces de río, puntos de precipitación clave, presas) y existen una gran cantidad de estaciones hidrometeorológicas, generalmente más de una por municipio. El periodo de captación de



los datos es diario.

Figura 3.1. Mapa de estaciones meteorológicas operativas de Sinaloa (Fuente: Elaboración propia)

En la figura 3.1 se pueden encontrar las 50 estaciones operativas en el estado de Sinaloa que están comprendidas dentro de esta investigación. Además, se encuentran localizados los flujos principales de agua y la altitud de la zona en rangos de colores. Las estaciones fueron

colocadas estratégicamente a lo largo de los ríos, desde su origen hasta su salida por el mar, esto con el fin de buscar la mejor calidad de datos que se pueda obtener.

### **3.2.1 Generación de datos sintéticos para sensores hidrometeorológicos**

La generación de datos sintéticos aplicada a sensores hidrometeorológicos es de gran importancia debido a la cantidad de datos faltantes y la inconsistencia de las captaciones de estos, el simular datos sería un proceso beneficioso para cualquier modelo predictivo que se llegue a desarrollar posteriormente.

La investigación desarrollada por Dodig, et al. (2024), demuestra que los modelos predictivos mejoran tras la generación de datos sintéticos en datos hidrometeorológicos, el cuál utiliza un modelo de Redes Generativas Adversariales (GAN) y otros modelos de en el cual se intenta predecir la calidad del agua que fluye de un río, en el cuál, utilizando un modelo de Redes Neuronales Adversarias para series de tiempo aplicadas en el mundo real con el fin de generar data sintética, es posible mejorar la predicción de los modelos de Redes Neuronales de Grafo (GNN), aumentando su desempeño hasta un 43.4%, y el de las LSTM en un 15%.

Taesam, et al. (2021) desarrollaron dentro de su investigación un modelo que utiliza Redes Neuronales Convolucionales (CNN) y LSTM aplicado en la hidrometeorología y ciencias ambientales. Además, proporcionan la metodología que utilizaron y diferentes derivaciones de algoritmos de DL para estimar parámetros por medio de retropropagación y aplicaciones en conjuntos de datos reales.

# Capítulo 4

## 4 Desarrollo del proyecto

### 4.1 Área de Estudio

Está delimitada dentro del estado de Sinaloa, México (tal y como se muestra en la Figura 4.1). La economía principal del estado de Sinaloa está basada en la agricultura intensiva, ganadería y la acuicultura. Todas estas actividades tienen como característica común el uso intensivo del agua. Razón por la cual es necesario el procurar una mayor capacidad de análisis con el fin de predecir, prevenir y gestionar sequías traería consigo un beneficio a nivel nacional, minimizando las pérdidas.

Para la realización del análisis, se utilizarán datos proporcionados por las 50 estaciones meteorológicas operativas dentro del territorio del estado.



Figura 4.1. Mapa de estaciones meteorológicas operativas de Sinaloa (Fuente: Elaboración propia)

Estas estaciones proporcionan un registro histórico de los siguientes índices hidrometeorológicos con un periodo de actualización diario:

- Precipitación (en mm)
- Evaporación (en mm)
- Temperatura máxima (en °C)
- Temperatura mínima (en °C)

## **4.2 Metodología de desarrollo**

En el diagrama presentado en la figura 4.2 se define la metodología creada para el desarrollo de este proyecto, a la cual se le fue denominada ETL-A por cada una de sus fases: Extracción, Transformación, Carga (Load) y Análisis, el cual sigue un proceso compuesto por 7 pasos abarcados en 4 etapas: Extracción, en donde se descargan los datos en lote desde el sitio de la CONAGUA. Transformación, que genera los archivos CSV normalizándolos, limpiándolos y procesándolos, además de generar concentrados diarios y mensuales. Carga, que consiste en cargar los archivos ya procesados en algún software para el análisis y/o explotación de datos. Y análisis, que se realiza una vez cargados los datos.

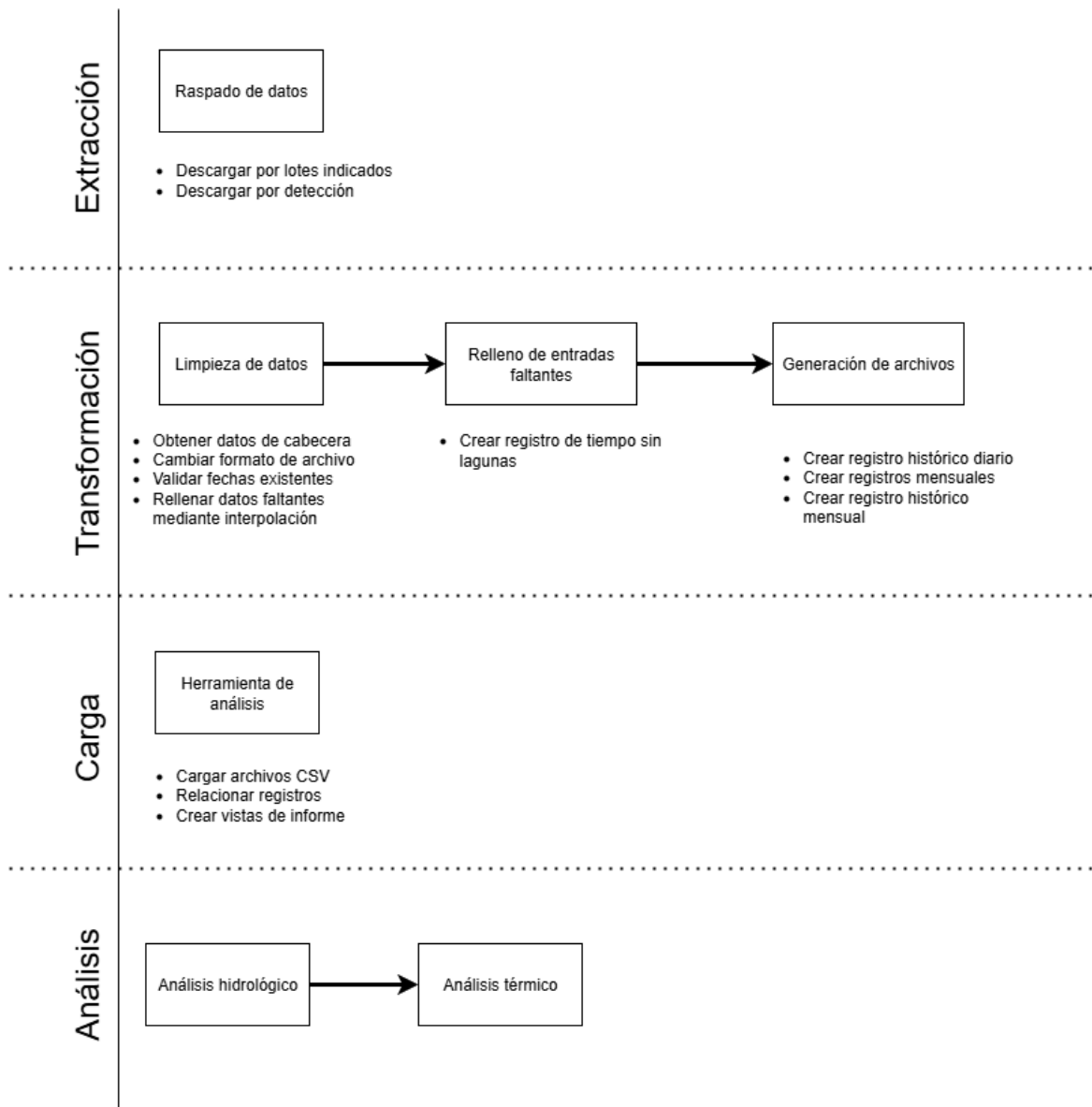


Figura 4.2. Diagrama de metodología ETL-A (Fuente: Elaboración propia)

### 4.2.1 Etapa 1: Extracción (datascraper.py)

En esta etapa se extraen los datos desde la plataforma de la CONAGUA utilizando web scraping, sugerida debido a su funcionalidad en la investigación de Thomas & Mathur (2019), en la cual utilizan un scraper de datos desarrollado en Python para realizar un posterior análisis de la información obtenida, los resultados reportados fueron satisfactorios debido a la utilización del scraper, que fue hecho a la medida de la información que necesitaban obtener, cuya extracción manual resultaba tediosa. Los datos necesarios se encuentran en formato de texto plano, albergados en el servidor de la CONAGUA, cuya URL tiene la siguiente estructura:

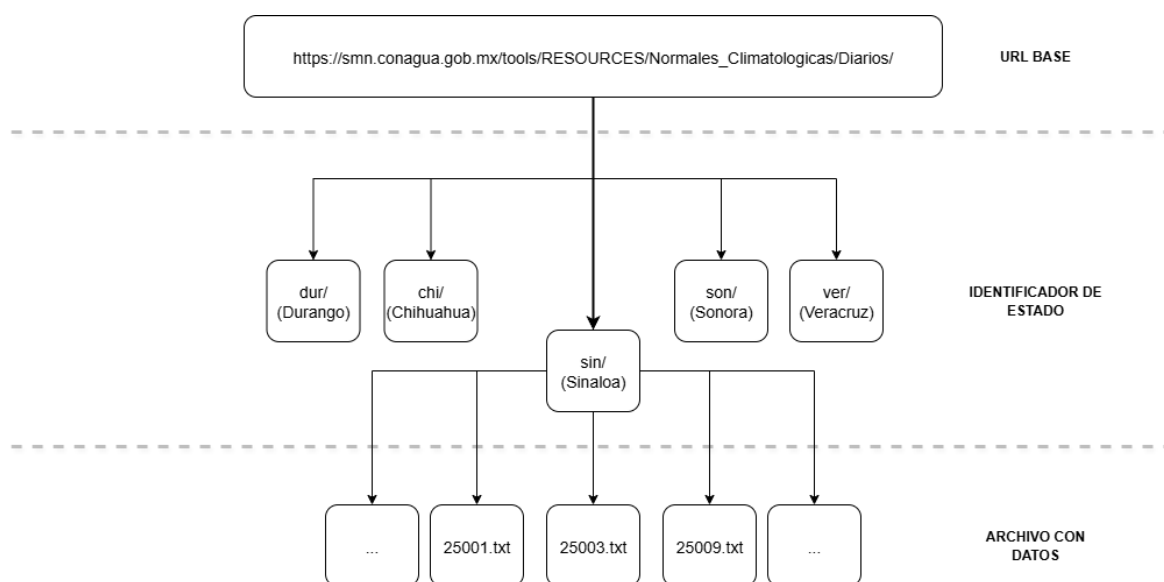


Figura 4.3. Estructura de página de CONAGUA (Fuente: Elaboración propia)

Como se puede apreciar en la figura 4.3, es posible acceder a los archivos individuales que contienen los datos hidrometeorológicos de las estaciones de Sinaloa a partir de un nivel en el árbol de directorios del repositorio de la CONAGUA. Esto reduce la búsqueda de los archivos a dos métodos diferentes:

- **Fuerza bruta (explore\_download):** El método `explore_download` de la figura 4.4, indica la búsqueda en el estado seleccionando estaciones dentro de un rango del 1 a 999. En el caso de Sinaloa, debido a que su código de estado es 25, buscaría del 25001 al 25009.

```
def explore_download(start, end, output_folder, estado = 'sin'):
    """ explore_download(start, end, output_folder)
        This function explores every url in a given range and saves the existing .txt files into an output folder

        Input:
        - start: a number associated with a file
        - end: final number to search in a range
        - output_folder: string with the name of the folder where the found files will be saved
        - estado: string with the state code (default is 'sin')
    """
    for i in range(start, end):
        url = f"https://smn.conagua.gob.mx/tools/RECURSOS/Normales_Climatologicas/Diarios/{estado}/dia{i}.txt"
        filename = f"{output_folder}/{i}.txt"
        download(url, filename)
```

Figura 4.4. Código de búsqueda por fuerza bruta. (Fuente: Elaboración propia)

- **Búsqueda específica (bulk\_download):** Se le proporcionan las estaciones de las cuales extraer datos. En este proyecto se utilizará esta forma de extracción de datos, debido a que las estaciones operativas son solamente cincuenta.

```
def bulk_download(stations, output_folder, estado = 'sin'):
    """ bulk_download(output_folder)

        This function downloads a range of files from the SMN website.

        Input:
        - output_folder: string with the name of the folder where the files will be saved
        - stations: list of strings with the code of the stations to download
        - estado: string with the state code (default is 'sin')

        Output: None
    """
    for estacion in stations:
        url = f"https://smn.conagua.gob.mx/tools/RECURSOS/Normales_Climatologicas/Diarios/{estado}/dia{estacion}.txt"
        filename = f"{output_folder}/{estacion}.txt"
        download(url, filename)
```

Figura 4.5. Descarga por búsqueda en lote. (Fuente: Elaboración propia)

Para iniciar con el proceso de extracción de la información, se deben seguir los siguientes pasos:

1. Se elige si la descarga va a ser por fuerza bruta o mediante una lista de estaciones determinadas.
2. Se descargan los archivos desde el sitio de la CONAGUA hacia el ordenador para su posterior procesamiento.

```
def download(url, filename):  
    '''download(url, filename)  
  
    This function downloads a file from a specific URL and saves it to a determined filename.  
  
    Input:  
    - url: string with the URL of the file to download  
    - filename: string with the name of the file to save the downloaded file  
  
    Output: None  
    ...  
    try:  
        urllib.request.urlretrieve(url, filename)  
        print(f"Downloaded {filename}")  
    except Exception as e:  
        print(f"Error downloading {url}: {e}. File does not exist")
```

Figura 4.6. Función que se referencia en ambos métodos de descarga (Fuente: Elaboración propia)

#### **4.2.2 Etapa 2: Transformación**

El procesamiento o transformación de los datos es la forma en la que se facilitará el posterior análisis. De acuerdo con El-Sappagh (2011), el proceso de Transformación en cualquier modelo de ETL tiende a ser una etapa que tiende a hacer limpieza conforme a los datos que recibe, con el fin de obtener datos correctos, completos, consistentes y no ambiguos.

Dentro de la metodología propuesta en este documento, la etapa de transformación es un proceso que se encarga de corregir el formato de los datos de texto extraídos del portal de CONAGUA, añadir información relevante a estos de forma correcta, es decir, como una columna más dentro del archivo CSV, rellenar espacios en el tiempo con el fin de eliminar las lagunas temporales y generar datos nuevos utilizando estrategias estadísticas, en este caso imputación mediante interpolación temporal de los datos mensuales. Método de imputación de datos comprobado por la investigación de Hernández (2013)

El proceso se define en cinco pasos que se detallan a continuación.

### 1.A.1.1. Paso 1: Limpiar archivo en formato de texto

Para la limpieza de los datos se definió una función llamada *limpiar\_txt\_a\_csv*, que recibe como parámetro de entrada un archivo de texto y el directorio en donde se depositará el archivo convertido en CSV.

```
def limpiar_txt_a_csv(archivo_txt, directorio_salida):
    """limpiar_txt_a_csv(archivo_txt, directorio_salida):

    This function receives a .txt file with a specific structure, then it cleans and processes it and saves it as a .csv file.
    Input:
    - archivo_txt (File): A .txt file with a specific structure.
    - directorio_salida (String): The directory where the .csv file will be saved.
    Return:
    None
    """

    codificacion = "utf-8"
    regex = r'\d{4}-\d{2}-\d{2}'
    linea_inicio = 25

    with open(archivo_txt, 'r', encoding=codificacion) as file:
        lineas = file.readlines()

    print(f'Procesando archivo {archivo_txt}')
    print(f'Número de líneas: {len(lineas)}')

    target_lines = [10, 11, 12, 13, 14, 16, 17, 18]
    etiquetas = []
    valores = []
    estacion = "Desconocida"
```

Figura 4.7. Inicio de paso 1 dentro del código de Python (Fuente: Elaboración propia)

La información que el código anterior (figura 4.7) extrae de la cabecera del archivo previo a eliminarlas es la siguiente:

- Número de estación
- Nombre de la estación
- Estado al que pertenece
- Municipio
- Condición de la estación (Operativa o Clausurada)
- Coordenadas geográficas
  - Latitud
  - Longitud
- Altitud

El algoritmo que se sigue durante este paso del proceso es el siguiente:

**1. Abrir un archivo del directorio en donde se depositaron los datos extraídos**

**2. Extraer datos relevantes de la cabecera del archivo**

**3. Limpiar archivo**

Se eliminan la cabecera del archivo, caracteres inválidos, tabulaciones y espacios que puedan ocasionar problemas al cambiar el formato del archivo.

**4. Transformar el archivo de formato de texto (.txt) a formato CSV (.csv)**

**5. Añadir columnas nuevas**

Se añaden las columnas para los datos relevantes que se extrajeron de la cabecera

**6. Validar formatos de fecha**

Se valida que el formato de fecha sea consistente en todo el archivo (DD-MM-AAAA o AAAA-MM-DD) en todos los registros existentes

**7. Rellenar datos nulos**

Si dentro de un mes existe algún dato nulo en uno o más días, pero dicho mes cuenta con datos captados, se obtiene la media de ese mes en ese año, para todos los campos índices. En caso de no contar con datos en el mismo mes y año, se coloca 'null' como valor.

Esta técnica es llamada interpolación temporal por promedio mensual, y es respaldada por diferentes aplicaciones dentro de investigaciones, como en la de Hernández (2013) y López, et al. (2010).

**Guardar archivo en el sistema**

El formato resultante es guardado en un archivo con el formato CSV (.csv).

### 1.A.1.2. Paso 2: Rellenar lagunas temporales

A partir del nuevo archivo generado en el paso anterior, se llenan los saltos que hay en los registros, probablemente producidos en lapsos de tiempo en los que el sensor se encontraba totalmente apagado o inactivo por mantenimiento, el algoritmo que se utiliza para este proceso se define a continuación

#### 1. Cargar archivo

Se carga el archivo en formato CSV.

#### 2. Contabilizar número de registros

#### 3. Delimitar espacio de tiempo de la serie

Se obtiene la primera y la última fecha contenida en los registros.

#### 4. Crear serie de tiempo auxiliar

Se crea un conjunto de datos nuevo, solo para el campo de fecha, y se colocan todas las fechas que hay dentro del rango dado por los registros.

```
def rellenar_fechas_faltantes(archivo_csv, umbral_faltantes=100):  
    """rellenar_fechas_faltantes(archivo_csv):  
  
    This function receives the processed csv file and fills in the missing dates by calculating the min-max date range.  
  
    Input:  
    - archivo_csv (File): csv file with missing dates.  
    - umbral_faltantes (int): Maximum percentage of missing dates allowed.  
  
    Return:  
    | None  
    """>  
  
    df = pd.read_csv(archivo_csv, dayfirst=True)  
    df['FECHA'] = pd.to_datetime(df['FECHA'], errors='coerce')  
  
    # Contar número de registros en el Dataframe df  
    total_registros_original = df.shape[0]  
  
    fecha_min = df['FECHA'].min()  
    fecha_max = df['FECHA'].max()  
    total_fechas = pd.date_range(start=fecha_min, end=fecha_max, freq='D')  
  
    all_dates = pd.DataFrame({'FECHA': total_fechas})
```

Figura 4.8. Código para primeras 4 instrucciones de paso 2 (Fuente: elaboración propia)

## 5. Calcular lagunas temporales

Se calcula el porcentaje de entradas faltantes a partir de una comparación del número de registros original y el del conjunto de datos auxiliar.

- Si el porcentaje de registros faltantes es mayor al umbral permitido definido antes de comenzar el preprocesamiento, el archivo CSV es descartado y eliminado del sistema.

## 6. Obtener serie de tiempo sin lagunas

Se fusiona el conjunto de datos original con el que contiene únicamente las fechas, con el fin de rellenar las brechas en la serie de tiempo.

## 7. Normalizar dataset

Se rellenan las demás columnas dentro de los registros añadidos con 'null'.

## 8. Guardar cambios

Se sobrescribe el archivo CSV existente dentro del sistema.

```
df_completo = all_dates.merge(df, on='FECHA', how='left')
total_registros_generados = df_completo.shape[0]
fechas_faltantes = total_registros_generados - total_registros_original
porcentaje_faltante = (fechas_faltantes / total_registros_generados) * 100

## Si el porcentaje de fechas faltantes es mayor al umbral, se elimina el archivo y se detiene el proceso
if porcentaje_faltante > umbral_faltantes:
    faltantes_msg = f"Entradas faltantes en {archivo_csv}\nTotal de entradas faltantes: {fechas_faltantes}\nPorcentaje de entradas faltantes: {porcentaje_faltante:.2f}%"
    print(faltantes_msg)
    print(f"El porcentaje de fechas faltantes es mayor al umbral permitido ({umbral_faltantes}%)")
    os.remove(archivo_csv)
    print(f"Archivo descartado: {archivo_csv}")
    return
else:
    df_completo['MONTH'] = df_completo['FECHA'].dt.month
    df_completo['YEAR'] = df_completo['FECHA'].dt.year

    if not df_completo.empty:
        if 'ESTACION' in df.columns:
            constantes = ['ESTACION', 'NOMBRE', 'ESTADO', 'PRINCIPIO', 'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
        else:
            constantes = ['ESTACIÓN', 'NOMBRE', 'ESTADO', 'PRINCIPIO', 'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
        for const in constantes:
            df_completo[const] = df[const].iloc[0] if const in df.columns and not df[const].isna().all() else "Desconocida"
        for column in ['PRECIP', 'EVAP', 'THAX', 'THIN']:
            df_completo[column] = df_completo[column].astype(str).replace(nan, 'null')
            df_completo[column] = df_completo[column].astype(str).replace('nan', 'null')
        df_completo = df_completo.astype(str).replace('nan', 'null')
        df_completo.to_csv(archivo_csv, index=False, encoding='utf-8')
    log_msg = f"Entradas faltantes rellenas en {archivo_csv}\nTotal de entradas faltantes: {fechas_faltantes}\nPorcentaje de entradas faltantes: {porcentaje_faltante:.2f}%"
    print(log_msg)
```

Figura 4.9. Código para últimos 4 pasos del paso 2 (Fuente: elaboración propia)

### 1.A.1.3. Paso 3: Generar archivos concentrados

Este paso consiste en concatenar los archivos ya procesados, de tal forma que se tiene un registro histórico de todas las estaciones.

Para facilitar el acceso a información requerida, se concentrados a diferentes escalas: diario y mensual. Como ya se tienen los datos diarios de cada estación por separado, generar un solo archivo que contenga los datos de todas las estaciones consiste en concatenarlos, sin embargo, generar concentrados mensuales implica una mayor labor.

Para generar un registro histórico diario se siguieron los siguientes pasos:

#### 1. Cargar datos

Se carga la carpeta que contiene todos los archivos ya procesados.

#### 2. Crear concentrados diarios

Se crea un conjunto de datos nuevo, al cual se le añaden los datos que contienen cada uno de los archivos procesados.

#### 3. Guardar archivo de concentrados

Se guarda el dataset en un archivo CSV llamado 'CD.csv'

```
def generar_concentrados_diarios(directorio_entrada):
    """generar_concentrados_diarios(directorio_entrada):

    This function receives the directory of the raw files, locates the clean files folder and generates a single file containing all the daily summaries.

    Input:
        directorio_entrada (String): folder where the RAW files are located
    Return: None
    """
    directorio_limpios = os.path.join(directorio_entrada, "limpio")
    archivos_csv = [os.path.join(directorio_limpios, f) for f in os.listdir(directorio_limpios) if f.endswith(".csv")]
    df_final = pd.concat([pd.read_csv(f) for f in archivos_csv], ignore_index=True)
    for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
        df_final[column] = df_final[column].astype(str).replace(nan, 'null')
        df_final[column] = df_final[column].astype(str).replace('nan', 'null')
    archivo_final = os.path.join(directorio_limpios, 'CD.csv')
    df_final.to_csv(archivo_final, index=False, encoding='utf-8')
    print(f'Archivo final de concentrado diario guardado como {archivo_final}')
```

Figura 4.10. Código para paso 3 (Fuente: elaboración propia)

La obtención de los concentrados mensuales es una tarea más compleja que la anterior debido a que implica la realización de cálculos con los registros del conjunto de datos, agrupándolo por año y mes, para esto se definen los siguientes pasos, que se detallan a continuación:

### 1. definir ruta de carpeta de todos los archivos procesados.

```
def generar_concentrados_mensuales(directorio_entrada):
    """generar_concentrados_mensuales(directorio_entrada):

    (this function receives the directory of the raw files, then locates the processed files folder and generates monthly summaries for each station.

    Input:
    | directorio_entrada (String): folder where the original files are located
    Return:
    | None
    """
    directorio_limpios = os.path.join(directorio_entrada, "limpio")
    directorio_concentrado = os.path.join(directorio_entrada, "CM")
    os.makedirs(directorio_concentrado, exist_ok=True)
    archivos_csv = [os.path.join(directorio_limpios, f) for f in os.listdir(directorio_limpios) if f.endswith(".csv")]
```

Figura 4.11. Código para creación de datos mensuales (Fuente: elaboración propia)

### 2. Procesamiento en lote de archivos:

1. Se crea un nuevo dataset.
2. Se agrupa el conjunto de datos a procesar por Año y Mes, tal que:
  - a. El registro de precipitación sea la suma mensual de las precipitaciones.
  - b. El registro de evaporación sea la media mensual de evaporaciones.
  - c. El registro de temperatura máxima sea la temperatura más alta registrada en ese año y mes.
  - d. El registro de temperatura mínima sea la temperatura más baja registrada ese año y mes.

### 3. Adjuntar valores constantes

Se le adjuntan los valores que son constantes al nuevo dataset (es decir, estación, nombre, estado, municipio, condición de la estación, coordenadas geográficas y altitud).

### 4. Se colocan los valores ‘null’ al archivo.

### 5. Se guarda el archivo con el formato ‘CM\_{estacion}.csv’

Se repite el proceso hasta generar los concentrados mensuales de todas las estaciones que contenga el directorio especificado.

Dicho proceso queda definido en el código de la figura 4.12.

```
for archivo in archivos_csv:
    df = pd.read_csv(archivo)

    df_mensual = df.groupby(['YEAR', 'MONTH']).agg({
        'PRECIP': 'sum',
        'EVAP': 'mean',
        'TMAX': 'max',
        'TMIN': 'min'
    }).reset_index()

    if not df_mensual.empty:
        if 'ESTACION' in df.columns:
            constantes = ['ESTACION', 'NOMBRE', 'ESTADO', 'MUNICIPIO', 'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
        else:
            constantes = ['ESTACIÓN', 'NOMBRE', 'ESTADO', 'MUNICIPIO', 'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
        for const in constantes:
            df_mensual[const] = df[const].iloc[0] if const in df.columns and not df[const].isna().all() else "Desconocida"

        for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
            df_mensual[column] = df_mensual[column].astype(str).replace(nan, 'null')
            df_mensual[column] = df_mensual[column].astype(str).replace('nan', 'null')

        if 'ESTACION' in df.columns:
            archivo_concentrado = os.path.join(directorio_concentrado, f'CM_{df["ESTACION"].iloc[0]}.csv')
        else:
            archivo_concentrado = os.path.join(directorio_concentrado, f'CM_{df_mensual["ESTACIÓN"].iloc[0]}.csv')
        df_mensual.to_csv(archivo_concentrado, index=False, encoding='utf-8')
        print(f'Archivo de concentrado mensual guardado como {archivo_concentrado}')
```

Figura 4.12. Resto del código para la crear archivo mensual (Fuente: elaboración propia)

Una vez generados los archivos mensuales para cada estación, se sigue el mismo proceso utilizado para obtener el registro histórico diario, pero con los nuevos archivos.

## Paso 1. Cargar carpeta que contiene los concentrados mensuales

## Paso 2. Concatenar datos

Se crea un conjunto de datos nuevo, al cual se le concatenan todos los registros de los archivos dentro del directorio.

## Paso 3. Guardar archivo de concentrado

Se guarda el archivo concatenado en una carpeta nueva llamada FINAL, con el nombre 'FINAL.csv'.

```
def concatenar_concentrados(directorio_entrada):
    """concatenar_concentrados(directorio_entrada):

    This function receives the directory of the raw files, then locates the monthly summaries folder and concatenates them into a single .csv file.

    Input:
    | directorio_entrada (String): folder where the raw files are located
    Return:
    | None
    """
    directorio_concentrado = os.path.join(directorio_entrada, "CM")
    directorio_final = os.path.join(directorio_entrada, "FINAL")

    os.makedirs(directorio_final, exist_ok=True)

    archivos_csv = [os.path.join(directorio_concentrado, f) for f in os.listdir(directorio_concentrado) if f.endswith(".csv")]

    df_final = pd.concat([pd.read_csv(f) for f in archivos_csv], ignore_index=True)
    for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
        df_final[column] = df_final[column].astype(str).replace(nan, 'null')
        df_final[column] = df_final[column].astype(str).replace('nan', 'null')
    archivo_final = os.path.join(directorio_final, 'FINAL.csv')
    df_final.to_csv(archivo_final, index=False, encoding='utf-8')
    print(f'Archivo final concatenado guardado como {archivo_final}')
```

Figura 4.13. Código para concatenar concentrados mensuales (Fuente: Elaboración propia)

### 4.2.3 Etapa 3 Transferencia de archivos

Una vez terminada la fase de procesamiento de los datos, es posible exportar en lote de forma directa los archivos procesados hacia cualquier aplicación de análisis que soporte el formato CSV. En este caso se utilizará PowerBI para realizar la carga de los datos nuevos. Para cargar los datos en formato CSV, se selecciona la opción “Obtener datos de otros orígenes” de la pantalla principal de Power BI:



Figura 4.14. Pantalla de inicio PowerBI (Fuente: Elaboración propia)

Se selecciona como fuente “Texto o CSV” y se hace clic en la opción Conectar.

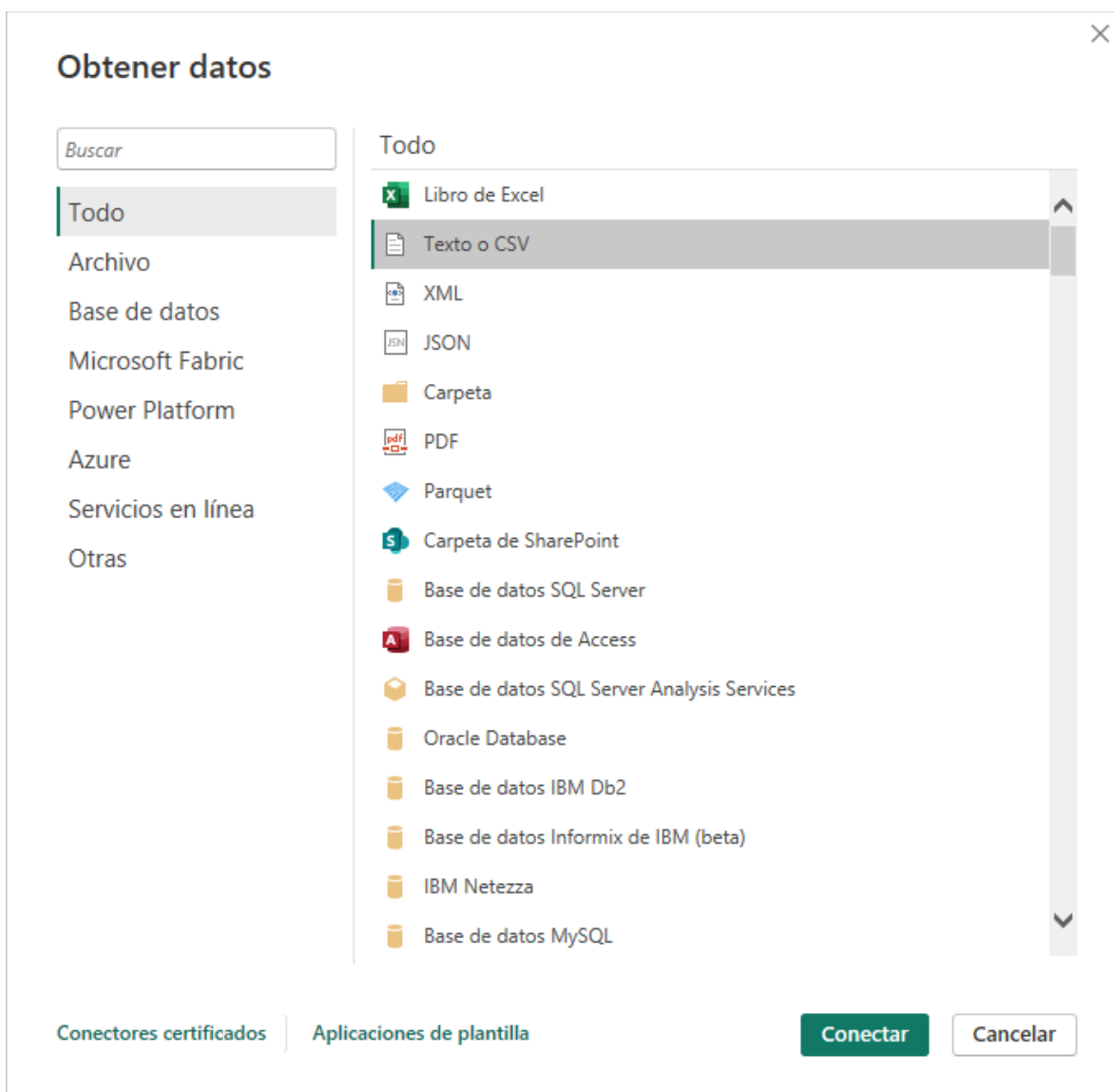


Figura 4.15. Pantalla de selección de fuentes de datos PowerBI (Fuente: Elaboración propia)

Aparecerá una ventana en la cual se permite elegir uno o varios archivos en formato CSV, si no se cambió de lugar ningún archivo generado previamente, el archivo con los datos históricos diarios de todas las estaciones debería de estar en la ruta seleccionada en la herramienta de limpieza de datos, dentro de la carpeta “limpios”, nombrado CD.csv.

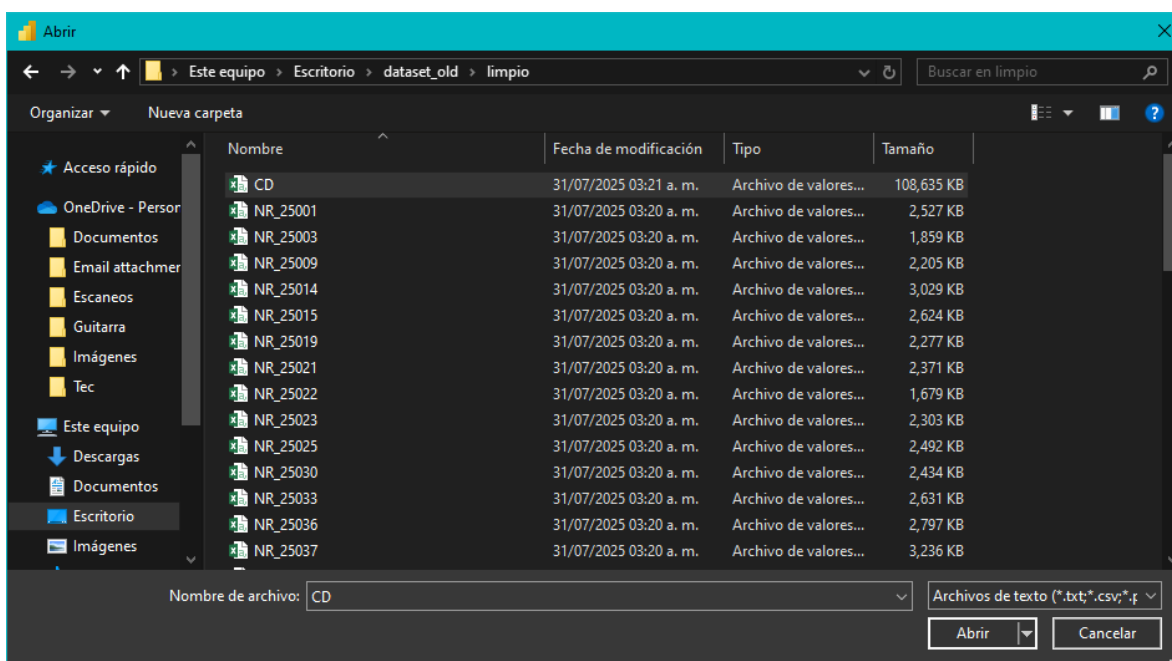


Figura 4.16. Seleccionando archivo a cargar (Fuente: elaboración propia)

Una vez se pulse el botón “Abrir”, se presentará otra pantalla, con una muestra pequeña de los datos que contiene el archivo que se desea cargar, en este caso, como no es requerida la transformación del archivo, se verifica que dentro de las opciones el Origen de archivo esté en UTF-8 y que el delimitador de cada valor sea la coma.

CD.csv

Origen de archivo

65001: Unicode (UTF-8)

Delimitador

Coma

Detección del tipo de datos

Basado en las primeras 200 filas

| FECHA      | PRECIP | LVAP | TMAX | TMIN | ESTACIÓN | NOMBRE   | ESTADO  | MUNICIPIO   | SITUACIÓN | LATITUD     | LONGITUD     |
|------------|--------|------|------|------|----------|----------|---------|-------------|-----------|-------------|--------------|
| 01/01/1961 | 0      | null | 25   | 13   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 02/01/1961 | 0      | null | 27.5 | 14   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 03/01/1961 | 0      | null | 31   | 11   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 04/01/1961 | 0      | null | 32   | 13   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 05/01/1961 | 0      | null | 29   | 14   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 06/01/1961 | 0      | null | 24   | 14.5 | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 07/01/1961 | 0      | null | 28.5 | 10   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 08/01/1961 | 0      | null | 32   | 12   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 09/01/1961 | 0      | null | 33.5 | 11   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 10/01/1961 | 0      | null | 32.5 | 9.5  | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 11/01/1961 | 0      | null | 30   | 9    | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 12/01/1961 | 0      | null | 33   | 9.5  | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 13/01/1961 | 0      | null | 32   | 10   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 14/01/1961 | 0      | null | 31   | 10   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 15/01/1961 | 0      | null | 32   | 8.5  | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 16/01/1961 | 0      | null | 31   | 11   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 17/01/1961 | 0.1    | null | 28   | 14.5 | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 18/01/1961 | 0.4    | null | 37   | 17   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 19/01/1961 | 13.5   | null | 23   | 19.5 | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |
| 20/01/1961 | 24     | null | 23   | 20   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.09722222 | -106.6666666 |

Extraer tabla mediante ejemplos

Cargar

Transformar datos

Cancelar

Figura 4.17. Seleccionando archivo a cargar (Fuente: Elaboración propia)

Una vez cargados los datos, PowerBI redirigirá a vista de informe, en donde el archivo cargado será mostrado del lado derecho de la pantalla.

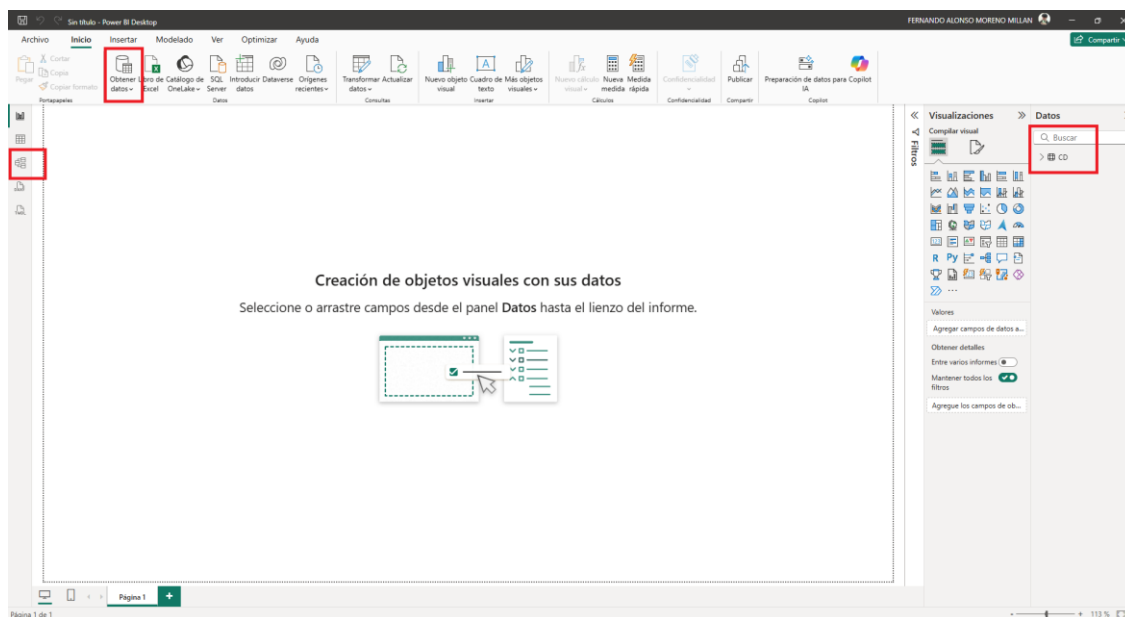


Figura 4.18. Vista de informe en blanco (Fuente: Elaboración propia)

El elemento señalado por un cuadro rojo en la parte superior de la figura 4.18 permite cargar más fuentes de datos, se repetirá el proceso, ahora para cargar el archivo con los datos históricos mensuales, localizado en la carpeta FINAL generada durante la transformación de los datos.

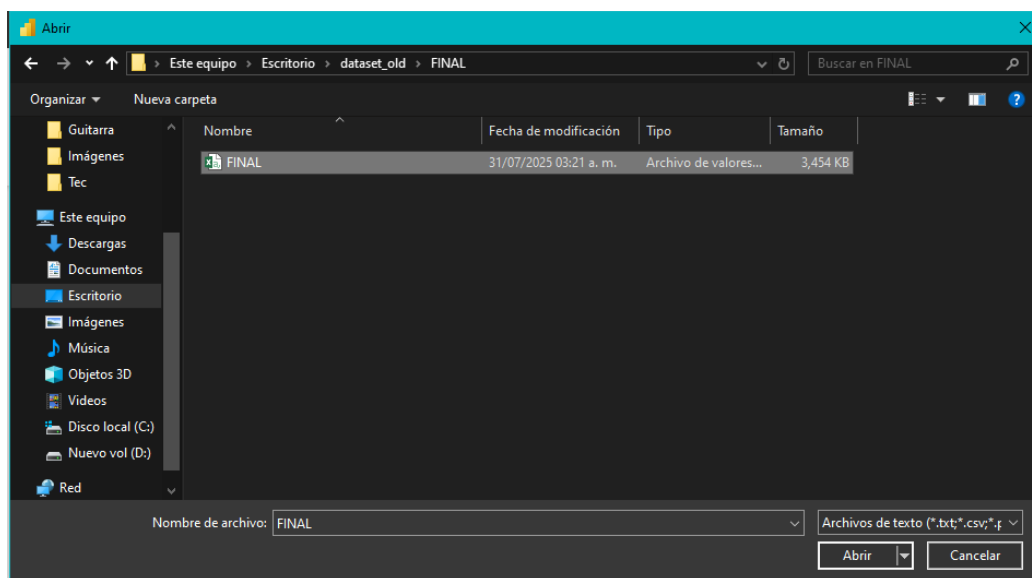


Figura 4.19. Cargar archivo de datos históricos mensuales (Fuente: elaboración propia)

Al hacer clic en la parte señalada a la izquierda en la figura 4.18 PowerBI desplegará la vista de modelo. Aquí nos aparecerá la información de lo que se haya cargado, en forma de entidades diferentes.

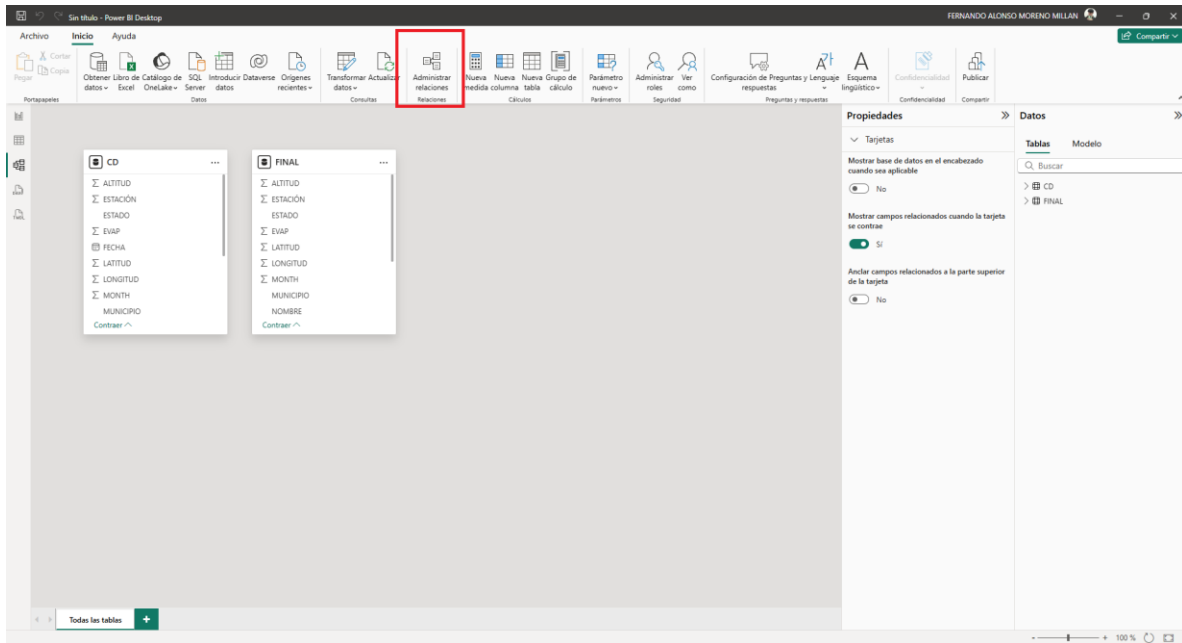


Figura 4.20. Vista de modelo PowerBI (Fuente: elaboración propia)

Es posible relacionar diferentes datos por medio de atributos en común, para hacerlo, es necesario hacer clic en el botón “Administrar relaciones” señalado en rojo en la figura 4.20.

## Administrar relaciones



+ Nueva relación ⚡ Detección automática ✎ Editar 🗑 Eliminar ☰ Filtro ▾

| <input type="checkbox"/> | Desde: tabla (columna) | ↑ | Relación | A: tabla (columna) | Estado |
|--------------------------|------------------------|---|----------|--------------------|--------|
|--------------------------|------------------------|---|----------|--------------------|--------|

Aún no se ha definido ninguna relación.

Close

Figura 4.21. Ventana emergente administrar relaciones (Fuente: elaboración propia)

Dentro de la pantalla presentada en la figura 4.21 se pueden ver las relaciones existentes entre las diferentes fuentes de datos cargadas e información de estas.

←

Nueva relación

×

Permite seleccionar tablas y columnas relacionadas.

Desde la tabla

CD

| IPIO | NOMBRE     | PRECIP | SITUACIÓN | TMAX | TMIN | YEAR |
|------|------------|--------|-----------|------|------|------|
| LÁN  | EL QUEMADO | 0      | OPERANDO  | 35   | 23   | 1979 |
| LÁN  | EL QUEMADO | 0      | OPERANDO  | 35   | 23   | 1980 |
| LÁN  | EL QUEMADO | 0      | OPERANDO  | 35   | 23   | 1980 |

A la tabla

FINAL

| IPIO | NOMBRE         | PRECIP | SITUACIÓN | TMAX | TMIN | YEAR |
|------|----------------|--------|-----------|------|------|------|
| LÁN  | MAZATLAN (...) | 0      | OPERANDO  | null | null | 1933 |
| LÁN  | MAZATLAN (...) | 0      | OPERANDO  | null | null | 1933 |
| LÁN  | MAZATLAN (...) | 0      | OPERANDO  | null | null | 1933 |

Cardinality

Varios a varios (\*:\*)

Dirección de filtro cruzado

Ambas

☒ Activar esta relación
 ☒ Aplicar filtro de seguridad en ambas direcciones

☐ Asumir integridad referencial

⚠

Esta relación tiene una cardinalidad de muchos para muchos. Esto solo debe usar si se espera que ninguna columna (YEAR y YEAR) contenga valores únicos y que se entienda el comportamiento significativamente diferente de las relaciones de muchos a muchos.
 [Más información](#)

Guardar

Cancelar

Figura 4.22. Creando relación (Fuente: elaboración propia)

Una vez relacionadas las fuentes de datos, es posible generar informes gráficos personalizados manejando la información de las diferentes fuentes de datos (véase Anexo B).

#### **4.2.4 Etapa 4: Análisis de datos**

Una vez cargados los datos en PowerBI, es posible llevar a cabo todas las tareas de análisis y explotación de datos que esta herramienta proporciona. Se pueden crear medidas personalizadas para realizar diferentes cálculos, se pueden programar funciones condicionales y crear vistas de informe para exponer la información de forma clara mediante medios gráficos.

Considerando los índices que proporcionan las estaciones de CONAGUA, es posible realizar análisis hidrológicos y térmicos.

##### **4.2.4.1 Análisis hidrológico**

La realización de un análisis hidrológico basado en la utilización de parámetros como la precipitación mínima, máxima y media es válida, como lo es demostrado dentro del análisis de Luque, et al. (2022). En el que se utilizaron estos parámetros para evaluar riesgos geológicos en la región de la costa de Granada.

##### **4.2.4.2 Análisis térmico**

Realizar un análisis térmico puede probarse útil para ver los efectos del cambio climático a largo plazo, como se ve demostrado en el trabajo de Del Saltos (2014). En el cual, mediante series de tiempo térmicas comprueba la existencia de variación de temperaturas provocadas por el cambio climático

## 5 Pruebas y Resultados

En el presente capítulo, se desarrolló un manual de usuario para la aplicación de escritorio desarrollada con el fin de facilitar el uso de la herramienta desarrollada con la metodología establecida (véase Anexo A para el código fuente).

También se ponen a prueba los datos generados, por medio de la comparación información durante años con ocurrencia de fenómenos climatológicos inusuales (inundaciones o sequías) contra años meteorológicamente normales. Estos eventos climatológicos adversos serán comparados con el año 2020, que, de acuerdo con datos recopilados por *Conselva* (2025), fue la última ocasión en que Sinaloa permaneció sin zonas con algún nivel de sequía.

### 5.1 Etapa de extracción de los archivos de datos de CONAGUA

Para esta etapa, se diseñó una interfaz gráfica para facilitar al usuario la extracción de los y transformación de los datos, utiliza técnicas de data scraping, para, esto se implementa la función *datascraper.py* (Anexo A.1).

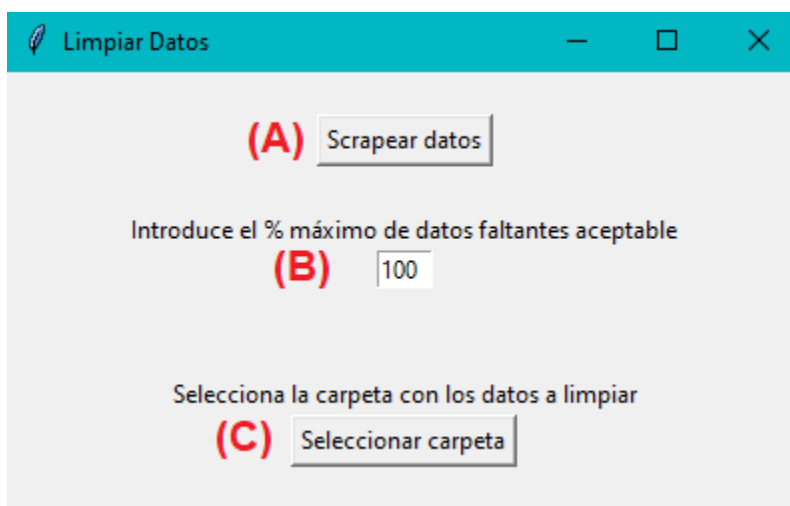


Figura 5.1. Interfaz de Usuario ETL. (Fuente: elaboración propia)

### 5.1.1 Configurar porcentaje de datos faltantes aceptable

Antes de comenzar el proceso de extracción de datos, es importante mirar al botón (B) de la figura 5.1, puesto que es posible configurar un umbral de porcentaje máximo de datos faltantes permitidos, al ingresar un porcentaje dentro de la caja de texto, se vuelve posible ajustar el máximo de datos faltantes aceptado para los archivos a procesar. Su función radica en ser un filtro, si es que se desean utilizar conjuntos de datos más completos, es decir, con menos lagunas temporales.

### 5.1.2 Extraer datos

En la interfaz representada en la figura 5.1, el botón (A) llama a la clase ‘datascraper.py’, mandándole un array de cadenas de texto que contienen todas las estaciones operativas para el estado de Sinaloa.

Al pulsar el botón de Extraer Datos, se despliega una ventana emergente que permite al usuario seleccionar una carpeta de su computadora:

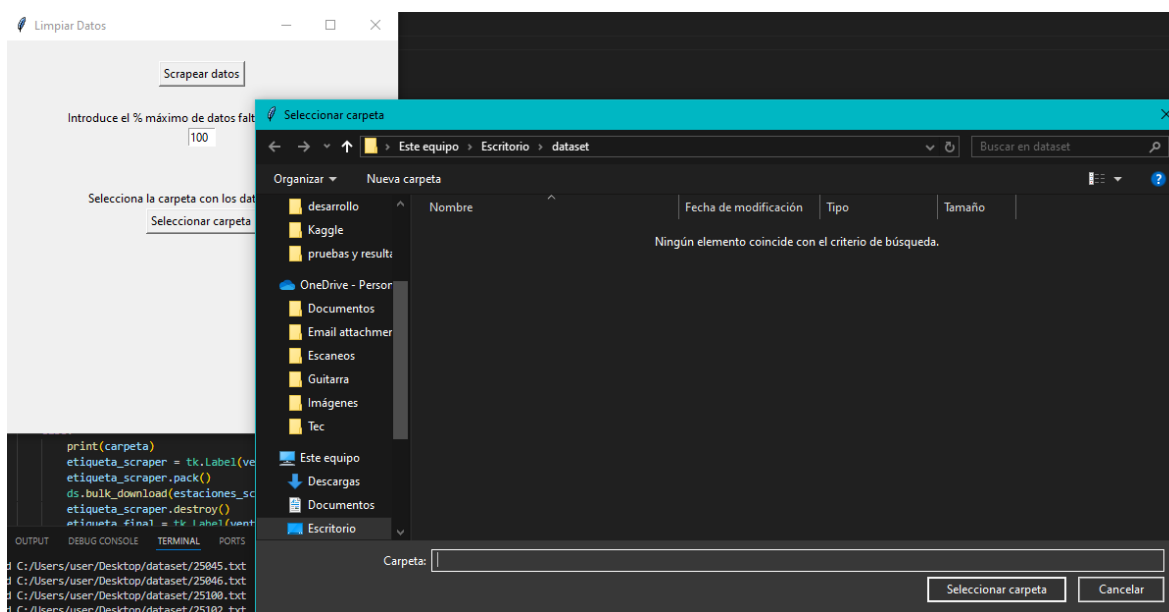


Figura 5.2. Ventana emergente: Seleccionar carpeta (Fuente: elaboración propia)

Una vez el usuario elija la carpeta en la que se extraerán los datos y pulse el botón “Seleccionar carpeta” de la ventana emergente, iniciará el proceso de scraping, indicado en la consola de Python.

```
PS C:\Users\user> python -u "c:\Users\user\Desktop\code_old\GUI.py"
C:/Users/user/Desktop/dataset
Downloaded C:/Users/user/Desktop/dataset/25001.txt
Downloaded C:/Users/user/Desktop/dataset/25003.txt
Downloaded C:/Users/user/Desktop/dataset/25009.txt
Downloaded C:/Users/user/Desktop/dataset/25014.txt
Downloaded C:/Users/user/Desktop/dataset/25015.txt
Downloaded C:/Users/user/Desktop/dataset/25019.txt
Downloaded C:/Users/user/Desktop/dataset/25021.txt
Downloaded C:/Users/user/Desktop/dataset/25022.txt
Downloaded C:/Users/user/Desktop/dataset/25023.txt
Downloaded C:/Users/user/Desktop/dataset/25025.txt
Downloaded C:/Users/user/Desktop/dataset/25050.txt
Downloaded C:/Users/user/Desktop/dataset/25062.txt
Downloaded C:/Users/user/Desktop/dataset/25064.txt
Downloaded C:/Users/user/Desktop/dataset/25074.txt
Downloaded C:/Users/user/Desktop/dataset/25077.txt
Downloaded C:/Users/user/Desktop/dataset/25078.txt
Downloaded C:/Users/user/Desktop/dataset/25080.txt
Downloaded C:/Users/user/Desktop/dataset/25081.txt
```

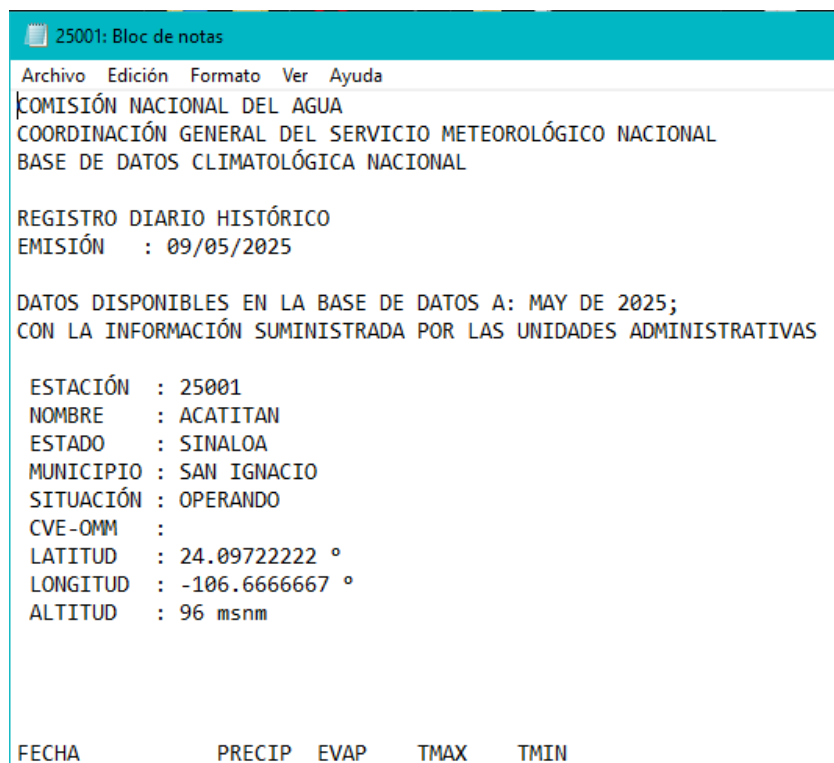
Figura 5.3. Log de extracción de datos (Fuente: Elaboración propia)

El resultado de llevar a cabo la extracción de los datos es, en este caso, 50 archivos en formato de texto plano (.txt).



Figura 5.4. Archivos extraídos (Fuente: Elaboración propia)

Los ficheros de texto extraídos cuentan con el siguiente formato:



25001: Bloc de notas

Archivo Edición Formato Ver Ayuda

COMISIÓN NACIONAL DEL AGUA  
COORDINACIÓN GENERAL DEL SERVICIO METEOROLÓGICO NACIONAL  
BASE DE DATOS CLIMATOLÓGICA NACIONAL

REGISTRO DIARIO HISTÓRICO  
EMISIÓN : 09/05/2025

DATOS DISPONIBLES EN LA BASE DE DATOS A: MAY DE 2025;  
CON LA INFORMACIÓN SUMINISTRADA POR LAS UNIDADES ADMINISTRATIVAS

ESTACIÓN : 25001  
NOMBRE : ACATITAN  
ESTADO : SINALOA  
MUNICIPIO : SAN IGNACIO  
SITUACIÓN : OPERANDO  
CVE-OMM :  
LATITUD : 24.09722222 °  
LONGITUD : -106.6666667 °  
ALTITUD : 96 msnm

FECHA                      PRECIP   EVAP        TMAX        TMIN

Figura 5.5. Ejemplo de la cabecera de un archivo de texto (Fuente: Elaboración propia)

## 5.2 Transformación de archivos

Una vez terminada la fase de extracción, se pulsa el botón “Seleccionar Carpeta” mostrado en la figura 5.1 (C), que despliega una ventana emergente igual a la representada en la figura 5.2, y se selecciona la carpeta que contiene los archivos de texto extraídos.

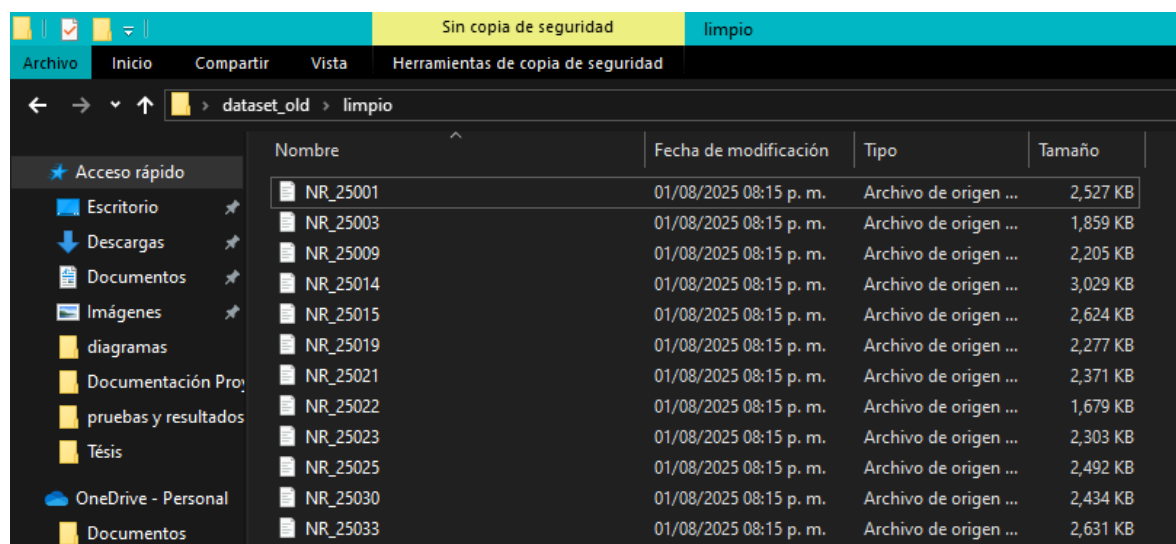
### 5.2.1 Limpiar archivos

Por cada archivo que encuentre generará un registro individual en la consola del desarrollador.

```
PS C:\Users\user> python -u "c:\Users\user\Desktop\Residencias Profesionales\RP\code\GUI.py"
C:/Users/user/Desktop/dataset
Procesando archivo C:/Users/user/Desktop/dataset\25001.txt
Número de líneas: 21395
Estación detectada: 25001
Archivo CSV guardado como C:/Users/user/Desktop/dataset\limpio\NR_25001.csv
Entradas faltantes rellenadas en C:/Users/user/Desktop/dataset\limpio\NR_25001.csv
Total de entradas faltantes: 2065
Porcentaje de entradas faltantes: 8.81%
```

Figura 5.6. Registro de consola mostrando archivo siendo procesado (Fuente: Elaboración propia)

Dicho registro quedará guardado en un subdirectorio creado automáticamente, llamado “limpio” con el formato NR\_{número de estación}.csv.



| Nombre   | Fecha de modificación  | Tipo                  | Tamaño   |
|----------|------------------------|-----------------------|----------|
| NR_25001 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,527 KB |
| NR_25003 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 1,859 KB |
| NR_25009 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,205 KB |
| NR_25014 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 3,029 KB |
| NR_25015 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,624 KB |
| NR_25019 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,277 KB |
| NR_25021 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,371 KB |
| NR_25022 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 1,679 KB |
| NR_25023 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,303 KB |
| NR_25025 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,492 KB |
| NR_25030 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,434 KB |
| NR_25033 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,631 KB |

Figura 5.7. Archivos limpios en el directorio de salida (Fuente: Elaboración propia)

Los archivos procesados tienen la siguiente estructura:

|    | A          | B         | C    | D    | E    | F        | G        | H       | I           | J         | K         | L          | M       | N     | O    |
|----|------------|-----------|------|------|------|----------|----------|---------|-------------|-----------|-----------|------------|---------|-------|------|
| 1  | FECHA      | PRECIP    | EVAP | TMAX | TMIN | ESTACION | NOMBRE   | ESTADO  | MUNICIPIO   | SITUACION | LATITUD   | LONGITUD   | ALTITUD | MONTH | YEAR |
| 2  | 01/01/1961 | 0 null    |      | 25   | 13   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 3  | 02/01/1961 | 0 null    |      | 27.5 | 14   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 4  | 03/01/1961 | 0 null    |      | 31   | 11   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 5  | 04/01/1961 | 0 null    |      | 32   | 13   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 6  | 05/01/1961 | 0 null    |      | 29   | 14   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 7  | 06/01/1961 | 0 null    |      | 24   | 14.5 | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 8  | 07/01/1961 | 0 null    |      | 28.5 | 10   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 9  | 08/01/1961 | 0 null    |      | 32   | 12   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 10 | 09/01/1961 | 0 null    |      | 33.5 | 11   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 11 | 10/01/1961 | 0 null    |      | 32.5 | 9.5  | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 12 | 11/01/1961 | 0 null    |      | 30   | 9    | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 13 | 12/01/1961 | 0 null    |      | 33   | 9.5  | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 14 | 13/01/1961 | 0 null    |      | 32   | 10   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 15 | 14/01/1961 | 0 null    |      | 31   | 10   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 16 | 15/01/1961 | 0 null    |      | 32   | 8.5  | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 17 | 16/01/1961 | 0 null    |      | 31   | 11   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 18 | 17/01/1961 | 0.1 null  |      | 28   | 14.5 | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 19 | 18/01/1961 | 0.4 null  |      | 37   | 17   | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |
| 20 | 19/01/1961 | 13.5 null |      | 23   | 19.5 | 25001    | ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.66667 | 96      | 1     | 1961 |

Figura 5.8. Primeros 20 registros de NR\_25001.csv (Fuente: Elaboración propia)

Como se puede observar, contienen datos que sirven como índices de medición (fecha, evaporación, evaporación, tmax, tmin) así como datos de referencia que servirán al momento de llevar a cabo cualquier tipo de análisis utilizando software de explotación de datos (número de estación, nombre, estado, municipio, ubicación geográfica, mes, año).

## 5.2.2 Generar registros de datos mensuales

Una vez obtenidos los archivos limpios referenciados en la figura 5.7, se procesan de nuevo, para obtener estadísticas mensuales de interés ya concentradas: la suma de la precipitación, la media de evaporación, así como la temperatura máxima más alta y mínima más baja. Los registros de datos mensuales son separados por su estación correspondiente, y guardados dentro de una carpeta creada automáticamente llamada “CM” (Concentrados Mensuales) dentro de la carpeta que contiene los archivos en formato de texto.

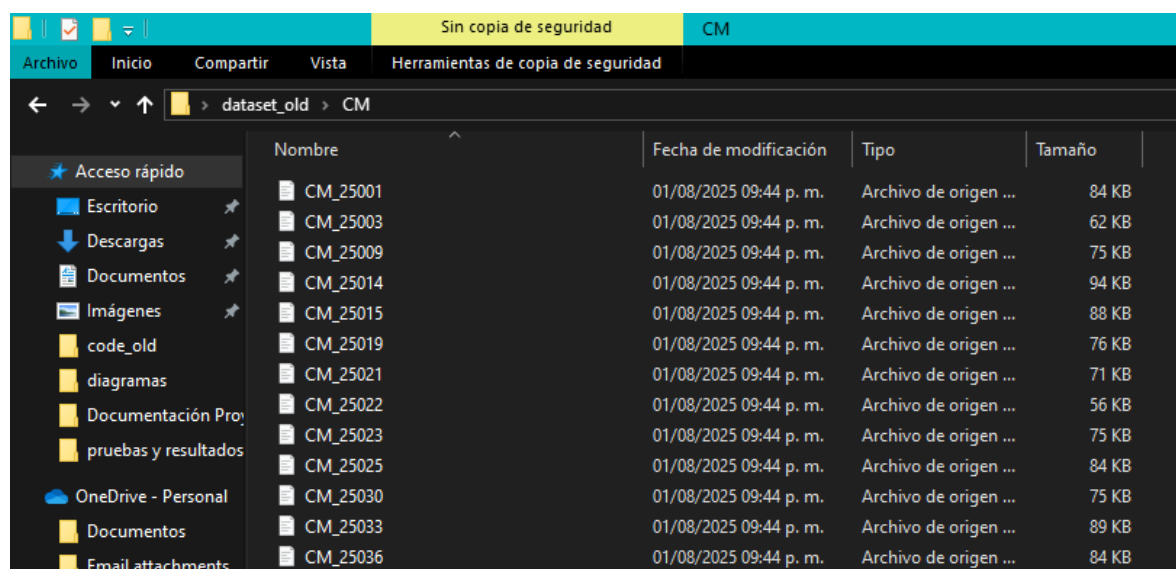


Figura 5.9. Carpeta que contiene los registros mensuales (Fuente: Elaboración propia)

Los archivos de concentrado mensual tienen la siguiente estructura básica:

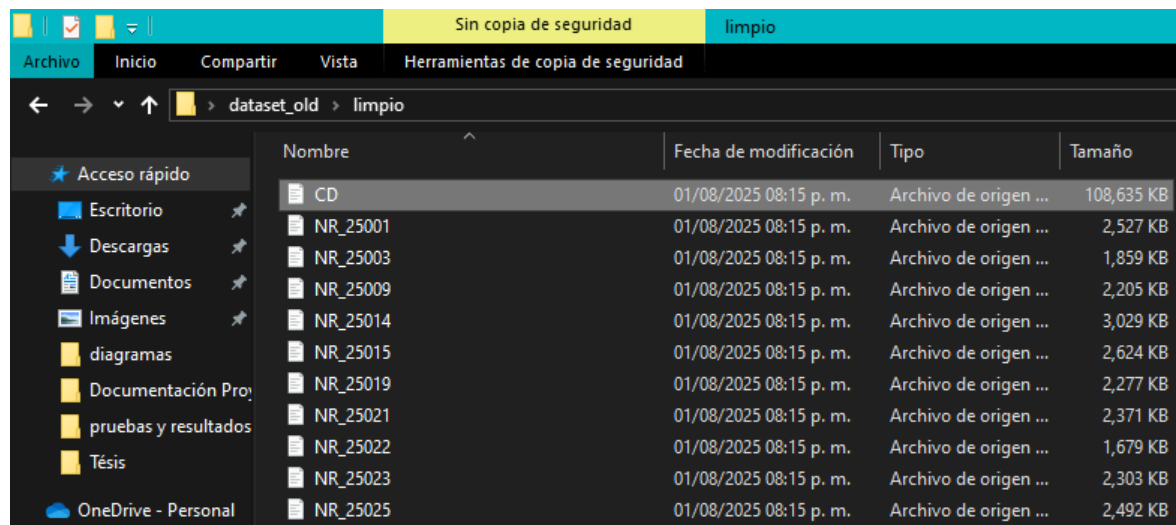
|    | A    | B     | C      | D         | E    | F    | G              | H       | I           | J         | K         | L           | M        | N       |
|----|------|-------|--------|-----------|------|------|----------------|---------|-------------|-----------|-----------|-------------|----------|---------|
| 1  | YEAR | MONTH | PRECIP | EVAP      | TMAX | TMIN | ESTACION       | NOMBRE  | ESTADO      | MUNICIPIO | SITUACION | LATITUD     | LONGITUD | ALTITUD |
| 2  | 1961 | 1     | 97.1   | null      | 37   | 8.5  | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 3  | 1961 | 2     | 0      | null      | 37.5 | 6    | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 4  | 1961 | 3     | 0      | 5.4633333 | 39   | 6    | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 5  | 1961 | 4     | 9.6    | 7.062069  | 42.5 | 9    | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 6  | 1961 | 5     | 0      | 8.3612903 | 41.5 | 13   | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 7  | 1961 | 6     | 44.1   | 7.2833333 | 40.5 | 19.5 | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 8  | 1961 | 7     | 301.2  | 4.216129  | 38   | 22.5 | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 9  | 1961 | 8     | 145.7  | 4.4290323 | 37.5 | 22   | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 10 | 1961 | 9     | 169.9  | 3.77      | 36   | 21   | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 11 | 1961 | 10    | 38.6   | 3.8354839 | 38.5 | 17   | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 12 | 1961 | 11    | 0      | 3.3833333 | 36.5 | 8.5  | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 13 | 1961 | 12    | 106.91 | 2.5032258 | 35   | 8    | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 14 | 1962 | 1     | 5      | 2.8451613 | 34.5 | 4.5  | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 15 | 1962 | 2     | 0      | 4.0285714 | 34.5 | 6.5  | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 16 | 1962 | 3     | 1.4    | 5.8677419 | 37.5 | 4.5  | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 17 | 1962 | 4     | 0      | 7.9       | 42   | 5.5  | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |
| 18 | 1962 | 5     | 0      | 8.7645161 | 41.5 | 14   | 25001 ACATITAN | SINALOA | SAN IGNACIO | OPERANDO  | 24.097222 | -106.666667 | 96       |         |

Figura 5.10. Primeros 20 registros de CM\_25001.csv (Fuente: Elaboración propia)

Es similar a los datos diarios, pero ya no contiene un campo de FECHA, y los valores de las columnas índices tienen representaciones diferentes.

### 5.2.3 Generar concentrado de registros diarios

Para manejar todos los datos de todas las estaciones al mismo tiempo, y reducir el tiempo que toma la carga de cada uno de los archivos en un software de análisis, se genera un solo archivo que contiene la información en escala diaria de todas las estaciones, esto se lleva a cabo concatenando todos los archivos contenidos en la carpeta de la figura 5.7, y guardados en un solo fichero llamado CD.csv, el cual es guardado junto a los demás archivos diarios.



| Nombre   | Fecha de modificación  | Tipo                  | Tamaño     |
|----------|------------------------|-----------------------|------------|
| CD       | 01/08/2025 08:15 p. m. | Archivo de origen ... | 108,635 KB |
| NR_25001 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,527 KB   |
| NR_25003 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 1,859 KB   |
| NR_25009 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,205 KB   |
| NR_25014 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 3,029 KB   |
| NR_25015 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,624 KB   |
| NR_25019 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,277 KB   |
| NR_25021 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,371 KB   |
| NR_25022 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 1,679 KB   |
| NR_25023 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,303 KB   |
| NR_25025 | 01/08/2025 08:15 p. m. | Archivo de origen ... | 2,492 KB   |

Figura 5.11. Archivo de Concentrados Diarios (CD) (Fuente: Elaboración propia)

### 5.2.4 Generar concentrado de registros mensuales

Con el mismo objetivo de la creación de un solo archivo que contenga la información diaria de todas las estaciones, se crea un archivo concentrando la información mensual de todas las estaciones, usando la información de los archivos ilustrados en la figura 5.9. El archivo es guardado en una carpeta de nombre FINAL, creada automáticamente en el directorio que contiene los archivos de texto usados para generar los CSV diarios.

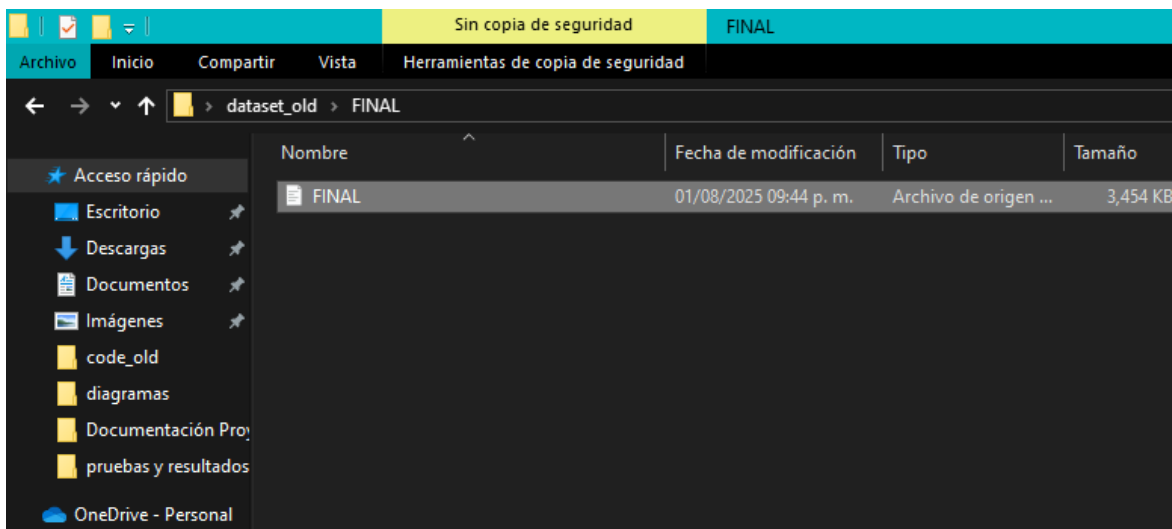


Figura 5.12. Archivo de concentrado mensual (FINAL.csv) (Fuente: Elaboración propia)

El archivo FINAL.CSV contiene la misma estructura básica vista en la figura 5.10, sin embargo, los datos que incluye son de todas las estaciones dentro del área de estudio.

### 5.3 Informes gráficos de información hidrometeorológica

PowerBI facilita la interpretación de la información por medio de la creación de material gráfico a partir de los datos, por medio de informes, en los que se pueden dejar botones, campos seleccionables y datos relevantes para facilitarle incluso a personal no experimentado la obtención y manejo de los datos.

Se utilizó este software de análisis de datos para obtener información estadística de forma rápida, elaborando diferentes vistas de informe (revisar Anexo B para más información).

#### **5.4 Análisis de los datos obtenidos**

Se utilizarán los diferentes índices de las series de tiempo obtenidas durante el desarrollo de esta metodología para realizar análisis estadísticos de temperatura y precipitación. A continuación, se propone como caso de estudio general la comparación entre dos periodos de tiempo largos: 1974-1999 y 2000-2025, con el fin de realizar un análisis comparativo entre dos periodos de tiempo extensos. Además, se proponen dos casos de estudio específicos: el huracán Manuel de 2013 y la sequía hidrológica del 2024. Estos dos eventos serán comparados con el año 2020, en el cual, de acuerdo con Conselva (2024) e informes del SMN (2020), fue el año en que ningún municipio dentro del estado de Sinaloa presentó algún tipo de sequía hidrológica, con el fin de identificar la manera en que los fenómenos meteorológicos afectan las diversas variables bioclimáticas.

### 5.4.1 Caso de estudio general: comparativa entre periodos 1974-1999 y 2000-2025

Para nuestra primera serie tenemos los siguientes datos hidrológicos y térmicos:

- **Suma de precipitación:** 684,401.81mm
- **Suma de evaporación:** 1,118,729.04mm
- **Promedio de precipitación:** 1.98mm
- **Promedio de evaporación:** 5.28mm
- **Mediana de precipitación:** 0.00mm
- **Mediana de evaporación:** 5.00mm
- **Varianza de precipitación:** 79.25mm
- **Varianza de evaporación:** 6.34mm
- **Desviación estándar de precipitación:** 8.9mm
- **Desviación estándar de evaporación:** 2.52mm



Figura 5.13. Informe hidrológico diario 1974-1999 (Fuente: Elaboración propia)

- **Temperatura máxima más alta en serie:** 55°C
- **Temperatura mínima más alta en serie:** 37°C
- **Temperatura máxima más baja en serie:** 6°C
- **Temperatura mínima más baja en serie:** -9°C
- **Mediana de temperaturas máximas:** 33°C

- **Mediana de temperaturas mínimas:** 16.5°C
- **Desviación estándar de temperaturas máxima:** 4.78°C
- **Desviación estándar de temperaturas mínima:** 6.13°C
- **Varianza de temperatura máxima:** 22.84°C
- **Varianza de temperatura mínima:** 37.53°C

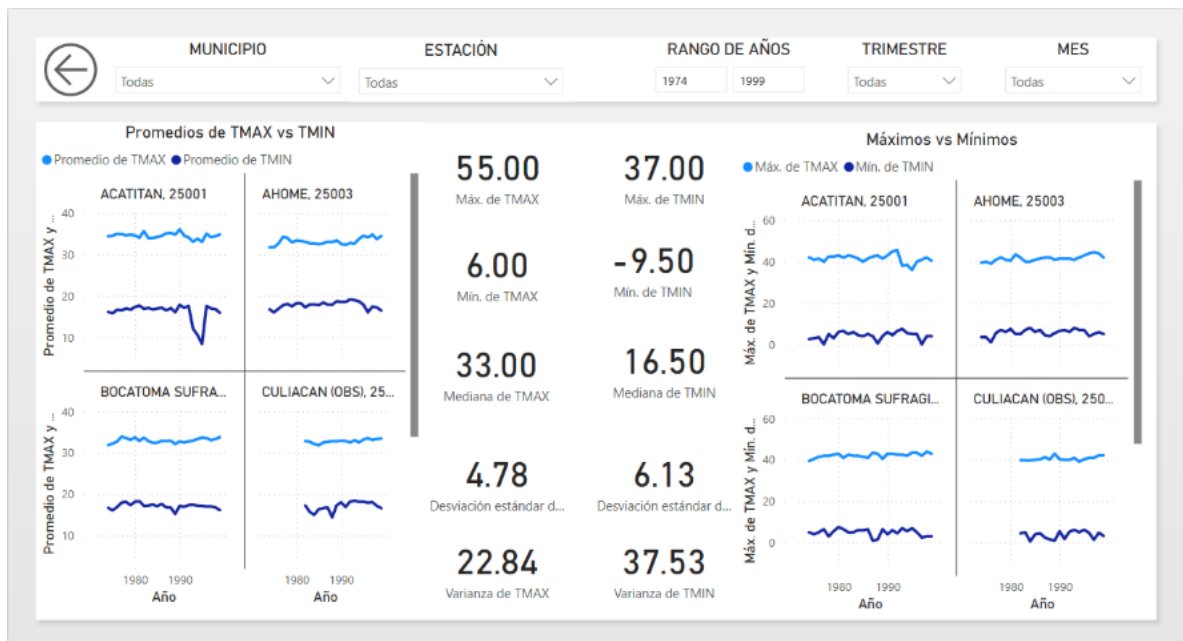


Figura 5.14. Informe térmico diario 1974-1999 (Fuente: Elaboración propia)

Mientras que para nuestra serie más reciente tenemos la siguiente información:

- **Suma de precipitación:** 690,345.154mm
- **Suma de evaporación:** 1,123,627.63mm
- **Promedio de precipitación:** 2.04mm
- **Promedio de evaporación:** 5.28mm
- **Mediana de precipitación:** 0.00mm
- **Mediana de evaporación:** 5.00mm
- **Varianza de precipitación:** 86.52mm
- **Varianza de evaporación:** 5.93mm
- **Desviación estándar de precipitación:** 9.3mm
- **Desviación estándar de evaporación:** 2.44mm

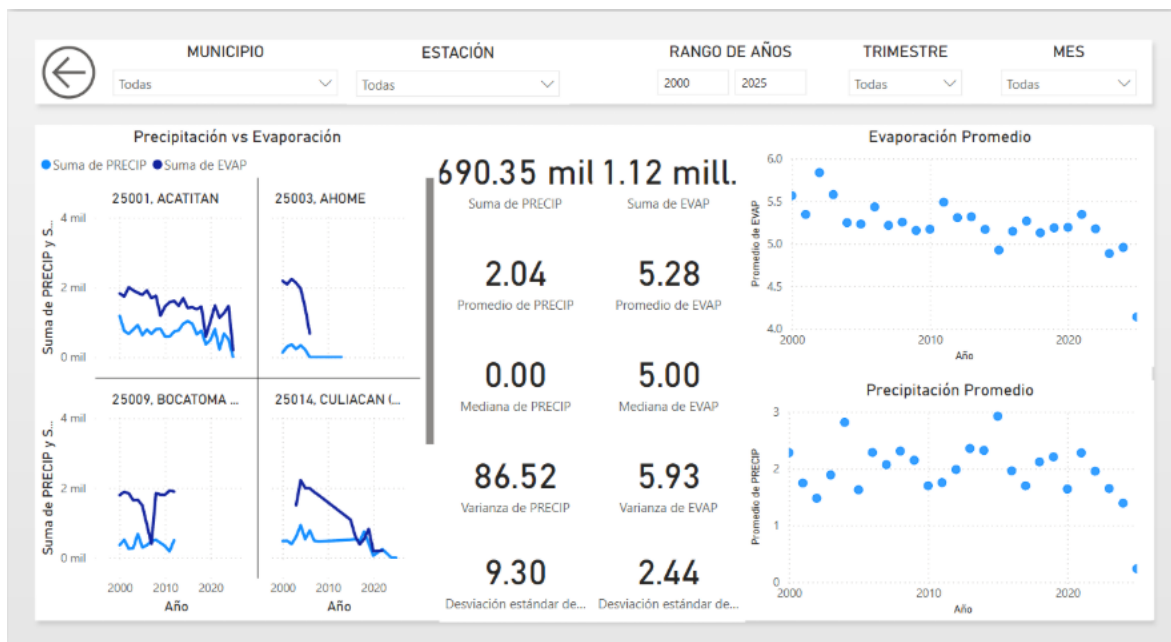


Figura 5.15. Informe hidrológico diario 2000-2025 (Fuente: Elaboración propia)

- **Temperatura máxima más alta en serie: 48.4°C**
- **Temperatura mínima más alta en serie: 35.4°C**
- **Temperatura máxima más baja en serie: 6°C**
- **Temperatura mínima más baja en serie: -10°C**
- **Mediana de temperaturas máximas: 34°C**
- **Mediana de temperaturas mínimas: 17°C**
- **Desviación estándar de temperaturas máxima: 4.62°C**
- **Desviación estándar de temperaturas mínima: 6.18°C**
- **Varianza de temperatura máxima: 21.33°C**
- **Varianza de temperatura mínima: 38.23°C**

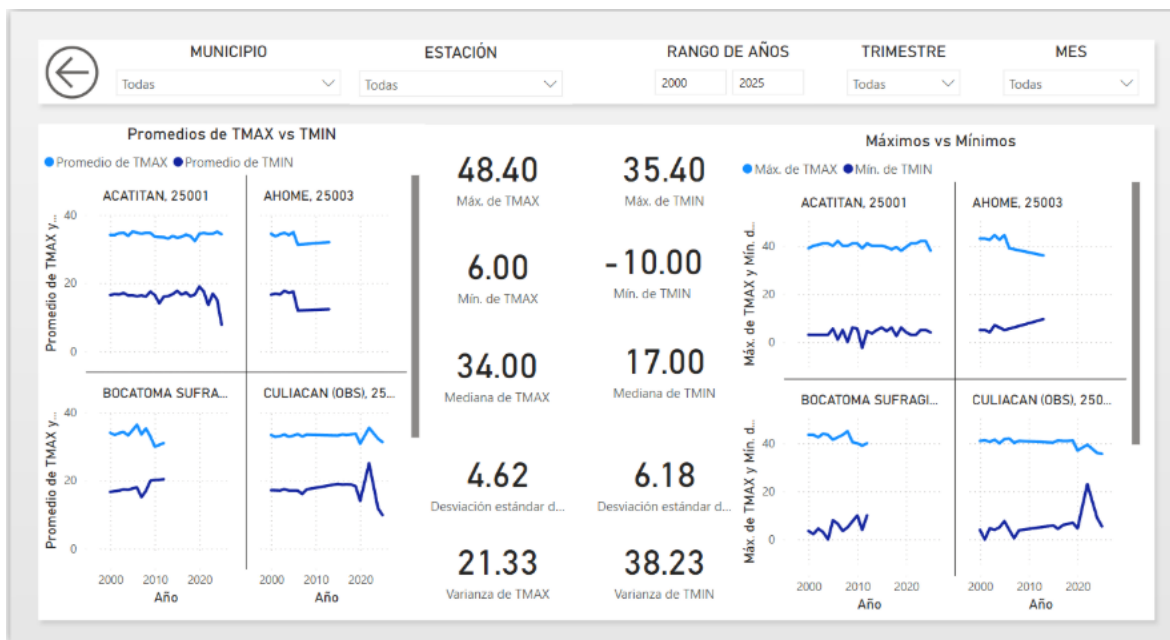


Figura 5.16. Informe térmico diario 2000-2025 (Fuente: Elaboración propia)

#### 5.4.1.1 Análisis hidrológico 1974-1999 vs 2000-2025

Se utilizó la información procesada para realizar análisis de tendencia. Dentro de Sinaloa se obtuvieron las siguientes estadísticas hidrológicas diarias:

Tabla 1. Análisis hidrológico

| Métrica             | 1974-1999    | 2000-2025     | Tipo de cambio |
|---------------------|--------------|---------------|----------------|
| Suma total          | 684,401.81mm | 690,345.154mm | Aumento        |
| Promedio diario     | 1.98mm       | 2.04mm        | Aumento        |
| Mediana             | 0.00mm       | 0.00mm        | Sin cambio     |
| Varianza            | 79.25mm      | 86.52mm       | Aumento        |
| Desviación estándar | 8.9mm        | 9.3mm         | Aumento        |

Se puede observar que las lluvias se han visto aumentadas, al igual que el promedio diario de estas, lo que sugiere que llueve más a menudo. Sin embargo, la mediana sigue siendo cero, con esta información se puede interpretar que la temporada de lluvias es corta, lo cual es correcto para el área de análisis.

Observando la tendencia en aumento de la desviación estándar y la varianza, es posible interpretar que hay un aumento en la variabilidad y la intensidad de las precipitaciones, lo cual quiere decir que los monzones, chaparrones o las lluvias se pueden llegar a presentar de forma más súbita e intensa.

#### 5.4.1.2 Análisis térmico 1974-1999 vs 2000-2025

Dentro de los informes generados en esta investigación, es posible obtener los siguientes datos térmicos para las series de tiempo:

Tabla 2. Análisis térmico

| Métrica                 | 1974-1999 | 2000-2025 | Tipo de cambio   |
|-------------------------|-----------|-----------|------------------|
| Máxima más alta         | 55°C      | 48.4°C    | Disminución      |
| Máxima más baja         | 6°C       | 6°C       | Sin cambios      |
| Mínima más alta         | 37°C      | 35.4°C    | Disminución leve |
| Mínima más baja         | -9°C      | -10°C     | Disminución      |
| Mediana de máximas      | 33°C      | 34°C      | Aumento          |
| Mediana de mínimas      | 16.5°C    | 17°C      | Aumento          |
| Desviación est. máximas | 4.78°C    | 4.62°C    | Disminución leve |
| Desviación est. mínimas | 6.13°C    | 6.18°C    | Aumento leve     |
| Varianza de máximas     | 22.84°C   | 21.33°C   | Disminución leve |
| Varianza de mínimas     | 37.53°C   | 38.23°C   | Aumento leve     |

La información que nos proporcionan los datos contrastados permite ver cosas importantes, como lo es el aumento de la mediana de máximas y mínimas, lo cual indica días más calientes y noches menos frías. Esto, aunado a la disminución de la variabilidad de máximas, indican días cálidos más constantes, mientras que el aumento de la variabilidad de temperaturas mínimas indica una mayor dispersión de estas a lo largo del tiempo, lo cual indica un comportamiento ligeramente más errático del clima durante la noche.

Otros datos recabados nos dicen que la temperatura más alta registrada fue menor en la serie de tiempo reciente, pero hubo una temperatura mínima extrema más baja, lo cual podría ser (o no) un evento aislado.

#### **5.4.2 Análisis hidrometereológico específico I (Caso de estudio: Huracán Manuel)**

De acuerdo con el Diario Oficial de la Federación (2013), se declaró estado de desastre natural en Angostura, Badiraguato, Culiacán, Navolato, Rosario, Mocorito, Sinaloa, Escuinapa y Salvador Alvarado del Estado de Sinaloa, debido a la presencia del Huracán Manuel los días 18 y 19 de septiembre. Además, de acuerdo con información de un artículo reciente del periódico *Debate* (2025), fue uno de los huracanes más fuertes presentes en México en tiempos modernos.

### 5.4.2.1 Recopilación de datos de análisis

Tabla 3. Datos hidrometeorológicos 2013, 2020

| Métrica                     | 2013         | 2020         |
|-----------------------------|--------------|--------------|
| <b>TEMPERATURAS</b>         |              |              |
| Máxima más alta             | 44°C         | 48°C         |
| Máxima más baja             | 10°C         | 9°C          |
| Mínima más alta             | 29°C         | 34°C         |
| Mínima más baja             | -4°C         | -7°C         |
| Mediana de máximas          | 33°C         | 34°C         |
| Mediana de mínimas          | 16.5°C       | 17°C         |
| Desviación estándar máximas | 4.52°C       | 4.98°C       |
| Desviación estándar mínimas | 6.17°C       | 6.12°C       |
| Varianza de máximas         | 20.39°C      | 24.81°C      |
| Varianza de mínimas         | 38.03°C      | 37.45°C      |
| <b>PRECIPITACIÓN</b>        |              |              |
| Acumulado                   | 30,378.33mm  | 18,049.84 mm |
| Promedio                    | 2.36 mm      | 1.64 mm      |
| Mediana                     | 0.00 mm      | 0.00 mm      |
| Varianza                    | 159.15 mm    | 66.18 mm     |
| Desviación estándar         | 12.62 mm     | 8.14 mm      |
| <b>EVAPORACIÓN</b>          |              |              |
| Acumulado                   | 46,338.58 mm | 27,622.94 mm |
| Promedio                    | 5.32 mm      | 5.19 mm      |
| Mediana                     | 5.08 mm      | 4.96 mm      |
| Varianza                    | 6.24 mm      | 6.34 mm      |
| Desviación estándar         | 2.50 mm      | 2.52 mm      |

#### **5.4.2.2 Análisis comparativo 2013-2020**

En el año 2013 el acumulado anual captado por los sensores fue de aproximadamente 30,378mm, superando las cifras del 2020 por 12,328.16mm, lo cual es evidencia de la presencia de precipitaciones anormales, corroborado por la acumulación de evaporación de 46,338.58mm, evidencia de la mayor presencia de humedad consecuencia de las lluvias.

### **5.4.3 Análisis hidrometeorológico II (Caso de estudio: Sequia extrema 2024)**

Según información de Conselva (2025) y una nota de Debate (2024), fue la primera vez en diez años que una sequía prevaleció después de la temporada de lluvias del año.

### 5.4.3.1 Recopilación de datos de análisis 2020-2024

Tabla 4. Datos hidrometeorológicos 2020-2024

| Métrica                     | 2020         | 2024         |
|-----------------------------|--------------|--------------|
| <b>TEMPERATURAS</b>         |              |              |
| Máxima más alta             | 48°C         | 46°C         |
| Máxima más baja             | 9°C          | 12°C         |
| Mínima más alta             | 34°C         | 30.5°C       |
| Mínima más baja             | -7°C         | -5°C         |
| Mediana de máximas          | 34°C         | 35°C         |
| Mediana de mínimas          | 17°C         | 15°C         |
| Desviación estándar máximas | 4.98°C       | 4.96°C       |
| Desviación estándar mínimas | 6.12°C       | 6.64°C       |
| Varianza de máximas         | 24.81°C      | 24.58°C      |
| Varianza de mínimas         | 37.45°C      | 44.14°C      |
| <b>PRECIPITACIÓN</b>        |              |              |
| Acumulado                   | 18,049.84 mm | 10,187.06 mm |
| Promedio                    | 1.64 mm      | 1.39 mm      |
| Mediana                     | 0.00 mm      | 0.00 mm      |
| Varianza                    | 66.18 mm     | 58.47 mm     |
| Desviación estándar         | 8.14 mm      | 7.65 mm      |
| <b>EVAPORACIÓN</b>          |              |              |
| Acumulado                   | 27,622.94 mm | 16,172.87 mm |
| Promedio                    | 5.19 mm      | 4.96 mm      |
| Mediana                     | 4.96 mm      | 4.79 mm      |
| Varianza                    | 6.34 mm      | 4.55 mm      |
| Desviación estándar         | 2.52 mm      | 2.13 mm      |

#### **5.4.3.2 Análisis comparativo 2020-2024**

En el año 2024 se presentó una acumulación de lluvias extremadamente menor, descendiendo a 10,187mm, aproximadamente un 43.56% menos con respecto a 2020, signo de una sequía extrema. De igual forma, se pueden comprobar las afectaciones hidrológicas por la disminución en la varianza y la desviación estándar presentada en 2024, signo de escasez de lluvias y variación en la presencia de estas. De acuerdo con los datos recopilados, el promedio de precipitación disminuyó en 0.25mm, la varianza en 7.71mm y la desviación estándar en 0.49mm, lo cual indica que hubo menos cantidad de lluvia a lo largo del año, con menor presencia de precipitaciones diarias y que durante esas precipitaciones la cantidad de lluvia también disminuyó. La sequía también se puede corroborar claramente en la temperatura, puesto que el encontrar la mínima más alta y baja mayor que en los otros dos años a comparar podría ser indicador de la falta de humedad consecuencia de este fenómeno climatológico.

## Capítulo 6

### 6 Conclusiones y trabajo futuro

Como se vio en la sección anterior, existe una tendencia hacia un clima más cálido, con precipitaciones más intensas y erráticas, las cuales significan un periodo de sequías más duradero e incierto, la estabilización de la evaporación puede deberse a un equilibrio entre el aumento de la temperatura y otros factores como la nubosidad, humedad o el cambio de uso de suelo provocado por la expansión de las actividades económicas del estado. Lo que representan todos estos datos son una indudable transformación de los patrones climáticos a largo plazo, lo cual demuestra los efectos del cambio climático sobre los eventos meteorológicos que acontecen en el estado de Sinaloa.

## **6.1 Más trabajo a futuro**

Este proyecto de investigación tiene el potencial de facilitar en gran medida el desarrollo de futuras investigaciones y experimentos dentro de diferentes áreas de aplicación como por ejemplo serían:

### **6.1.1 Ampliación del alcance del proyecto**

Es posible incorporar más variables de diferentes sensores de CONAGUA para complementar y mejorar los resultados de los análisis de datos, como lo podrían ser el porcentaje de llenado de presas, humedad, densidad de nubes, entre otros factores que podrían ser clave.

### **6.1.2 Aplicaciones estadísticas**

Utilizando, manipulando y complementando los datos y los informes que se pueden generar con los datos procesados, es posible llevar a cabo procesos estadísticos más complejos, como lo podría ser el análisis de tendencias.

### **6.1.3 Aplicaciones en la IA**

La creación del proceso de ETL sobre estos datos para generar datasets pueden ser de gran ayuda a la hora de entrenar modelos de ML enfocados en diferentes áreas tales como la creación de modelos predictivos o la creación de modelos para la generación de data sintética.

## 7 Bibliografía

*Historia del Servicio Meteorológico Nacional (SMN) y su edificio.* (2021). gob.mx. Recuperado el 6 de mayo de 2025, de <https://www.gob.mx/smn/videos/historia-del-servicio-meteorologico-nacional-y-su-edificio>

Moreno, M, Torres, M. (2017). Primer observatorio astronómico y meteorológico nacional de la Ciudad de México. *Historia y arqueología. Arqueología*, (26), 113-126.

Sosa-Baldivia, A., & Ruíz-Ibarra, G. (2017). La disponibilidad de alimentos en México: un análisis de la producción agrícola de 35 años y su proyección para 2050. *Papeles De Población*, 23(93), 207-230. Consultado de <https://rppoblacion.uaemex.mx/article/view/9111>

Secretaría de Agricultura y Desarrollo Rural (SADER). (2021). *Cinco estados con cinco estrellas en producción agrícola.* gob.mx. Recuperado el 8 de mayo de 2025, de <https://www.gob.mx/agricultura/articulos/cinco-estados-con-cinco-estrellas-en-produccion-agricola>

Secretaría de Agricultura y Desarrollo Rural (SADER). (2022). *Infografía alimentaria Sinaloa.* <https://estadisticas.sinaloa.gob.mx/documentos/Infografiasagroalimentarias/Sinaloa-Infografia-Agroalimentaria-2022.pdf>

Reuters. (2024, junio 12). *Producción de maíz blanco de México caería 9% en 2024 por la sequía.* El Economista. <https://www.eleconomista.com.mx/economia/Produccion-de-maiz-blanco-de-Mexico-caeria-9-en-2024-por-la-sequia-20240612-0088.html>

Amabilis-Sosa, L., Roé-Sosa, A. & Mendivil, K. (2023). *Preservación del agua en Sinaloa, México, a través del monitoreo instrumentado de la calidad del agua. Estudio de Caso: Río Presidio.* 11. 273-285.

Flores, L., Arzola-González, J., Ramírez-Soto, M., & Osorio-Pérez, A. (2012). Repercusiones del cambio climático global en el estado de sinaloa, México. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 21(1), 115-129. Retrieved May 16, 2025, from [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0121-215X2012000100009&lng=en&tlng=es](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-215X2012000100009&lng=en&tlng=es).

Cervantes-Osornio, R., Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Cevamex, Texcoco, Estado de México, México, Arteaga-Ramírez, R., Vázquez-Peña, M. A., Ojeda-Bustamante, W., Universidad Autónoma Chapingo, Departamento de Irrigación, Estado de México, México, Universidad Autónoma Chapingo, Departamento de Irrigación, Estado de México, México, & Instituto Mexicano de Tecnología del Agua, Jiutepec, Morelos, México. (2023). Calidad de datos de estaciones meteorológicas automáticas ubicadas en tres estados de México. *Tecnología y ciencias del agua*, 14(1), 432-469. <https://doi.org/10.24850/j-tyca-14-01-10>

Lobato-Sánchez, René. (2016). El monitor de la sequía en México. *Tecnología y ciencias del agua*, 7(5), 197-211. Recuperado en 08 de mayo de 2025, de [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2007-24222016000500197&lng=es&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-24222016000500197&lng=es&tlng=es).

Secretaría de Educación Pública (SEP). (2025). La diversidad en México: regiones naturales. Gob.mx. Recuperado el 8 de mayo de 2025, de <https://nuevaescuelamexicana.sep.gob.mx/contenido/coleccion/la-diversidad-en-mexico-regiones-naturales-2/>

World Wildlife Fund (WWF). (2024). *¿Qué es la deforestación y cómo impacta a nuestro planeta?* wwf.org.co. <https://www.wwf.org.co/?389690/que-es-la-deforestacion>

Monjardín-Armenta, S. A., Pacheco-Angulo, C. E., Plata-Rocha, W., & Corrales-Barraza, G. (2017). La deforestación y sus factores causales en el estado de Sinaloa, México. *Madera y bosques*, 23(1), 7. <https://doi.org/10.21829/myb.2017.2311482>

Real Academia Española. (RAE). (2024) Sequía. En *Diccionario del estudiante*. Recuperado en 11 de mayo de 2025, de <https://dle.rae.es/sequia>

Coronel, A. (2013). *Sequía: concepto e índices de monitoreo. Propuesta de un nuevo índice*. fcagr.unr.edu.ar. <https://fcagr.unr.edu.ar/Extension/Agromensajes/37/1AM37.pdf>

United Nations Development Programme. (UNDP). (2024). *Quedándonos secos: Abordando el estrés hídrico en América Latina y el Caribe*. Undp.org. <https://www.undp.org/es/latin-america/blog/quedandonos-secos-abordando-el-estres-hidrico-en-america-latina-y-el-caribe#:~:text=El%20estrés%20hídrico%20refleja%20la,los%20niveles%20de%20estrés%20hídrico.>

Magaña, V., & Neri, C. (2012). Cambio climático y sequías en México. *Revista Ciencia*, 26–35.

Velázquez, A., Mas, J. F., DíazGallegos, J. R., MayorgaSaucedo, R., Alcántara, P. C., Castro, R., Fernández, T., Bocco, G., Ezcurra, E., & Palacio, J. L. (2002). Patrones y tasas de cambio de uso del suelo en México. *Gaceta Ecológica*, (62), 21-37.

Ley General del Equilibrio Ecológico y Protección al Ambiente [L.G.E.E.P.A], Reformada, Diario Oficial de la Federación [D.O.F.], 1 de abril de 2024.

Ley Federal de Responsabilidad Ambiental [L.F.R.A.], Reformada, Diario Oficial de la Federación [D.O.F.] 20 de mayo de 2021

Ley General de Desarrollo Forestal Sustentable [L.G.D.F.S.], Reformada, Diario Oficial de la Federación [D.O.F.] 1 de abril de 2024

Comisión Nacional del Agua (CONAGUA). (2025). *Información Estadística Climatológica*. Servicio Meteorológico Nacional (SMN). Recuperado el 11 de mayo de 2025. <https://smn.conagua.gob.mx/es/climatologia/informacion-climatologica/informacion-estadistica-climatologica>

Secretaría del Medio Ambiente. (SEDEMA). (2025). *Programas*. Secretaría del Medio Ambiente. Recuperado el 11 de mayo de 2025, de <https://www.sedema.cdmx.gob.mx/programas>

- Vassiliadis, P., Simitsis, A., & Baikousi, E. (2009). A taxonomy of ETL activities. *Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP*.
- Antúnez, C. (2011, enero 30). ANÁLISIS DE SERIES DE TIEMPO. *Contribuciones a la Economía*, 9(1), 46.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3a ed.). OTexts.
- Servicio Meteorológico Nacional (SMN). (2023). *Índice Estandarizado de Precipitación (SPI)*. Gob.mx. Recuperado el 8 de mayo de 2025, de <https://smn.conagua.gob.mx/es/climatologia/monitor-de-sequia/spi>
- Alley, W. M. (1984). The Palmer drought severity index: Limitations and assumptions. *Journal of Climate and Applied Meteorology*, 23(7), 1100–1109. [https://doi.org/10.1175/1520-0450\(1984\)023<1100:tpdsil>2.0.co;2](https://doi.org/10.1175/1520-0450(1984)023<1100:tpdsil>2.0.co;2)
- Ramírez, J., & JR Engineering Company. (2024, mayo 22). *Validación Serie de Tiempo*. <https://rpubs.com/JREngineering/1187127>
- Comisión Nacional del Agua (CONAGUA). (2025). *Hidrometeorología*. Gobierno de México. Recuperado el 8 de mayo de 2025, de <https://www.gob.mx/conagua/acciones-y-programas/hidrometeorologia>
- Comisión Nacional del Agua (CONAGUA), & Servicio Meteorológico Nacional (SMN). (2025). *Monitor de Sequía de América del Norte (NADM)*. <https://smn.conagua.gob.mx>. <https://smn.conagua.gob.mx/es/climatologia/monitor-de-sequia/monitor-de-sequia-de-america-del-norte>
- Universidad Nacional de Córdoba. (2023). *Licenciatura en Hidrometeorología*. <https://www.famaf.unc.edu.ar>. <https://www.famaf.unc.edu.ar/academica/grado/licenciatura-en-hidrometeorologia/>
- Vilchis-Mata, I., Bâ, K. M., Franco-Plata, R., & Díaz-Delgado, C. (2015). Modelación hidrológica con base en estimaciones de precipitación con sensores hidrometeorológicos. *Tecnología y Ciencias del Agua*, VI(4), 45-60.
- Servicio Meteorológico Nacional (SMN). (2025). *Estaciones Meteorológicas Automáticas (EMA's)*. <https://smn.conagua.gob.mx>. Recuperado el 11 de mayo de 2025, de <https://smn.conagua.gob.mx/es/observando-el-tiempo/estaciones-meteorologicas-automaticas-ema-s>
- Servicio Meteorológico Nacional (SMN). (2025). *¿Qué es un Radar Meteorológico?* <https://smn.conagua.gob.mx>. Recuperado el 11 de mayo de 2025, de [https://smn.conagua.gob.mx/tools/GUI/visor\\_radares\\_v3/index.html?kche=110525040524#/info](https://smn.conagua.gob.mx/tools/GUI/visor_radares_v3/index.html?kche=110525040524#/info)
- Dodig, A., Stankovic, V., Stankovic, L., & Stojkovic, M. (2024). *Enhancement of hydrological time series prediction with real-world time series generative adversarial network-based synthetic data and deep learning models*. Ssrn.com. Recuperado el 12 de mayo de 2025, de [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5022039](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5022039)

- Khayyat, M., Laabidi, K., Almalki, N., & Al-zahrani, M. (2021). Time series Facebook prophet model and python for COVID-19 outbreak prediction. *Computers, Materials & Continua*, 67(3), 3781–3793. <https://doi.org/10.32604/cmc.2021.014918>
- Brodeur, Z. P., & Steinschneider, S. (2021). A multivariate approach to generate synthetic short-to-medium range hydro-meteorological forecasts across locations, variables, and lead times. *Water Resources Research*, 57(6). <https://doi.org/10.1029/2020wr029453>
- Alzantot, M., Chakraborty, S., & Srivastava, M. (2017). SenseGen: A deep learning architecture for synthetic sensor data generation. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*.
- Lee, T., Singh, V. P., & Cho, K. H. (2021). *Deep learning for hydrometeorology and environmental science* (2021a ed.). Springer Nature.
- Shafranovich, Y. (2005). *RFC 4180: Common format and MIME type for Comma-Separated Values (CSV) files*. IETF Datatracker. Recuperado el 14 de mayo de 2025, de <https://datatracker.ietf.org/doc/html/rfc4180>
- Arreguín-Cortés, F., López-Pérez, M., & Marengo-Mogollón, H. (2016). Las inundaciones en un marco de incertidumbre climática. *Tecnología y ciencias del agua*, 7(5), 5-13. Recuperado en 16 de mayo de 2025, de [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S2007-24222016000500005&lng=es&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-24222016000500005&lng=es&tlng=es).
- Ruiz-Jiménez, C., Téllez-Valdés, O., & Luna-Vega, I. (2012). Clasificación de los bosques mesófilos de montaña de México: afinidades de la flora. *Revista mexicana de biodiversidad*, 83(4), 1110-1144. <https://doi.org/10.7550/rmb.29383>
- Hernandez, N. (2023). Variables bioclimáticas que determinan la distribución de la hormiga myrmecocystus Mexicanus en México. *CIENCIA E INNOVACIÓN AGROALIMENTARIA DE LA UNIVERSIDAD DE GUANAJUATO*, 1(1), 48–56. <http://repositorio.ugto.mx/handle/20.500.12059/11647>
- Coordinación Nacional de Protección Civil de México (2018). *Inundaciones en México* [Diapositiva de PowerPoint]. Gobierno de México. [https://www.gob.mx/cms/uploads/attachment/file/336984/1.\\_180503\\_SRI\\_Inundaciones.pdf](https://www.gob.mx/cms/uploads/attachment/file/336984/1._180503_SRI_Inundaciones.pdf)
- Mustapha, S., Man, M., Wan Abu Bakar, W. A., Yusof, M. K., & Ahmad Sabri, I. A. (2024). Demystified overview of data scraping. *International Journal of Data Science and Advanced Analytics*, 6(6), 290–296. <https://doi.org/10.69511/ijdsaa.v6i6.205>
- Python Software Foundation. (2013). *About Python*. Python.org. Recuperado el 21 de mayo de 2025, de <https://www.python.org/about/>
- Microsoft Corporation. (2015). *Power BI: Visualización de datos*. Microsoft.com. Recuperado el 21 de mayo de 2025, de <https://www.microsoft.com/es-mx/power-platform/products/power-bi>

Cuesta, F., & Cabezas, A. M. (2002). Unas definiciones polémicas: medio ambiente y gasto ambiental. *Técnica Contable*, 54, 565-574.

Sánchez-Ortiz, E. (2013). El uso de la disponibilidad relativa de agua como indicador en la cuenca del río San Pedro, México. *Tecnología y ciencias del agua*, 4(5), 199-206.

Botero Botero, S., & Cano Cano, J. A. (2008). Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia. *Cuadernos de Economía*, 27(48), 173-208.

VALLVÉ, M. B., & VIDE, J. M. (1994). ALGUNOS MÉTODOS CUANTITATIVOS EN CLIMATOLOGÍA HISTÓRICA: LAS CORRELACIONES LINEALES CON DESFASE EN EL ANÁLISIS DE SERIES. *PERFILES ACTUALES DE LA GEOGRAFÍA CUANTITATIVA EN ESPAÑA*, pp. 39 – 48.

Ogata, K. (2003). *Ingeniería de control moderna*. Pearson educación.

Burbano Pesántez, J. D. (2023). *Determinación de estacionalidad en función de los datos meteorológicos en la ciudad de Cuenca* (Bachelor's thesis, Universidad del Azuay).

Bocco, G., Mendoza, M., & Masera, O. R. (2001). La dinámica del cambio del uso del suelo en Michoacán: Una propuesta metodológica para el estudio de los procesos de deforestación. *Investigaciones geográficas*, (44), 18-36.

Palacio, J. L., Velázquez, A., Mas, J. F., Gallegos, J. D., Saucedo, R. M., Alcántara, P. C., ... & Ezcurra, E. (2002). Patrones y tasas de cambio de uso del suelo en México. *Gaceta ecológica*, (62), 21-37.

San José Albacete, A. (2011). Procesamiento de datos LiDAR con ArcGIS Desktop 10.

Núñez-Cuadra, O., Talavera-Bustamante, I., Porro-Muñoz, D., Hernández-Fernández, N., & Bustio-Martínez, L. (2011). Experiencias en el uso del software Quimiometrix para el procesamiento multivariado de datos químicos y bioquímicos. *Revista CENIC. Ciencias Químicas*, 42(1), 3-16.

Madueño, A., López, M., Estévez, J., & Giráldez, J. V. (2016). Validación de datos hidrológicos con redes neuronales artificiales. Aplicación a niveles en ríos. In *II Simposio Nacional de Ingeniería Hortícola: Automatización y TICS en agricultura*.

Rojas, A. R., Velastegui, A. H., Solórzano, F. C., Fiallos, F. M., Medina, D. M., & Paredes-Beltran, B. (2024). Aplicación de Redes Neuronales Artificiales para la Estimación de Pre-cipitaciones: Caso de Estudio de la Cuenca del Río Pastaza, Ecuador: Application of Artificial Neural Networks for Rainfall Estimation: Case Study in the Pastaza River Basin, Ecuador. *Revista Científica FINIBUS-Ingeniería, Industria y Arquitectura*, 7(14), 131-146.

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3).

Thomas, D. M., & Mathur, S. (2019, June). Data analysis by web scraping using python. In *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 450-454). IEEE.

Castañeda, M. B. (2010). *Procesamiento de datos y análisis estadísticos utilizando SPSS: Un libro práctico para investigadores y administradores educativos*. Edipucrs.

Luque, J. A., Cantalejo, M., Millares, A., Mateos, R. M., Montserrat, O., Sarro, R., Reyes-Carmona, C., Peña, E., Cuevas-González, M., Galve, J. P. & Barra, A. (2022). ANÁLISIS COMPARATIVOS DE SERIES DE PRECIPITACIONES Y MODELOS DE CAMBIO CLIMÁTICO COMO HERRAMIENTA PARA EVALUAR RIESGOS GEOLÓGICOS.

Del Saltos Arrieta, M. A. (2014). Estrategia de mitigación al cambio climático mediante un análisis comparativo de las temperaturas del suelo y ambiental de la granja experimental Udla - Nono (Tesis de pregrado). Universidad de las Américas, Quito.

Moyroud, N., & Portet, F. (2018). Introduction to QGIS. *QGIS and generic tools*, 1, 1-17.

Khan, S., & Mohiuddin, K. (2018). Evaluating the parameters of ArcGIS and QGIS for GIS Applications. *Int. J. Adv. Res. Sci. Eng*, 7, 582-594.

Programa de las Naciones Unidas para el Desarrollo. (2024). *Quedándonos secos: Abordando el estrés hídrico en América Latina y el Caribe*. Undp.org. <https://www.undp.org/es/latin-america/blog/quedandonos-secos-abordando-el-estres-hidrico-en-america-latina-y-el-caribe>

Ragunathan, T. E. (2021). Synthetic data. *Annual review of statistics and its application*, 8(1), 129-140.

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., ... & Weller, A. (2022). Synthetic Data--what, why and how?. *arXiv preprint arXiv:2205.03257*.

Conselva. (2024). *Y la sequía sigue aquí; a pesar de las lluvias la crisis hídrica se prolongará en Sinaloa*. Conselva, Costas y Comunidades. <https://conselva.org/y-la-sequia-sigue-aqui-a-pesar-de-las-lluvias-la-crisis-hidrica-se-prolongara-en-sinaloa/>

Lozano, J. (2024). Por primera vez en 10 años, la sequía en Sinaloa permanece tras las lluvias. *Debate*. <https://www.debate.com.mx/sinaloa/culiacan/Por-primera-vez-en-10-anos-la-sequia-en-Sinaloa-permanece-tras-las-lluvias-20241015-0072.html>

Osorio, K. (2025). ¡Los 9 huracanes más fuertes que han golpeado a México en los últimos años! *Debate*. <https://www.debate.com.mx/Los-9-huracanes-mas-fuertes-que-han-golpeado-a-Mexico-en-los-ultimos-anos-t202506180014.html>

El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University-Computer and Information Sciences*, 23(2), 91-104.

Hernández, F. L. (2013). Interpolación temporal de temperatura media mensual a partir de mapas trazados por expertos, mediante el modelo propuesto por Charles-Edwards.

Diario Oficial de la Federación. (2013, 26 de septiembre). *Declaratoria de Desastre Natural por la ocurrencia de lluvia severa provocada por el huracán “Manuel”, los días 18 y 19 de septiembre de 2013, en 9 municipios del Estado de Sinaloa*. [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5315364&fecha=26/09/2013](https://www.dof.gob.mx/nota_detalle.php?codigo=5315364&fecha=26/09/2013)

Anzules, Í. D. C. P., & Castro, D. W. M. (2022). Contaminación ambiental. *Recimundo*, 6(2), 93-103.

Domínguez, M. C. (2015). La contaminación ambiental, un tema con compromiso social. *Producción+ limpia*, 10(1), 9-21.

Campos, M. Y. (2022). La contaminación ambiental: un espacio de reflexión y sus consecuencias. *Paidagogo*, 4(2), 70-80.

Servicio Meteorológico Nacional (SMN). (2020). *Monitor de Sequía de México*. Gob.mx. <https://www.gob.mx/cms/uploads/attachment/file/598545/MS-15-Abril2020.pdf>

López, N. G., Verduras, C. L., & García, J. P. (2010). Estimación de los flujos de transporte de mercancías interregionales trimestrales mediante técnicas de interpolación temporal. *Estudios de economía aplicada*, 28(3), 1-38.

Kufel, J., Bargiel-Lączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., ... & Gruszczyńska, K. (2023). What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. *Diagnostics*, 13(15), 2582.

Lu, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., & Wei, W. (2023). Machine learning for synthetic data generation: a review. *arXiv preprint arXiv:2302.04062*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

## A. Código fuente

### A.1 datascraper.py

```

1. import urllib.request
2.
3. def download(url, filename):
4.     '''download(url, filename)
5.
6.     This function downloads a file from a specific URL and saves it to a
       determined filename.
7.
8.     Input:
9.         - url: string with the URL of the file to download
10.        - filename: string with the name of the file to save the
       downloaded file
11.
12.        Output: None
13.        '''
14.        try:
15.            urllib.request.urlretrieve(url, filename)
16.            print(f"Downloaded {filename}")
17.        except Exception as e:
18.            print(f"Error downloading {url}: {e}. File does not exist")
19.
20.
21.    def explore_download(start, end, output_folder, estado = 'sin'):
22.        ''' explore_download(start, end, output_folder)
23.        This function explores every url in a given range and saves
       the existing .txt files into an output folder
24.
25.        Input:
26.            - start: a number associated with a file
27.            - end: final number to search in a range
28.            - output_folder: string with the name of the folder
       where the found files will be saved
29.            - estado: string with the state code (default is 'sin')
30.        '''
31.        for i in range(start, end):
32.            url =
       f"https://smn.conagua.gob.mx/tools/RECURSOS/Normales_Climatologicas/Diar
       ios/{estado}/dia{i}.txt"
33.            filename = f"{output_folder}/{i}.txt"
34.            download(url, filename)
35.

```

```

36.     def bulk_download(stations, output_folder, estado = 'sin'):
37.         ''' bulk_download(output_folder)
38.
39.         This function downloads a range of files from the SMN website.
40.
41.         Input:
42.             - output_folder: string with the name of the folder where
              the files will be saved
43.             - stations: list of strings with the code of the stations
              to download
44.             - estado: string with the state code (default is 'sin')
45.
46.         Output: None
47.         '''
48.
49.         for estacion in stations:
50.             url =
              f"https://smn.conagua.gob.mx/tools/RECURSOS/Normales_Climatologicas/Diar
              ios/{estado}/dia{estacion}.txt"
51.             filename = f"{output_folder}/{estacion}.txt"
52.             download(url, filename)

```

## A.2 limpiar\_datos.py

```
1. import pandas as pd
2. import re
3. import os
4. from numpy import nan
5.
6. def procesar_lote_txt_a_csv(directorio_entrada, umbral_faltantes=100):
7.     """procesar_lote_txt_a_csv(directorio_entrada):
8.
9.         This function receives a directory, detects the .txt files in it and processes
           them using the limpiar_txt_a_csv function.
10.
11.         Input:
12.             directorio_entrada (String): The directory where the .txt files are located.
13.             umbral_faltantes (int): Maximum percentage of missing dates allowed.
14.         Return:
15.             None
16.         """
17.         directorio_salida = os.path.join(directorio_entrada, "limpio")
18.         os.makedirs(directorio_salida, exist_ok=True)
19.         for archivo in os.listdir(directorio_entrada):
20.             if archivo.endswith(".txt"):
21.                 archivo_txt = os.path.join(directorio_entrada, archivo)
22.                 limpiar_txt_a_csv(archivo_txt, directorio_salida)
23.                 rellenar_fechas_faltantes(os.path.join(directorio_salida,
24. f'NR_{archivo.replace(".txt", "")}.csv'), umbral_faltantes)
25. def limpiar_txt_a_csv(archivo_txt, directorio_salida):
26.     """limpiar_txt_a_csv(archivo_txt, directorio_salida):
27.
28.         This function receives a .txt file with a specific structure, then it cleans
           and processes it and saves it as a .csv file.
29.         Input:
30.             - archivo_txt (File): A .txt file with a specific structure.
31.             - directorio_salida (String): The directory where the .csv file will be saved.
32.         Return:
33.             None
34.         """
35.
36.         codificacion = "utf-8"
37.         # REGEX para detectar la fecha en el formato YYYY-MM-DD
38.         regex = r'\d{4}-\d{2}-\d{2}'
39.         # Después de la línea 25 termina la cabecera, se espera que los datos comiencen
40.         linea_inicio = 25
41.
42.         with open(archivo_txt, 'r', encoding=codificacion) as file:
43.             lineas = file.readlines()
44.
45.         print(f'Procesando archivo {archivo_txt}')
46.         print(f'Número de líneas: {len(lineas)}')
47.
48.         target_lines = [10, 11, 12, 13, 14, 16, 17, 18]
49.         etiquetas = []
50.         valores = []
51.         estacion = "Desconocida"
52.
53.         for i in target_lines:
54.             if i < len(lineas) and ':' in lineas[i]:
55.                 clave, valor = map(str.strip, re.split(r':\s*', lineas[i], maxsplit=1))
```

```

56.         valor = valor.replace(' ', '').replace('.', '').replace('-', '')
57.         if clave == 'ALTITUD':
58.             valor = re.sub(r'\s*msnm', '', valor)
59.             etiquetas.append(clave)
60.             valores.append(valor)
61.
62.         if "ESTACIÓN" in lineas[i]: # Detección explícita de la estación
63.             match = re.search(r'ESTACIÓN\s*:\s*(\d+)', lineas[i])
64.             if match:
65.                 estacion = match.group(1)
66.         if "ESTACION" in lineas[i]:
67.             match = re.search(r'ESTACION\s*:\s*(\d+)', lineas[i])
68.             if match:
69.                 estacion = match.group(1)
70.
71.     print(f'Estación detectada: {estacion}')
72.     datos = []
73.
74.     for linea in lineas[linea_inicio:]:
75.         if re.match(regex, linea):
76.             valores_linea = re.split(r'\t+', linea.strip()) if '\t' in linea else
re.split(r'\s+', linea.strip())
77.             if len(valores_linea) >= 5:
78.                 datos.append(valores_linea[:5])
79.
80.     if not datos:
81.         print("⚠ Advertencia: No se encontraron datos climáticos en el archivo.")
82.         return
83.
84.     columnas = ['FECHA', 'PRECIP', 'EVAP', 'TMAX', 'TMIN'] + etiquetas
85.     df = pd.DataFrame(datos, columns=columnas[:len(datos[0])])
86.
87.     for i, etiqueta in enumerate(etiquetas):
88.         df[etiqueta] = valores[i] if i < len(valores) else ""
89.
90.     df.replace('Nulo', None, inplace=True)
91.     for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
92.         df[column] = pd.to_numeric(df[column], errors='coerce')
93.
94.     df['FECHA'] = pd.to_datetime(df['FECHA'], format='%Y-%m-%d')
95.     df = df.dropna(subset=['FECHA'])
96.     df['MONTH'] = df['FECHA'].dt.month
97.     df['YEAR'] = df['FECHA'].dt.year
98.
99.     for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
100.         df[column] = df[column].groupby(['YEAR',
'MONTH']).transform(lambda x: x.fillna(x.mean() if not pd.isna(x.mean())
else nan))
101.
102.     for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
103.         df[column] = df[column].astype(str).replace(nan, 'null')
104.
105.     archivo_salida = os.path.join(directorio_salida, f'NR_{estacion}.csv')
106.     df.to_csv(archivo_salida, index=False, encoding='utf-8')
107.     print(f'Archivo CSV guardado como {archivo_salida}')
108.
109.     def rellenar_fechas_faltantes(archivo_csv, umbral_faltantes=100):
110.
111.         """rellenar_fechas_faltantes(archivo_csv):
112.

```

```

113.         This function receives the processed csv file and fills in the missing
           dates by calculating the min-max date range.
114.
115.         Input:
116.             - archivo_csv (File): csv file with missing dates.
117.             - umbral_faltantes (int): Maximum percentage of missing dates allowed.
118.
119.         Return:
120.             None
121.         """
122.
123.         df = pd.read_csv(archivo_csv, dayfirst=True)
124.         df['FECHA'] = pd.to_datetime(df['FECHA'], errors='coerce')
125.
126.         # Contar número de registros en el Dataframe df
127.         total_registros_original = df.shape[0]
128.
129.         fecha_min = df['FECHA'].min()
130.         fecha_max = df['FECHA'].max()
131.         total_fechas = pd.date_range(start=fecha_min, end=fecha_max, freq='D')
132.
133.         all_dates = pd.DataFrame({'FECHA': total_fechas})
134.
135.         df_completo = all_dates.merge(df, on='FECHA', how='left')
136.         total_registros_generados = df_completo.shape[0]
137.         fechas_faltantes = total_registros_generados - total_registros_original
138.         porcentaje_faltante = (fechas_faltantes / total_registros_generados) *
100
139.
140.         ## Si el porcentaje de fechas faltantes es mayor al umbral, se elimina
           el archivo y se detiene el proceso
141.         if porcentaje_faltante > umbral_faltantes:
142.             faltantes_msg = f"Entradas faltantes en {archivo_csv}\nTotal de
           entradas faltantes: {fechas_faltantes}\nPorcentaje de entradas faltantes:
           {porcentaje_faltante:.2f}%"
143.             print(faltantes_msg)
144.             print(f'El porcentaje de fechas faltantes es mayor al umbral
           permitido ({umbral_faltantes}%)')
145.             os.remove(archivo_csv)
146.             print(f'Archivo descartado: {archivo_csv}')
147.             return
148.         else:
149.             df_completo['MONTH'] = df_completo['FECHA'].dt.month
150.             df_completo['YEAR'] = df_completo['FECHA'].dt.year
151.
152.             if not df_completo.empty:
153.                 if 'ESTACION' in df.columns:
154.                     constantes = ['ESTACION', 'NOMBRE', 'ESTADO', 'MUNICIPIO',
           'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
155.                 else:
156.                     constantes = ['ESTACIÓN', 'NOMBRE', 'ESTADO', 'MUNICIPIO',
           'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
157.                 for const in constantes:
158.                     df_completo[const] = df[const].iloc[0] if const in
           df.columns and not df[const].isna().all() else "Desconocida"
159.                 for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
160.                     df_completo[column] =
161.                     df_completo[column].astype(str).replace(nan, 'null')
162.                     df_completo[column].astype(str).replace('nan', 'null')
162.                 df_completo = df_completo.astype(str).replace('nan', 'null')

```

```

163.         df_completo.to_csv(archivo_csv, index=False, encoding='utf-8')
164.         log_msg = f"Entradas faltantes rellenas en {archivo_csv}\nTotal de
entradas faltantes: {fechas_faltantes}\nPorcentaje de entradas faltantes:
{porcentaje_faltante:.2f}%"
165.         print(log_msg)
166.
167.     def generar_concentrados_diarios(directorio_entrada):
168.         """generar_concentrados_diarios(directorio_entrada):
169.
170.             This function receives the directory of the raw files, locates the
clean files folder and generates a single file containing all the daily summaries.
171.
172.             Input:
173.                 directorio_entrada (String): Folder where the RAW files are located
174.             Return: None
175.             """
176.             directorio_limpios = os.path.join(directorio_entrada, "limpio")
177.             archivos_csv = [os.path.join(directorio_limpios, f) for f in
os.listdir(directorio_limpios) if f.endswith(".csv")]
178.             df_final = pd.concat([pd.read_csv(f) for f in archivos_csv],
ignore_index=True)
179.             for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
180.                 df_final[column] = df_final[column].astype(str).replace(nan,
'null')
181.                 df_final[column] = df_final[column].astype(str).replace('nan',
'null')
182.             archivo_final = os.path.join(directorio_limpios, 'CD.csv')
183.             df_final.to_csv(archivo_final, index=False, encoding='utf-8')
184.             print(f'Archivo final de concentrado diario guardado como
{archivo_final}')
185.
186.     def generar_concentrados_mensuales(directorio_entrada):
187.         """generar_concentrados_mensuales(directorio_entrada):
188.
189.             This function receives the directory of the raw files, then locates
the processed files folder and generates monthly summaries for each station.
190.
191.             Input:
192.                 directorio_entrada (String): folder where the original files are
located
193.             Return:
194.                 None
195.             """
196.             directorio_limpios = os.path.join(directorio_entrada, "limpio")
197.             directorio_concentrado = os.path.join(directorio_entrada, "CM")
198.             os.makedirs(directorio_concentrado, exist_ok=True)
199.             archivos_csv = [os.path.join(directorio_limpios, f) for f in
os.listdir(directorio_limpios) if f.endswith(".csv")]
200.
201.             for archivo in archivos_csv:
202.                 df = pd.read_csv(archivo)
203.
204.                 df_mensual = df.groupby(['YEAR', 'MONTH']).agg({
205.                     'PRECIP': 'sum',
206.                     'EVAP': 'mean',
207.                     'TMAX': 'max',
208.                     'TMIN': 'min'
209.                 }).reset_index()
210.
211.                 if not df_mensual.empty:
212.                     if 'ESTACION' in df.columns:

```

```

213.         constantes = ['ESTACION', 'NOMBRE', 'ESTADO', 'MUNICIPIO',
'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
214.         else:
215.             constantes = ['ESTACIÓN', 'NOMBRE', 'ESTADO', 'MUNICIPIO',
'SITUACIÓN', 'LATITUD', 'LONGITUD', 'ALTITUD']
216.             for const in constantes:
217.                 df_mensual[const] = df[const].iloc[0] if const in
df.columns and not df[const].isna().all() else "Desconocida"
218.
219.             for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
220.                 df_mensual[column] =
df_mensual[column].astype(str).replace(nan, 'null')
221.                 df_mensual[column] =
df_mensual[column].astype(str).replace('nan', 'null')
222.
223.             if 'ESTACION' in df.columns:
224.                 archivo_concentrado = os.path.join(directorio_concentrado,
f'CM_{df["ESTACION"].iloc[0]}.csv')
225.             else:
226.                 archivo_concentrado = os.path.join(directorio_concentrado,
f'CM_{df_mensual["ESTACIÓN"].iloc[0]}.csv')
227.                 df_mensual.to_csv(archivo_concentrado,
index=False,
encoding='utf-8')
228.                 print(f'Archivo de concentrado mensual guardado como
{archivo_concentrado}')
229.
230.     def concatenar_concentrados(directorio_entrada):
231.         """concatenar_concentrados(directorio_entrada):
232.
233.         This function receives the directory of the raw files, then locates
the monthly summaries folder and concatenates them into a single .csv file.
234.
235.         Input:
236.             directorio_entrada (String): folder where the raw files are located
237.         Return:
238.             None
239.         """
240.         directorio_concentrado = os.path.join(directorio_entrada, "CM")
241.         directorio_final = os.path.join(directorio_entrada, "FINAL")
242.
243.         os.makedirs(directorio_final, exist_ok=True)
244.
245.         archivos_csv = [os.path.join(directorio_concentrado, f) for f in
os.listdir(directorio_concentrado) if f.endswith(".csv")]
246.
247.         df_final = pd.concat([pd.read_csv(f) for f in archivos_csv],
ignore_index=True)
248.         for column in ['PRECIP', 'EVAP', 'TMAX', 'TMIN']:
249.             df_final[column] = df_final[column].astype(str).replace(nan,
'null')
250.             df_final[column] = df_final[column].astype(str).replace('nan',
'null')
251.         archivo_final = os.path.join(directorio_final, 'FINAL.csv')
252.         df_final.to_csv(archivo_final, index=False, encoding='utf-8')
253.         print(f'Archivo final concatenado guardado como {archivo_final}')

```

## A.3 GULpy

```
1. import limpiar_datos as ld
2. import datascraper as ds
3. import tkinter as tk
4. from tkinter import filedialog
5. from tkinter import *
6.
7. #Este array se puede cambiar luego por algo menos hard-coded
8. estaciones_scrapeables = [
9.     '25001', '25003', '25009', '25014', '25015', '25019', '25021', '25022',
10.    '25023', '25025',
11.    '25050', '25062', '25064', '25074', '25077', '25078', '25080', '25081',
12.    '25087', '25093',
13.    '25030', '25033', '25036', '25037', '25038', '25041', '25042', '25044',
14.    '25045', '25046',
15.    '25100', '25102', '25103', '25105', '25106', '25107', '25110', '25115',
16.    '25119', '25150',
17.    '25155', '25158', '25161', '25171', '25172', '25173', '25176', '25178',
18.    '25183', '25186'
19. ]
20.
21. if __name__ == "__main__":
22.     # Crear la ventana
23.     ventana = tk.Tk()
24.     ventana.title("Limpiar Datos")
25.     ventana.geometry("400x400")
26.
27.     espacio = tk.Label(ventana, text="")
28.     espacio.pack()
29.
30.     def scrapear_datos():
31.         carpeta = tk.filedialog.askdirectory()
32.         if not carpeta:
33.             return
34.         else:
35.             print(carpeta)
36.             etiqueta_scraper = tk.Label(ventana, text="Scrapeando datos")
37.             etiqueta_scraper.pack()
38.             ds.bulk_download(estaciones_scrapeables, carpeta)
39.             etiqueta_scraper.destroy()
40.             etiqueta_final = tk.Label(ventana, text="Datos scrapeados")
41.             etiqueta.after(1000, lambda: etiqueta_final.pack())
42.             etiqueta_final.after(2000, lambda: etiqueta_final.destroy())
43.             print("Datos scrapeados")
44.
45.     boton_scrapear = tk.Button(ventana, text="Scrapear datos",
46.                                command=scrapear_datos)
47.     boton_scrapear.pack()
48.
49.     espacio = tk.Label(ventana, text="")
50.     espacio.pack()
51.
52.     label_umbral = tk.Label(ventana, text="Introduce el % máximo de datos
53.                                faltantes aceptable")
54.     label_umbral.pack()
55.     porcentaje = tk.StringVar()
56.     porcentaje.set("100")
57.     entrada_umbral = tk.Entry(ventana, textvariable=porcentaje, width=4)
58.     entrada_umbral.pack()
```

```

52.
53.     espacio = tk.Label(ventana, text="")
54.     espacio.pack()
55.
56.
57.     espacio = tk.Label(ventana, text="")
58.     espacio.pack()
59.
60.     # Crear la etiqueta
61.     etiqueta = tk.Label(ventana, text="Selecciona la carpeta con los datos a
limpiar")
62.     etiqueta.pack()
63.
64.     # Crear la entrada mediante un botón que deje seleccionar una carpeta
65.     def seleccionar_carpeta():
66.         carpeta = tk.filedialog.askdirectory()
67.         if not carpeta:
68.             return
69.         else:
70.             print(carpeta)
71.             etiqueta_txt_csv = tk.Label(ventana, text="Limpiando archivos .txt
a .csv")
72.             etiqueta_txt_csv.pack()
73.             # Validar que el porcentaje sea un número
74.             try:
75.                 porcentaje = int(entrada_umbral.get())
76.             except ValueError:
77.                 etiqueta_error = tk.Label(ventana, text="Porcentaje no válido")
78.                 etiqueta_error.pack()
79.                 etiqueta_error.after( 1000, lambda: etiqueta_error.destroy())
80.                 return
81.
82.             #ld.procesar_lote_txt_a_csv(carpeta, porcentaje)
83.             ld.procesar_lote_txt_a_csv(carpeta, porcentaje)
84.             etiqueta_txt_csv.destroy()
85.             etiqueta_concentrados = tk.Label(ventana, text="Generando concentrados
mensuales")
86.             etiqueta_concentrados.pack()
87.             ld.generar_concentrados_mensuales(carpeta)
88.             ld.generar_concentrados_diarios(carpeta)
89.             ld.concatenar_concentrados(carpeta)
90.             etiqueta_concentrados.destroy()
91.             etiqueta_concatenados = tk.Label(ventana, text="Concatenando
concentrados")
92.             etiqueta_concatenados.pack()
93.             etiqueta_concatenados.after( 1000, lambda:
etiqueta_concatenados.destroy())
94.             etiqueta_final = tk.Label(ventana, text="Datos limpiados")
95.             etiqueta.after(1000, lambda: etiqueta_final.pack())
96.             etiqueta_final.after( 2000, lambda: etiqueta_final.destroy())
97.             print("Datos limpiados")
98.
99.             boton = tk.Button(ventana, text="Seleccionar carpeta",
command=seleccionar_carpeta)
100.             boton.pack()
101.
102.             # Mostrar la ventana
103.             ventana.mainloop()

```

## B. Manual de Usuario para informes

### B.1 Tablero inicial

En esta etapa del desarrollo, ya se cuenta con la información procesada y lista para su análisis, se puede observar que la interfaz inicial o informe (Figura 5.11) que se tienen funciones como por ejemplo el filtrado de estaciones (Figura 5.11, opción C), estas funcionalidades, se describen a continuación:

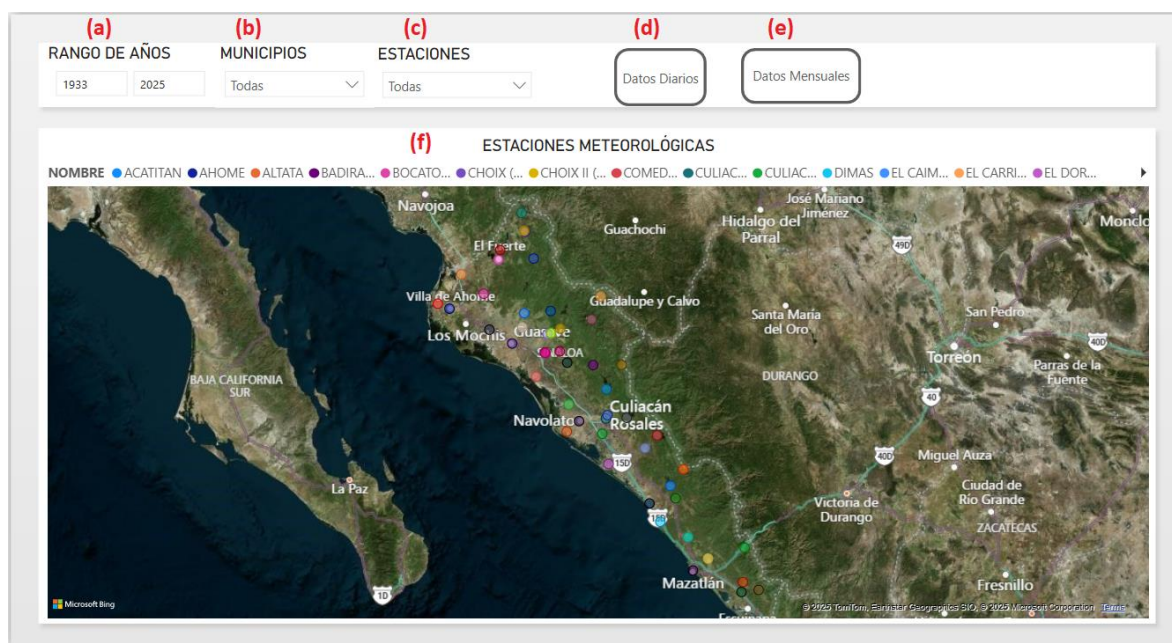


Figura 7.1. Informe de tablero inicial en PowerBI (Fuente: Elaboración propia)

En este informe es posible encontrar:

- a. **Rango de años:** El año inicial por defecto es el año más antiguo del que se tienen datos y el final el más reciente.
- b. **Municipios:** En este campo aparecen los municipios que contienen estaciones meteorológicas del estado de Sinaloa. Es posible filtrar la información de uno o más municipios al mismo tiempo, pero inicialmente se toman en cuenta todos los municipios.
- c. **Estaciones:** Es una lista que muestra los identificadores y nombres de las estaciones, y que permite filtrar una o más de estas. Por defecto se utilizan datos de todas las estaciones.

- d. **Datos diarios:** Al hacerle clic pulsando la tecla Ctrl, lleva al tablero que muestra la información diaria de todas las estaciones.
- e. **Datos mensuales:** Utilizando la combinación de teclas Ctrl + Clic lleva al tablero que presenta la información de los datos mensuales de todas las estaciones.

## B.2 Filtro superior para datos diarios

El filtro superior para este tablero y los que le siguen en su jerarquía es un poco diferente al tablero inicial, debido a que la estructura de la cabecera es la misma en todos los datos diarios y para evitar redundancias, solo se colocará una vez su explicación aquí:

Figura 7.2. Filtro superior para informes diarios (Fuente: Elaboración propia)

Dentro de esta sección superior podemos encontrar las siguientes funcionalidades:

- a. **Botón regresar:** Como su nombre lo indica, permite regresar al informe anterior.
- b. **Municipio:** Permite filtrar la información por municipio.
- c. **Estación:** Hace posible seleccionar una o más estaciones a analizar.
- d. **Rango de años:** Sirve para delimitar los años que tomen en cuenta los componentes gráficos.
- e. **Trimestres:** Es un filtro más especializado; permite filtrar entre los 4 trimestres del año.
- f. **Mes:** Para aún más precisión y poder llevar a cabo un mejor análisis de estacionalidad, se pueden seleccionar meses a analizar.

### B.3 Tablero de datos diarios

Los datos diarios presentan la información tal y como fue captada por los sensores, y aparte de poder seleccionar un rango de años, es posible magnificarla por trimestre, mes y día.

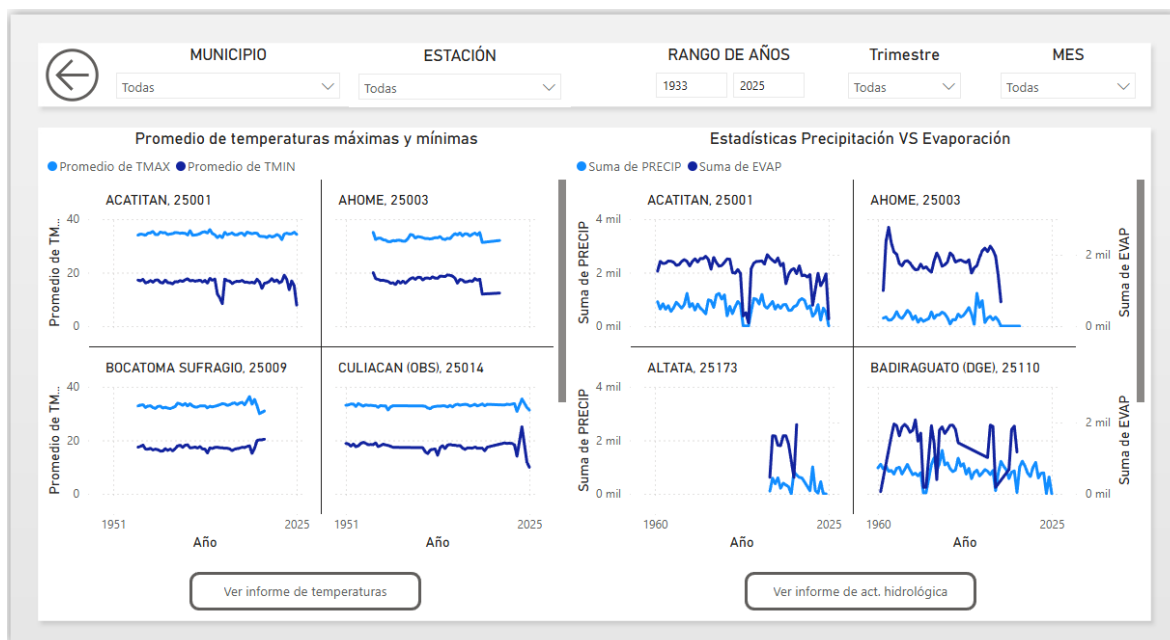


Figura 7.3. Tablero central para informes hidrometeorológicos diarios (Fuente: Elaboración propia)

## B.4 Cuerpo del informe de datos diarios

En el cuerpo de este informe podemos encontrar nuestras las tablas con información de cada una de las estaciones capturadas (o seleccionadas), en las cuales podremos aumentar o disminuir el rango de tiempo que toma en cuenta la gráfica, siendo conformado por los siguientes componentes:

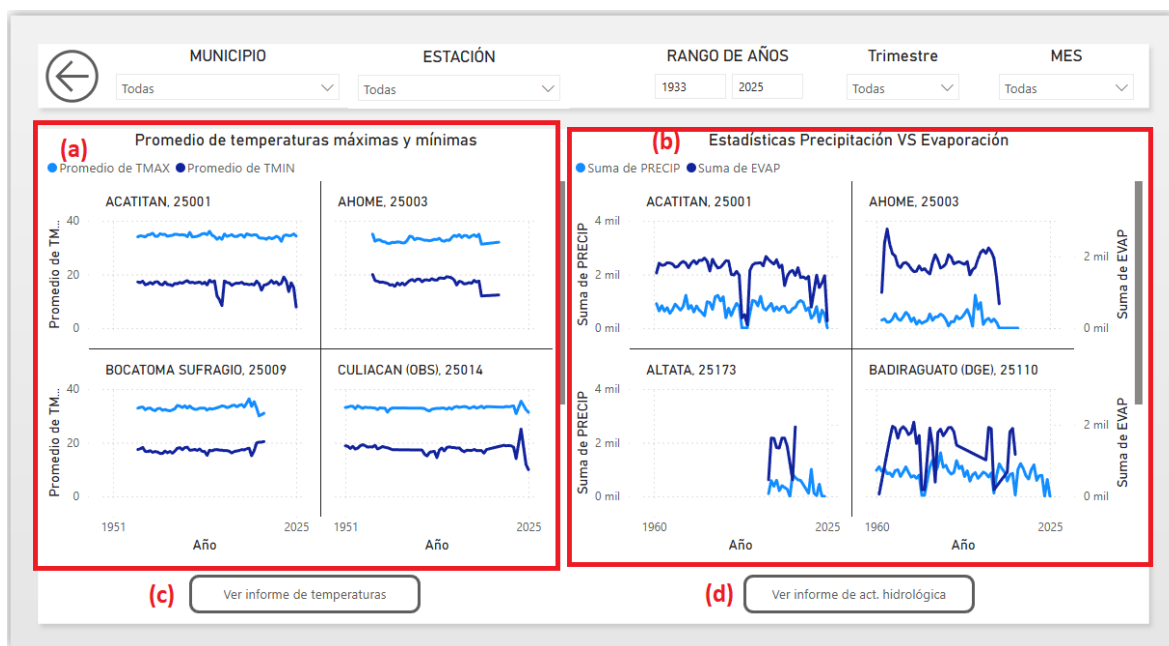


Figura 7.4. Cuerpo del informe del informe general para datos diarios. (Fuente: Elaboración propia)

- Gráficas de promedios de temperaturas:** Estas gráficas muestran el promedio del de temperaturas más altas registradas (en color azul claro) dentro del rango de tiempo seleccionado, y también muestra el promedio de temperaturas más bajas registradas (en azul oscuro).
- Gráficas de precipitación y evaporación:** Estas gráficas permiten visualizar la suma de las precipitaciones (en azul claro) y la suma de evaporaciones (en azul oscuro).
- Botón para informes de temperatura:** Haciendo Ctrl + Clic sobre este botón permite ver información más detallada sobre las temperaturas.
- Botón para informe de actividad hidrológica:** Lleva hacia la pantalla con información sobre la actividad hidrológica.

## B.5 Informe de temperaturas diarias

En este informe se pueden apreciar datos más específicos con respecto a las temperaturas.

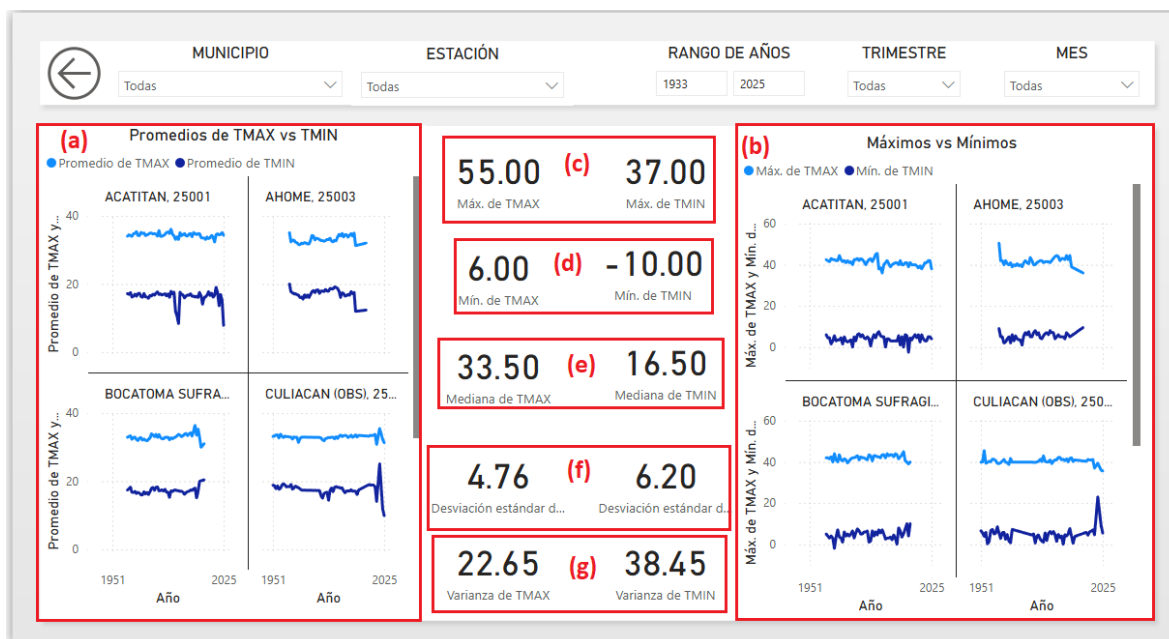


Figura 7.5 Cuerpo del informe de temperaturas diarias (Fuente: Elaboración propia)

El informe de temperaturas diarias permite ahondar más en datos específicos, y se compone por:

- Promedios de temperatura máxima y mínima:** Se mantienen la gráfica de la sección anterior con el fin de servir como una referencia gráfica más clara.
- Máximos vs mínimos:** Se toman las temperaturas más altas (en azul claro) y más bajas (azul oscuro) registradas en la serie de tiempo y las estaciones definidas.
- Sección de máximos:** De los campos de temperatura máxima y temperatura mínima, obtiene los valores más altos registrados dentro del rango de tiempo y estaciones especificadas.
- Sección de mínimos:** De las entradas de temperatura máxima y mínima, se obtienen las temperaturas más bajas registradas.
- Medianas:** Se obtienen los valores correspondientes a la mediana para las temperaturas máximas y mínimas.
- Desviación estándar:** Se obtiene la desviación de las temperaturas sobre el rango de tiempo seleccionado.

- g. **Varianza:** Calcula las varianzas de las temperaturas más altas y bajas para el rango de tiempo definido.

## B.6 Informe de actividad hidrológica diaria

Este informe, como su nombre lo indica, permite ver la actividad hidrológica utilizando diferentes parámetros como filtros, y obteniendo diferentes métricas de estos.

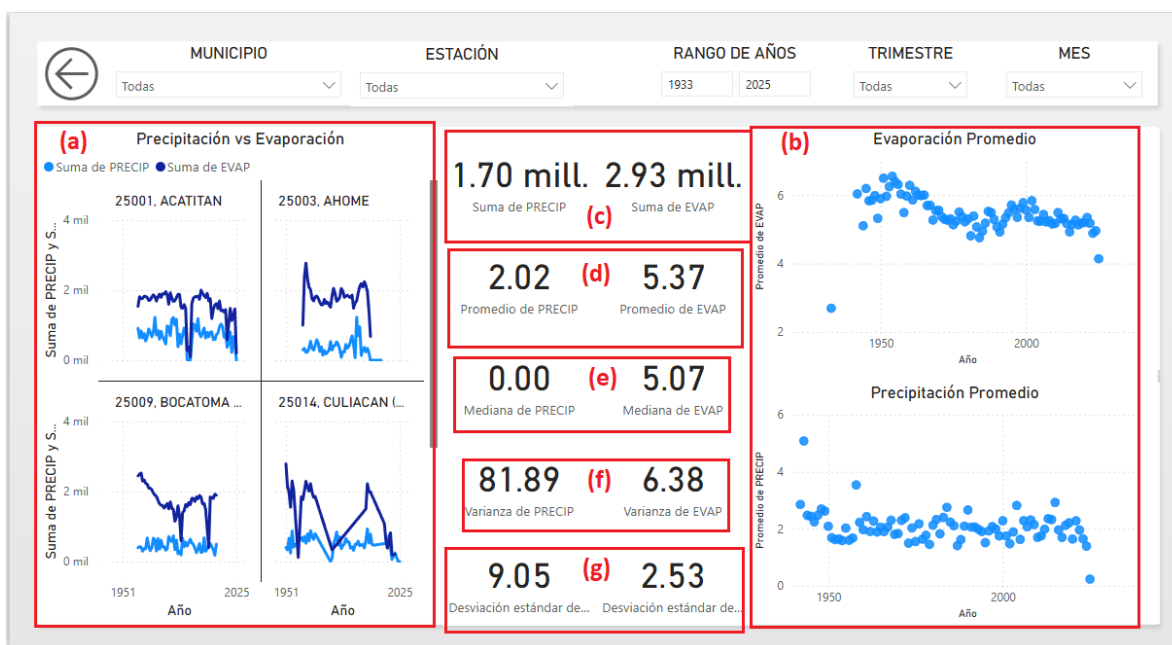


Figura 7.6. Cuerpo del informe de actividad hidrológica diaria (Fuente: Elaboración propia)

Con los datos obtenidos podemos visualizar las métricas en los siguientes componentes:

- Precipitación vs evaporación:** En estas gráficas es posible ver la suma de las precipitaciones (azul claro) y la suma de todas las evaporaciones (azul oscuro) de forma gráfica.
- Promedios hidrológicos:** En los gráficos de dispersión se pueden ver los promedios de evaporación y precipitación.
- Datos acumulados:** Es posible ver la acumulación de precipitación y evaporación.
- Promedios:** Permite visualizar los promedios de precipitación y evaporación.

- e. **Mediana hidrológica:** Se puede ver la tendencia de los datos. Debido a que Sinaloa es un clima relativamente seco, en donde las precipitaciones se presentan en un periodo corto del año, la mediana de estas tiende a ser cero.
- f. **Varianzas:** Se puede apreciar la varianza de precipitación y evaporación. Debido a lo mencionado anteriormente la varianza de las precipitaciones es mucho mayor.
- g. **Desviación estándar:** Muestra la desviación de precipitación y evaporación.

## B.7 Informes de datos mensuales

Tomando en consideración que los datos mensuales son de índole más general que los datos diarios, es posible utilizar un mayor número de elementos gráficos para representarlos y analizarlos.



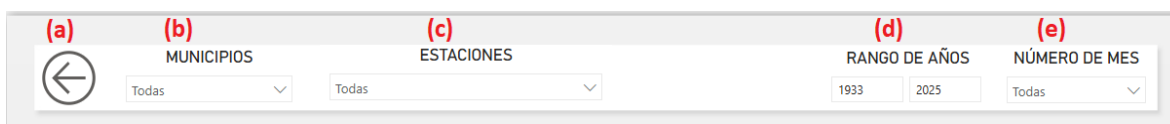
Figura 7.7. Informe concentrado mensual (Fuente: Elaboración propia)

Para la obtención de estos datos se les condensó de la siguiente manera:

- **Precipitaciones:** Son la suma mensual de las precipitaciones diarias.
- **Evaporaciones:** Es el promedio mensual de las evaporaciones diarias.
- **Temperatura máxima:** Es representado por la temperatura más alta del mes.
- **Temperatura mínima:** Representada por la temperatura más baja del mes.

## B.8 Filtro superior para datos mensuales

El filtro para los concentrados mensuales es similar al filtro para los datos diarios:



The image shows a horizontal filter bar with five labeled components: (a) a circular back button with a left-pointing arrow; (b) a dropdown menu labeled 'MUNICIPIOS' with 'Todas' selected; (c) a dropdown menu labeled 'ESTACIONES' with 'Todas' selected; (d) two input fields for 'RANGO DE AÑOS' containing '1933' and '2025'; and (e) a dropdown menu labeled 'NÚMERO DE MES' with 'Todas' selected.

Figura 7.8. Filtro superior de datos mensuales

- a. **Botón regresar:** Permite regresar al tablero inicial
- b. **Selector municipio:** Permite elegir uno o más municipios
- c. **Selector estaciones:** Permite seleccionar una o más estaciones
- d. **Rango de años:** Permite elegir entre un rango de años o un año en particular
- e. **Selector de número de mes:** Da la posibilidad de elegir uno o más meses

### C. Lista de estaciones meteorológicas

| ESTACIÓN | MUNICIPIO            | NOMBRE                                | LATITUD        | LONGITUD    | ALTITUD |
|----------|----------------------|---------------------------------------|----------------|-------------|---------|
| 25001    | SAN IGNACIO          | ACATITAN                              | 24.097222<br>2 | -106.666667 | 96      |
| 25003    | AHOME                | AHOME                                 | 25.918888<br>9 | -109.172222 | 10      |
| 25009    | EL FUERTE            | BOCATOMA<br>SUFRAGIO                  | 26.070833<br>3 | -108.784444 | 31      |
| 25014    | CULIACÁN             | CULIACAN<br>(OBS)                     | 24.634722<br>2 | -107.440556 | 22      |
| 25015    | CULIACÁN             | CULIACAN<br>(DGE)                     | 24.791944<br>4 | -107.398056 | 60      |
| 25019    | CHOIX                | CHOIX II (DGE)                        | 26.71          | -108.326111 | 239     |
| 25021    | SAN IGNACIO          | DIMAS                                 | 23.723888<br>9 | -106.781667 | 19      |
| 25022    | AHOME                | EL CARRIZO                            | 26.268611<br>1 | -109.037222 | 8       |
| 25023    | EL FUERTE            | EL FUERTE                             | 26.416388<br>9 | -108.617778 | 85      |
| 25025    | EL FUERTE            | PRESA MIGUEL<br>HIDALGO Y<br>COSTILLA | 26.515833<br>3 | -108.597222 | 144     |
| 25030    | ANGOSTURA            | EL PLAYON                             | 25.222222<br>2 | -108.190278 | 6       |
| 25033    | BADIRAGUATO          | EL VAREJONAL                          | 25.094444<br>4 | -107.393056 | 119     |
| 25036    | COSALÁ               | GUADALUPE DE<br>LOS REYES             | 24.268888<br>9 | -106.519722 | 696     |
| 25037    | SALVADOR<br>ALVARADO | GUAMUCHIL<br>(AARSP)                  | 25.469444<br>4 | -108.091667 | 44      |
| 25038    | GUASAVE              | GUASAVE (DGE)                         | 25.5625        | -108.461111 | 22      |
| 25041    | BADIRAGUATO          | GUATENIPA                             | 25.343055<br>6 | -107.220833 | 512     |
| 25042    | AHOME                | HIGUERA DE<br>ZARAGOZA                | 25.97          | -109.304444 | 10      |
| 25044    | CHOIX                | HUITES                                | 26.892777<br>8 | -108.354167 | 269     |
| 25045    | SAN IGNACIO          | IXPALINO                              | 23.968055<br>6 | -106.608333 | 69      |
| 25046    | SINALOA              | JAINA                                 | 25.894444<br>4 | -108.025    | 130     |

|       |                      |                               |                |             |      |
|-------|----------------------|-------------------------------|----------------|-------------|------|
| 25050 | ELOTA                | LA CRUZ                       | 23.913888<br>9 | -106.904167 | 9    |
| 25062 | MAZATLÁN             | MAZATLAN<br>(OBS)             | 23.216666<br>7 | -106.410556 | 4    |
| 25064 | MOCORITO             | MOCORITO<br>(DGE)             | 25.482777<br>8 | -107.921111 | 87   |
| 25074 | CONCORDIA            | POTRERILLOS                   | 23.453611<br>1 | -105.825833 | 1572 |
| 25077 | MOCORITO             | ROSA MORADA                   | 25.366666<br>7 | -107.8375   | 134  |
| 25078 | ROSARIO              | ROSARIO                       | 22.991944<br>4 | -105.861389 | 32   |
| 25080 | GUASAVE              | RUIZ CORTINEZ                 | 25.704444<br>4 | -108.719444 | 20   |
| 25081 | CULIACÁN             | SANALONA II                   | 24.8           | -107.1625   | 104  |
| 25087 | COSALÁ               | SANTA CRUZ DE<br>AYALA        | 24.4875        | -106.954167 | 109  |
| 25093 | BADIRAGUATO          | SURUTATO                      | 25.808333<br>3 | -107.560278 | 1460 |
| 25100 | CHOIX                | YECORATO<br>(CFE)             | 26.429444<br>4 | -108.217778 | 400  |
| 25102 | CULIACÁN             | ESCUELA DE<br>BIOLOGIA (UAS)  | 24.8275        | -107.380278 | 69   |
| 25103 | SINALOA              | EL CAIMAN                     | 25.872222<br>2 | -108.323611 | 74   |
| 25105 | SINALOA              | EL MEZQUITE                   | 25.661111<br>1 | -108.06     | 115  |
| 25106 | NAVOLATO             | EL TIGRE                      | 24.936111<br>1 | -107.82     | 17   |
| 25107 | BADIRAGUATO          | SACA DE AGUA                  | 26.043055<br>6 | -107.456667 | 1105 |
| 25110 | BADIRAGUATO          | BADIRAGUATO<br>(DGE)          | 25.341944<br>4 | -107.543056 | 191  |
| 25115 | SALVADOR<br>ALVARADO | GUAMUCHIL<br>(DGE)            | 25.466666<br>7 | -108.083333 | 44   |
| 25119 | MAZATLÁN             | SIQUEROS                      | 23.340555<br>6 | -106.240556 | 40   |
| 25150 | ROSARIO              | LAS TORTUGAS                  | 23.095833<br>3 | -105.845833 | 65   |
| 25155 | CHOIX                | CHOIX (OBS)                   | 26.7           | -108.316667 | 224  |
| 25158 | SINALOA              | EL PALMAR DE<br>LOS SEPULVEDA | 25.715277<br>8 | -107.916389 | 262  |

|       |          |             |                |             |     |
|-------|----------|-------------|----------------|-------------|-----|
| 25161 | CULIACÁN | EL DORADO   | 24.323611<br>1 | -107.367778 | 10  |
| 25171 | NAVOLATO | NAVOLATO    | 24.765555<br>6 | -107.703056 | 16  |
| 25172 | SINALOA  | SAN JOAQUIN | 25.667777<br>8 | -108.02     | 144 |
| 25173 | NAVOLATO | ALTATA      | 24.658333<br>3 | -107.842222 | 4   |
| 25176 | MAZATLÁN | EL QUEMADO  | 23.5625        | -106.466667 | 50  |
| 25178 | GUASAVE  | ZOPILOTE    | 25.730555<br>6 | -108.345833 | 37  |
| 25183 | COSALÁ   | COMEDERO    | 24.618055<br>6 | -106.816667 | 310 |
| 25186 | ROSARIO  | OTATITAN    | 23.013888<br>9 | -105.666667 | 93  |

#### D. Carta de no inconveniente de titulación

La carta presentada a continuación da autorización al Instituto Tecnológico de Culiacán de utilizar el documento de investigación actual como evidencia para la realización de trámites de titulación profesional.

-

RFC: TNM140723GFA,  
AV. JUAN DE DIOS BÁTIZ 310 PTE. GUADALUPE 80220 CULIACÁN SINALOA

Culiacán, Sin., a 20 de junio de 2025.

Asunto: Carta de no inconveniencia.

**ZURIEL DATHAN MORA FELIX**

Asesor interno del Instituto tecnológico de Culiacán

Por medio de la presente, se hace constar que el(la) C. **FERNANDO ALONSO MORENO MILLAN**, alumno(a) del Instituto Tecnológico de Culiacán de la carrera de **ING. EN SISTEMAS COMPUTACIONALES** y con número de control **20170759**, ha concluido satisfactoriamente su Proyecto de Residencia Profesional y aspira a obtener el título profesional con el informe del proyecto denominado **ANÁLISIS TEMPORAL DE DATOS METEOROLÓGICOS Y SU RELACIÓN CON EL CAMBIO CLIMÁTICO Y SEQUÍA EN LA ZONA CENTRO DE SINALOA**, que realizó en esta empresa, un período **Enero-Junio del 2025**.

Por lo que no se tiene ningún inconveniente en el uso de la información contenida en dicho informe para que se utilice únicamente para los fines antes mencionados y se concede la autorización para su uso ante el Instituto Tecnológico de Culiacán, con el fin de que [el/la] interesado(a) alcance su objetivo.

Atentamente

TECNOLÓGICO NACIONAL DE MÉXICO/INSTITUTO TECNOLÓGICO DE CULIACAN  
TNM140723GFA

  
\_\_\_\_\_  
**DR. JESÚS GABRIEL RANGEL PERAZA**  
PROFESOR INVESTIGADOR DE TIEMPO  
COMPLETO

Figura 7.9. Carta de no inconveniente de titulación