

Adquisición y análisis de datos hidrometeorológicos open source.

Acquisition and analysis of open source hydrometeorological data.

Sorangela Cárdenas-Rodríguez¹, Aldo Pardo-García², Orlando Maldonado-Bautista³

^{1,2,3}*Universidad de Pamplona, Pamplona - Colombia*

ORCID: ¹[0000-0002-1770-2741](https://orcid.org/0000-0002-1770-2741), ²[0000-0003-2040-9420](https://orcid.org/0000-0003-2040-9420), ³[0000-0002-2281-8546](https://orcid.org/0000-0002-2281-8546)

Recibido: 04 de marzo 2024.

Aceptado: 19 de abril de 2024.

Publicado: 1 de mayo de 2024.

Resumen- En el presente artículo se realiza una revisión enfocada a la evaluación de diversas fuentes de datos hidrometeorológicos con el potencial de ser incorporadas en un sistema de alerta temprana para la prevención de inundaciones en la cuenca del río Arauca. En el análisis se abordan estaciones terrestres, radares y estaciones aéreas, detallando las particularidades de los datos recopilados y los procedimientos necesarios para su acceso. Además, se profundiza en consideraciones relacionadas con la calidad y confiabilidad de estos datos, al tiempo que se exponen los desafíos inherentes a los procesos de recolección y análisis de estos. Este trabajo contribuye significativamente a la comprensión y la implementación de un sistema efectivo de alertas tempranas para la mitigación de inundaciones en la cuenca del río Arauca.

Palabras clave: datos, predicción, precipitación, procesamiento de datos, imágenes satelitales.

Abstract— This review article focuses on assessing various sources of hydrometeorological data for their utilization in an early warning system for flood prevention in the Arauca River basin. The analysis covers ground stations, radars, and aerial stations, detailing the specifics of the collected data and the procedures required for access. Additionally, it delves into considerations related to data quality and reliability, while outlining the challenges inherent in the data collection and analysis processes. This work significantly contributes to the understanding and implementation of an effective early warning system for flood mitigation in the Arauca River basin.

Keywords: data, prediction, precipitation, data processing, satellite images.

*Autor para correspondencia.

Correo electrónico: sorangela.cardenas@unipamplona.edu.co (Sorangela Cárdenas Rodríguez).

La revisión por pares es responsabilidad de la Universidad de Santander.

Este es un artículo bajo la licencia CC BY (<https://creativecommons.org/licenses/by/4.0/>).

Como citar este artículo: S. Cárdenas-Rodríguez, A. Pardo-García y O. Maldonado-Bautista, "Adquisición y análisis de datos hidrometeorológicos open source", *Aibi revista de investigación, administración e ingeniería*, vol. 12, no. 2, pp. 196-203 2024, doi: [10.15649/2346030X.3852](https://doi.org/10.15649/2346030X.3852)

I. INTRODUCCIÓN

Las estadísticas hidrometeorológicas como la precipitación, evaporación, nivel del río, son fundamentales para la predicción de inundaciones [1], [2]. El propósito de la adquisición de datos y las estadísticas posteriores es estimar la probabilidad de eventos reales, la gravedad de los mismos y ubicarlos en el espacio y tiempo [3]. Los datos meteorológicos se requieren en tiempo real para la previsión de pronósticos y advertencias de fenómenos naturales como: inundaciones [4], [5], y posibles deslizamientos de tierra teniendo en cuenta las condiciones del suelo [6].

Las inundaciones se pueden monitorear [7] y predecir mediante una combinación de observaciones y modelados [8], en los últimos años se han involucrado en la gestión de recursos hídricos mediante técnicas de aprendizaje de máquina, la interpretación de imágenes satelitales, variación de la cubierta vegetal, índices de pluviosidad, nivel de los ríos, entre otros [9]. También se recomienda obtener información desde el punto de vista de la observación que hacen en forma manual los pobladores de la región en estudio [10].

Otro uso de las bases de datos libres está ligado a la producción del campo, usando herramientas para el seguimiento de datos meteorológicos, y mejorando los cultivos para mejorar la cadena agrícola y ser competitivos cada día [11].

La teledetección por imágenes en el aprendizaje automático (Machine Learning ML) ha mostrado su potencial para la clasificación efectiva, aunque la implementación de una clasificación no es sencilla, existen variedad de métodos dependiendo de la aplicación, que van desde medición a variables meteorológicas hasta la obtención de imágenes avanzadas para la detección de fallas en procesos industriales [12].

Las fortalezas del aprendizaje automático incluyen la capacidad de manejar datos de alta dimensionalidad y mapear clases con características muy complejas [13], que permiten agilizar tareas como el procesamiento de datos, evaluación de peticiones hasta el reconocimiento de personas [14], siendo las redes neuronales convolucionales unas de las más usadas para tratamiento de imágenes [15]. Dentro de los métodos más conocidos están máquinas de soporte vectorial, árboles de decisión (DT), bosques aleatorios, redes neuronales artificiales y k-vecinos más cercanos (k-NN) [16], [17]. Además, el uso de técnicas de aprendizaje profundo ha expandido las aplicaciones, brindando soluciones eficientes para problemas complejos en el tratamiento de datos hidrometeorológicos [30].

Pero a su vez existen diferentes retos frente a la obtención de datos con calidad, los datos incompletos en las series de tiempo, diferentes autores proponen técnicas en [18], [19], [20] para aplicar y obtener una data completa y confiable. En la sección dos se describen las principales fuentes disponibles para la recolección de datos hidrometeorológicos, igualmente se aborda el tema de la calidad y fiabilidad de los datos, y se analizan los desafíos y posibles soluciones en la recolección de datos y se sintetizan algunas conclusiones sobre la revisión realizada.

El objetivo de este artículo es realizar una revisión de la literatura y compilar las publicaciones relacionados con adquisición de datos hidrometeorológicos de acceso libre, satelitales y de estaciones físicas, tomados de revistas científicas, tesis de grado, y manuales.

Con el fin de contar con información científica que de respuesta a la pregunta de investigación ¿El utilizar datos open source tomados de satélites y estaciones físicas provee información certera y confiable para prevenir inundaciones? Con ello en un futuro estos permiten comprender mejor los procesos climáticos y desarrollar con un margen de incertidumbre menor modelos eficientes, ya que estos datos serán los valores de las variables que se procesan en los modelos hidrodinámicos seleccionados, que proporcionarían pronósticos sobre la ocurrencia y magnitud de inundación que afecten a la población ribereña.

II. METODOLOGÍA

La investigación tiene como objetivo recolectar información de fuentes primarias y secundarias y realizar un análisis crítico, que responda a la pregunta de investigación, ¿El uso de datos open source tomados de satélites y estaciones físicas provee información certera y confiable para prevenir inundaciones?

En este artículo las herramientas utilizadas para la recolección de datos incluyen la recopilación documental a través de la revisión del estado del arte en diferentes lugares del mundo sobre datos de acceso libre y bases de datos incluyendo manuales de uso [21]. Se consultaron 57 fuentes de las bases de datos Scielo, Springer, ScienceDirect, Elsevier, así como libros, tesis doctorales, manuales.

Para la búsqueda de la información se realizó a través de Google Académico, usando palabras claves como: datos satelitales, hidrometeorológicas, adquisición de datos, estaciones terrestres, predicción de inundaciones. La combinación de términos como “datos satelitales + predicción inundaciones”, “descargar satélites + hidrometeorológicos”, resultó en información relevante para la investigación. Por otro lado en la descarga de datos se usaron las técnicas de recolección de datos dinámicas [22], constituidas por las bases de datos de las estaciones meteorológicas que están en el cauce del río y afluentes.

Descripción de las fuentes de datos hidro meteorológicos

La información y la adquisición de los datos han experimentado hoy en día una gran transformación tecnológica, pasando de nivel local a uno global, a través de internet mediante bases de datos de libre acceso [23]. La adquisición de datos hidrometeorológicos implica la recolección de variables climáticas, e hidrológicas como caudal y nivel de ríos mediante la utilización de una variedad de instrumentos y tecnologías [24], como estaciones tradicionales hasta tecnologías satelitales avanzadas. Estos datos se obtienen tanto a nivel terrestre con estaciones meteorológicas, hidrométricas y radares meteorológicos, como a nivel aéreo a través de satélites equipados con sensores remotos [4]. A pesar de los avances tecnológicos, la adquisición de datos sigue enfrentándose a desafíos como la calidad de los datos [25], la cobertura geográfica y el costo de la infraestructura necesaria [3].

III. RESULTADOS, ANÁLISIS E INTERPRETACIÓN

A continuación, una descripción de cómo se recopilan los datos hidrometeorológicos de acuerdo con la ubicación de las fuentes; terrestres y aéreas.

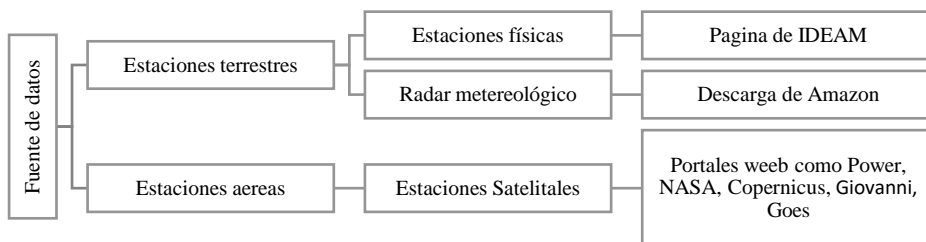


Figura 1: Clasificación de las fuentes de datos hidrometeorológicos según su ubicación.
Fuente: Elaboración propia.

1. Estaciones terrestres

Dentro de las estaciones terrestres físicas se tienen las estaciones meteorológicas, conformadas por componentes destinados a medir y registrar diversas variables meteorológicas que miden la precipitación, la temperatura, el viento, los niveles de los ríos, entre otras variables [7]. De otro lado, están las estaciones hidrométricas, ubicadas en un lugar fijo y específico del río, donde se toman y se realizan diversas operaciones con el propósito de determinar el caudal circulante en un momento y tiempo determinado [26].

El Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) en Colombia juega un papel fundamental en la adquisición y análisis de datos para la predicción de inundaciones, cuenta con una amplia red de estaciones meteorológicas distribuidas en todo el país [27]. Estas estaciones recolectan datos en históricos y tiempo real sobre la temperatura, la presión mínima, la velocidad y dirección del viento, la precipitación entre otros. Los datos se utilizan para monitorear las condiciones locales y regionales, así como para calibrar y validar los datos satelitales [28].

El IDEAM tiene 1735 estaciones hidrológicas en todo el país [29], las cuales se pueden visualizar en la plataforma web, (<http://fews.ideam.gov.co>). Para el estudio de caso el río Arauca, en la zona oriente se tiene en cuenta cinco estaciones hidrometeorológicas que conforman el caso de estudio, ubicadas en lugares estratégicos de la vertiente del río Arauca ver tabla 1.

Tabla 1: Estaciones hidrológicas y/o meteorológicas en el cauce del río Arauca.

Estación	Tipo de estación	Tecnología	Dpto. Municipio	Instalación estación
San Bernardo	Meteorológica pluviométrica	Convencional	N. Santander. Toledo	15/10/1973
Peña de los Micos	Hidrométrica Limnigráfica	Automática telemetría		
Alcaraván				Arauca Arauquita
Arauquita	Meteorológica pluviométrica	Convencional		15/06/1972
Puente Internacional	Hidrométrica Limnigráfica		Arauca	15/11/1997
Arauca	Meteorológica pluviométrica		Arauca	15/05/1983

Fuente: Adaptado del catálogo de estaciones del IDEAM [28], [29].

Los Radares meteorológicos se encuentran en las estaciones terrestres, son un sistema de medición remoto y eficiente que realiza mediciones en cuatro dimensiones (4D), tres espaciales y una temporal: largo, alto y ancho respectivamente, y la cuarta dimensión que visualiza la evolución de los fenómenos de precipitación en el tiempo. Para las observaciones en la predicción del tiempo este instrumento aporta información relevante [30] ya que puede medir grandes áreas y tomar miles de datos en minutos [31].

En Colombia el IDEAM cuenta con 10 radares meteorológicos, de los cuales dos de ellos tienen influencia en el departamento de Arauca, el radar Carimagua que se encuentran en el departamento del Meta, con un rango de acción de 240 km, alcanzando a Casanare y la zona sur del departamento de Arauca, y en el 2019 [32] fue instalado en Barrancabermeja el radar meteorológico para la emisión de alertas tempranas, que beneficiara a los municipios cercanos en un rango de 240 km, con un alcance en los departamentos de Santander, Norte de Santander y la zona norte de Arauca [33]. La información de los radares se encuentra disponible en Amazon, para descargar se debe acceder a Amazon S3 y especificar día, mes, año, nombre del radar, y dato a descargar [34].

2. Estaciones áreas

Los satélites artificiales son equipados con sensores remotos de tipo activos y/o pasivos, que pueden captar información para diferentes bandas o regiones del espectro, que obtienen imágenes durante el día y noche, siendo la teledetección de radar la más ampliamente utilizada, por capturar imágenes de noche y sin importar el clima [4]. Siendo esencial para seguimiento en el tiempo mediante procesamiento de imágenes en diferentes procesos [35].

Como afirma Bonilla en [25] los sensores pasivos miden la energía radiante reflejada, esta energía se convierte en cantidades biogeo-físicas como temperatura, precipitación, humedad del suelo, como ejemplo los satélites MODIS, Landsat, TM y ETM, entre otros [25]. Los sensores activos, lanzan rayos de radiación sobre el sistema de la tierra atmosfera y miden la radiación retro dispersada, la que se convierte en parámetros geofísicos, como el radar de precipitación, LIDAR [36]. Los satélites pueden ser Geoestacionarios u Órbita terrestre baja (LEO) dependiendo de la altitud y de tipo polar o no polar de acuerdo con la dirección de rotación [37].

En [38] se explica como una imagen satelital presenta cuatro dimensiones: el tamaño en píxel por área en metros o resolución espacial, las bandas o anchura de las regiones del espectro que puede captar el sensor o resolución espectral, la frecuencia con que un satélite observa la misma área de la tierra o resolución temporal [26].

En la tabla se presenta la descripción de los satélites que captan información para la gestión de recursos hídricos se describe: año de lanzamiento, tipo de sensores abordo, resolución espacial y temporal, entre otros parámetros.

Tabla 2: Descripción de algunos de los satélites hidrometeorológicos para América.

Satélite / Año	Espacial	Temporal	Espectral	Tipo orbita / Altitud
Gpm 2014	5.2 km	3 horas	Ku 13.6 GHz, Ka 35.5 GHz	LEO / 407 km
Goes 2016	500 m IR 1 km	10 min	16 bandas	GEO / 35786 km
Sentinel 2 A 2B 3 / 2015-17	10, 20 y 60 m	2, 5, 6 días	13, 21 Bandas	LEO / 786 km
Lansat 9/ 2021	15, 30 y 80 m	8 días	11 bandas	LEO / 705 km

Fuente: Adaptado con base a los aportes de [39], [40], [41], [42], [43].

Existe una relación directa entre el clima y las variaciones en las fuentes hídricas, compuestas por varios elementos meteorológicos como temperatura, humedad, precipitación, velocidad del viento, radiación solar, presión atmosférica, entre otros, estos elementos presentan variación espacial y temporal [44].

A continuación, se mencionan los portales con información satelital para la gestión de recursos hídricos, se describe el uso y los enlaces para acceder a las bases de datos.

1. Power Nasa: 1981 – actualidad.

Pertenece a la Nasa y contiene información sobre precipitación, flujos solares, velocidad del viento, radiación solar, temperatura, humedad, presión del viento, enlace <https://power.larc.nasa.gov/>.

2. Giovanni: 2016 – actualidad

Sistemas de asimilación de datos terrestres mundiales y norteamericanos. Contiene información sobre análisis meteorológico, radiación solar superficial, Precipitación, textura del suelo, clasificación de la vegetación e índice de área foliar, topografía, enlace <https://giovanni.gsfc.nasa.gov/giovanni/>.

3. Copernicus: 2014 – actualidad

Programa de Observación de la Tierra de la Unión Europea, contiene variables medio ambientales, corrientes marinas, monitoreo de inundaciones, vapor de agua, topografía, superficie del mar, temperatura, biogeoquímica, índice UV. Satélite sentinel. Enlace <https://scihub.copernicus.eu/dhus/#/home>.

4. Goes 16: 2016 – actualidad.

Programa de observación de la NASA, NOAA, contiene información sobre nubosidad, vapor de agua, temperatura, de Temperatura de la superficie, aerosoles, viento, incendios, precipitación acumulada, seguimiento de ceniza volcánica, descargas eléctricas. Enlace <http://www.pronosticosyalertas.gov.co/Goes-portlet/index.html>.

2.1. Adquisición de datos satelitales

Los datos son tomados del portal Power NASA, con las coordenadas de las estaciones del IDEAM y se procede a establecer estaciones virtuales para tomar puntos de variables como precipitación, humedad relativa, pendiente, velocidad del viento, temperatura y tipo de suelo.

Para descargar la información se debe realizar los siguientes pasos:

Se ingresa al portal a través del enlace <https://power.larc.nasa.gov/data-access-viewer/>.

1. Se selecciona la comunidad, ejemplo energías renovables.
2. Se selecciona el promedio temporal, dependiendo de la variable, por horas, diario, mensual.
3. Ingresar la ubicación de donde se toma los datos, latitud y longitud
4. Seleccionar la temporalidad de los datos (fecha).
5. Seleccionar el formato para descargar (archivo csv).
6. Por último, se selecciona el parámetro de búsqueda clasificados en grupos: flujos solares y relacionados, parámetros solares, temperaturas, humedad/precipitación, presión del viento.
7. Después de seleccionado los datos se procede a descargarlos.

2.2. Adquisición de imágenes satelitales, Satélite GOES 16

América cuenta con los satélites Geoestacionario GOES 16 y 18, con 16 bandas que permite visualizar el espectro visible, infra rojo IR cercano hasta el IR lejano, mediante estos se puede tener Información de Temperatura de la superficie, detección de aerosoles, tope de nubes de presión, tope de nubes temperatura, vientos derivados por satélite, desplazamiento de la nubes, puntos calientes/incendios, estimativos de precipitación acumulada, cantidad de agua precipitable que tiene la atmosfera (datos vapor de agua), seguimiento de ceniza volcánica y detección de descargas eléctricas [42], [43], [45]. Los datos satelitales cuentan con preprocesamiento y filtrado realizado por programas especializados para corrección.

La metodología para realizar la descarga de imágenes se describe a continuación:

1. Se accede a través del enlace del servidor del IDEAM <http://181.225.72.46:45128/almacen/interno/satelite/> o también directamente desde la página de la NOAA <https://www.star.nesdis.noaa.gov/GOES/sector.php?sat=G16§or=nsa#>.

2. Se visualiza la información por departamentos y se selecciona el de interés, se selecciona el mes y día. (para Canal 13 se encuentran imágenes satelitales cada 10 minutos en formato JPG) del año 2023.
4. Si se requiere información de años anteriores se encuentran alojadas en Amazon y deben ser descargadas a través de otros enlaces o solicitando información a través de un PQR en la página del IDEAM.

2.3. Calidad y fiabilidad de los datos

Después de obtener los datos históricos se revisa la información recolectada, y pasa al preprocesamiento con el fin de analizar y hacer estimaciones más precisas [46].

La utilidad de un sistema de pronóstico de inundaciones depende en gran medida de la integridad y continuidad de los datos disponibles. La incertidumbre en los datos de entrada puede deberse a datos erróneos o faltantes [4], por esto se hace necesario la homogenización o preparación de datos [8] de las variables que influyen en el modelo. Esto se puede realizar de tres formas:

1. Eliminar los datos faltantes.
2. Rellenar los huecos de la serie de tiempo a partir de técnicas como interpolación.
3. Rellenar los huecos de la serie de tiempo creando datos con técnicas de inteligencia artificial.

La validación de los datos se realiza a través de técnicas para detectar los errores comunes que pueden ocurrir, dependiendo de la densidad de la red de observaciones en relación con la variación espacial de la variable [3]. Las comprobaciones en su mayoría se basan en comprobaciones relativas y físico estadísticas usando series temporales, correlación, regresión múltiple y técnicas de ajuste de superficie, también existen otras como verificación absoluta y verificación relativa [26].

En la literatura reciente, se han explorado diversas metodologías para abordar la problemática de los datos faltantes en estudios hidrológicos. Una investigación sobre el análisis de componentes principales bayesiano mostró que este método superó al KNN en casos de grandes brechas faltantes continuas [47]. Otro estudio sobre metodologías estadísticas y de aprendizaje automático aplicadas a la imputación de anomalías y datos faltantes, incluyendo imputación de media aritmética, análisis de componentes principales, métodos basados en regresión y métodos de imputación múltiple, destacó que los avances metodológicos en otros campos de investigación son relevantes para un uso más intensivo de estos métodos en hidrología [48]. En el estudio de Yang [49], se utilizaron métodos de imputación de la media de puntos cercanos para completar los datos faltantes del conjunto de datos integrado del nivel del agua y datos atmosféricos.

2.4. Metodología para la adquisición de datos del IDEAM

Paso 1: Se procede a descargar los datos de la página web del IDEAM como archivo “*.cvs” y se organiza en columnas en “*.xls”, correspondiente a las estaciones hidrológicas y meteorológicas de las variables independientes como precipitación, nivel y la variable dependiente caudal. Estas variables se encuentran registradas en los datos históricos organizados por años, meses y días, de las que se hace referencia a los años en los que existe más información en meses de invierno y registro de crecientes, como se observa en la Tabla 3.

Tabla 3: Presentación de los datos históricos del 2018 de la estación Peña de los Micos.

FECHA	CAUDAL	NIVEL (cm)	PRECIPITACIÓN (mm)
17/04/2018 0:00	118,00783	206,583333	0,1
18/04/2018 0:00	179,56247	241,875	22,8
19/04/2018 0:00	345,52103	309,895833	1
20/04/2018 0:00	191,57467	248,895833	1,3
21/04/2018 0:00	147,46897	224,979167	2,3
22/04/2018 0:00	126,60229	212,291667	0,1

Fuente: Elaboración propia.

En la Tabla 3 se observan los datos de la estación hidrométrica Peña de los Micos y estación meteorológica San Bernardo, en las cuales están referenciados la variable de precipitación, nivel y caudal del año 2018.

Paso 2: Se reorganiza la información, se ubican las columnas de las variables independientes que se encuentran en el mismo espacio con respecto a la variable dependiente de caudal, y se procede a homogenizar la data, con el propósito de trabajar de forma más eficiente.

Paso 3: Al homogenizar la información obtenida de las estaciones hidrometeorológicas, se procede a obtener el modelo de tendencia lineal, y de potencia, a través del uso de la herramienta Excel, utilizando el módulo de datos estadísticos de regresión y el de relaciones potenciales se realiza por el método de convergencia.

En la Figura 2, se visualiza la estación de hidrométrica de Peña de los Micos del año 2018, información de nivel del río en metros del año 2018.

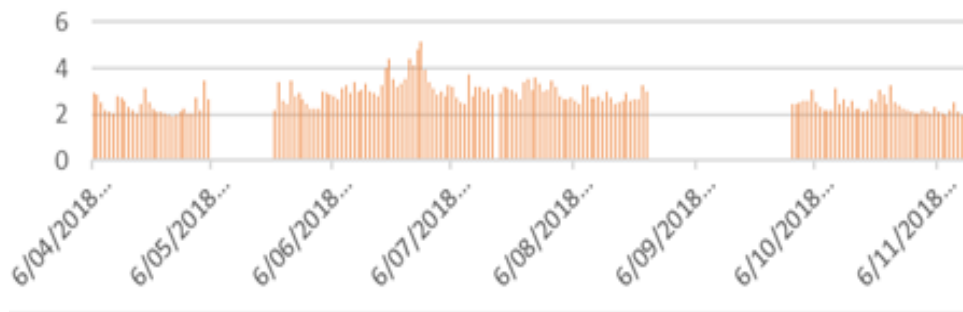


Figura 2: Información sin homogenizar.
Fuente: Elaboración propia.

Se observa que existen fechas en las que no hay datos por daños en los sensores o falta de energía, lo que genera una data discontinua que debe ser homogenizada.

3. Normalización de la información

La normalización de los datos que consiste en re-escalar los valores dentro de un rango específico entre 0 a 1, comúnmente llamado información por unidad, que está descrita por la ecuación 1.

$$Valor\ unidad = \frac{(V_{Medido} - V_{Mínimo})}{(V_{Máximo} - V_{Mínimo})} \quad \text{Ecuación 1}$$

4. Importancia de la calidad y la fiabilidad de los datos recopilados

La información hidrológica básica es usada para mejorar la comprensión de los procesos de monitoreo realizados a los ríos. La garantía de la calidad de información es un componente muy importante, por eso hay que tener en cuenta la recolección de datos y el mantenimiento continuo de los instrumentos hasta la difusión de los datos e información, del cual son preparados los dispositivos automáticos para la captura de los datos codificados, reducidos y normalizados [50].

La calidad y fiabilidad de los datos es fundamental para la predicción de inundaciones [51] y difusión de alertas, con esta información los entes gubernamentales pueden tomar decisiones, sirve de insumo a los planes de gestión del riesgo para disminuir el impacto de las inundaciones, y se realiza la planificación y respuesta a las necesidades de la comunidad.

Para realizar predicciones es fundamental que los datos sean de buena calidad y confiables, ya que estos datos serán los valores de las variables que se procesan en los modelos hidrodinámicos seleccionados, que proporcionarían pronósticos sobre la ocurrencia y magnitud de inundación, información que debe ser difundida a las autoridades y comunidades para prepararse ante eventos extremos.

También los datos hidrometeorológicos de alta calidad son una fuente importante para la investigación, estos permiten comprender mejor los procesos climáticos y desarrollar con un margen de incertidumbre menor modelos sofisticados. Como se realizó en la investigación realizada en río Kikuletwa en la cuenca del río Pangani, en el norte de Tanzania, donde se recopilaban los datos mediante sensores de distancia ultrasónicos [52]. El conjunto de datos consta de mediciones por hora de los niveles de agua de los ríos, lo que proporciona una serie temporal de alta resolución que se puede utilizar para estudiar las tendencias en los cambios del nivel del agua y desarrollar sistemas de alerta temprana de inundaciones más precisos.

5. Desafíos y las soluciones en la recolección de datos hidrometeorológicos

El mayor desafío en la recolección de datos, son las grandes áreas donde no se encuentran estaciones hidrometeorológicas, debido a los altos costos en de las estaciones, falta de infraestructura y falta de recursos económicos lo que lleva a datos inexistentes. Por otra parte, el mantenimiento garantiza la vida útil de los equipos de las estaciones meteorológicas e hidrométricas y el escaso mantenimiento conlleva a estaciones fuera de servicio o datos de series de tiempos incompletas [53].

Es por todo lo anterior que se deben tener varias fuentes para la toma de datos hidrometeorológico, terrestres y aéreos, físicos y sensores satelitales remotos, que adicionalmente se encuentran de acceso libre como lo es sentinel 1 y 2, Jason 1 y 2 que mediante técnicas de altimetría por radar [54], se ha logrado monitorear el nivel del mar, humedales, lagos y ríos, a través de series temporales, en puntos específicos creando estaciones virtuales [55]. También se tiene información de satélites en plataformas como Power Nasa, que recopila y combina información de satélites, para una mayor precisión y con solo conocer la coordenada o ubicación del lugar se descargan datos como precipitación, humedad relativa, temperatura, velocidad del viento, en formatos csv.

Igualmente, en Colombia a través del IDEAM se pueden descargar las series de tiempo de las estaciones que son afluentes del río Arauca, donde se presenta la información en históricos previamente validados por la institución y en tiempo real a través de un archivo json [56].

Es deseable que, mediante el desarrollo de algoritmos utilizando software libre y otras herramientas de código abierto, se puedan encontrar soluciones a los desafíos propuestos a través de técnicas de inteligencia artificial. Estos algoritmos tienen el objetivo de procesar datos de calidad que permitan el análisis efectivo mediante modelos hidrodinámicos, contribuyendo así a la resolución de problemáticas existentes [57].

IV. CONCLUSIONES

Los desafíos en la cobertura geográfica, calibración, interoperabilidad y almacenamiento pueden afectar la calidad y confiabilidad de los datos hidrometeorológicos. Estos son parte fundamental para un análisis preciso. En esta investigación se consideran variables como humedad relativa, temperatura, velocidad del viento, nivel del río, pluviosidad, caudal, las cuales se obtienen de fuentes como: el IDEAM, Amazon y diversas plataformas satelitales.

El procesamiento previo al análisis busca generar datos de calidad, adecuados al tipo de información, ya sea numérica o de imagen. En el caso de las imágenes satelitales, se aplican técnicas de preprocesamiento y filtrado utilizando programas especializados para corrección. Para los datos numéricos, diversos estudios recomiendan completar las series de tiempo mediante métodos de interpolación y técnicas de inteligencia artificial, como redes neuronales, máquinas de soporte vectorial, árboles de decisión, bosques aleatorios y k-vecinos más cercanos (k-NN).

En lo que respecta al caso de estudio que motiva esta revisión, existen diversidad de fuentes que proporcionan información hidrometeorológica del departamento de Arauca, existen diferentes técnicas encontradas en la literatura que permiten lidiar con los casos faltantes, y es posible que se mejoren los resultados de predicción mediante la implementación de un modelo que permita incorporar datos heterogéneos provenientes de las diversas fuentes disponibles. El diseño e implementación de dicho modelo es el resto que prosigue en el proyecto que ocupa a los autores.

V. REFERENCIAS

- [1] V. Bernal, J. Gomez, and C. Pardo, "Prediccion del fenomeno de la precipitacion ambiental en el municipio de Aquitania," *Reivsta Colomb. Tecnol. Av.*, vol. 2, 2023, doi: <https://doi.org/10.24054/rcta.v2i42.2649>.
- [2] J. Ayala-García and K. Ospino-Ramos, "Desastres naturales en Colombia: un análisis regional," Bogotá, Colombia, Colombia, Jul. 2023. doi: [10.32468/dtseru.317](https://doi.org/10.32468/dtseru.317).
- [3] World Meteorological organization WMO, *Manual on Flood Forecasting and Warning*, Segunda., no. 1072. Suiza, 2011.
- [4] S. K. Jain et al., "A Brief review of flood forecasting techniques and their applications," *Int. J. River Basin Manag.*, vol. 16, no. 3, pp. 329–344, 2018, doi: [10.1080/15715124.2017.1411920](https://doi.org/10.1080/15715124.2017.1411920).
- [5] J. Styweart, L. Giraldo, J. Styweart, and L. Giraldo, "Evaluación del riesgo urbano por inundaciones del río Supía Evaluación del riesgo urbano por inundaciones del río Supía," 2021.
- [6] R. J. G. Amaya, "Analisis de falla de un talud en un sector urbano del municipio de Ocaña, Colombia," *Rev. Colomb. Tecnol. Av.*, vol. 1, 2020, doi: <https://doi.org/10.24054/rcta.v1i35.35>.
- [7] IDEAM and PNUD, "Tercera Comunicación Nacional de Cambio Climático," 3ra Comun. Nac. Cambio Clim., p. 15, 2015, [Online]. Available: http://documentacion.ideam.gov.co/openbiblio/bvirtual/022965/documento_nacional_regional.pdf.
- [8] R. G. Bartual and E. C. Gimenez, "Predicción de caudales en tiempo real en grandes cuencas utilizando Redes Neuronales Artificiales," Universidad Politecnica de Valencia, 2009.
- [9] W. S. Vargas-Crispin, E. Montes-Raymundo, M. Castrejón-Valdez, and R. A. Hinojosa-Benavides, "Machine Learning como Herramienta para Determinar la Variación de los Recursos Hídricos," *Sci. Res. J. CIDI*, vol. 1, no. 1, pp. 56–69, Jul. 2021, doi: [10.53942/srjicidi.v1i1.46](https://doi.org/10.53942/srjicidi.v1i1.46).
- [10] J. Ayala and K. Ospino, "Desastres naturales en Colombia: un análisis regional," Bogotá, Colombia, Jul. 2023. doi: [10.32468/dtseru.317](https://doi.org/10.32468/dtseru.317).
- [11] A. C. Alfaro, H. G. Ramos, J. P. Roza, I. A. Cacunubo, and F. J. Lagos, "Relacion de la cadena de valor del sector agricola, sus beneficios financiero - medioambientales y el turismo sostenible," *Rev. Colomb. tecnologías Av.*, vol. 2, 2020, doi: <https://doi.org/10.24054/16927257.v36.n36.2020.4015>.
- [12] J. A. Gomez, H. Y. Jaramillo, and L. A. C. Rojas, "Sistema para detección de fallos críticos en tuberías horizontales," *Rev. Colomb. Tecnol. Av.*, vol. 1, no. 1692–7257, 2020, doi: <https://doi.org/10.24054/rcta.v1i35.41>.
- [13] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018, doi: [10.1080/01431161.2018.1433343](https://doi.org/10.1080/01431161.2018.1433343).
- [14] A. Bahamon and J. P. Barrero, "Regulación o no para la IA? propuesta de la regulación híbrida de la IA en Colombia," *Rev. Colomb. Tecnol. Av.*, vol. 2, 2020, doi: <https://doi.org/10.24054/rcta.v2i36.17>.
- [15] I. Sergio, L. Beleño, C. Viviana, M. Luna, I. Fernando, and M. Bernal, "Inspección de aisladores en líneas de transmisión eléctrica usando inteligencia artificial," *Rev. Colomb. Tecnol. Av.*, vol. 2, 2020, doi: <https://doi.org/10.24054/16927257.v36.n36.2020.4018>.
- [16] S. López and J. Fernando, "Vehículo clasificación and counting through image analysis . Clasificación y conteo vehicular mediante análisis de imágenes . Resumen," vol. 6, no. 4, pp. 313–334, 2022, doi: <https://doi.org/10.56048/MQR20225.6.4.2022.313-334>.
- [17] D. Marín and I. Pineda, "Modelo predictivo Machine Learning aplicado a análisis de datos Hidrometeorológicos para un SAT en Represas," 2019.
- [18] S. Wang, "Uncertainties in the Assessment of Individual and Compound Flooding from River Discharge and Coastal Water Levels under Climate Change," 2020.
- [19] W. F. Rafael Miño, P. V. R. Vilcherres Lizárraga, S. P. Muñoz Pérez, V. A. Tuesta Monteza, and H. I. Mejía Cabrera, "Modelamiento de procesos hidrológicos aplicando técnicas de inteligencia artificial: una revisión sistemática de la literatura," *Iteckne*, vol. 19, no. 1, 2021, doi: [10.15332/iteckne.v19i1.2645](https://doi.org/10.15332/iteckne.v19i1.2645).
- [20] F. García, J. Rojas, A. Vásquez, D. Parra, and E. Castro, "Estimating missing data in historic series of global radiation through neural network algorithms," *Sist. y Telemática*, vol. 14, no. 37, pp. 9–22, 2016, doi: [10.18046/syt.v14i37.2239](https://doi.org/10.18046/syt.v14i37.2239).
- [21] D. Aragon, L. E. Gomez, G. Rivera, and W. Cuevas, "Estudio de la tecnología de TVWS en Colombia y comparación con el mundo," *Rev. Colomb. Tecnol. Av.*, vol. 2, no. 1692–7257, 2019.
- [22] D. P. L. Ramirez and D. fernando G. Rojas, "Determinación De La Cota De Inundación Del Río Arauca En La Vereda Barrancones, Municipio De Arauca Mediante Un Modelo Hidrodinamico," 2019.
- [23] Y. Gonzales, M. Peñaranda, and O. Manzano, "Innovaciones tecnologicas en las practicas academicas virtuales," *Rev. Colomb. Tecnol. Av.*, vol. 1, 2019, doi: <https://doi.org/10.24054/rcta.v1i33.87>.
- [24] M. A. A. Coll, "Sistema de Alerta Temprana para la predicción del nivel de peligrosidad en inundaciones pluviales," 2018.
- [25] A. Gupta and R. S. Govindaraju, "Uncertainty quantification in watershed hydrology: Which method to use?," *J. Hydrol.*, vol. 616, no. November 2022, p. 128749, 2023, doi: [10.1016/j.jhydrol.2022.128749](https://doi.org/10.1016/j.jhydrol.2022.128749).

- [26] A. López, “Diseño de Herramienta Hidrológica como Apoyo a Los Sistemas De Alerta Temprana Ante Inundaciones,” Universidad Autonoma Estado de Mexico, 2017. [Online]. Available: <http://hdl.handle.net/20.500.11799/67762>.
- [27] IDEAM, “Zonificación y Codificación de Cuencas Hidrográficas,” Doc. IDEAM, p. 46, 2013, [Online]. Available: <http://documentacion.ideam.gov.co/openbiblio/bvirtual/022655/MEMORIASMAPAZONIFICACIONHIDROGRAFICA.pdf>.
- [28] IDEAM, “DHIME - Manual de Usuario Consulta y Descarga de datos hidrometeorológicos - IDEAM,” Manual, p. 53, 2019.
- [29] IDEAM, “Catálogo Nacional de Estaciones del IDEAM.” <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Cat-logo-Nacional-de-Estaciones-del-IDEAM/hp9r-jxuu/data>.
- [30] J. Wu, L. Lin, T. Li, Q. Cheng, C. Zhang, and H. Shen, “Fusing Landsat 8 and Sentinel-2 data for 10-m dense time-series imagery using a degradation-term constrained deep network,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 108, no. December 2021, 2022, doi: [10.1016/j.jag.2022.102738](https://doi.org/10.1016/j.jag.2022.102738).
- [31] P. J. Copete, “Uso de radares meteorológicos como herramienta para la Gestión del Riesgo y monitoreo dentro de los Sistemas de Alerta Temprana (SAT),” 2017.
- [32] IDEAM, “Seguimos apostándole a la prevención: Minambiente.” p. 3, 2019. [Online]. Available: http://www.ideam.gov.co/web/sala-de-prensa/noticias/-/asset_publisher/LdWW0ECY1uxz/content/minambiente-y-el-ideam-inauguran-radar-meteorologico-en-barrancaberme-2%0Ahttp://www.ideam.gov.co/documents/24277/83134001/COMUNICADO+RADAR/fed8bd8a-40f2-4f5a-8cc3-.
- [33] IDEAM, “Plataforma IDEAM radar,” 2023. <http://www.pronosticosvalertas.gov.co/archivos-radar>.
- [34] J. J. Helmus and S. M. Collis, “The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language,” *J. Open Res. Softw.*, vol. 4, no. 1, p. 25, 2016, doi: [10.5334/jors.119](https://doi.org/10.5334/jors.119).
- [35] J. A. S. Duarte, M. A. Contreras, and J. A. Torres, “Caracterización geotécnica del subsuelo en el relleno sanitario regional ‘la cortada’, pamplona (norte de santander) a partir de datos geofísicos,” *Rev. Colomb. Tecnol. Av.*, vol. 2, 2020, doi: <https://doi.org/10.24054/rcta.v2i36.15>.
- [36] OEA. Irlanda., “Manual para el Diseño e Implementación de un Sistema de Alerta Temprana de Inundaciones en Cuencas Menores,” Manual. Washington, D.C, p. 55, 2001.
- [37] Organizacion Mundial Metereologica OMM, Manual sobre predicción y avisos de crecidas. Ginebra, Suiza, 2011.
- [38] A. Florez, C. Durán, and J. Carrillo, “Data processing from electrical signals acquired by an E-nose system used for quality control of cocoa,” *J. Phys. Conf. Ser.*, vol. 1704, no. 1, 2020, doi: [10.1088/1742-6596/1704/1/012013](https://doi.org/10.1088/1742-6596/1704/1/012013).
- [39] N. F. B. Morales, Teledeteccion espacial Landsat, Sentinel2, Aster L1T y Modis. 2017. [Online]. Available: https://acolita.com/wp-content/uploads/2018/01/Teledeteccion_espacial_ArcGeek.pdf.
- [40] NASA ARSET, “Fundamentos de la Teledetección (Percepción Remota).” 2021.
- [41] K. Saylor, “Landsat 9 Data Users Handbook Version 1.0,” Nasa, no. February, p. 107, 2022, [Online]. Available: https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/media/files/LSDS-2082_L9-Data-Users-Handbook_v1.pdf.
- [42] J. A. Martínez and P. A. Arias, “Estimación y pronóstico de radiación solar en el valle de Aburrá – Colombia Estimation and forecast of solar radiation in the Aburrá Resumen Introducción,” vol. 46, no. 179, pp. 529–549, 2022.
- [43] X. Wang, J. Gong, Y. Zhang, and P. M. Atkinson, “Science of Remote Sensing Near real-time surface water extraction from GOES-16 geostationary satellite ABI images by constructing and sharpening the green-like band,” *Sci. Remote Sens.*, vol. 5, no. May, p. 100055, 2022, doi: [10.1016/j.srs.2022.100055](https://doi.org/10.1016/j.srs.2022.100055).
- [44] L. F. Gopar-Merino and A. Velázquez, “Componentes del paisaje como predictores de cubiertas de vegetación: estudio de caso del estado de Michoacán, México,” *Investig. Geogr.*, vol. 2016, no. 90, pp. 75–88, 2016, doi: [10.14350/riig.46688](https://doi.org/10.14350/riig.46688).
- [45] Administracion Nacional Oceanica Atmosferica NOAA, “GOES IMAGE VIEWER,” Plataforma online, 2023. <https://www.star.nesdis.noaa.gov/GOES/sector.php?sat=G16§or=nsa#>.
- [46] L. A. Lasso, E. Rincon, and G. Estrada, “INTRODUCTION TO CAPACITY EVALUATION: A THEORETICAL REVIEW,” *Rev. Colomb. Tecnol. Av.*, vol. 2, 2020, doi: <https://doi.org/10.24054/rcta.v2i36.18>.
- [47] W. Y. Lai and K. K. Kuok, “A Study on Bayesian Principal Component Analysis for Addressing Missing Rainfall Data,” *Water Resour. Manag.*, vol. 33, no. 8, pp. 2615–2628, 2019, doi: [10.1007/s11269-019-02209-8](https://doi.org/10.1007/s11269-019-02209-8).
- [48] Y. Gao, C. Merz, G. Lischeid, and M. Schneider, “A review on missing hydrological data processing,” *Environ. Earth Sci.*, vol. 77, no. 2, p. 47, 2018, doi: [10.1007/s12665-018-7228-6](https://doi.org/10.1007/s12665-018-7228-6).
- [49] J. Yang, C. Cheng, and C. Chan, “A Time-Series Water Level Forecasting Model Based on Imputation and Variable Selection Method,” vol. 2017, 2017.
- [50] K. Z. Vallejo, “Sistema Electrónico Inalámbrico De Alerta Temprana Y Monitoreo Del Comportamiento Del Nivel De Los Ríos De Bajo Costo.,” Tesis Pregr., p. 117, 2013.
- [51] T. Khampungson and W. Wang, “Novel Methods for Imputing Missing Values in Water Level Monitoring Data,” *Water Resour. Manag.*, vol. 37, no. 2, pp. 851–878, 2023, doi: [10.1007/s11269-022-03408-6](https://doi.org/10.1007/s11269-022-03408-6).
- [52] L. Mdegela, Y. De Bock, E. Luhanga, J. Leo, and E. Mannens, “Monitoring Kikuletwa river levels in northern Tanzania : A data set unlocking insights for effective flood early warning systems,” vol. 49, 2023, doi: [10.1016/j.dib.2023.109395](https://doi.org/10.1016/j.dib.2023.109395).
- [53] I. Silva, M. Rodriguez, R. Acosta, and P. Gomez, “Diseño de plan de mejoramiento preventivo para los talleres del centro CIES SENA regional Norte de Santander utilizando metodología AMEF,” *Rev. Colomb. Tecnol. Av.*, vol. 3, no. 1692–7257, pp. 143–154, 2020, doi: <https://doi.org/10.24054/rcta.v3iEspecial.880>.
- [54] F. Fichtner, N. Mandery, M. Wieland, S. Groth, and S. Martinis, “International Journal of Applied Earth Observation and Geoinformation Time-series analysis of Sentinel-1 / 2 data for flood detection using a discrete global grid system and seasonal decomposition,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 119, no. May, p. 103329, 2023, doi: [10.1016/j.jag.2023.103329](https://doi.org/10.1016/j.jag.2023.103329).
- [55] C. Schwatke, D. Dettmering, W. Bosch, and F. Seitz, “DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry,” pp. 4345–4364, 2015, doi: [10.5194/hess-19-4345-2015](https://doi.org/10.5194/hess-19-4345-2015).
- [56] S. Cardenas, C. Vides, and A. Pardo, “Sistema de alerta temprana de inundaciones para el río Arauca basado en técnicas de inteligencia artificial,” vol. 12, pp. 315–326, 2022.
- [57] D. Echeverría, “Aplicacion del indice de vulnerabilidad a la inundacion en la cuenta del rio IPA en la region del chaco Bolivariano,” 2019.