

CGIAR DIGITAL INNOVATION
Digital Twin

TYPE
Report

FUNDERS
CGIAR Digital Innovation
Initiative, Leona M. and Harry B.
Helmsley Charitable Trust

PARTNERS
LIMCOM, IWMI, Hidromod,
AWARD, Digital Earth Africa,
AWS

Operational SWAT+ Model: Advancing Seasonal Forecasting in the Limpopo River Basin



INITIATIVE ON
Digital Innovation



INITIATIVE ON
Digital Innovation

THE LEONA M. AND HARRY B.
HELMSLEY
CHARITABLE TRUST

Operational SWAT+ Model: Advancing Seasonal Forecasting in the Limpopo River Basin

Authors Pedro Chambel-Leitão¹, Flávio Santos¹, Débora Barreiros¹, Hélio Santos¹, Paulo Silva², Thilina Madushanka², Karthikeyan Matheswaran², Lal Muthuwatta², Keerththanan Vickneswaran², Hugo Retief³, Chris Dickens², Mariangel Garcia Andarcia²

1 Hidromod, Lisbon, Portugal (www.hidromod.com)

2 International Water Management Institute (IWMI), Colombo, Sri Lanka

3 Association for Water and Rural Development (AWARD), South Africa.

Date 09/01/2024

Work package WP3

Partners LIMCOM, IWMI, Hidromod, AWARD, Digital Earth Africa, AWS



Abstract

This "Operational SWAT+ Limpopo River Basin Seasonal Forecasting System" report outlines the development and implementation of an automated hydrological forecasting system using the Soil and Water Assessment Tool Plus (SWAT+). This system leverages publicly available global datasets and open-source modeling tools integrated within a custom developed automated system to predict seasonal water availability in the Limpopo River Basin (LRB). Key components include integrating the CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data.) and ECMWF (European Centre for Medium-Range Weather Forecasts) precipitation data, comprehensive database management, and real-time monitoring scripts. The system provides accurate and timely water availability forecasts within the LRB to support operational decision making. Future directions focus on improving model calibration, incorporating additional weather variables, better representation of large reservoirs and irrigated areas, applying database optimization procedures, and transitioning to a Docker-based deployment on Amazon Web Services (AWS) for improved scalability and reliability. This SWAT+ operational seasonal forecasting system for the LRB marks a significant step towards bridging a key knowledge gap in the basin to support better decision making on multiple water uses and users including provision for environmental flows. This seasonal forecasting system as a part of the larger river basin Digital Twin is designed to influence effective water resource management in the Southern African region.

Table of Contents

Abstract.....	ii
Introduction.....	1
Implementation.....	3
SWAT model setup.....	5
Seasonal forecast.....	8
Operational Seasonal Forecasting Framework.....	10
Architecture and information flow.....	16
Data integration.....	19
Data Sources.....	20
Data Formats.....	20
Data Handling and System Operations.....	21
Results.....	23
Breakdown of Flow Types.....	24
SWAT flow estimates.....	25
Water Availability Forecast.....	26
Discharges.....	31
Reservoirs.....	32
Monitoring.....	35
Server and infrastructure.....	38
Conclusion.....	41
Next Steps and Future Direction.....	42
Acknowledgement.....	43
References.....	44
Annex I – Data integration.....	48
Data formats.....	48

Annex II Database management	52
Discharges	56
Reservoirs	59
SQL queries to extract data from the Operational DB.....	62

List of Figures

Figure 1 - Comparative Analysis of SWAT Model Performance Using CHIRPS Data for Channel 215 Against Observed Flow Data at the Limpopo River (2008-2015).....	6
Figure 2 - Comparison of Monthly Evapotranspiration Estimates from SWAT and WAPOR for HRU ID 2 (2018-2024)	7
Figure 3 - Linear Regression Analysis of Evapotranspiration: SWAT Model versus WAPOR Data for HRU ID 2	7
Figure 4 – ECMWF daily precipitation for the control and for the 10 th , 50 th and 90 th percentile simulations of the seasonal forecast, at station ID s24067s31692e (corresponding to channel 844).....	10
Figure 5 - Operational Workflow of the SWAT Model for Seasonal Water Availability Forecasting in the Limpopo River Basin.....	11
Figure 6 - Repository Overview of the Operational SWAT Model scripts.....	12
Figure 7 - Integrated SWAT Operational Framework for Automated Water Resource Forecasting in the Limpopo River Basin: A Modular Approach Encompassing Data Acquisition, Pre-Processing, Modeling, and Monitoring.....	17
Figure 8 - Automated SWAT Limpopo Workflow in Windows Task Scheduler: Scheduled Data and Model Operations	19

Figure 9 - Limpopo Digital Twin Operational Database Status 22

Figure 10 - Flow out and accumulated precipitation for Channel 844 for the SWAT model run with control, pc10, pc50 and pc90 seasonal forecast precipitation. This is the prediction made in 1 of May 2024..... 25

Figure 11 - Methodology and Example of Water Availability Forecasting with SWAT Data 27

Figure 12 - Discharge Monitoring Network and example Time-Series Data for the Limpopo River Basin. Black Line: SWAT Model Output | Blue Line: Verified Station Discharge | Orange Line: Unverified Station Discharge..... 31

Figure 13 - Overview of Operational Waterbody Monitoring System for the Limpopo River Basin. 33

Figure 14 - Integrated Reservoir Monitoring from Earth Observation and Gauge Stations 34

Figure 15 - Storage Allocation and System Resources of Tlaloc Server 40

Figure 16 - Folder Structure of WAPOR Containing FieldStats, Rasters, NullMasks, and TimeSeries 50

Figure 17 - WAPOR export format in CSV file..... 50

Figure 18 - The water availability format provided is CSV file 52

Figure 19 - Database Structure for the Limpopo River Basin Operational SWAT+ System. 52

Figure 20 - Database Schema of the 'eflow_iwmi' SWAT Operational Database with Entity Relationship Diagram..... 55

Figure 21 - Database Schema for Discharge Monitoring..... 59

Figure 22 - Database Schema for Reservoir and Waterbody Data Integration. 61

List of Tables

Table 1 - Geographic and Update Metadata for River Station LUVU-A91K-OUTPO.....	28
Table 2 - Chronological Water Availability Forecast and Ecological Flow Assessment for River Station LUVU-A91K-OUTPO for the control run	28
Table 3 - 90th Percentile Water Availability Forecast for River LUVU-A91K-OUTPO	29
Table 4 - 10th Percentile Water Availability Forecast for River LUVU-A91K-OUTPO	30

Acronym Meaning

API	Application Programming Interface
AWS	Amazon Web Services
CDS	Copernicus Climate Data Store
CDS	Copernicus Climate Data Store.
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
CSV	Comma-Separated Values
DEA	Digital Earth Africa
DNS	Domain Name System
DWS	Department of Water and Sanitation
ECMWF	European Centre for Medium-Range Weather Forecasts
ET	Evapotranspiration
FAO	Food and Agriculture Organization
FTP	File Transfer Protocol
GEE	Google Earth Engine
GeoTiff	Geospatial Tagged Image File Format
GFS	Global Forecast System
GFS025	Global Forecast System (0.25-degree resolution)
HDD	Hard Disk Drive
HRU	Hydrological Response Unit
INWARDS	Integrated Water Resource Decision Support
IWMI	International Water Management Institute
LIMCOM	Limpopo Watercourse Commission.
LRB	Limpopo River Basin
NetCDF	Network Common Data Form

Acronym Meaning

pc10	10th Percentile of Seasonal Forecast
pc50	50th Percentile of Seasonal Forecast
pc90	90th Percentile of Seasonal Forecast
RAID	Redundant Array of Independent Disks
RAM	Random Access Memory
RDS	Relational Database Service
SQL	Structured Query Language
SSD	Solid-State Drive
SWAT+	Soil and Water Assessment Tool Plus
WAPOR	FAO's Water Productivity Open-access Portal
WRF	Weather Research and Forecasting Model.

Introduction

The Limpopo River Basin (LRB) is a region of immense environmental, socio-economic significance, extending across Botswana, South Africa, Zimbabwe, and Mozambique. It encompasses a diverse range of ecosystems, including forests, wetlands, and savannahs, supporting a rich variety of flora and fauna (Mosase, Ahiablame and Srinivasan 2019). The river and its tributaries provide essential water resources for agriculture, livestock, and fisheries, supporting millions of livelihoods in the Southern African region (Nhassengo, Somura, & Wolfe, 2021; African Development Bank, 2014; Busari, 2007). Despite these invaluable resources, the LRB faces challenges such as hydro-climatic extremes, excessive abstractions, land degradation, poor water quality, urbanisation, and population growth, which reduce river discharge, threaten biodiversity and weaken local communities' resilience to climate change. The basin also faces increasing competition for access to limited water resources and arable land, with women and youth disproportionately affected (Brito, et al. 2009, Zhu and Ringler 2012, Reason, et al. 2005)

The LRB riparian countries and stakeholders are eager to address these pressing challenges through informed and data driven decision making. To support these efforts, IWMI is creating a Digital Twin of the Limpopo River Basin (Garcia Andarcia, et al. 2024, Gurusinghe, et al. 2024, Botai, et al. 2023). The LRB Digital Twin will increasingly serve as a virtual representation of the basin's socio-ecological system, by utilizing multiple modelling and mapping toolsets including the Soil and Water Assessment Tool Plus (SWAT+) model for the representation of LRB hydrological process (Bieger, et al. 2017) to forecast seasonal water availability and support sustainable water management practices. This approach mirrors other Digital Twin initiatives, such as the Digital Twin for the Mediterranean Basin (Brocca, et al. 2024), which integrates Earth observation (EO) data with hydrological modeling. This

Mediterranean Digital Twin demonstrates how EO, advanced modeling, and AI can be combined to forecast hydrologic variables, highlighting the potential for large-scale implementation of Digital Twin's in water management.

Operational hydrological models have been implemented in various regions, demonstrating their importance in water resource management. In Portugal, the "AquaSafe Douro" system provides daily streamflow predictions for the Douro River hydropower cascade, optimizing hydroelectric production and flood control. This system integrates operational SWAT models, supported by meteorological data from WRF and GFS models, ensuring reliable streamflow forecasting (P. Chambel-Leitão 2016). In Brazil, an operational platform was established for the Cubatão River watershed, providing near real-time monitoring of fluviometric and rainfall data. This platform includes SWAT modeling and integrates data from various meteorological sources to forecast river flow and support flood mitigation efforts (Ribeiro, et al. 2022). Similarly, in Australia, SWAT was operationalized for the Namoi Watershed to predict flow and soil moisture, incorporating multiple meteorological datasets for accurate forecasting (Chambel-Leitão, Alderlieste and Jackson 2023). These case studies highlight the versatility and reliability of SWAT and other hydrological models in diverse geographical and environmental contexts.

The **Introduction** sets the stage by outlining the significance of the LRB and the need for the seasonal water availability forecasting system. The **Implementation** section details the practical steps taken to set up and operationalize the SWAT+ model, including the data choices, modelling framework and overall integration for the operationalization of the forecasting system. **Data Integration** explains how diverse data sources and formats are harmonized within the operational framework to ensure the model's accuracy and reliability. **Database Management** focuses on the organization, maintenance, and querying of the operational database to ensure robust data handling and retrieval. The

Results chapter showcases the key outputs of the seasonal operational system, illustrating the efficacy of the model in forecasting the water availability across the LRB. **Monitoring** describes the system's capability to provide real-time updates to ensure continuous operation. **Server and Infrastructure** outlines the technical setup that supports the system's robust performance and reliability. Finally, the **Conclusion** summarizes the key outputs and takeaways from the operational seasonal forecasting system and the **Next Steps and Future Directions** highlight potential improvements and future expansions of the SWAT+ model and overall forecasting system, ensuring its long-term relevance and effectiveness in the LRB water resource management.

Implementation

This chapter provides a detailed description of how the SWAT+ model was set up and operationalized for the LRB. It covers the steps taken to automate the seasonal forecasting system, including data acquisition from ECMWF server, processing, automated model execution, and result dissemination. This chapter highlights the technical infrastructure and tools used to achieve these tasks.

The SWAT model's processing chain for the Limpopo region has been refined to automate seasonal water availability forecasting. This sophisticated system is engineered to generate monthly outputs, execute the SWAT model daily using the most up-to-date data, and generate timely results. The data processing and model execution chain ensure that the outputs are consistent over time and space, allowing for seamless integration and automatic dissemination. The development was done using a cloud windows server called Tlaloc.

The model makes use of data starting January 2001 until the present. To enhance the model's accuracy, the system includes an automated processing script that assimilates the

previous month's data, utilizing observations from globally and regionally accessible datasets, like CHIRPS, to inform the upcoming month's forecast. The automation extends to downloading weather data, transforming it to fit the SWAT model's input requirements, and integrating seasonal forecasts. The system adeptly manages spatial resolution discrepancies and temporal misalignments between datasets and model time steps. Robust error handling and validation processes are in place to address any data input inconsistencies.

Operationally, the model consistently outputs results from January 2023 to the present, ensuring that there is a sufficient warm-up period for the model to generate reliable results. The automation scripts are precisely tuned to initialize the SWAT model with forecasted conditions for optimal performance. Moreover, specialized scripts are deployed to extract pertinent output variables from the SWAT results. The SQL database writing process facilitates this integration, with the results also being accessible via AWS.

In addition to the operational system, mechanisms have been established to benchmark the SWAT model outcomes against WAPOR observations, fostering continual model calibration and validation. The entire processing system's source code has been documented, complete with a README file that provides exhaustive instructions for system modification and execution. This transfer of knowledge and resources guarantees that in the future it is possible to adapt and maintain the SWAT-based forecasting system with a high degree of autonomy and technical proficiency.

This chapter provides an in-depth overview of the setup and validation of the SWAT operational model for the Limpopo basin. It details the integration of CHIRPS daily precipitation data from Google Earth Engine and validates the model's accuracy through comparisons with observed data. Moreover, it encompasses the presentation of the database system and operational scripts vital for the SWAT model's functionality.

The operational database, referred to numerous times throughout this document, is called `eflow_iwmi`. This database plays a central role in storing and managing the data required for the SWAT model's operational framework. Structured as a MySQL database, it ensures efficient handling of large datasets such as precipitation, evapotranspiration, water availability, and SWAT model outputs. Detailed information on its structure and table relationships can be found in Annex II Database management, where the organization of the database is thoroughly described.

SWAT model setup

The SWAT operational model for the Limpopo basin has been prepared for deployment. This model uses CHIRPS daily precipitation data sourced from Google Earth Engine (Funk, et al. 2015), spanning from 2001 to the present.

We prepared an operational SWAT model based on the existing model implementation called `LimpopoFull` (Gurusinghe, et al. 2024).

Using the CHIRPS data, we conducted comparative analyses of different SWAT model setups by evaluating the output flows from channel 215 against observed data from the Limpopo River at Beit Bridge station (A7H008). Initially, we compared the results from the model run (executed using Revision 60.5.4 of the SWAT model without calibration) with those from the same model utilizing the operational CHIRPS data fetched from GEE. This allows to confirm that both CHIRPS datasets, the one used for calibration and the one used for the operational workflow, are similar.

As illustrated in Figure 1, the comparison highlights the consistency of the precipitation data source. The grey solid line represents the original model run with Revision 60.5.4 using static CHIRPS data, while the blue dashed line represents the model run with operationally downloaded CHIRPS data from GEE. Both results are closely aligned, with minor differences attributed to a 1-day shift between the datasets. This demonstrates the reliability and

accuracy of the operational CHIRPS data integration into the SWAT model. Both results do not include calibration and so they are both different from observed flow.

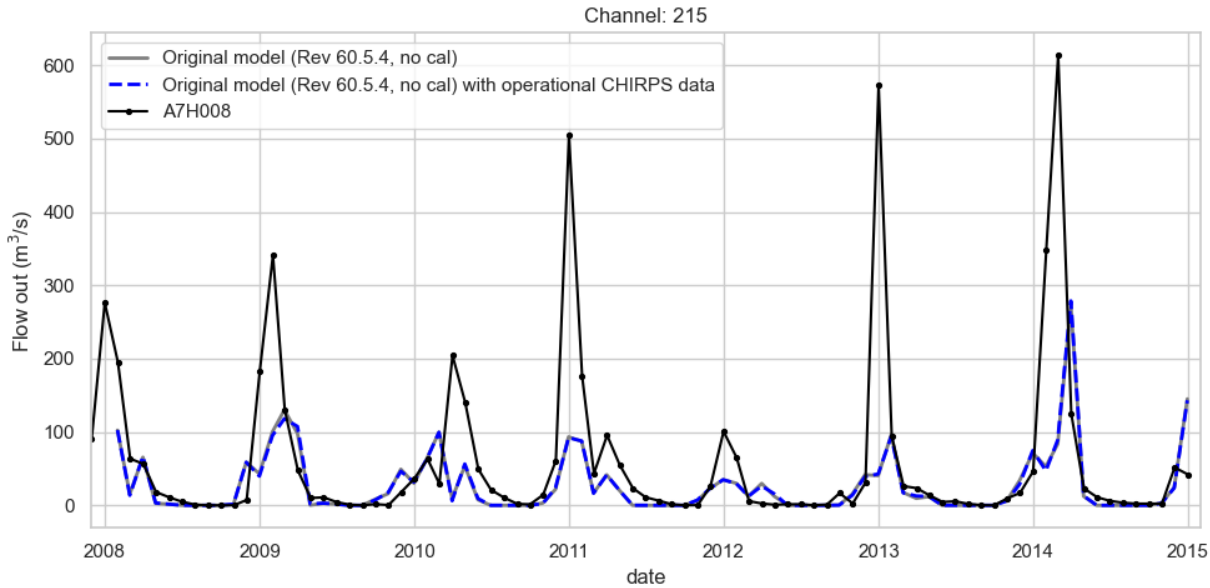


Figure 1 - Comparative Analysis of SWAT Model Performance Using CHIRPS Data for Channel 215 Against Observed Flow Data at the Limpopo River (2008-2015)

Comparing SWAT and WAPOR Evapotranspiration

The historical evapotranspiration estimates from the LRB SWAT+ model for selected Hydrological Response Unit (HRU) was compared with data obtained from the WAPOR portal (FAO 2020). To perform this comparison, we used the operational SWAT run with starting date 2001-01-01. Specifically, we used the monthly 'et' results in the output file hru_wb_mon.txt.

In the plot of Figure 2, we show the evapotranspiration time series obtained from the SWAT run (blue line) for a HRU and the corresponding WAPOR data averaged over the HRU's area.

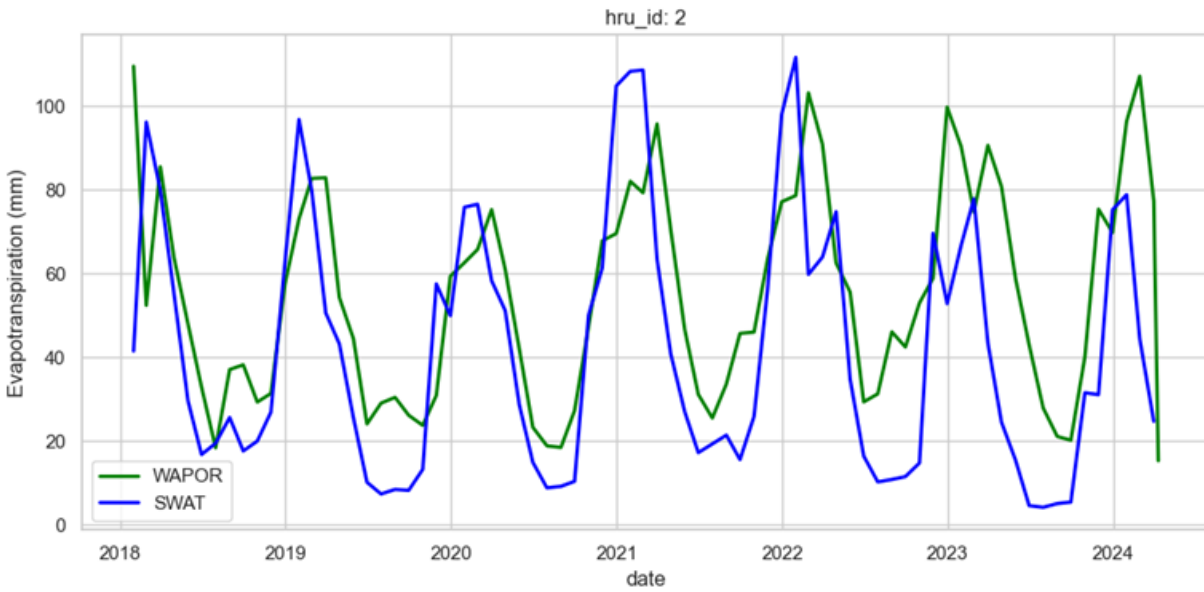


Figure 2 - Comparison of Monthly Evapotranspiration Estimates from SWAT and WAPOR for HRU ID 2 (2018-2024)

Figure 3 illustrates the linear function that best fits the SWAT and WAPOR results from which we see that in general there is a good correlation between the two ETs. We verified that for most channels, SWAT monthly results and WAPOR data do correlate well.

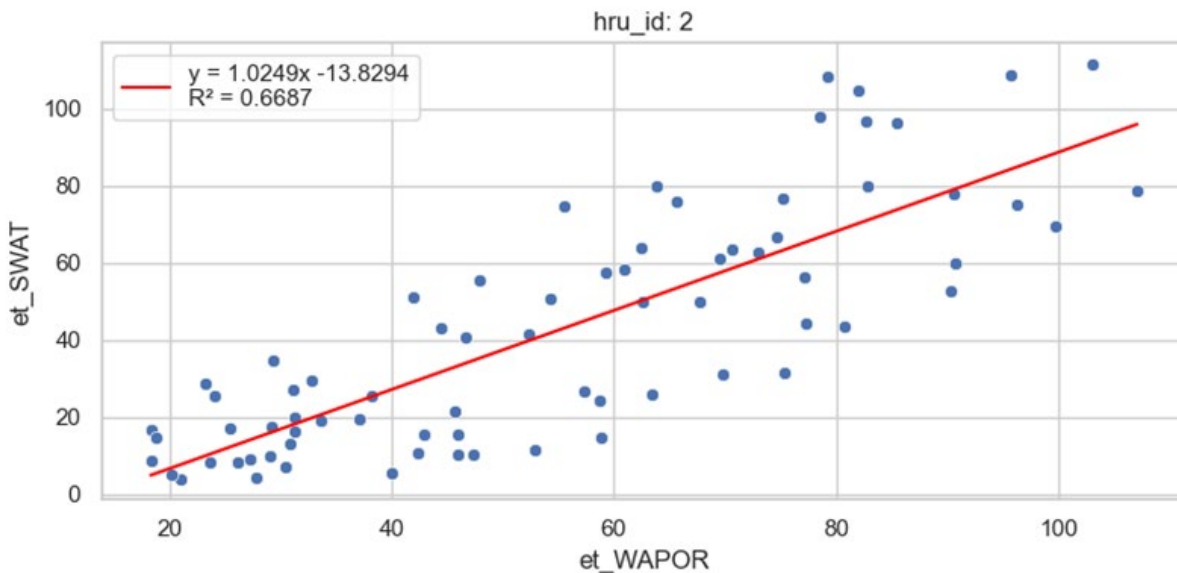


Figure 3 - Linear Regression Analysis of Evapotranspiration: SWAT Model versus WAPOR Data for HRU ID 2

Seasonal forecast

The seasonal forecast precipitation data used within this hydrological database system is sourced from the publicly available European Centre for Medium-Range Weather Forecasts (ECMWF) via the Copernicus Climate Data Store (CDS) API (Copernicus Climate Change Service, 2018).

ECMWF seasonal forecast data provides long-range weather predictions, covering six months ahead, and is widely used for climate and hydrological modeling. The ECMWF seasonal forecast system relies on a large ensemble of simulations, typically consisting of 51 ensemble members. Each ensemble member represents a slightly different version of the initial conditions or model physics to capture uncertainties in weather patterns, which allows for a more comprehensive assessment of potential outcomes. The precipitation data from the ECMWF seasonal forecast is segmented into different percentiles that represent varying hydrometeorological conditions:

- **pc10:** This represents drier conditions and is based on the 10th percentile of accumulated precipitation data, indicating a lower probability of precipitation that might lead to drier scenarios.
- **pc50:** This corresponds to average conditions and is derived from the 50th percentile of the accumulated precipitation data, reflecting median forecast outcomes.
- **pc90:** This indicates wetter conditions, based on the 90th percentile of the accumulated precipitation data, suggesting a higher likelihood of significant precipitation events.

Choosing these percentiles reduces the computational burden and storage needs of running all 51 members while still providing valuable insight into the range of possible

weather outcomes. These percentiles help decision-makers assess different risk levels and plan for various potential future conditions

The database's operational workflow is programmed to leverage the CDS API to check for new forecast data releases regularly. Once available, it automatically downloads and post-processes this data to label forecast ensemble members appropriately, including the "control" run and the specified percentiles. The pc10, pc50, and pc90 simulations are chosen among all the ensemble members by:

1. Calculating the mean over time of the accumulated precipitation value for all ensemble members in the forecast period (the default value is 215 days).
2. Calculate the 10th, 50th and 90th percentiles of all mean accumulated precipitation values.
3. Select the simulations where the mean accumulated precipitation is closer to the 10th, 50th and 90th percentile values. These will correspond to the simulations labelled as pc10, pc50 and pc90, respectively.

The workflow further processes this data to distil precipitation metrics for designated SWAT precipitation stations. For each precipitation station, the corresponding forecast data is extracted from the ECMWF NetCDF file based on the geographical coordinates where the station is located. This is done by identifying the grid cell in the ECMWF data that overlaps with the station's location. Since ECMWF data covers large areas, a single grid cell's value is assumed to represent the conditions around that station. In Figure 4, we show an example of the fetched ECMWF precipitation for the station with ID s24067s31692e (corresponding to channel 844, according to the rainfall_station table in the database).

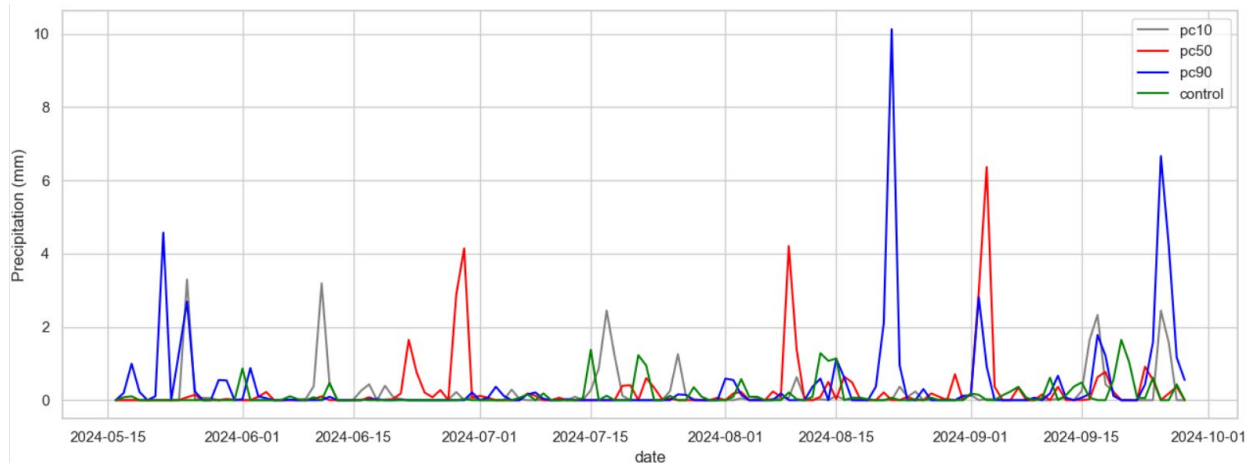


Figure 4 – ECMWF daily precipitation for the control and for the 10th, 50th and 90th percentile simulations of the seasonal forecast, at station ID s24067s31692e (corresponding to channel 844).

In this case, the accumulated precipitation values for the pc10, pc50 and pc90 curves in the forecast period are respectively 27.06 mm, 37.18 mm, and 55.03 mm, justifying the association of these simulations with possible drier, median, and wetter precipitation scenarios.

The processed control data is then inserted into the table rainfall_data which serves as input to the operational model and the SWAT+ run results based on the percentile precipitation data are updated to tables swat_output_2 and water_availability_2, using the labels (pc10, pc50, pc90) to inform water availability calculations and river flow forecasts in the context of climate variability and change.

Operational Seasonal Forecasting Framework

The "Operational Seasonal Forecasting Framework" outlines the automated system used to generate seasonal water availability forecasts for the Limpopo River Basin using the SWAT model. This framework integrates data collection, processing, and model execution into a seamless workflow, ensuring timely and accurate forecasts.

A comprehensive overview of the scripting framework that constitutes the operational backbone of the SWAT model for the Limpopo River Basin is shown in Figure 5. The central

objective of this operational run is to produce seasonal forecasts of water availability within the Limpopo region. The process is structured into a multi-stage operational workflow, encompassing different functions.

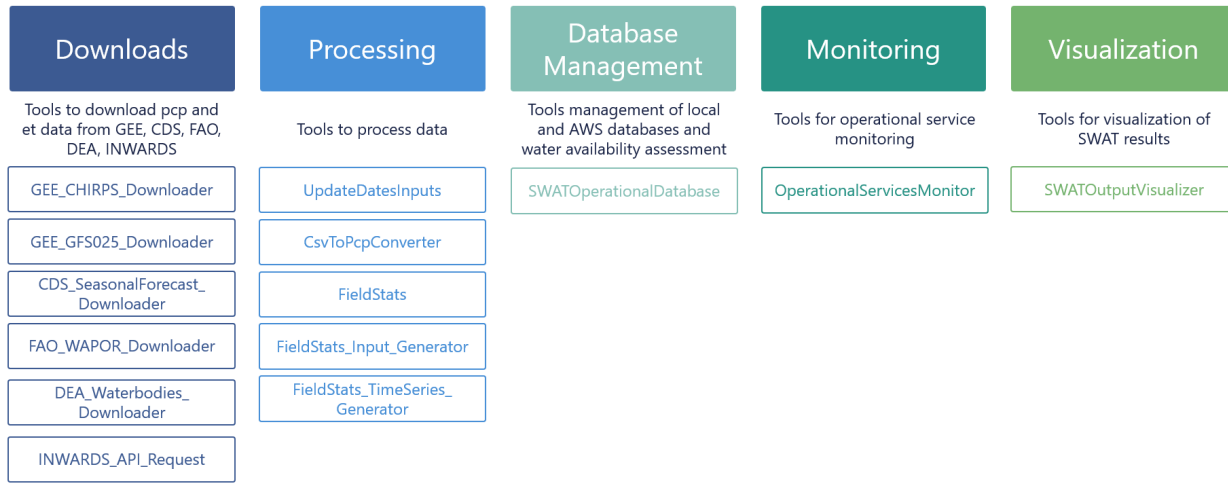


Figure 5 - Operational Workflow of the SWAT Model for Seasonal Water Availability Forecasting in the Limpopo River Basin

In Figure 6, a snapshot of a GitHub repository, specifically structured for managing the components of an Operational SWAT Model shown. The main page showcases a list of directories and files, indicating a repository with sections dedicated to various aspects of the SWAT model operation scripts as defined in Figure 5. This repository acts as the central hub for version control and distribution of the SWAT model components. The inclusion of example runs within these folders serve as valuable resources for users to understand the operational workflow or to replicate the model runs for their purposes.

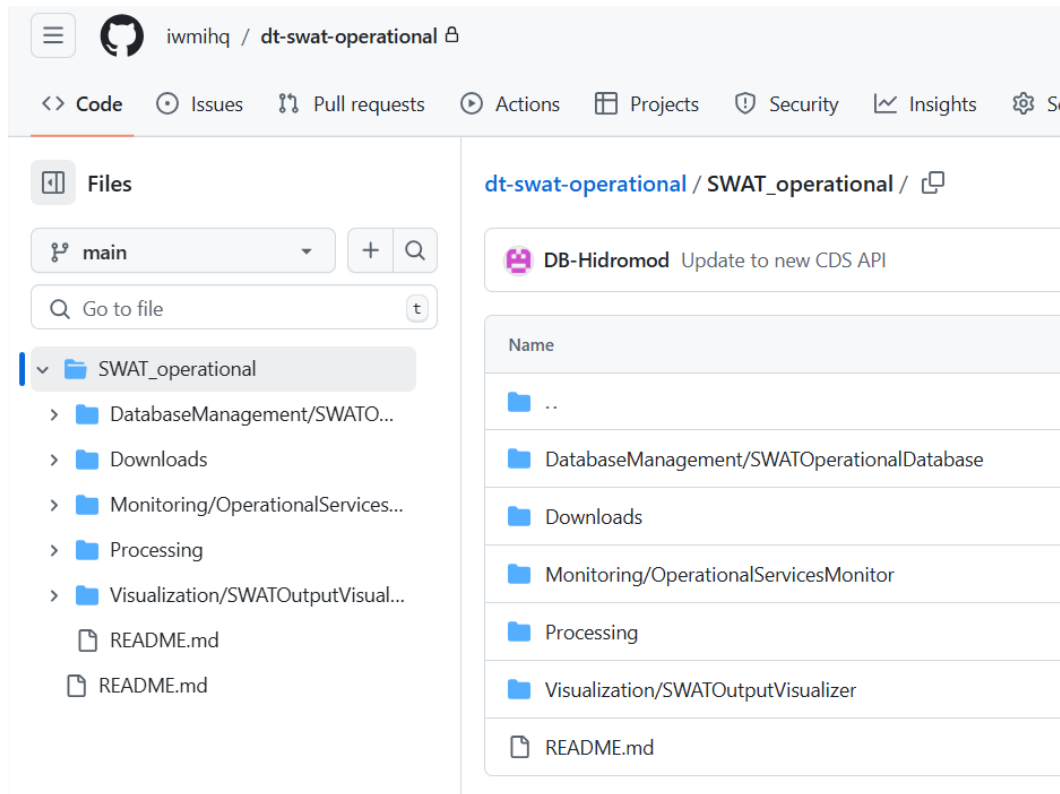


Figure 6 - Repository Overview of the Operational SWAT Model scripts

Downloads

The Downloads category is a suite of scripts designed to efficiently acquire hydrological and meteorological data, which are crucial inputs for the SWAT+ model. These scripts are tailored to interface with various data repositories, such as the Google Earth Engine (GEE), the Copernicus Climate Data Store (CDS), the Food and Agriculture Organization (FAO), Digital Earth Africa (DEA) and the INWARDS API. They are responsible for automated retrieval of data such as precipitation from CHIRPS, short-term forecast data from GFS025, and long-term seasonal forecasts from ECMWF, as well as evapotranspiration data from WAPOR, waterbody water area from DEA and Department of Water Affairs and Sanitation (DWS) discharge station data facilitated by the INWARDS API. By automating the download process, these scripts ensure that the most current and relevant datasets are always on hand for initiating the SWAT model's operational runs.

- **GEE_CHIRPS_Downloader:** Imports CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) daily precipitation data from GEE for a defined time interval based on the required number of hindcast days, and for given precipitation station locations.
- **GEE_GFS025_Downloader:** Imports GFS 0.25 degree daily precipitation data from GEE for a defined time interval based on the required number of hindcast and forecast days, and for given precipitation station locations.
- **CDS_SeasonalForecast_Downloader:** Imports grided values of daily seasonal forecast precipitation data from the Copernicus CDS API for a defined time interval based on the required number of forecast days.
- **FAO_WAPOR_Downloader:** Imports rasters containing evapotranspiration data from the FAO GIS data API for a given spatial extent and for a defined time interval based on the required number of hindcast days.
- **DEA_Waterbodies_Downloader:** Imports water area from the Digital Earth Africa Waterbodies Monitoring Service for a defined time interval and for given waterbodies.
- **INWARDS_API_Request:** Imports verified and unverified discharge stations and respective discharge timeseries.

Processing

In the Processing category, scripts like `CsvToPcpConverter` and the `FieldStats` series prepare the downloaded data for the SWAT+ model. These scripts transform raw data into SWAT-compatible formats and generate time-series statistics necessary for accurate modeling. The processing tools take care of nuances such as format consistency and data integrity, providing pre-processed inputs primed for simulation. They play a crucial role in ensuring that the inputs are not only accurate but also aligned with the model's requirements for efficient execution.

- **UpdateDatesInputs:** Automates the configuration of time-related parameters in input files (e.g., SWAT model run, `CsvToPcpConverter`, `SWATOutputVisualizer`, `FieldStats`) by calculating updated start and end dates based on the present-day date and the required number of hindcast and/or forecast days.
- **CsvToPcpConverter:** Converts precipitation data from .csv to the .pcp format required by the SWAT model. It combines precipitation data from different sources

by determining the corresponding time intervals to be used across multiple sources according to a given order of priorities.

- **FieldStats:** Derives statistics for a property from image data for a given polygonal region. This code takes as input a raster file containing IDs (representing polygons) and a list of IDs, which should correspond to the IDs present in the raster file.
- **FieldStats_Input_Generator:** Generates input files for running FieldStats, given a list of FieldStats configuration parameters and paths for the folders containing input rasters and the folders receiving the output .csv files.
- **FieldStats_TimeSeries_Generator:** Processes FieldStats output files, extracting and converting data to time series format.

Database Management

The Database Management scripts, were called *SWATOperationalDatabase*, and these scripts ensure the structured storage and timely data update within both local and cloud-based databases. This category of scripts manages the intricacies of data flow from the model outputs to the database entries, encompassing the maintenance of records related to rainfall, water availability, SWAT results, and validation datasets for evapotranspiration. By handling both the ingress and egress of data, these scripts maintain the database as a reliable source for both current operational needs and historical reference.

- **SWATOperationalDatabase:** Performs management operations on the operational database including connection to the local (using Docker) or the AWS databases; deletion of outdated SWAT, precipitation, and evapotranspiration data based on user-defined hindcast months and subsequent insertion of updated data; assessment of water availability using SWAT outputs and update of the obtained results in the database.

Each of the datasets mentioned below provides a distinct type of data crucial for hydrological modeling. CHIRPS offers observed precipitation data, GFS provides short-term weather forecasts, and ECMWF delivers long-term seasonal forecasts. We utilize a "best data source" approach for precipitation data in our database. This means we prioritize specific datasets based on their perceived accuracy. Operationally, there are cases where forecast data may be required to fill gaps in the hindcast period. In such instances, we prioritize GFS in relation to ECMWF seasonal data. This approach ensures that the best

available data is always used—whether observational or forecasted—ensuring more accurate hydrological predictions and continuity in the SWAT model's operation.

1. **CHIRPS:** The publicly available global gridded rainfall product CHIRPS was selected for inclusion in the database based on the comparison with gauge datasets with LRB.
2. **GFS:** If CHIRPS data is unavailable, we use the Global Forecast System (GFS) data as the next best option.
3. **Seasonal ECMWF:** Lastly, for longer-term forecasts, data from the Seasonal European Centre for Medium-Range Weather Forecasts (ECMWF) model.

Data Freshness and Historical Updates:

To ensure data reliability and reflect the latest information, we follow a two-step process:

1. **Data Cutoff:** We maintain a rolling window of precipitation data in the RDS database. This means we typically erase data older than two months from the current date.
2. **Historical and Forecast Updates:** When CHIRPS or other data becomes available for past dates (including those erased) or for upcoming forecasts, we completely overwrite (re-write) the existing historical and forecast precipitation data in the "rainfall_data" table.

This approach ensures our database reflects the most reliable and up-to-date information based on our current understanding of data quality.

Visualization

The visualization category is embodied by the SWATOutputVisualizer, a tool that converts complex model output data into intuitive and informative visual formats. This script allows stakeholders to easily comprehend and analyze the SWAT results, facilitating evidence-based decision-making. By presenting data in a user-friendly manner, the visualizations help to bridge the gap between sophisticated model computations and practical applications, allowing for the clear communication of water availability forecasts and other key hydrological metrics.

- **SWATOutputVisualizer:** Processes and visualizes time series data from SWAT output files.

Architecture and information flow

The SWAT operational model of the Limpopo River Basin is a comprehensive data management and simulation system that uses a structured, modular approach to water resource management and forecasting. The architecture is divided into three primary components or "Scheduled Packages," each focusing on specific tasks and workflows within the larger system (Figure 7).

Scheduled Package #1 involves the data download process, where the system defines the hindcast and forecast periods necessary for operational monitoring. The download process utilizes various data sources, such as FAO, Google Earth Engine, and the Copernicus Client, integrating diverse datasets, including raster images and time series data. Once the relevant data is collected, a Cell Extraction Tool processes this information to extract the needed granularity for further analysis. This stage is vital for initializing the system with the necessary historical and current data, and the outcomes are recorded in a local log file for traceability and quality control.

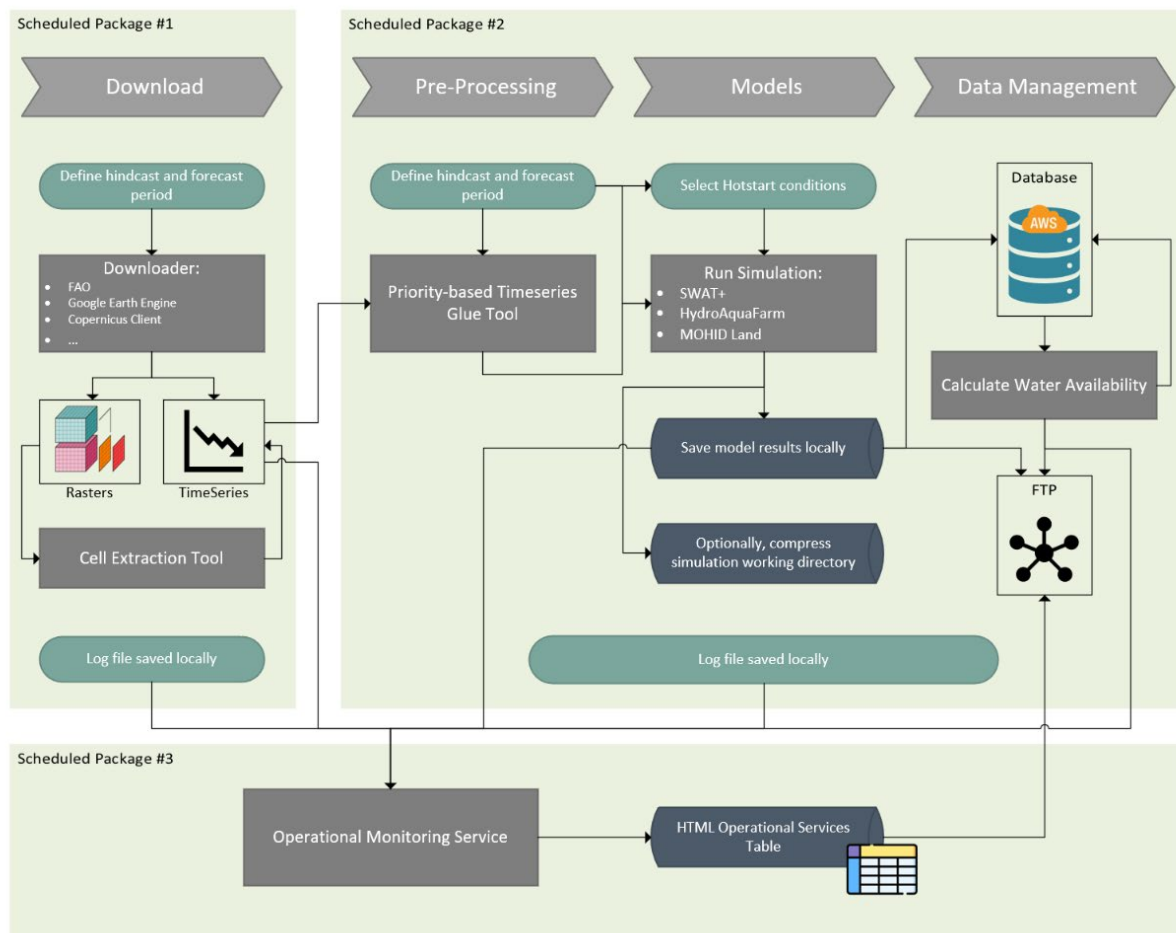


Figure 7 - Integrated SWAT Operational Framework for Automated Water Resource Forecasting in the Limpopo River Basin: A Modular Approach Encompassing Data Acquisition, Pre-Processing, Modeling, and Monitoring

Scheduled Package #2 shifts the focus to pre-processing, where the system again defines the relevant hindcast and forecast periods, this time to prepare the data for model input. It utilizes a Priority-based Timeseries Glue Tool, possibly a custom tool designed to integrate and align time series data based on predefined priorities. After the pre-processing stage, the system selects Hotstart conditions to initialize the models effectively, which include SWAT+, HydroAquaFarm, and MOHID Land. These models simulate various aspects of the hydrological cycle and other relevant processes. Once the simulations are run, the model results are saved locally for immediate use. There's also an option to compress the

simulation working directory, which could be for efficient storage or archiving purposes. The results of these processes are also documented in a local log file.

Scheduled Package #3 and beyond involve utilizing the processed and simulated data for operational monitoring services. These services include real-time monitoring and assessment of water resources, providing information for decision-making. The model results and data are managed through an efficient storage solution, with AWS (Amazon Web Services) databases for secure and scalable cloud storage, and an FTP (File Transfer Protocol) service for data distribution. Additionally, the architecture integrates HTML Operational Services Tables, which serve as a user-friendly interface for visualizing and interacting with the data outputs. These components collectively contribute to a robust framework that supports water availability calculations and other critical analyses for the Limpopo basin, ensuring that stakeholders have access to timely and accurate information for managing water resources effectively.

Windows Task Scheduler, which is a component of Microsoft Windows that provides the ability to schedule the launch of programs or scripts at pre-defined times or after specified time intervals. It's a critical tool for automating routine tasks on a computer. Here's an explanation of how scheduling is set up. The screenshot in Figure 8 showcases the series of tasks within the Windows Task Scheduler tailored to the SWAT Limpopo River Basin operational workflow. The tasks are arranged to automate the downloading and processing of environmental data as well as the execution of various hydrological models. For instance, tasks labelled with "Download - Limpopo" followed by the data type (like "WAPOR 300m Month", "GFS025_hourly", "CHIRPS") are scheduled to automatically retrieve datasets pertinent to the Limpopo region's weather, climate, and hydrology at specific times each day. This ensures that the most current data feeds into subsequent analyses and model simulations without the need for manual intervention.

Name	Status	Triggers	Next Run Ti...	Last Run Time
MicrosoftEdgeUpdateTaskMachineUA	Ready	At 22:36 every day - After trigg...	22/04/2024 10:36:01	22/04/2024 09:36:03
GoogleUpdateTaskMachineUA(6457DA25-0E95-4A51-B3C6-274FC...	Ready	At 02:51 every day - After trigg...	22/04/2024 10:51:25	22/04/2024 09:51:27
MicrosoftEdgeUpdateTaskMachineCore	Ready	Multiple triggers defined	22/04/2024 23:06:01	21/04/2024 23:06:02
GoogleUpdateTaskMachineCore(B0A2BC7B-9A4A-4EFB-9F80-D7F...	Ready	Multiple triggers defined	23/04/2024 02:51:25	22/04/2024 02:51:26
Download - Limpopo - WAPOR 300m Month	Ready	At 05:23 every day	23/04/2024 05:23:00	22/04/2024 05:23:01
Download - Limpopo - GFS025_hourly	Ready	At 05:25 every day	23/04/2024 05:25:00	22/04/2024 05:25:01
Download - Limpopo - CHIRPS	Ready	At 05:30 every day	23/04/2024 05:30:00	22/04/2024 05:30:01
Download - Limpopo - ECMWF_C-10-50-90	Ready	At 05:32 every day	23/04/2024 05:32:00	22/04/2024 05:32:02
Model - Limpopo - FieldStats WAPOR + Timeseries	Ready	At 05:40 every day	23/04/2024 05:40:00	22/04/2024 05:40:01
Model - Limpopo - SWAT_Limpopo_control	Ready	At 06:00 every day	23/04/2024 06:00:00	22/04/2024 06:00:02
Model - Limpopo - SWAT_Limpopo_pc10	Ready	At 06:05 every day	23/04/2024 06:05:00	22/04/2024 06:05:01
Model - Limpopo - SWAT_Limpopo_NatFlows_control	Ready	At 06:08 every day	23/04/2024 06:08:00	22/04/2024 06:08:01
Model - Limpopo - SWAT_Limpopo_pc50	Ready	At 06:10 every day	23/04/2024 06:10:00	22/04/2024 06:10:01
Model - Limpopo - SWAT_Limpopo_pc90	Ready	At 06:15 every day	23/04/2024 06:15:00	22/04/2024 06:15:01
Publications - Limpopo - FTP Water Availability Update	Ready	At 06:20 every day	23/04/2024 06:20:00	22/04/2024 06:20:01
Publications - Limpopo - SWAT Limpopo Plots	Ready	At 06:20 every day	23/04/2024 06:20:00	22/04/2024 06:20:01
Publications - Limpopo - SWAT Limpopo Natural Flow Plots	Ready	At 06:21 every day	23/04/2024 06:21:00	22/04/2024 06:21:01
Publications - Limpopo - SWAT WAPOR Plots	Ready	At 06:22 every day	23/04/2024 06:22:00	22/04/2024 06:22:01
Publications - Limpopo - FTP PNG Plots Update	Ready	At 06:25 every day	23/04/2024 06:25:00	22/04/2024 06:25:01
Publications - Limpopo - Services Monitoring	Ready	At 06:30 every day	23/04/2024 06:30:00	22/04/2024 06:30:01

Figure 8 - Automated SWAT Limpopo Workflow in Windows Task Scheduler: Scheduled Data and Model Operations

In the next phase of the workflow, tasks beginning with "Model - Limpopo" indicate automated runs of the SWAT model, each representing different scenario simulations. These model-related tasks are scheduled sequentially, ensuring that each simulation uses the latest available data. Complementing these are tasks aimed at updating publications and water availability metrics, which are executed shortly after model simulations conclude. The "Next Run Time" and "Last Run Time" columns provide a clear schedule and historical log of each task's execution, facilitating a reliable and systematic update cycle that feeds into the operational monitoring services of the SWAT operational Limpopo River Basin system.

Data integration

Different types of data, such as precipitation and evapotranspiration, are incorporated to provide a comprehensive dataset for hydrological modeling.

Data Sources

The Limpopo Operational SWAT Digital Twin integrates data from various sources to generate comprehensive hydrological insights for the basin. Daily precipitation data comes from CHIRPS and GFS025, retrieved through Google Earth Engine (GEE). Seasonal forecasts are obtained via the Copernicus Climate Data Store (CDS) API, leveraging data from the European Centre for Medium-Range Weather Forecasts (ECMWF). The database maintains a rolling window of precipitation data, typically erasing data older than two months. When new data becomes available (including historical or forecasts), it overwrites existing data in the database.

Evapotranspiration data is sourced from the WAPOR dataset, accessible through the FAO (Food and Agriculture Organization) GIS data API. This combined approach ensures the Digital Twin incorporates the most up-to-date and reliable information for accurate hydrological modeling.

Data Formats

The Limpopo Operational SWAT Digital Twin incorporates a wealth of data formats to provide a holistic understanding of hydrological conditions within the Limpopo River Basin. Precipitation data forms a cornerstone, with CHIRPS and GFS025 offering daily precipitation measurements in CSV format. These files capture the amount of precipitation at specific locations and times. Additionally, ECMWF seasonal forecasts include daily precipitation data alongside percentiles (pc10, pc50, pc90) to account for the range of potential hydrometeorological conditions. For evapotranspiration, the Digital Twin leverages raster data from WAPOR with a 300-meter resolution. This data offers valuable information on the combined effect of soil evaporation, plant transpiration, and intercepted rainfall across the basin. Finally, water availability data provides a detailed picture of water flow and stress conditions within various river channels. This information, stored in CSV format, includes

historical data, flow percentiles, and calculated water availability metrics, along with alert flags to monitor water stress situations.

For more detailed information on the data formats used in the database, refer to Annex I – Data integration. This annex provides comprehensive details on the various data formats used, including CSV files for CHIRPS precipitation data, WAPOR evapotranspiration in GeoTiff and CSV formats, and water availability data. It also outlines how these data formats are structured, stored, and processed to support the hydrological modeling and forecasting capabilities of the Limpopo Operational SWAT Digital Twin.

Data Handling and System Operations

The operational seasonal forecast database for water availability is a specialized hydrological and meteorological information system designed for the storage and analysis of river flow and water availability data. It incorporates a set of interrelated tables containing time-series data, statistical summaries, and metadata related to various aspects of hydrological modeling. This system includes the `swat_output_2` and `water_availability_2` tables, which are tailored to store results from specific model runs using the 10th, 50th, and 90th percentile ECMWF Seasonal Forecasts. These tables have an added `RunLabel` column to differentiate between the forecast runs and are intended to supplement the baseline or 'Control' run data. They operate on a cycle of deletion and updating to ensure that only the most current forecasts are represented, with a monthly data labelling system that tracks from the last day of the current month to the last simulated date. Furthermore, changes to the `swat_output` table have been made to streamline data columns and improve clarity. Tables on evapotranspiration, discharges and reservoirs were also added.

The Figure 9 provides a comprehensive overview of the various numbers on the datasets integrated within the operational framework of the Limpopo River Basin project as of July

5th, 2024. This snapshot highlights the extensive data collection and management efforts that support the SWAT model and other hydrological analyses essential for the region.

Limpopo Digital Twin Operational Database Status

Gridded Rainfall Stations

303 locations
8586 historical dates
150 forecast dates
2647008 records
65448 recent records updated*
Sourced by CHIRPS, GFS025, ECMWF

Verified Discharge Stations

362 stations
192 associated SWAT channels
897931 records
635725 good quality dates
20 recently updated records*
Sourced by DWS through INWARDS API

Unverified Discharge Stations

183 stations
113 associated SWAT channels
96637 records
27 recently updated records*
Sourced by DWS through INWARDS API

Waterbodies

1424 waterbodies
96 connected reservoirs
1766361 records
4062 recently updated records*
Sourced by Digital Earth Africa

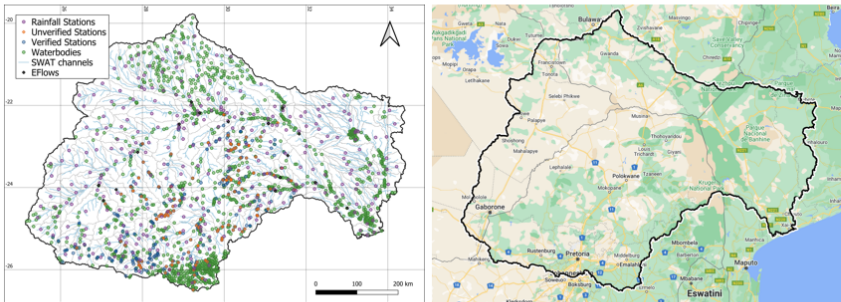
SWAT Flows

1408 channels
282 historical (monthly) dates
5 forecast (monthly) dates per forecast scenario
4 seasonal forecast scenarios (c, pc10, pc50, pc90)
Sourced by IWMI 425216 records

EFlows

16 locations
91 years of data
96637 eflow records
4018 water availability messages
648 alert messages (level 1-4, warning)
Sourced by IWMI

*previous 5 days



Status on the 5th of July 2024

Figure 9 - Limpopo Digital Twin Operational Database Status

The Gridded Rainfall Stations section in Figure 9 reveals that data has been collected from 303 locations, encompassing 8,586 historical dates and 150 forecast dates. This has resulted in a total of 2,647,008 records, with 65,448 recent records updated. These data sources include CHIRPS, GFS025, and ECMWF.

In terms of discharge stations, the database includes verified data from 362 stations, with 192 associated SWAT channels. This segment consists of 897,931 records, 635,725 of which are marked as good quality dates. There have been 20 recent updates sourced from DWS through the INWARDS API. Additionally, there are 183 unverified discharge stations with 113 associated SWAT channels, contributing 96,637 records and 27 recent updates.

Waterbodies are another critical component, with 1,424 monitored waterbodies and 96 connected reservoirs. The database holds 1,766,361 records for these waterbodies, including 4,062 recent updates sourced by Digital Earth Africa. The SWAT Flows data includes information on 1,408 channels, covering 282 historical monthly dates and five forecast monthly dates per forecast scenario across four seasonal forecast scenarios (control, pc10, pc50, and pc90). This section comprises 425,216 records, sourced by IWMI.

The Water Availability (EFlows) section highlights data for 14 locations, with 282 historical and forecast monthly dates for the same four seasonal forecast scenarios. This part of the database contains 4,228 water availability records and 692 alert messages, indicating the system's capacity to monitor and provide warnings for water availability and environmental flow requirements. In the near future, we plan to expand this coverage to include all channels simulated by SWAT, adding the missing sites to the system. This will enhance the database's capacity, enabling more comprehensive monitoring and forecasts across the entire Limpopo River Basin, thereby increasing the overall accuracy and reliability of water availability assessments.

The detailed information on the data available in the database, its organization, and how to retrieve it can be found in Annex II Database management. This annex outlines the structure of the database, including tables for evapotranspiration, discharges, reservoirs, and waterbodies. It also provides guidance on how to access the data through SQL queries, ensuring users can efficiently extract the necessary information for hydrological modeling and forecasting tasks.

Results

This chapter is intended to illustrate the results of the workflow presented in the previous chapters. We start by showing some outputs of the SWAT model. Then we explore the water

availability forecasts, evaluating fluctuations in flow against ecological thresholds using percentile-based scenarios from the ECMWF Seasonal Forecast. Additionally, SQL queries are presented to extract operational data from the SWAT model database, facilitating informed decision-making and resource management.

Breakdown of Flow Types

Statistic Flows (Based on historical measurements, these flows are only available for a limited number of river sections and a specific historical period)

- Pitman_study Natural Flow: This represents the historical flow of the watershed assuming minimal human impact. This is based on historical data and doesn't account for current land use changes, reservoirs, or water withdrawals.
- Statistic Eflow (Environmental Flow): This is a predetermined flow level considered necessary to maintain a healthy ecosystem within the watershed, based on ecological studies presented here (<https://limpopo-eflows.iwmi.org/>).

SWAT Flows (Modeled - Can be forecast and exists for all Limpopo reaches)

- SWAT Present Flow (Present Land and water use): This is the simulated flow generated by the SWAT model under current conditions, including existing land use, reservoirs, and but not including water withdrawals. This is the flow that presently is being operationally forecasted with different seasonal forecasts.
- SWAT Natural Flow (Pristine Land Use): This is a scenario where the SWAT model simulates the flow assuming a natural state for the land use, with no agriculture. It helps assess the impact of human activity on the flow. This is the flow that presently is available in the operational database for historical values. These values can be updated in the future as we gain more knowledge of the system.

SWAT flow estimates

The SWAT results for the control and percentile runs are depicted in the database tables `swat_output` and `swat_output_2`. To exemplify, we present in Figure 10 the monthly estimated flow out and accumulated precipitation for channel 844 (upper and lower plots, respectively). The results are presented for a SWAT model that has run with control, pc10, pc50 and pc90 seasonal forecast precipitation (in orange, blue, green, and red, respectively).



Figure 10 - Flow out and accumulated precipitation for Channel 844 for the SWAT model run with control, pc10, pc50 and pc90 seasonal forecast precipitation. This is the prediction made in 1 of May 2024.

From the upper plot in Figure 10 we can see that for channel 844, the largest flow is obtained using the 50th percentile of the seasonal forecast precipitation data. This is due to the observed peak in the accumulated precipitation of June (see green line in the

precipitation plot, which can be compared to the red curve, representing the fetched daily precipitation for the station associated with this channel). By calculating the accumulated precipitation value over the seasonal forecast period, we confirm that the 10th, 50th and 90th percentile curves correspond to overall drier, median, and wetter forecast scenarios, with accumulated values of 98820 m³, 125798 m³, and 176988 m³, respectively. This reinforces the importance of looking at daily/monthly precipitation input data to better understand the SWAT flow estimates.

Water Availability Forecast

Water availability forecast is calculated based on SWAT estimated flow data and is based on relevant data retrieved from a MySQL database, including flow measurements and statistics from hydrologic stations and associated channels in the SWAT model (Figure 11). Using natural flow and environmental flow (**Eflow**) statistics, it calculates water availability by determining the closest percentile match and computing the difference between SWAT flow and environmental flow. Subsequently, it calculates the actual percentage of water availability relative to eflow. Alert types are assigned based on predefined conditions regarding the actual percentage, and corresponding messages are generated.

Water Availability Forecast

- **Water availability forecast** is calculated using:
 - **SWAT** estimated flow data
 - **Natural flow** and **environmental flow (eflow)** statistics from hydrologic stations associated to the SWAT channels
- Obtained by determining the closest percentile match and computing the difference between SWAT flow and eflow, and the **relative percentage** compared to the eflow.
- **Alert types** are assigned based on predefined conditions regarding the water availability percentage, and corresponding messages are generated.

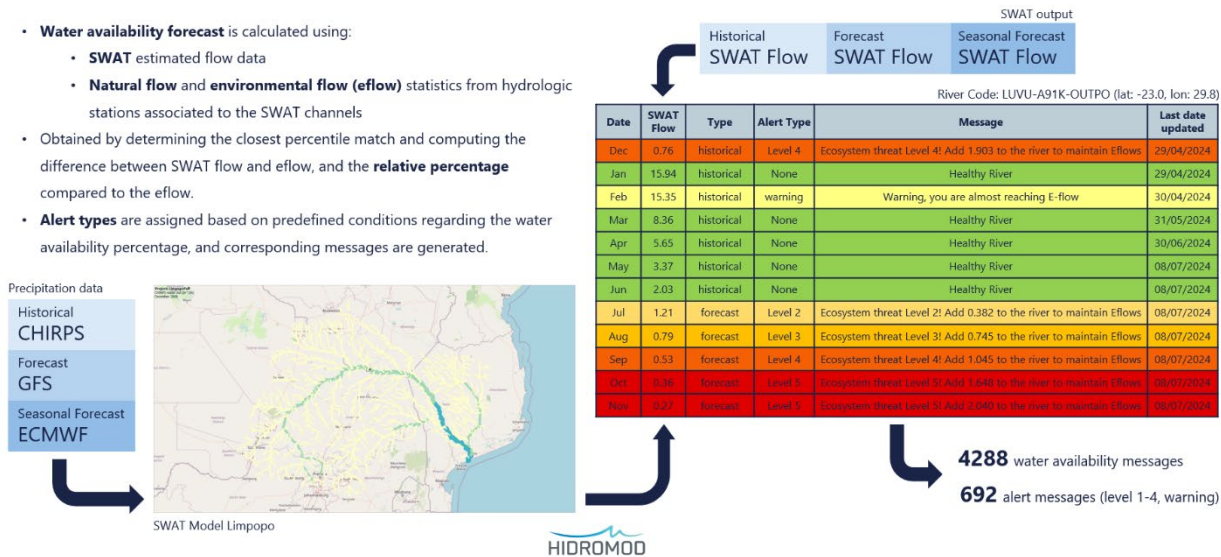


Figure 11 - Methodology and Example of Water Availability Forecasting with SWAT Data

Below we provide a water availability forecast example for the gage station LUVU-A91K-OUTPO, specifying its location coordinates and the date of the last update. It presents a chronological record of flow data, including dates, flow measurements, types (historical or forecasted), corresponding percentiles, natural flow values, Eflow values, computed water availability, actual percentage of water availability relative to Eflow, assigned alert types, and messages. The dataset indicates variations in flow over time, with some instances showing warnings or healthy river conditions based on the comparison between observed flow and environmental flow thresholds. For instance, on 30/04/2024, the forecasted flow of 5.2 m³/s exceeds the 95th percentile of natural flow (5.1 m³/s) and remains above the environmental flow of 3.1 m³/s, indicating a healthy river condition. However, by 31/07/2024, the forecasted flow decreases to 1.1 m³/s, falling below both natural flow and environmental flow thresholds, triggering a Level 2 ecosystem threat alert with a corresponding message advising the addition of 0.488 m³/s to maintain environmental flows.

Example of water availability forecast for river LUVU-A91K-OUTPO.

Table 1 - Geographic and Update Metadata for River Station LUVU-A91K-OUTPO

RiverCode	latitude	longitude
LUVU-A91K-OUTPO	-23.0	29.8

Table 2 - Chronological Water Availability Forecast and Environmental Flow Assessment for River Station LUVU-A91K-OUTPO for the control run

Date	SWAT flow	Type	% Percentile	Natural Flow	Eflow	Water Availability	Actual Percentage	Alert Type	message
29/02/2024	15.4	historical	60.0	16.3	14.9	0.5	3.1	warning	Warning, you are almost reaching E-flow
31/03/2024	8.2	historical	80.0	7.6	6.9	1.3	19.4	None	Healthy River
30/04/2024	5.2	forecast	95.0	5.1	3.1	2.1	67.4	None	Healthy River
31/05/2024	3.1	forecast	99.0	2.6	2.0	1.1	54.7	None	Healthy River
30/06/2024	1.9	forecast	99.9	2.3	1.8	0.1	5.7	warning	Warning, you are almost reaching E-flow
31/07/2024	1.1	forecast	99.9	1.6	1.6	-0.5	-30.7	Level 2	Ecosystem threat Level 2! Add 0.488 to the river to maintain Eflows
31/08/2024	0.7	forecast	99.9	1.6	1.5	-0.8	-54.2	Level 3	Ecosystem threat Level 3! Add 0.834 to the river to maintain Eflows

The water availability forecasts based on the 10th and 90th percentiles from the ECMWF Seasonal Forecast depict varying scenarios for river LUVU-A91K-OUTPO, as described in

tables below. In the 90th percentile scenario, starting from April 30, 2024, the river exhibits relatively healthy surplus water availability at 67.4% compared to the ecosystem flow, gradually declining to a deficit by July and August, triggering alerts of increasing severity, up to Level 3, indicating a critical threat to the ecosystem, necessitating water additions to maintain the Eflows. Conversely, in the 10th percentile scenario, while initially comparable to the 90th percentile, the water availability rapidly decreases by June 30, 2024, reaching a warning level due to the imminent risk of falling below the Eflow threshold, followed by further decline and escalating alert levels by July and August, signalling the urgent need for water additions to prevent ecosystem collapse.

Table 3 - 90th Percentile Water Availability Forecast for River LUVU-A91K-OUTPO

Date	SWAT flow	Type	% Percentile	Natural Flow	Eflow	Water Availability	Actual Percentage	Alert Type	message
30/04/2024	5.2	forecast	95.0	5.1	3.1	2.1	67.4	None	Healthy River
31/05/2024	3.2	forecast	99.0	2.6	2.0	1.2	59.0	None	Healthy River
30/06/2024	2.4	forecast	99.0	2.4	1.8	0.6	30.5	None	Healthy River
31/07/2024	1.6	forecast	99.9	1.6	1.6	0.0	-2.2	Level 1	Drought risk collapse! Add 0.035 to the river to maintain Eflows
31/08/2024	0.9	forecast	99.9	1.6	1.5	-0.6	-41.3	Level 3	Ecosystem threat Level 3! Add 0.635 to the river to maintain Eflows

Table 4 - 10th Percentile Water Availability Forecast for River LUVU-A91K-OUTPO

Date	SWAT flow	Type	% Percentile	Natural Flow	Eflow	Water Availability	Actual Percentage	Alert Type	message
30/04/2024	5.2	forecast	95.0	5.1	3.1	2.1	67.4	None	Healthy River
31/05/2024	3.1	forecast	99.0	2.6	2.0	1.1	54.7	None	Healthy River
30/06/2024	1.9	forecast	99.9	2.3	1.8	0.1	4.7	warning	Warning, you are almost reaching E-flow
31/07/2024	1.1	forecast	99.9	1.6	1.6	-0.5	-30.5	Level 2	Ecosystem threat Level 2! Add 0.485 to the river to maintain Eflows
31/08/2024	0.7	forecast	99.9	1.6	1.5	-0.8	-54.2	Level 3	Ecosystem threat Level 3! Add 0.834 to the river to maintain Eflows

From Table 2, Table 3 and Table 4 in 30/04/2024 the flow predicted is exactly the same for the control and the p10 and p90. The reason for this is that for April, the precipitation input (pcp) is derived from CHIRPS and the Global Forecast System (GFS), neither of which provides percentile-based precipitation data. The percentile values become relevant only when using precipitation from the ECMWF seasonal forecast, which is not the case here.

In May, we integrate precipitation data from both GFS and the ECMWF seasonal forecast. This results in initial discrepancies in water availability projections due to the different nature of the datasets. However, as May progresses, these differences will diminish as we transition to using the most reliable precipitation data available at the time, aligning with our strategy to employ the best available data for precipitation to refine and improve the

accuracy of our forecasts. This approach ensures that our water availability assessments are robust and adapt to the best available climatic data.

Discharges

The introduced discharge monitoring system for the Limpopo River Basin is demonstrated Figure 12, showcasing the integration of verified and unverified discharge stations. The map offers a comprehensive view of the basin, highlighting the spatial distribution of 362 verified stations and 183 unverified stations. This differentiation is crucial for ensuring data accuracy and reliability in hydrological analyses.

Discharge Stations

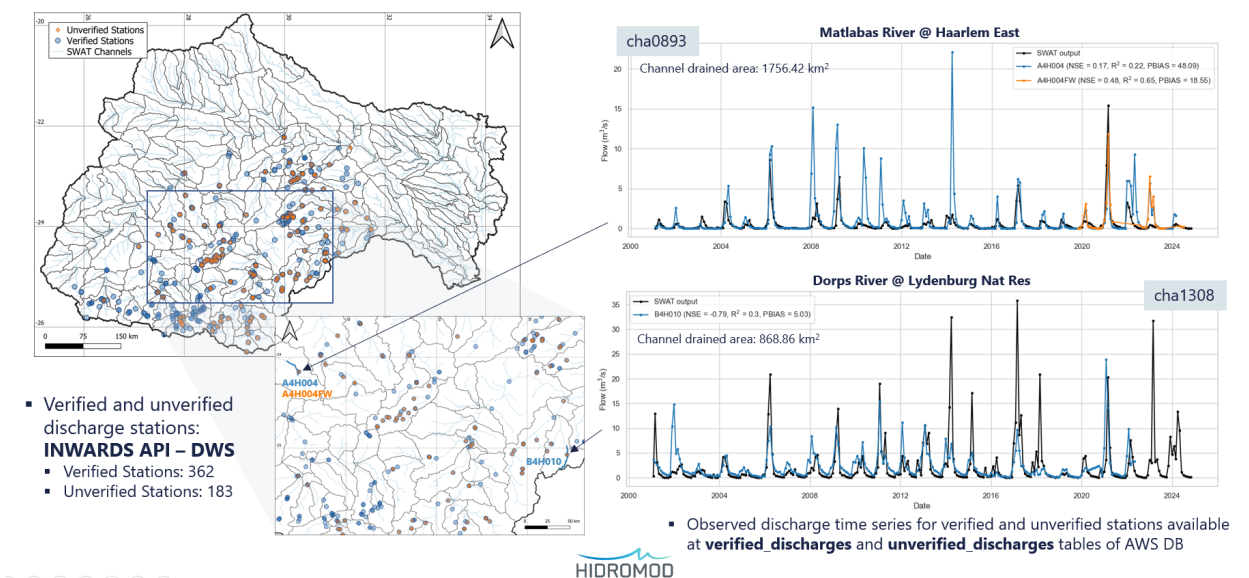


Figure 12 - Discharge Monitoring Network and example Time-Series Data for the Limpopo River Basin. Black Line: SWAT Model Output | Blue Line: Verified Station Discharge | Orange Line: Unverified Station Discharge

The detailed time-series graphs for some discharge stations, such as the Matlabas River at Haarlem East and the Dorps River at Lydenburg Nat Res, illustrate the high-resolution data captured through this monitoring network. These graphs display flow rates over time, providing insights into seasonal variations, peak flow events, and long-term trends. The

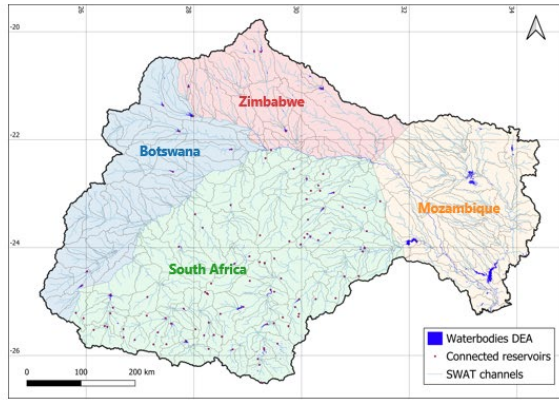
black line represents the SWAT model flow simulations, while the blue line shows the observed flow data from verified gage stations. The orange line corresponds to the observed flow from unverified gage stations. By comparing the SWAT model outputs with observed data, the system ensures the validation and calibration of hydrological models, enhancing predictive capabilities.

The integration of discharge data into the MySQL database, categorized into tables such as 'Verified Discharges' and 'Unverified Discharges', supports advanced data analysis through SQL queries. This structured approach allows users to generate custom visualizations and reports, facilitating more informed water resource management decisions. The system's ability to differentiate between verified and unverified data further strengthens the reliability of the analyses performed. The combination of spatial and temporal data facilitates a deeper understanding of hydrological patterns, ensuring that water management strategies are based on robust and up-to-date information.

Reservoirs

The operational waterbody monitoring system for the Limpopo River Basin is illustrated in Figure 13, which shows the spatial distribution of monitored waterbodies and connected reservoirs across Botswana, Mozambique, Zimbabwe, and South Africa. The system currently monitors 1,424 waterbodies, with 62 in Botswana, 313 in Mozambique, 231 in Zimbabwe, and 818 in South Africa. Additionally, 96 reservoirs have metadata associated within the network, ensuring a comprehensive coverage of the region's hydrological infrastructure.

Operational Waterbody Monitoring



Current database status:

- Monitored waterbodies: **1424**
 - Botswana: 62
 - Mozambique: 313
 - Zimbabwe: 231
 - South Africa: 818
- Connected reservoirs: 96
- Recently updated dates: 5666
- Total number of dates: **1765950**

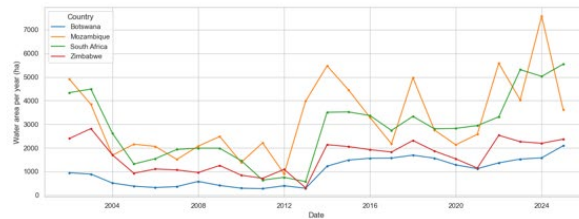


Figure 13 - Overview of Operational Waterbody Monitoring System for the Limpopo River Basin.

The database is robust, containing a total of 1,765,950 entries, with about 5000 updates being made every 5 days. The graph included in the image provides insights into the temporal dynamics of data collection, displaying the frequency and consistency of updates over time across different countries. This visual representation underscores the system's capability to manage and analyze extensive datasets, offering a detailed overview of waterbody conditions.

By integrating this dataset, the monitoring system supports the generation of detailed graphs and visualizations based on user-defined SQL queries. This enables users to analyze trends and patterns, facilitating informed decision-making for water resource management. The data from Digital Earth Africa (DEA) and other sources is seamlessly integrated, ensuring that the database remains up-to-date and accurate, reflecting the current state of waterbodies and reservoirs within the basin.

In Figure 14, we see a spatial distribution map of dams within the Limpopo River Basin alongside time-series graphs illustrating reservoir area and dam levels over time for selected reservoirs, such as the Groot-Marico River at Molatedi Dam and the Koster River at Kosterrivier Dam. These graphs demonstrate the compatibility of data sources, showing correlations between DEA observations and INWARDS API gauge data.

Reservoir Monitoring

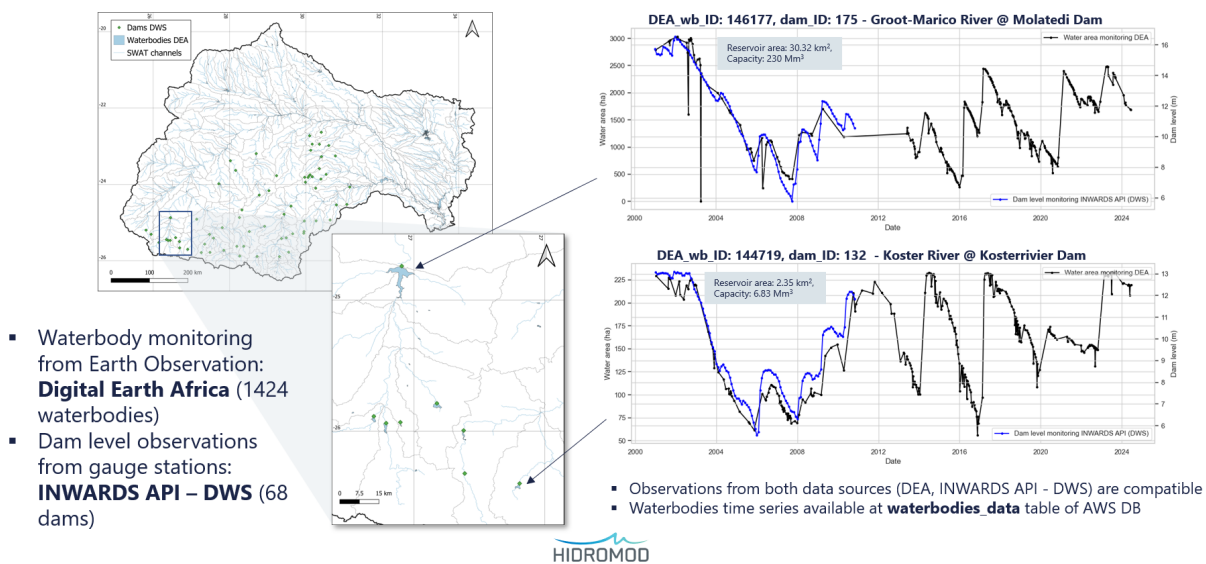


Figure 14 - Integrated Reservoir Monitoring from Earth Observation and Gauge Stations

Figure 14 highlights, dam level observations from gauge stations via the INWARDS API, accounting for 68 dams. The integration of these datasets allows monitoring of reservoir dynamics, supporting real-time water management decisions. By incorporating satellite imagery and ground-based observations, the system ensures robust and accurate reservoir monitoring. This dataset is useful for managing water resources effectively, offering insights into water storage trends, usage patterns, and potential areas for intervention, thereby complementing the overall hydrological analysis framework established in the operational waterbody monitoring.

Monitoring

The `OperationalServicesMonitor` script plays a vital role in overseeing the entire operational system. It continuously evaluates the health and performance of data downloads, processing tasks, database updates, and model executions. By identifying and addressing potential failures or bottlenecks in real-time, the script ensures the uninterrupted operation of the SWAT model's workflow. This vigilant oversight is essential for maintaining the high availability and reliability of the service.

`OperationalServicesMonitor` manages and monitors various operational services by assessing their timeliness and update status. It handles multiple types of data services, including timeseries data, log files, simple files, and SWAT model outputs. After processing all specified services, the script generates an HTML table that provides an easy-to-visualize status of each service.

The HTML table created by `OperationalServicesMonitor` offers a comprehensive overview of the status of various operational services within the SWAT model's ecosystem. This table is critical for real-time monitoring and management, ensuring each component of the system is functioning efficiently. This table is reviewed daily, and a ticket is generated for the team to address any issues that arise. It lists several key attributes for each service:

- **Service Name:** Identifies the specific service being monitored, ranging from downloading logs to the execution and results logging of the SWAT model.
- **Type:** Categorizes the type of data or process being logged, such as logs, SWAT output, simple files, or timeseries data.
- **Initial Date:** Indicates the start date for the data or service being monitored. For instance, the "DEA Waterbodies Download Log" began on 2024-06-28, while the SWAT output data spans from 2023-01-31.

- **End Date:** Shows the most recent date of the data or the latest update for the service. Many entries are current as of 2024-07-09, indicating the system's up-to-date status.
- **Current Lag:** Displays any delay in the service or data update, measured in days. A value of 0 signifies no delay, indicating the service is current.
- **Maximum Lag:** Records the maximum observed delay for the service, helping to understand past performance and identify any recurring delays.
- **On Time:** A Boolean indicator (True/False) showing whether the service is running on schedule. All services in this table are marked as "True," indicating that they are operating within expected timelines.

Each service plays a crucial role in ensuring the accurate simulation of water availability, river flows, and other hydrological variables for the Limpopo River Basin. Here is a description of each service monitored:

- **DEA Waterbodies Download Log:** Tracks the download status of waterbody data from Digital Earth Africa. The data is used for monitoring waterbodies and their areas within the Limpopo River Basin.
- **INWARDS API Discharge Request:** Monitors the status of requests to the INWARDS API, which retrieves discharge data for both verified and unverified stations within the basin.
- **SWAT Control Log:** Logs information related to the general control and execution of the SWAT model, including its operational runs and status checks.
- **SWAT Natural Flows Control Log:** Monitors the execution and status of SWAT model runs that simulate natural flow conditions, where minimal human impact is considered.

- **SWAT PC10 Log:** Tracks the execution of the SWAT model using the 10th percentile (pc10) precipitation data from seasonal forecasts, representing drier-than-average conditions.
- **SWAT PC50 Log:** Monitors the SWAT model runs with the 50th percentile (pc50) precipitation data, representing median forecast conditions.
- **SWAT PC90 Log:** Logs the SWAT model runs using the 90th percentile (pc90) precipitation data, representing wetter-than-average forecast scenarios.
- **WAPOR 300 Raster Download:** Tracks the download of 300-meter resolution raster data from the WAPOR dataset, which includes evapotranspiration data used in hydrological modeling.
- **SWAT Control Channel Results:** Logs the results of the SWAT model simulations based on control scenarios, which are baseline runs without percentile-based seasonal forecasts.
- **SWAT Natural Flows Control Channel Results:** Monitors and logs the results of SWAT simulations for natural flows, assuming pristine conditions in the watershed.
- **SWAT PC10 Channel Results:** Logs and tracks the results of SWAT model runs using the 10th percentile precipitation data, capturing drier scenarios for specific channels.
- **SWAT PC50 Channel Results:** Tracks the results of SWAT model runs using 50th percentile precipitation data, representing typical or median forecast conditions for specific channels.
- **SWAT PC90 Channel Results:** Logs the results of SWAT simulations using the 90th percentile precipitation data, representing wetter-than-average conditions for specific channels.
- **SWAT Flow Plots:** Tracks the generation of flow plots that visualize the SWAT model's flow output for various channels and scenarios.

- **SWAT Natural Flow Plots:** Monitors the creation of plots that visualize natural flow conditions simulated by the SWAT model, showing flow patterns without human impacts.
- **CHIRPS Raw Data:** Tracks the download and processing of raw precipitation data from the CHIRPS dataset, a critical input for SWAT model simulations.
- **ECMWF Data:** Monitors the retrieval and processing of seasonal forecast data from the European Centre for Medium-Range Weather Forecasts (ECMWF), used to simulate different hydrological conditions.
- **GFS Raw Data:** Tracks the download of raw short-term weather forecast data from the Global Forecast System (GFS), which provides daily updates for hydrological forecasting.
- **WAPOR Timeseries:** Monitors the processing of timeseries data from the WAPOR dataset, which includes evapotranspiration and other related environmental variables.

Server and infrastructure

All the scripts previously described form an integral part of the SWAT model operational workflow and are executed on the Tlloc server. This automation is facilitated by the Windows Task Scheduler, a tool native to the Windows operating system, following a sequence that adheres to the dependencies and requirements of the workflow.

For instance, data download scripts will be triggered to retrieve the latest datasets from specified sources like FAO or Google Earth Engine. Once the data is downloaded, the processing scripts take over, manipulating the data to fit the necessary input format for the SWAT models. Following data processing, model execution scripts launch the SWAT

simulations. The entire process is logged and monitored to ensure accuracy and efficiency. The use of Windows Task Scheduler is very important for maintaining the continuity and reliability of the operational SWAT model, enabling the system to function with minimal human intervention and providing timely water resource data for the Limpopo River Basin.

This dedicated server is configured with security measures to mitigate potential risks, ensuring a fortified digital environment. Embedded within its chassis are three distinct hard drives, each serving a purpose tailored to optimize system performance and data integrity. The System drive, labelled C:/, is reserved for the operating system, ensuring a segregated and uncluttered environment for the server's core functionalities. The Data drive, denoted as D:/ and constructed with HDD technology, offers expansive storage capabilities, safeguarded by a RAID 1 or mirror configuration, providing redundancy to protect against data loss. Lastly, the Working drive, E:/, boasts SSD technology, delivering superior speed and efficiency for programs that frequently engage in intensive I/O operations.

Named after the revered Aztec deity of rain, water, lightning, and agriculture, Tlaloc encapsulates the spirit of elemental forces and the nurturing essence of growth. It serves not only as a technological resource but also as an homage to cultural heritage, channelling the ancient god's providence into the modern realm of environmental modeling and analysis.

The server is powered by an Intel(R) Core(TM) i7-7700 CPU and is equipped with a substantial 64 GB of RAM, supporting complex computational tasks (Figure 15). The local storage is strategically partitioned across three drives: the System (C:) drive ensures a clean operating system environment with ample free space for system updates and core applications; the Data (D:) drive, with its vast 3.63 TB capacity, offers extensive data storage capabilities; and the Working (E:) drive utilizes SSD technology for high-performance tasks,

striking a balance between used space and free space to accommodate demanding processing workflows.

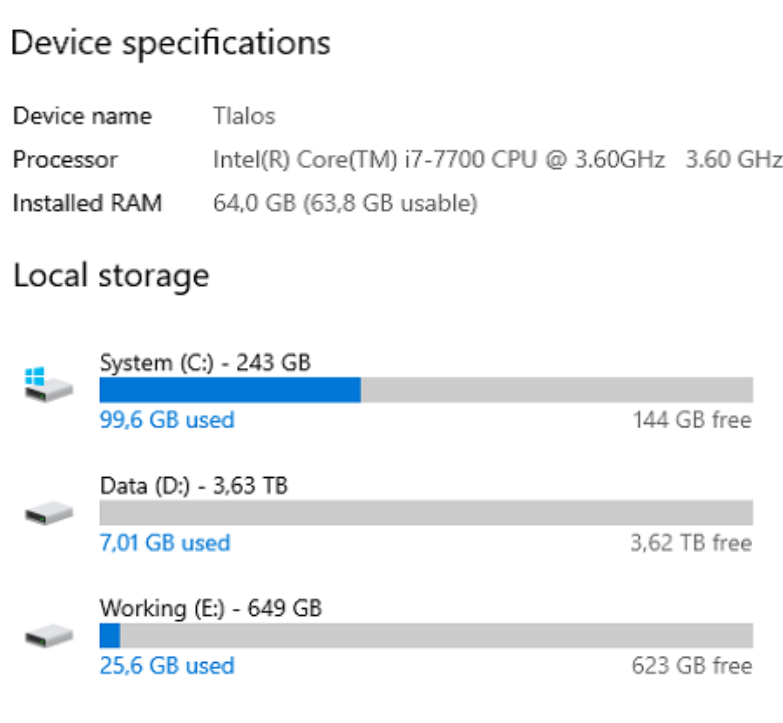


Figure 15 - Storage Allocation and System Resources of Tlaloc Server

The decision to utilize the Tlaloc server from a private cloud provider offers significant advantages over maintaining a local server. A private cloud server ensures higher availability, scalability, and robust security compared to on-premises infrastructure. By leveraging cloud services, the Tlaloc server benefits from professional data center management, redundant power supplies, and 24/7 monitoring, which enhance uptime and system reliability. Additionally, cloud hosting mitigates risks associated with local hardware failures and provides flexibility to scale resources as needed, making it a more resilient and future-proof solution for supporting the SWAT model's operational workflow.

Conclusion

The SWAT+ operational forecasting system LRB represents a significant contribution and knowledge for sustainable water resource management for the region. The project's main achievements include the establishment of a fully operational SWAT+ hydrological model that integrates multi-source datasets. This model's accuracy was validated against WAPOR evapotranspiration data, ensuring reliable outputs. Additionally, a sophisticated system was developed to automate the downloading and processing of seasonal forecast precipitation data from the ECMWF, accessed through the Copernicus Climate Data Store API. The cornerstone of the data management aspect is a custom-designed hydrological database hosted on Amazon Web Services, which guarantees high availability and reliable management of critical river flow and water availability data. High availability means always accessible and minimizes downtime, ensuring continuous access to critical river flow and water availability data. Reliable management ensures data integrity, security, performance, and scalability.

Complementing the model and data management framework is an operational scripting setup that automates the entire workflow, including data acquisition, processing, model execution, database updates, and monitoring. To aid in decision-making, a visualization tool was also established, transforming complex model outputs into accessible formats for stakeholders. The document details the SWAT model's configuration and validation, describes the operational SWAT of the Limpopo River Basin architecture, and explains how the Windows Task Scheduler is leveraged for system automation. It includes an illustrative forecast for river water availability, portraying various scenarios informed by ECMWF Seasonal Forecast percentiles. The document also provides SQL queries for extracting information from the Operational Database, demonstrating the project's commitment to transparency and utility.

Next Steps and Future Direction

Moving forward with the SWAT+ Operational Limpopo River Basin seasonal HydroForecast system requires a strategic focus on data fusion, model calibration, and the integration of alternative variables to enhance its predictive capabilities. Fusion of CHIRPS, local precipitation, and GFS data presents an opportunity to refine precipitation estimates and improve flow predictions. Calibration efforts should commence in upper catchments before extending to larger river systems, taking into account the complexities of reservoir operations and cumulative errors. By systematically calibrating the model and studying the impact of fused precipitation data, the system can achieve greater accuracy in forecasting water availability and river flows.

In addition to precipitation, integration of alternative weather variables such as temperature, wind, and radiation could further enhance model performance. In fact, the present simulation uses the SWAT weather generator to derive the daily values of these variables. Evaluating the effectiveness of incorporating these variables (using for example a sensitivity analysis) is important for improving seasonal rainfall forecasts and improving overall predictive accuracy. Furthermore, the integration of improved seasonal forecasts with other weather variables offers potential for reliable SWAT+ model predictions, thereby providing decision-makers with valuable insights for water resource management.

Operational validation and reporting are essential components of the system's development, requiring a comprehensive assessment of forecast accuracy and uncertainties. By addressing these aspects, the SWAT+ Operational Limpopo River Basin seasonal HydroForecast system can evolve into a robust decision support tool for effective water resource management in the region.

Additionally, to optimize the eflow_iwmi operational database, improve data accessibility, and enhance the clarity of its structure, the database should be revisited. This includes reorganizing tables to consolidate similar datasets, such as unifying precipitation data into a single table, and renaming features to provide a clearer understanding of each section. These changes aim to streamline data retrieval and improve overall efficiency in database management.

One further step aspect is the transformation of the operational system into a Docker-based system using Kubernetes, with the goal of running it on AWS infrastructure. This transition will provide scalability, flexibility, and reliability to the system, allowing for seamless deployment and management of resources. For that, all the scripts written in Python and the SWAT model itself need to be compiled to run efficiently on Linux, ensuring compatibility and optimization within the Docker environment. This step is essential for streamlining the deployment process and maximizing the system's performance on the AWS cloud platform.

Acknowledgement

We would like to express our sincere gratitude to the LIMCOM Member States and the **LIMCOM-UNDP/GEF** project team, supported by the Global Water Partnership Southern Africa (GWPSA), the United Nations Development Programme (UNDP) facilitated through the Global Environment Facility (GEF), for their vision, encouragement and support throughout this project. The invaluable contributions from all stakeholders and partners, whose experience, deep understanding of the system and commitment played a crucial role which cannot be overstated. Their involvement was instrumental in shaping the development process. We also extend our appreciation to the **Pegasys Consulting team**, whose guidance, sharing of data and technical experience were key to aligning and

advancing the modelling efforts. Additionally, we acknowledge the critical contributions of all data providers, particularly **Digital Earth Africa** and the **Department of Water and Sanitation, South Africa**, whose observed datasets were essential to the modelling process. A special thanks to **AWS** for their technical support and cloud computing credits, which significantly enhanced our capacity to deliver this project. Finally, we are deeply appreciative of the **Leona M. and Harry B. Helmsley Charitable Trust** for their generous grant on DIWASA project, which not only enabled this project but also provided an opportunity to collaboratively develop tools that will have a lasting impact on the Limpopo River Basin.

References

- African Development Bank. 2014. *Feasibility Studies for Building Climate Resilience of Limpopo Basin in Mozambique*. Mozambique: OWAS/AWF/GECL/DEPARTMENTS.
- Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. "LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT." *JAWRA Journal of the American Water Resources Association* 73-89. doi:10.1111/j.1752-1688.1998.tb05961.x.
- Bieger, Katrin, Jeffrey G. Arnold, Hendrik Rathjens, Michael White, David Bosch, Peter Allen, Martin Volk, and Raghavan Srinivasan. 2017. "Introduction to SWAT+, a completely restructured version of the soil and water assessment tool." *JAWRA Journal of the American Water Resources Association*, 53(1) 115-130.

- Botai, Joel O., Surajit Ghosh, Karthikeyan Matheswaran, Chris Dickens, Nkateko Langa, and Mariangel Garcia Andarcia. 2023. *Options for digital twin application in developing country river basin management: a review*. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Initiative on Digital Innovation., 20. <https://hdl.handle.net/10568/134763>.
- Brito, R., S. Famba, P. Munguambe, N. Ibraimo, and C. Julaia. 2009. "Profile of the Limpopo Basin in Mozambique, a contribution to the Challenge Program on Water and Food Project 17." *Integrated Water Resource Management for Improved Rural Livelihoods: Managing risk, mitigating drought and improving water productivity in the water scarce Limpopo Basin*. Harare: WaterNet.
- Brocca, L., S. Barbetta, S. Camici, L. Ciabatta, J. Dari, P. Filippucci, C. Massari, et al. 2024. "A Digital Twin of the terrestrial water cycle: a glimpse into the future through high-resolution Earth observations." *Front Sci* (2024) 1:1190191. doi:10.3389/fsci.2023.1190191.
- Busari, Ola. 2007. "Groundwater in the Limpopo Basin: occurrence, use and impact." *Environment, Development and Sustainability* 10 (2008): 943–957. doi:10.1007/s10668-007-9094-3.
- Chambel-Leitão, P. 2016. "Hydrologic Modelling for Portugal." *Doctoral dissertation, Instituto Superior Técnico, Lisbon*.
- Chambel-Leitão, Pedro, Marcel Alderlieste, and Brian Jackson. 2023. "Operational Hydrological Models for Water Management: Case Studies from Australia, Brazil, and Portugal." *SWAT Conference*. Aarhus, Denmark: Aarhus University, Department of Ecoscience.

- Copernicus Climate Change Service. 2018. "Seasonal forecast daily and subdaily data on single levels." *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. Accessed August 2, 2024. doi:10.24381/cds.181d637e.
- FAO. 2020. *WaPOR for monitoring agriculture in conflict areas*. Rome: FAO.
- Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, et al. 2015. "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes." *Sci Data* 2 (150066). doi:10.1038/sdata.2015.66.
- Garcia Andarcia, Mariangel, Chris Dickens, Paulo Silva, Karthikeyan Matheswaran, and J. Koo. 2024. *Digital Twin for management of water resources in the Limpopo River Basin: a concept*. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Initiative on Digital Innovation., 4. <https://hdl.handle.net/10568/151898>.
- Gurusinghe, Thilina, Lal Muthuwatta, Karthikeyan Matheswaran, and Chris Dickens. 2024. *Developing a foundational hydrological model for the Limpopo River Basin using the Soil and Water Assessment Tool Plus (SWAT+)*. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Initiative on Digital Innovation., 14. <https://hdl.handle.net/10568/151939>.
- Mosase, Esther, Laurent Ahiablame, and Raghavan Srinivasan. 2019. "Spatial and temporal distribution of blue water in the Limpopo River Basin, Southern Africa: A case study." *Ecohydrology & Hydrobiology* 19 (2): 252-265. doi:10.1016/j.ecohyd.2018.12.002.
- Nhassengo, Osvaldo, Hiroaki Somura, and June Wolfe. 2021. "Environmental flow sustainability in the Lower Limpopo River Basin, Mozambique." *Journal of Hydrology: Regional Studies* 36: 100843. doi:10.1016/j.ejrh.2021.100843.

- Reason, C. J. C., S. Hachigonta, and R. F. Phaladi. 2005. "Interannual variability in rainy season characteristics over the Limpopo region of southern Africa." *International Journal of Climatology* 25 (14): 1835-1853. doi:10.1002/joc.1228.
- Ribeiro, R. B., M. S. Ruiz, A. F. P. Sampaio, T. S. Santos, R. M. Ferreira, P. Chambel-Leitão, A. C. P. B. Santos, and J. Leitão. 2022. "Monitoramento fluviométrico em tempo quase-real e modelagem hidrológica na bacia hidrográfica do Rio Cubatão." *XXIV ENCOB 2022 - Encontro Nacional dos Comitês de Bacias Hidrográficas*. Foz do Iguaçu - Paraná.
- Williams, J. R., J. G. Arnold, J. R. Kiniry, P. W. Gassman, and C. H. Green. 2008. "History of model development at Temple, Texas." *Hydrological Sciences Journal* 948-960. doi:10.1623/hysj.53.5.948.
- Zhu, Tingju, and Claudia Ringler. 2012. "Climate Change Impacts on Water Availability and Use in the Limpopo River Basin." *Water* 4 (1): 63-84. doi:10.3390/w4010063.
- Brocca L.; Barbetta S.; Camici S.; Ciabatta L.; Dari J.; Filippucci P.; Massari C.; Modanesi S.; Tarpanelli A.; Bonaccorsi B.; Mosaffa H.; Wagner W.; Vreugdenhil M.; Quast R.; Alfieri L.; Gabellani S.; Avanzi F.; Rains D.; Miralles DG.; Mantovani S.; Briese C.; Domeneghetti A.; Jacob A.; Castelli M.; Camps-Valls G.; Volden E.; Fernandez D. (2024) A Digital Twin of the terrestrial water cycle: a glimpse into the future through high-resolution Earth observations. *Front Sci* (2024) 1:1190191. doi: 10.3389/fsci.2023.1190191

Annex I – Data integration

Data formats

CHIRPS format

This is a format used to store daily precipitation data, typically stored in a comma-separated values (CSV) file. Each row in the CSV file represents a precipitation measurement at a specific location and time. The first column in the file contains the date of the measurement, while the subsequent columns contain the location code, longitude, latitude, and amount of precipitation. This file contains data for five precipitation measurements. The first measurement was taken on January 2, 2001, at location code 1. The longitude for this measurement is 25.438 degrees east, and the latitude is -24.734 degrees south. The amount of precipitation for this measurement was 3.208 millimeters.

```

File Edit Search View Encoding Language Settings Tools Macro Run Plugins
data_update.py input.dat Limpopo_CHIRPS_20010102000000.csv
1 date_time, location code, long, lat, pcp_mm/24h
2 2001-01-02, 1, 25.438, -24.734, 3.2081825733184814
3 2001-01-02, 2, 25.513, -24.457, 3.1062018871307373
4 2001-01-02, 3, 25.514, -25.019, 0.0
5 2001-01-02, 4, 25.755, -23.417, 0.0
6 2001-01-02, 5, 25.755, -24.217, 3.9907383918762207
    
```

WAPOR

L1AETIM from WAPOR is a dataset containing monthly global data on actual evapotranspiration and interception at a 300-meter resolution. It provides information on the combined effect of soil evaporation, plant transpiration, and evaporation of intercepted rainfall. This data can be useful for various applications in hydrology, agriculture, and climate studies.

Here are the details of the folders and files displayed in L1AETIM_Limpopo (Figure 16):

Folders:

- "FieldStats"
- "NullMasks"
- "Rasters"
- "TimeSeries"

Here are some additional details about the contents of the folder:

- "FieldStats" contains the csv obtained from the downloaded WAPOR raster files.
- "Rasters" folders contain WAPOR GeoTiff files.
- "NullMasks" folder contain GeoTiff files that define areas of missing data.
- "TimeSeries" folder stores WAPOR over time, obtained based on the FieldStats csv.

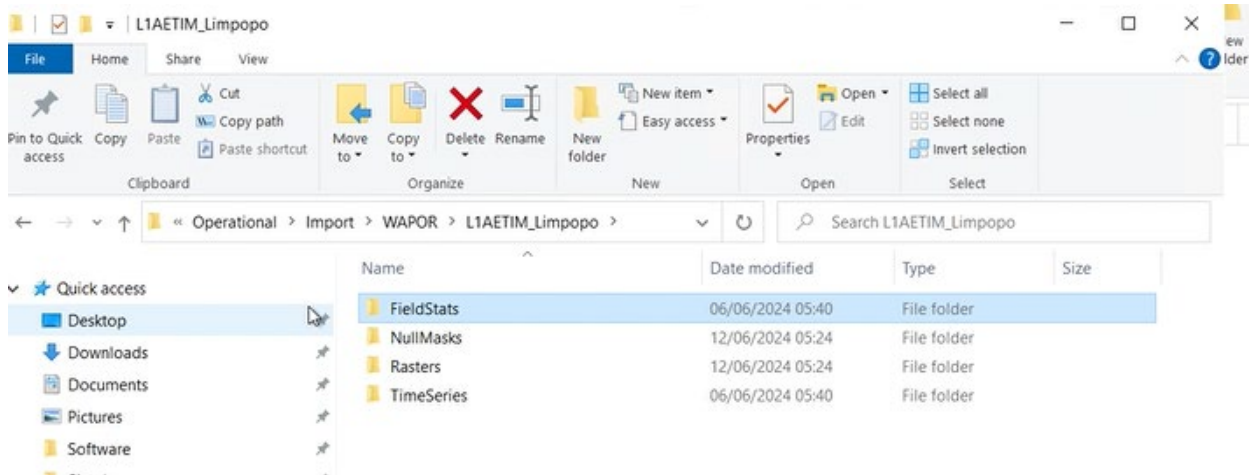


Figure 16 - Folder Structure of WAPOR Containing FieldStats, Rasters, NullMasks, and TimeSeries

WAPOR export format in CSV file with the following structure (**Error! Reference source not found.**):

- date_time
- location_code
- long (longitude)
- lat (latitude)
- et_mm/month (evapotranspiration in millimeters per month)

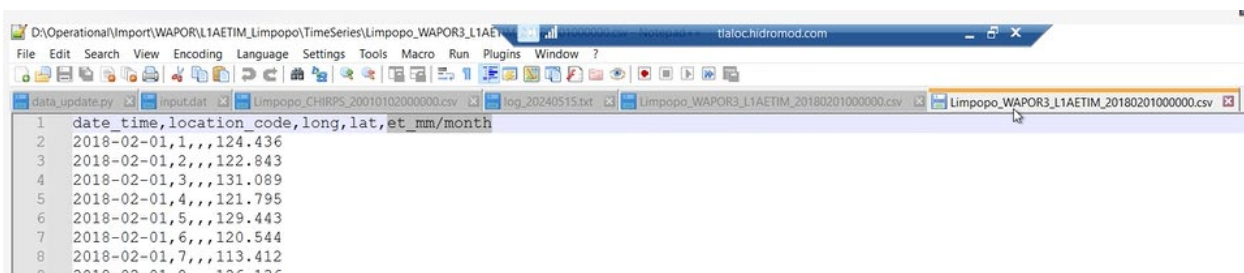


Figure 17 - WAPOR export format in CSV file

Water availability

This data format provides a detailed picture of water flow and availability for various river channels within the system. It includes historical data, flow percentiles, and calculated

water availability metrics. The "Alert Type" and "message" columns show the data can be used for monitoring water stress and triggering alerts.

```

D:\temp\water_availability_20240531_20241130_20240704.csv - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
water_availability_20240531_20241130_20240704.csv [2]
1 ChannelR,Subbasin,jday,month,day,yr,gis_id,name,Date,flo_out,last_date_updated,Type,RiverCode,latitude,longitude,% Percentile,Natural Flow,Eflow,Water Availability,Actual Percentage,Alert Type,message
2 221,63,152,5,31,2024,222,cha0222,2024-05-31,34.02,2024-07-04,historical,LIMP-A711-MAFUN,-22.760838932587345,28.18814227033181,15.0,32.719,11.571,22.449000000000005,194.011,None,Healthy River
3 235,51,152,5,31,2024,237,cha0237,2024-05-31,1.0,2024-07-04,historical,SHAS-V208-TULIB,-22.571653733721764,28.02126433172482,15.0,1.224,0.342,0.6579999999999999,192.398,None,Healthy River
4 535,106,152,5,31,2024,538,cha0538,2024-05-31,50.32,2024-07-04,historical,LIMP-Y300-PAFUR,-23.32192464590184,30.31749338460196,20.0,49.156,13.173,37.147,281.993,None,Healthy River
5 226,99,152,5,31,2024,370,cha0370,2024-05-31,21.74,2024-07-04,historical,LIMP-A36C-LIMFK,-23.057552470902014,27.76473893301556,15.0,21.558,8.762,12.977999999999999,148.117,None,Healthy River
6 407,92,152,5,31,2024,411,cha0411,2024-05-31,9.42,2024-07-04,historical,SAND-A71K-R508B,-22.999341640491079,28.802046226071056,0.1,6.505,0.227,5.193,4049.78,None,Healthy River
7 431,93,152,5,31,2024,432,cha0432,2024-05-31,3.37,2024-07-04,historical,LUVU-B31K-OUFPO,-23.01874525062806,29.822817840732213,95.0,3.771,2.008,1.362,67.829,None,Healthy River
8 592,147,152,5,31,2024,558,cha0558,2024-05-31,1.373,2024-07-04,historical,LEHA-A50H-SEKPO,-23.73806185818496,26.8878628622614,90.0,1.651,0.421,0.952,236.128,None,Healthy River
9 652,131,152,5,31,2024,653,cha0653,2024-05-31,0.7412,2024-07-04,historical,SHIN-B90H-FOACH,-23.892716190547949,30.077805860561292,15.0,0.71,0.414,0.3272,79.034,None,Healthy River
10 825,187,152,5,31,2024,826,cha0826,2024-05-31,10.81,2024-07-04,historical,LIMP-A41D-SFAMW,-24.45865481954325,26.01696932906818,20.0,11.242,6.868,3.542,57.397,None,Healthy River
11 837,182,152,5,31,2024,839,cha0839,2024-05-31,52.66,2024-07-04,historical,KIEP-Y30C-SIMOU,-24.61064974561628,30.64564995949481,30.0,56.28,16.499,36.163,218.171,None,Healthy River
12 999,206,152,5,31,2024,855,cha0855,2024-05-31,140.9,2024-07-04,historical,OLIF-B73C-MANBA,-24.81196222078745,29.803176313789667,0.1,99.133,8.349,132.55100000000002,1587.627,None,Healthy River
Normal text file length: 18,818 lines: 100 Ln: 2 Col: 35 Pos: 237 Windows (CR LF) UTF-8 INS

```

The water availability format provided is CSV file containing data on water flow and availability in a river system. Here's a breakdown of the column headers:

- **ChannelR:** unique identifier for SWAT river channel.
- **Subbasin:** The subbasin within which the river channel is located.
- **jday:** Julian day of the year (day of the year since January 1st).
- **month, day, yr:** Date in separate month, day, and year format.
- **gis_id:** Unique identifier assigned by a Geographic Information System (GIS)
- **name:** A name for the specific river reach or monitoring station.
- **Date:** Date in YYYY-MM-DD format, likely representing the date the data was collected or updated.
- **flo_out:** Outflow from the river channel (units not specified).
- **last_date_updated:** Date the data was last updated in YYYY-MM-DD format.
- **Type:** Data type, "historical" or "forecast".
- **RiverCode:** Unique identifier for the river.
- **latitude, longitude:** Geographic coordinates of the monitoring station.
- **% Percentile:** Percentile of flow compared to historical data.
- **Natural Flow:** Natural flow of the river (m3/s).
- **Eflow:** Environmental flow requirement (m3/s).
- **Water Availability:** Calculated water availability (m3/s).

- **Actual Percentage:** Percentage of available water currently present.
- **Alert Type:** Alert status related to water availability.
- **message:** Additional message related to the alert status.

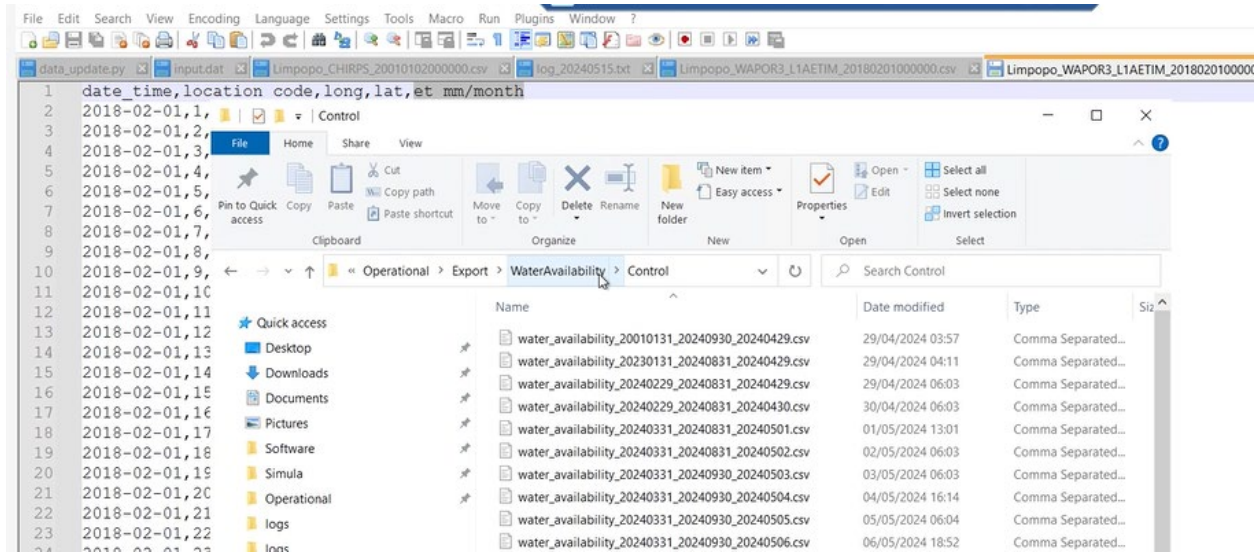


Figure 18 - The water availability format provided is CSV file

Annex II Database management

Based on the hierarchy and their relationships, the tables in the database can be listed as follows:

Figure 19 - Database Structure for the Limpopo River Basin Operational SWAT+ System

Category	Table Name	Description
Metadata	flowtype	Defines types of flow information related to hydrological modeling.

	channel_coordinate	Stores coordinates of river channels for mapping and model linkage.
	rainfall_station	Information about rainfall station locations and coordinates.
	discharge_quality_codes	Contains codes indicating the quality of discharge data.
	reservoirs	Metadata about reservoirs in the basin, including name, location, and capacity.
	waterbodies	Stores metadata for various water bodies, including geometry and surface area.
eFlow	statisticsummary	Summary statistics of environmental flows used for water availability and environmental assessments.
SWAT Flows	swat_output	Primary table containing the SWAT model output for flow data.
	swat_output_2	Complementary table to swat_output, storing percentile-based forecast data.
Water Availability	water_availability	Contains current water availability results from SWAT outputs.
	water_availability_2	Complementary table to water_availability, storing forecast data (pc10, pc50, pc90) with added RunLabel for scenario tracking.
Precipitation	rainfall_data	Time-series data on daily precipitation collected from various sources.
Evapotranspiration	et_validation	Contains validation data for evapotranspiration, from WAPOR.
Discharges	verified_stations	Contains details of verified discharge measurement stations.
	unverified_stations	Stores information on unverified discharge stations, with location and status.
	verified_discharges	Time-series data of discharge from verified stations.
	unverified_discharges	Time-series data of discharge from unverified stations, may include water level information.
Reservoirs	waterbodies_data	Time-series data on surface area and other metrics for waterbodies and reservoirs.

The database is hosted on Amazon Web Services (AWS) platform, utilizing the Amazon Relational Database Service (RDS). Located in the af-south-1 region, AWS's designation for the Africa (Cape Town) region, the database is accessed through the endpoint `iwmi-dt-db-operational.cluster-cd6gaumoede8.af-south-1.rds.amazonaws.com`. The hosting on AWS RDS ensures a high degree of availability and reliability.

The database, named `eflow_iwmi`, is the central repository for the SWAT operational data and is configured to facilitate secure connections via port 3306, which is the default for MySQL databases. Utilization of the JDBC (Java Database Connectivity) standard in the connection string suggests the possibility of database interactions being managed by Java-based applications or tools that support JDBC drivers. This setup not only facilitates reliable data management and analysis but also integrates seamlessly with the operational workflows of SWAT, ensuring that data-driven decisions are based on the latest and most accurate forecasts.

Discharges

These tables will facilitate the integration and management of hydrological data, specifically discharge data (INWARD API), for the Limpopo River Basin.

The new tables included in this update are:

- Discharge Quality Codes: Defines quality codes associated with discharge data.
- Verified Stations: Stores information about verified gauging stations for discharge measurement.
- Verified Discharges: Holds time-series discharge data for verified stations.
- Unverified Stations: Contains details on unverified gauging stations.
- Unverified Discharges: Stores time-series discharge data for unverified stations (may include water level data as well).

Discharge Quality Codes

Name: discharge_quality_codes

Columns:

- id: quality code unique identifier (primary key)
- symbol: 1-character symbol identifying quality code
- description: short description of quality code

Verified Stations

Name: verified_stations

Columns:

- id: A unique identifier for each verified station (primary key)
- station_name: Station code (string) identifying the station
- wma: The name of the water management area where the station is located.
- location: A descriptive location or the geographical place of the station.
- latitude: y-coordinates of verified station (EPSG: 4326)
- longitude: x-coordinates of verified station (EPSG: 4326)
- swat_channel: swat channel id associated to verified station (channel_coordinates primary key).

- latest_record: timestamp of last discharge timeseries update for station.
- date_modified: timestamp of last insertion or modification for verified station, automatically attributed by database, in UTC (in discussion whether these stations' parameters will be modified throughout).

Verified Discharges

Name: verified_discharges

Columns:

- id: A unique identifier for each timeseries value for each verified station (primary key, automatic incrementation)
- Station_id: identifier of verified station containing the timeseries value (verified_stations primary key)
- Date: date for station timeseries value
- Discharge: value of discharge for station at date (m³/s).
- Quality_code: code indicating the quality of the data presented for date, value, and station (discharge_quality_codes primary key)
- Last_date_updated: timestamp of timeseries insertion in database, automatically attributed by database, in UTC

Unverified Stations

Name: unverified_stations

Columns:

- id: A unique identifier for each unverified station (primary key)
- Station_name: Station code (string) identifying the station
- wma: The name of the water management area where the station is located.
- location: A descriptive location or the geographical place of the station.
- station_status: The operational status of the station (ACTIVE or INACTIVE).
- latitude: y-coordinates of unverified station (EPSG: 4326)
- longitude: x-coordinates of unverified station (EPSG: 4326)
- Verified_station: verified station id associated with unverified station, by proximity (verified_stations primary key)

- `swat_channel`: swat channel id associated to unverified station (`channel_coordinates` primary key).
- `latest_record`: timestamp of last discharge timeseries update for station.
- `date_modified`: timestamp of last insertion or modification for unverified station, automatically attributed by database, in UTC (in discussion whether these stations' parameters will be modified throughout).

Unverified Discharges

Name: `unverified_discharges`

Columns:

- `id`: A unique identifier for each timeseries value for each unverified station (primary key, automatic incrementation)
- `Station_id`: identifier of unverified station containing the timeseries value (`unverified_stations` primary key)
- `Date`: date for station timeseries value
- `Discharge`: value of discharge for station at date (m^3/s).
- `Level`: value of water level for station at date (m).
- `Last_date_updated`: timestamp of timeseries insertion in database, automatically attributed by database, in UTC

Diagram

The database schema diagram is shown in Figure 21 and illustrates the structure and relationships between tables in the `eflow_iwmi` MySQL database, focusing on the integration of discharge monitoring data. This diagram highlights the connections between tables storing verified and unverified discharge data, station information, and quality codes, ensuring a comprehensive and organized data repository.

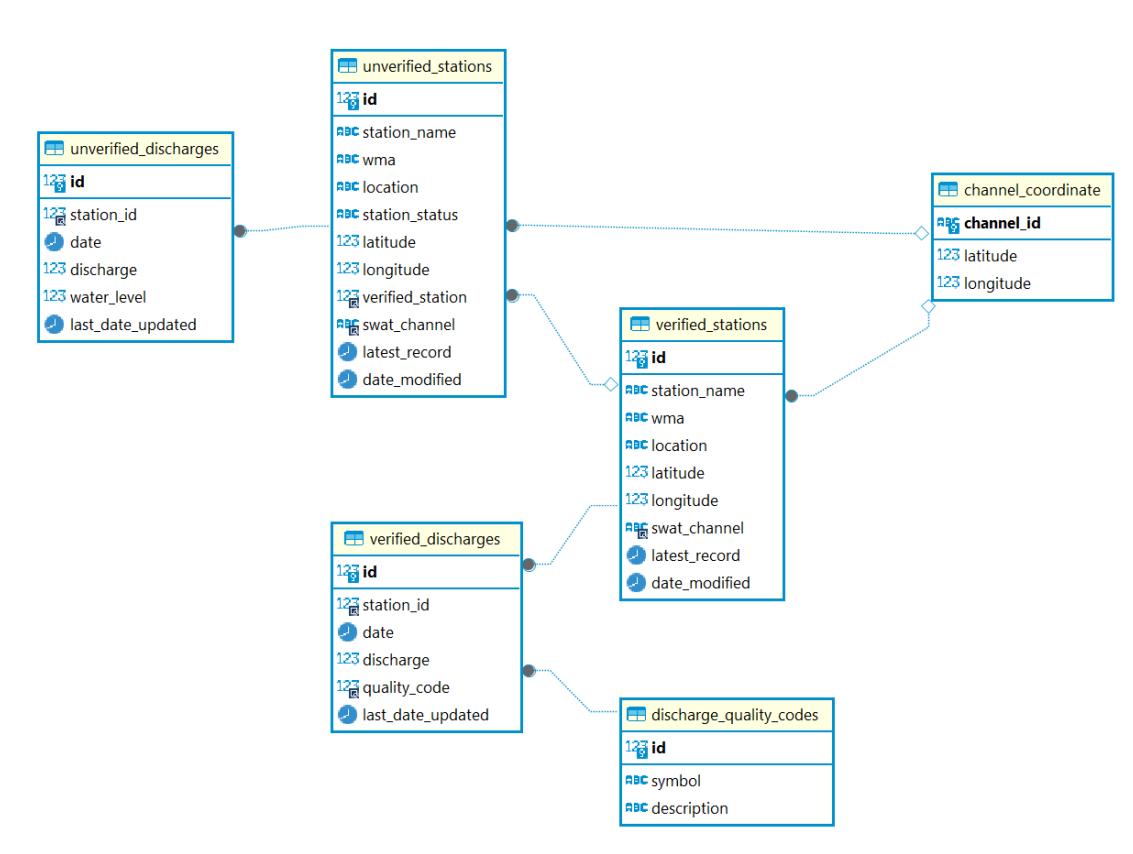


Figure 21 - Database Schema for Discharge Monitoring.

Reservoirs

These tables will facilitate the integration and management of hydrological data, specifically discharge data (INWARD API) and reservoir information (Digital Earth Africa DEA), for the Limpopo River Basin.

The new tables included in this update are:

- Reservoirs: Provides information on reservoirs within the basin.
- Waterbodies: Stores details on water bodies, potentially including reservoirs.
- Waterbodies Data: Holds time-series water body data, such as water area.

Reservoirs

Name: reservoirs

Columns:

- id: Unique reservoir identifier (primary key=
- name: Name of the reservoir
- description: Additional information of the reservoir
- country: Country where the reservoir is located
- latitude: Latitude of the reservoir
- longitude: Longitude of the reservoir
- area_ha: Water surface area in ha
- capacity_mcm: Reservoir capacity in million cubic meter
- impounded_river: River where the reservoir is located
- water_use: Use for the water stored in the reservoir
- data_sources: Source of available reservoir metadata
- waterbody_id: A unique identifier for each WB (waterbodies primary key)

Waterbodies

Name: waterbodies

Columns:

- id: A unique identifier for each WB (primary key)
- area_ha: Area of the water body in ha
- length_m: Length of the waterbody
- perimeter_m: Perimeter of the waterbody
- geometry: Geometry of the waterbody in EPSG:4326
- swat_channel: A foreign key referencing the SWAT ChannelID in the channel_coordinate table (channel_coordinates primery key)

Waterbodies data

Name: waterbodies_data

Columns:

- id: Unique timeseries row identifier (primary key)
- waterbody_id: A unique identifier linking the waterbody water area record to the waterbody.date: The date of the satellite image acquisition used to measure the water area of the waterbody
- water_area_ha: water area in ha in the reservoir
- percent_invalid: percentage of invalid area (not clearly observed in the satellite acquisition)
- last_date_updated: timestamp of timeseries insertion in database, automatically attributed by database, in UTC

Diagram

The database schema diagram in Figure 22 illustrates the integration and management of reservoir and waterbody data within the eflow_iwmi MySQL database. This structure includes detailed tables for reservoirs, waterbodies, and related time-series data, ensuring a comprehensive approach to hydrological data management.

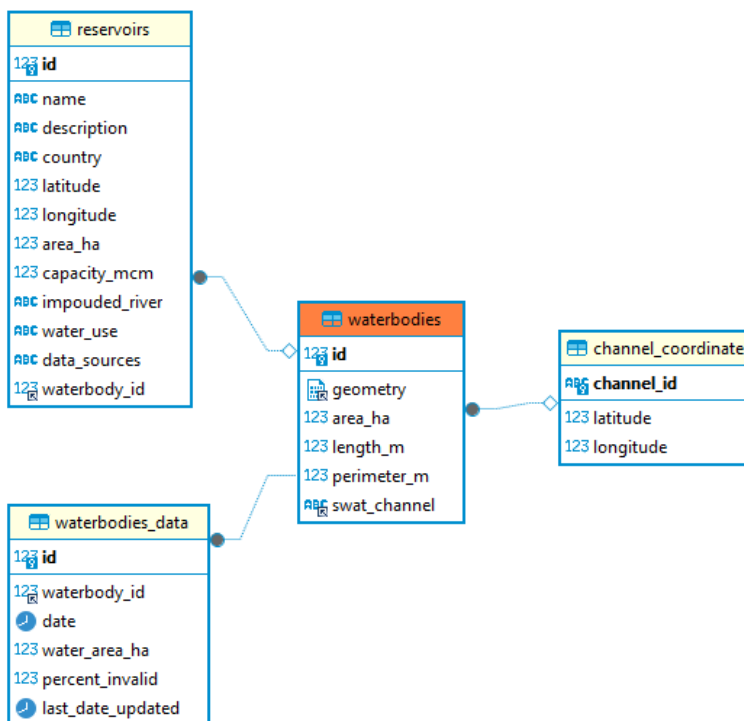


Figure 22 - Database Schema for Reservoir and Waterbody Data Integration.

SQL queries to extract data from the Operational DB

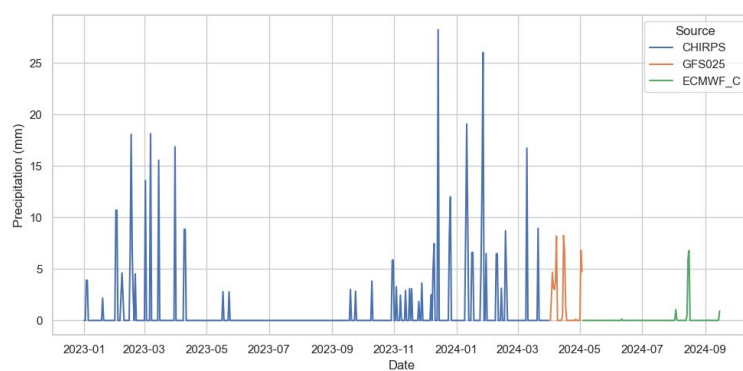
This chapter presents SQL queries essential for retrieving data from the Operational Database supporting the SWAT model. The queries are designed to extract datasets, such as daily precipitation figures and time-series of flow and water availability. These SQL statements are tailored to access the latest updated records from specific tables like `rainfall_data`, `swat_output`, and `water_availability`, corresponding to various stations and river codes such as LIMP-A71L-MAPUN. The intention is to provide a clear pathway for fetching relevant data that represents both control and forecasted scenarios across different percentiles, thereby facilitating detailed analysis and informed decision-making. Each query serves as a building block in the operational workflow, allowing for a seamless transition from raw data to actionable insights.

Daily precipitation data

Extract available **daily precipitation data** from the **rainfall_data** table:

```
SELECT x.date, x.precipitation, x.Source FROM eflow_iwmi.rainfall_data x
WHERE x.station = 's24734s25438e'
AND x.Source IN ('CHIRPS', 'GFS025', 'ECMWF_C')
```

	date	123 precipitation	ABC Source
1	2001-01-01	3.20818	CHIRPS
2	2001-01-02	0	CHIRPS
3	2001-01-03	0	CHIRPS
4	2001-01-04	0	CHIRPS
5	2001-01-05	0	CHIRPS
6	2001-01-06	0	CHIRPS
7	2001-01-07	0	CHIRPS
8	2001-01-08	0	CHIRPS
9	2001-01-09	0	CHIRPS
10	2001-01-10	0	CHIRPS
11	2001-01-11	0	CHIRPS
12	2001-01-12	6.94683	CHIRPS
13	2001-01-13	0	CHIRPS
14	2001-01-14	0	CHIRPS
15	2001-01-15	0	CHIRPS

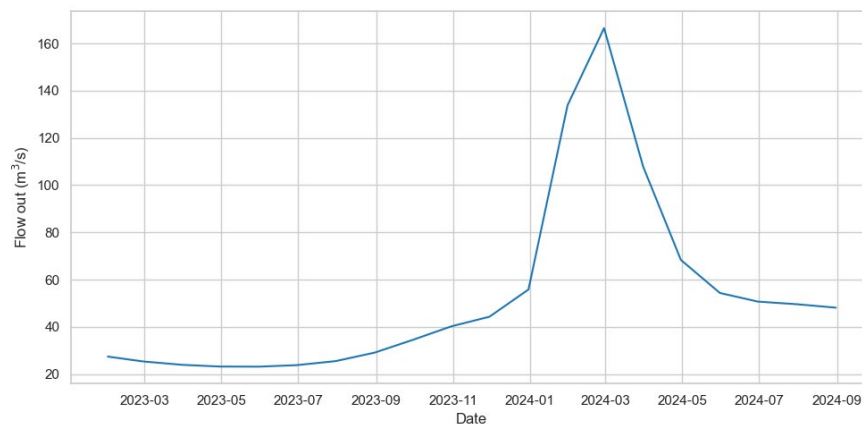


Flow timeseries (control seasonal forecast)

Extract lastly updated **flow timeseries** (control seasonal forecast) from the **swat_output** table for channel 222 associated to the flow station with RiverCode LIMP-A71L-MAPUN:

```
SELECT x.date, x.flo_out FROM eflow_iwmi.swat_output x
INNER JOIN eflow_iwmi.metadata m ON x.Channel = m.channel_id
WHERE m.RiverCode = 'LIMP-A71L-MAPUN'
```

	date	123 flo_out
1	2001-01-31	0.00008774
2	2001-02-28	0.009142
3	2001-03-31	0.1282
4	2001-04-30	0.2789
5	2001-05-31	0.2943
6	2001-06-30	0.2974
7	2001-07-31	0.2808
8	2001-08-31	0.2429
9	2001-09-30	0.2297
10	2001-10-31	0.1991
11	2001-11-30	0.2833
12	2001-12-31	0.3019
13	2002-01-31	0.2859
14	2002-02-28	0.2999
15	2002-03-31	0.564

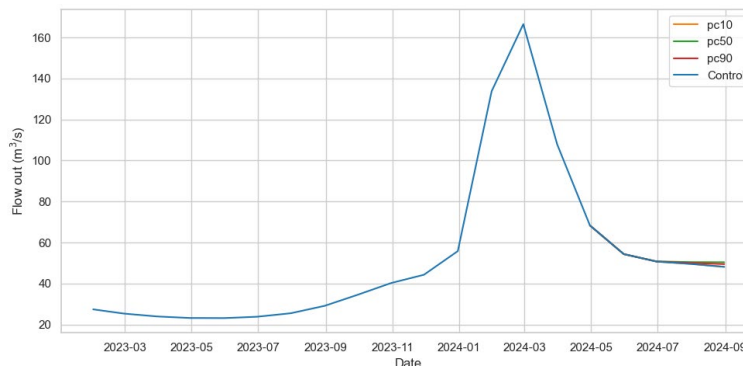
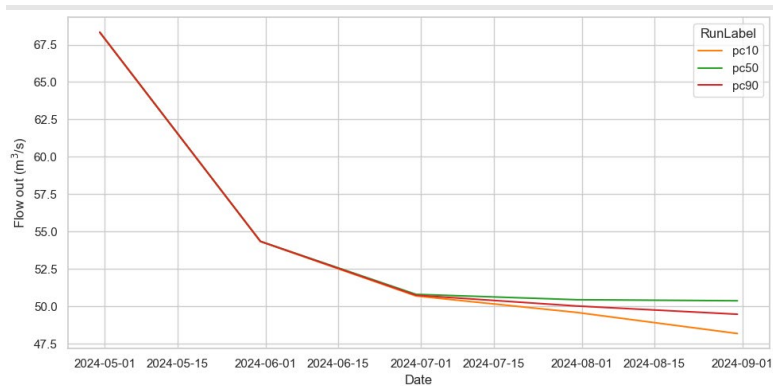


Flow timeseries (pc10, pc50, pc90 seasonal forecast)

Extract lastly updated **flow timeseries** (pc10, pc50, pc90 seasonal forecast) from the **swat_output_2** table for channel 222 associated to the flow station with RiverCode LIMP-A71L-MAPUN:

```
SELECT x.date, x.flo_out, x.RunLabel FROM eflow_iwmi.swat_output_2 x
INNER JOIN eflow_iwmi.metadata m ON x.Channel = m.channel_id
WHERE m.RiverCode = 'LIMP-A71L-MAPUN'
```

	date	123 flo_out	ABC RunLabel
1	2024-04-30	68.33	pc10
2	2024-05-31	54.34	pc10
3	2024-06-30	50.7	pc10
4	2024-07-31	49.6	pc10
5	2024-08-31	48.19	pc10
6	2024-04-30	68.33	pc50
7	2024-05-31	54.35	pc50
8	2024-06-30	50.81	pc50
9	2024-07-31	50.45	pc50
10	2024-08-31	50.38	pc50
11	2024-04-30	68.33	pc90
12	2024-05-31	54.35	pc90
13	2024-06-30	50.75	pc90
14	2024-07-31	50.03	pc90
15	2024-08-31	49.48	pc90

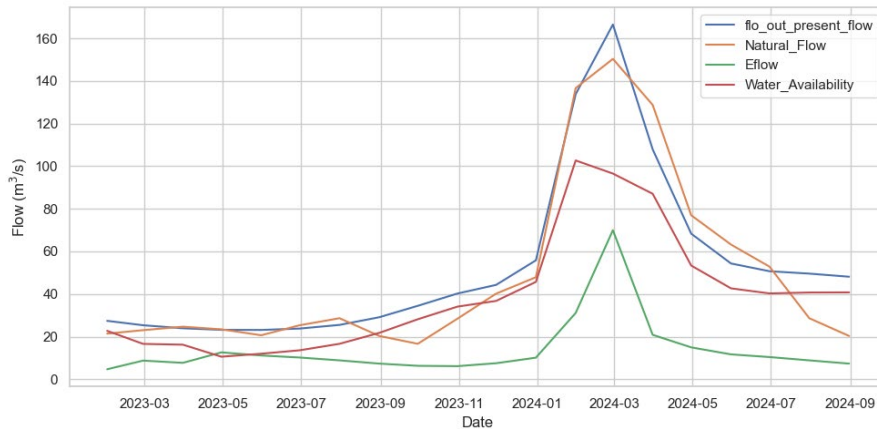


Water availability timeseries (control seasonal forecast)

Extract lastly updated **water availability timeseries** (control seasonal forecast) from the **water_availability** table for flow station with RiverCode LIMP-A71L-MAPUN:

```
SELECT x.date, x.flo_out_present_flow, x.Natural_Flow, x.Eflow, x.Water_Availability
FROM eflow_iwmi.water_availability x
WHERE x.river_code = 'LIMP-A71L-MAPUN'
```

	date	123 flo_out_present_flow	123 Natural_Flow	123 Eflow	123 Water_Availability
1	2001-01-31	0.00008774	7.06	4.685	-4.68491226
2	2001-02-28	0.009142	4.474	4.474	-4.464858
3	2001-03-31	0.1282	2.414	2.406	-2.2778
4	2001-04-30	0.2789	1.26	1.182	-0.9031
5	2001-05-31	0.2943	0.624	0.532	-0.2377
6	2001-06-30	0.2974	0.887	0.887	-0.5896
7	2001-07-31	0.2808	0.805	0.798	-0.5172
8	2001-08-31	0.2429	0.584	0.584	-0.3411
9	2001-09-30	0.2297	0.238	0.233	-0.0033
10	2001-10-31	0.1991	0.172	0.141	0.0581
11	2001-11-30	0.2833	0.379	0.302	-0.0187
12	2001-12-31	0.3019	1.529	1.391	-1.0891
13	2002-01-31	0.2859	7.06	4.685	-4.3991
14	2002-02-28	0.3000	4.474	4.474	-4.1741

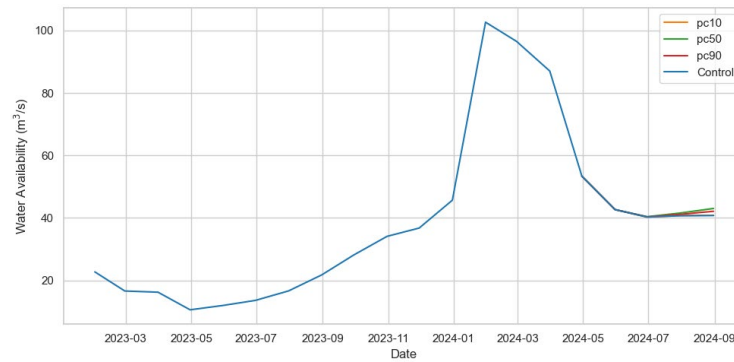


Water availability timeseries (p10, pc50, pc90 seasonal forecast)

Extract lastly updated **water availability timeseries** (p10, pc50, pc90 seasonal forecast) from the **water_availability_2** table for flow station with RiverCode LIMP-A71L-MAPUN:

```
SELECT x.date, x.flo_out_present_flow, x.Natural_Flow, x.Eflow, x.Water_Availability,
x.RunLabel
FROM eflow_iwmi.water_availability_2 x
WHERE x.river_code = 'LIMP-A71L-MAPUN'
```

	date	123 flo_out_present_flow	123 Natural_Flow	123 Eflow	123 Water_Availability	pc RunLabel
1	2024-04-30	68.33	76.968	14.963	53.367	pc10
2	2024-05-31	54.34	63.217	11.676	42.664	pc10
3	2024-06-30	50.7	52.855	10.412	40.288	pc10
4	2024-07-31	49.6	28.644	8.867	40.733	pc10
5	2024-08-31	48.19	20.318	7.344	40.846	pc10
6	2024-04-30	68.33	76.968	14.963	53.367	pc50
7	2024-05-31	54.35	63.217	11.676	42.674	pc50
8	2024-06-30	50.81	52.855	10.412	40.398	pc50
9	2024-07-31	50.45	28.644	8.867	41.583	pc50
10	2024-08-31	50.38	20.318	7.344	43.036	pc50
11	2024-04-30	68.33	76.968	14.963	53.367	pc90
12	2024-05-31	54.35	63.217	11.676	42.674	pc90
13	2024-06-30	50.75	52.855	10.412	40.338	pc90
14	2024-07-31	50.03	28.644	8.867	41.163	pc90
15	2024-08-31	49.48	20.318	7.344	42.136	pc90



Discharges

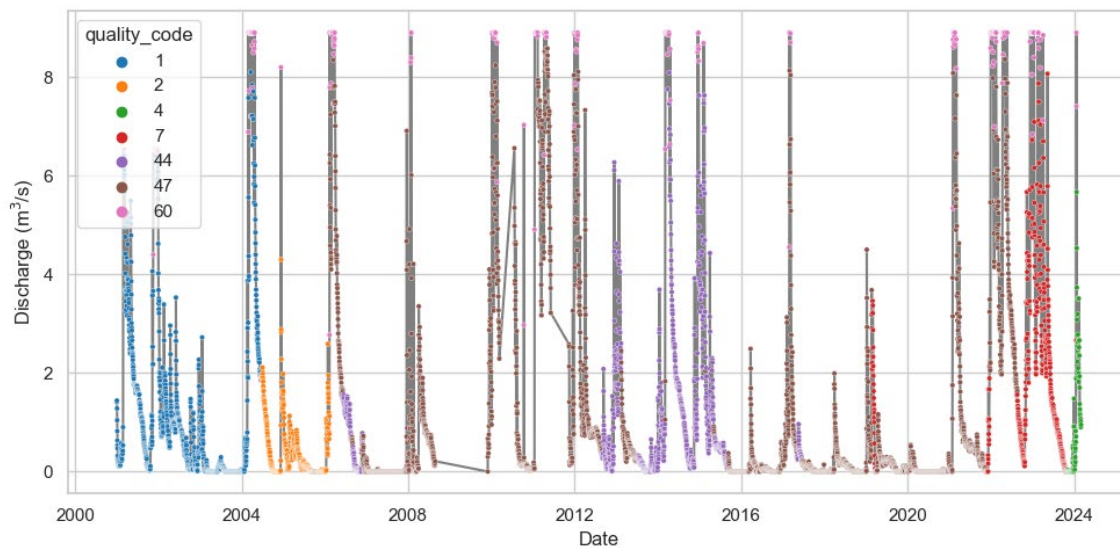
This chapter details SQL examples of the tables to the eflow_iwmi MySQL database, specifically discharge data (INWARD API) for the Limpopo River Basin.

Verified discharge

Query 1: Extract available **discharge station data** from the **verified_discharges** table for a given **station_name**

```
SELECT vs.station_name, vd.date, vd.discharge_m3s, vd.quality_code
FROM eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.station_name = 'A4H002'
```

	adc station_name	date	123 discharge_m3s	124 quality_code
1	A4H002	2001-01-01	1.41	1
2	A4H002	2001-01-02	1.44	1
3	A4H002	2001-01-03	1.28	1
4	A4H002	2001-01-04	1.04	1
5	A4H002	2001-01-05	0.82	1
6	A4H002	2001-01-06	0.63	1
7	A4H002	2001-01-07	0.53	1
8	A4H002	2001-01-08	0.47	1
9	A4H002	2001-01-09	0.32	1
10	A4H002	2001-01-10	0.24	1
11	A4H002	2001-01-11	0.2	1
12	A4H002	2001-01-12	0.19	1
13	A4H002	2001-01-13	0.2	1
14	A4H002	2001-01-14	0.21	1
15	A4H002	2001-01-15	0.21	1
16	A4H002	2001-01-16	0.2	1
17	A4H002	2001-01-17	0.18	1
18	A4H002	2001-01-18	0.16	1
19	A4H002	2001-01-19	0.15	1
20	A4H002	2001-01-20	0.11	1



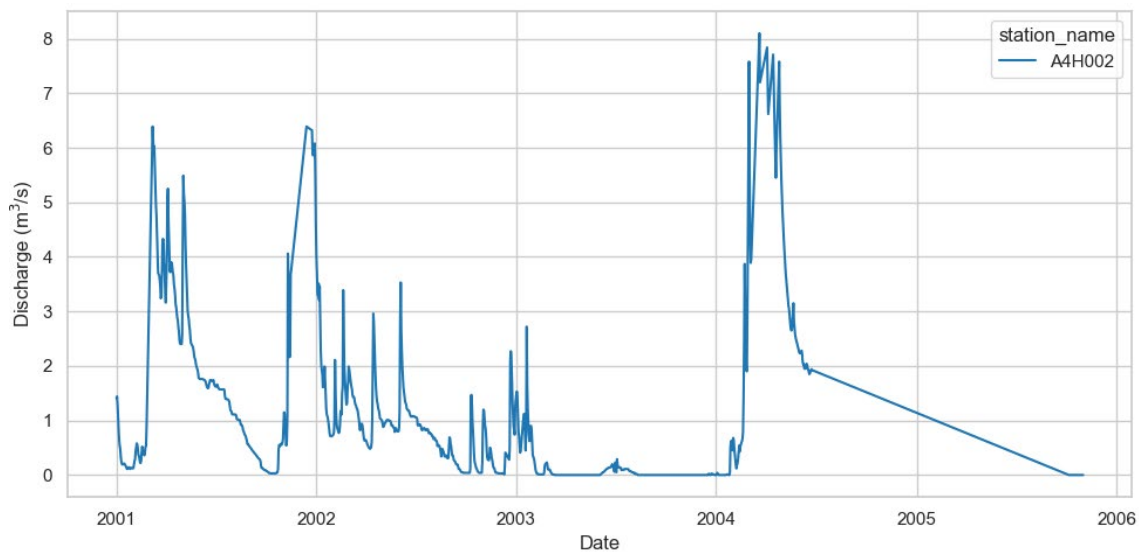
Quality codes: 1 – Good continuous data, 2 - Good edited data, 4 – Unaudited, 7 - Good edited unaudited, 44 - Checked - still unaudited, 47 - Edited and checked - still unaudited, 60 - Above Rating (discharge_quality_codes table)

Verified discharge with quality_code=1

Query 2: Extract available discharge station data from the verified_discharges table for a given station_name and quality_code

```
SELECT vs.station_name, vd.date, vd.discharge_m3s
FROM eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.station_name = 'A4H002' and vd.quality_code = 1
```

	RPC station_name	date	123 discharge_m3s
1	A4H002	2001-01-01	1.41
2	A4H002	2001-01-02	1.44
3	A4H002	2001-01-03	1.28
4	A4H002	2001-01-04	1.04
5	A4H002	2001-01-05	0.82
6	A4H002	2001-01-06	0.63
7	A4H002	2001-01-07	0.53
8	A4H002	2001-01-08	0.47
9	A4H002	2001-01-09	0.32
10	A4H002	2001-01-10	0.24
11	A4H002	2001-01-11	0.2
12	A4H002	2001-01-12	0.19
13	A4H002	2001-01-13	0.2
14	A4H002	2001-01-14	0.21
15	A4H002	2001-01-15	0.21
16	A4H002	2001-01-16	0.2
17	A4H002	2001-01-17	0.18
18	A4H002	2001-01-18	0.16
19	A4H002	2001-01-19	0.15
20	A4H002	2001-01-20	0.11
21	A4H002	2001-01-21	0.11
22	A4H002	2001-01-22	0.13

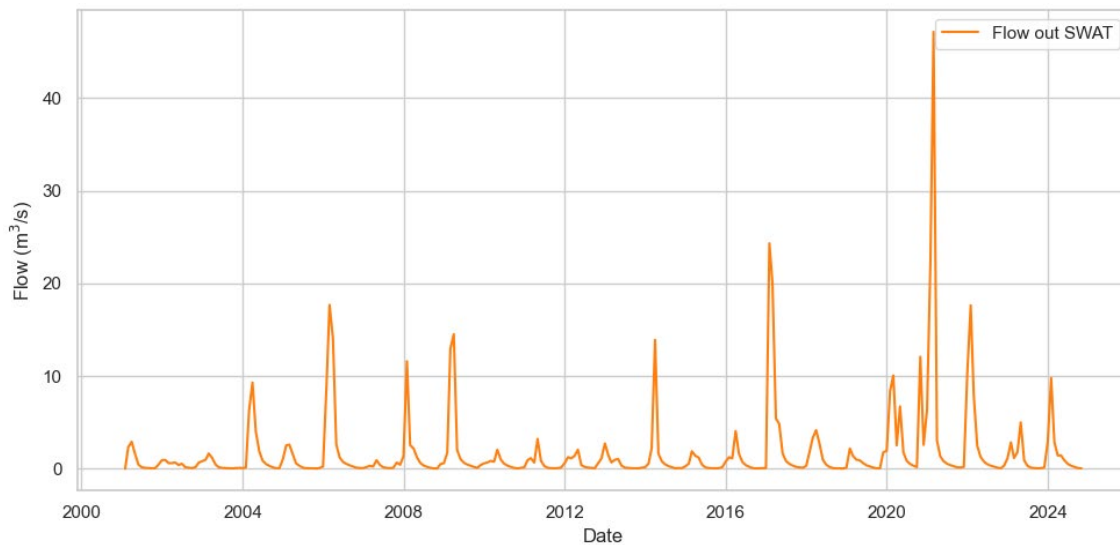


SWAT flow per gage station name

Query 3: Extract SWAT channel flow data from the `swat_output` table for a given `station_name`

```
SELECT vs.swat_channel, so.date, so.flo_out, so.flo_in
FROM eflow_iwmi.swat_output so
inner join verified_stations vs ON vs.swat_channel = so.name
where vs.station_name = 'A4H002'
```

	swat_channel	date	123 flo_out	123 flo_in
1	cha1034	2001-01-31	0.00000664	0.0006167
2	cha1034	2001-02-28	2.327	3.663
3	cha1034	2001-03-31	2.909	3.025
4	cha1034	2001-04-30	1.622	0.9664
5	cha1034	2001-05-31	0.4507	0.1853
6	cha1034	2001-06-30	0.1546	0.03647
7	cha1034	2001-07-31	0.09246	0.02306
8	cha1034	2001-08-31	0.07134	0.01945
9	cha1034	2001-09-30	0.05011	0.01659
10	cha1034	2001-10-31	0.06823	0.06712
11	cha1034	2001-11-30	0.456	0.7722
12	cha1034	2001-12-31	0.9353	1.16
13	cha1034	2002-01-31	0.9404	0.771
14	cha1034	2002-02-28	0.6024	0.5301
15	cha1034	2002-03-31	0.5984	0.551
16	cha1034	2002-04-30	0.6636	0.807
17	cha1034	2002-05-31	0.3965	0.2419
18	cha1034	2002-06-30	0.5346	0.5232
19	cha1034	2002-07-31	0.156	0.05047
20	cha1034	2002-08-31	0.1035	0.03817
21	cha1034	2002-09-30	0.07143	0.02929
22	cha1034	2002-10-31	0.1657	0.2846
23	cha1034	2002-11-30	0.6619	0.8017
24	cha1034	2002-12-31	0.8112	1.079
25	cha1034	2003-01-31	0.9669	0.8308

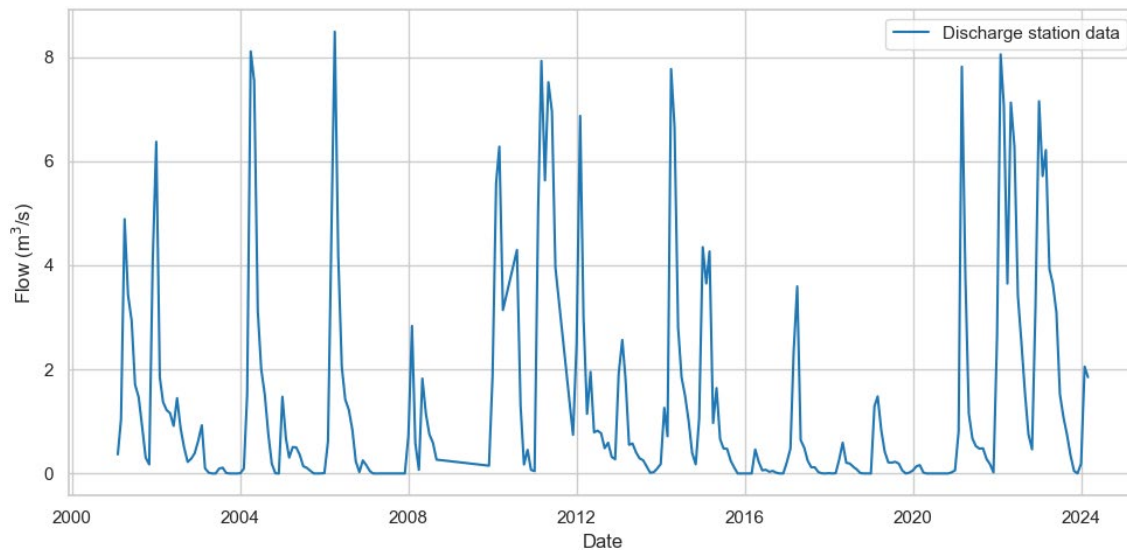


Verified discharge (dates with values)

Query 4: Extract monthly aggregated **discharge station data** from the **verified_discharges** table for a given **station_name** and excluding data with quality codes 255 and 151 (missing data)

```
SELECT vs.swat_channel, vs.station_name, AVG(vd.discharge_m3s) as "avg_discharge_m3s", last_day(vd.date)
as "date"
FROM eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.station_name = 'A4H002' and vd.quality_code not in (255, 151)
group by vs.swat_channel, vs.station_name, last_day(vd.date)
```

	swat_channel	noc station_name	avg_discharge_m3s	date
1	cha1034	A4H002	0.366774194	2001-01-31
2	cha1034	A4H002	1.04	2001-02-28
3	cha1034	A4H002	4.886451613	2001-03-31
4	cha1034	A4H002	3.429333333	2001-04-30
5	cha1034	A4H002	2.939354839	2001-05-31
6	cha1034	A4H002	1.709	2001-06-30
7	cha1034	A4H002	1.466451613	2001-07-31
8	cha1034	A4H002	0.891612903	2001-08-31
9	cha1034	A4H002	0.306666667	2001-09-30
10	cha1034	A4H002	0.173870968	2001-10-31
11	cha1034	A4H002	4.057666667	2001-11-30
12	cha1034	A4H002	6.372580645	2001-12-31
13	cha1034	A4H002	1.84516129	2002-01-31
14	cha1034	A4H002	1.375357143	2002-02-28
15	cha1034	A4H002	1.218709677	2002-03-31
16	cha1034	A4H002	1.159333333	2002-04-30
17	cha1034	A4H002	0.914516129	2002-05-31
18	cha1034	A4H002	1.447666667	2002-06-30
19	cha1034	A4H002	0.862903226	2002-07-31
20	cha1034	A4H002	0.49	2002-08-31
21	cha1034	A4H002	0.222666667	2002-09-30
22	cha1034	A4H002	0.28516129	2002-10-31
23	cha1034	A4H002	0.388666667	2002-11-30



Combining verified station data with SWAT flow

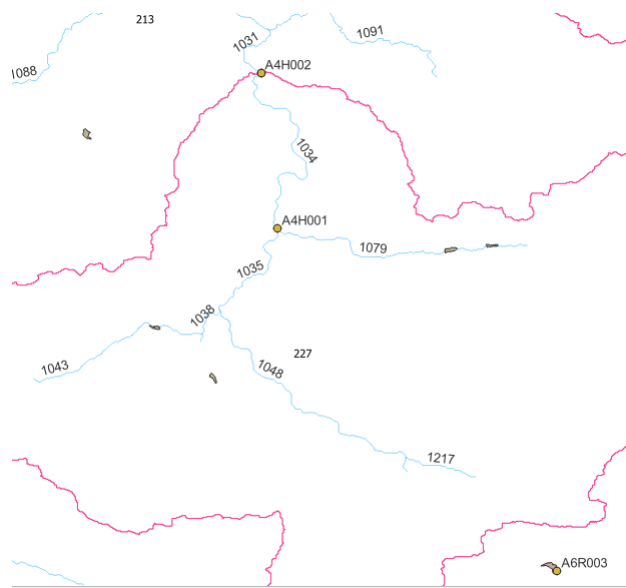
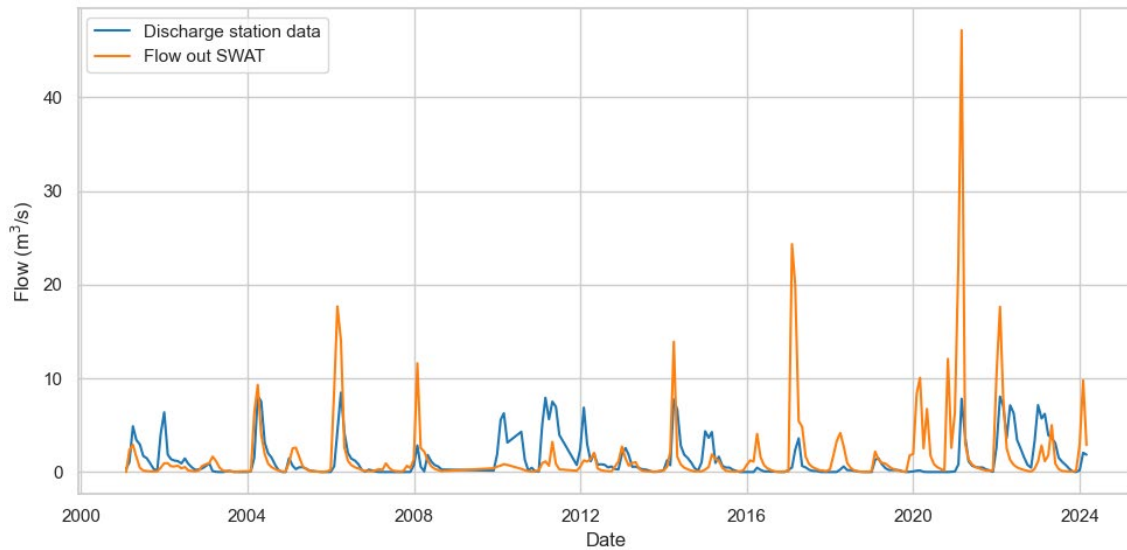
Query 4: Extract monthly aggregated discharge station data from the verified_discharges table and swat flow data from the swat_output table for a given station_name

```
SELECT discharge_data.swat_channel, discharge_data.station_name, discharge_data.avg_discharge_m3s,
discharge_data.date, swat_data.flo_out, swat_data.flo_in
FROM
(
  -- Subquery for aggregated discharge observations
  SELECT vs.swat_channel, vs.station_name, AVG(vd.discharge_m3s) AS avg_discharge_m3s,
LAST_DAY(vd.date) AS date
FROM eflow_iwmi.verified_discharges vd
INNER JOIN verified_stations vs ON vs.id = vd.station_id
WHERE vs.station_name = 'A4H002' AND vd.quality_code NOT IN (255, 151)
GROUP BY vs.swat_channel, vs.station_name, LAST_DAY(vd.date)
```

```

) AS discharge_data
RIGHT JOIN
(
-- Subquery for SWAT output data
SELECT so.date, so.flo_out, so.flo_in
FROM eflow_iwmi.swat_output so
INNER JOIN verified_stations vs ON vs.swat_channel = so.name
WHERE vs.station_name = 'A4H002'
) AS swat_data ON swat_data.date = discharge_data.date
    
```

	swat_channel	station_name	avg_discharge_m3s	date	flo_out	flo_in
12	cha1034	A4H002	6.372580645	2001-12-31	0.9353	1.16
13	cha1034	A4H002	1.84516129	2002-01-31	0.9404	0.771
14	cha1034	A4H002	1.375357143	2002-02-28	0.6024	0.5301
15	cha1034	A4H002	1.218709677	2002-03-31	0.5984	0.551
16	cha1034	A4H002	1.159333333	2002-04-30	0.6636	0.807
17	cha1034	A4H002	0.914516129	2002-05-31	0.3965	0.2419
18	cha1034	A4H002	1.447666667	2002-06-30	0.5346	0.5232
19	cha1034	A4H002	0.862903226	2002-07-31	0.156	0.05047
20	cha1034	A4H002	0.49	2002-08-31	0.1035	0.03817
21	cha1034	A4H002	0.222666667	2002-09-30	0.07143	0.02929
22	cha1034	A4H002	0.28516129	2002-10-31	0.1657	0.2846
23	cha1034	A4H002	0.388666667	2002-11-30	0.6619	0.8017
24	cha1034	A4H002	0.629354839	2002-12-31	0.8112	1.079
25	cha1034	A4H002	0.926129032	2003-01-31	0.9669	0.8308
26	cha1034	A4H002	0.099642857	2003-02-28	1.642	2.205
27	cha1034	A4H002	0.017096774	2003-03-31	1.165	0.7048
28	cha1034	A4H002	0	2003-04-30	0.4651	0.2475
29	cha1034	A4H002	0	2003-05-31	0.145	0.04257
30	cha1034	A4H002	0.093	2003-06-30	0.09033	0.02233
31	cha1034	A4H002	0.112258065	2003-07-31	0.06664	0.02234
32	cha1034	A4H002	0.01	2003-08-31	0.05034	0.01796
33	cha1034	A4H002	0	2003-09-30	0.03724	0.01467
34	cha1034	A4H002	0	2003-10-31	0.06957	0.08186
35	cha1034	A4H002	0	2003-11-30	0.08263	0.08457

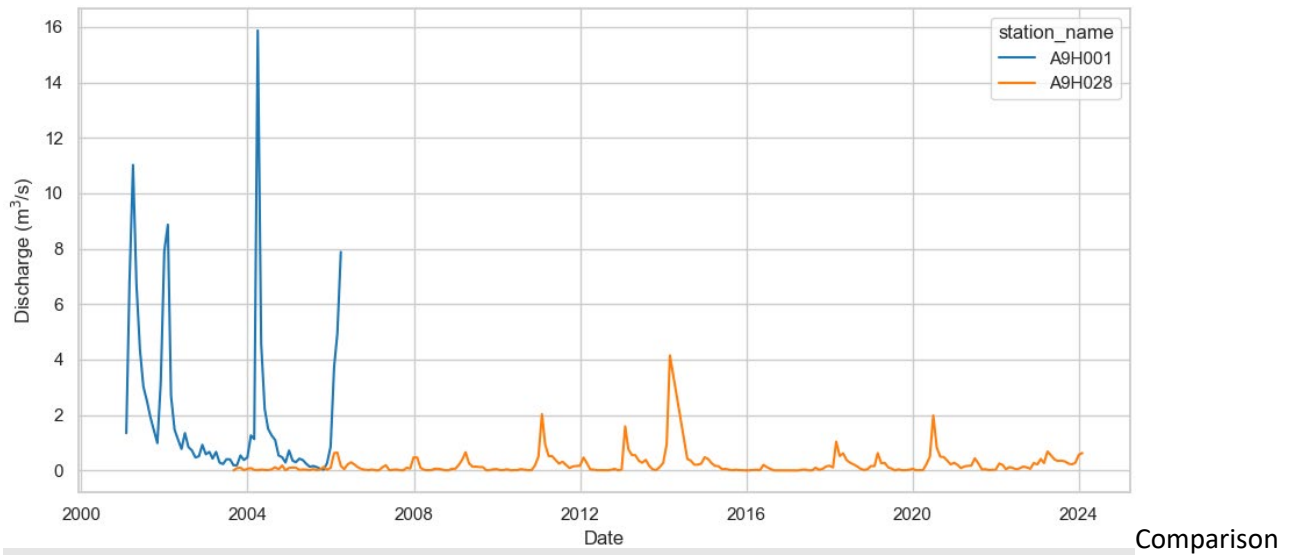


All verified discharges linked to a SWAT channel (dates with values)

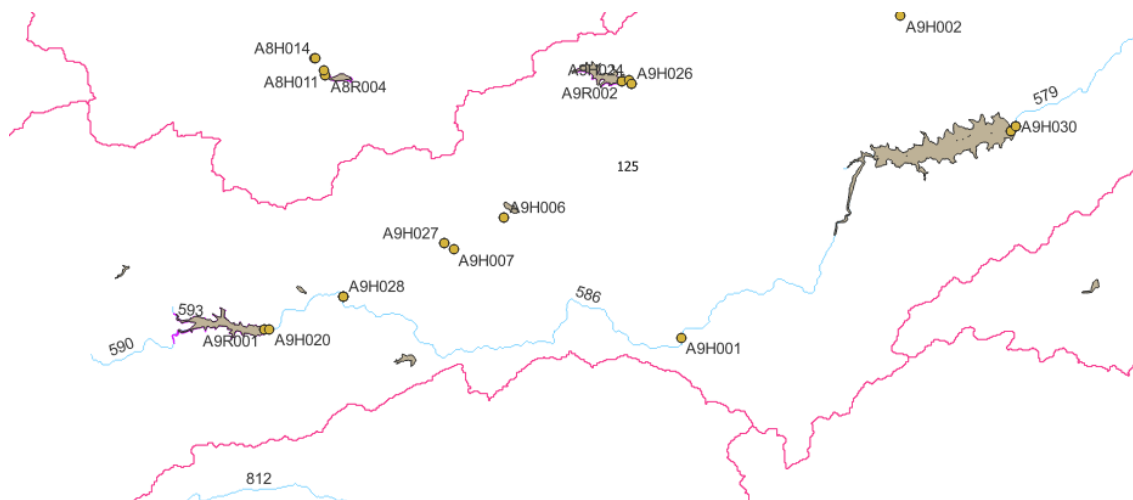
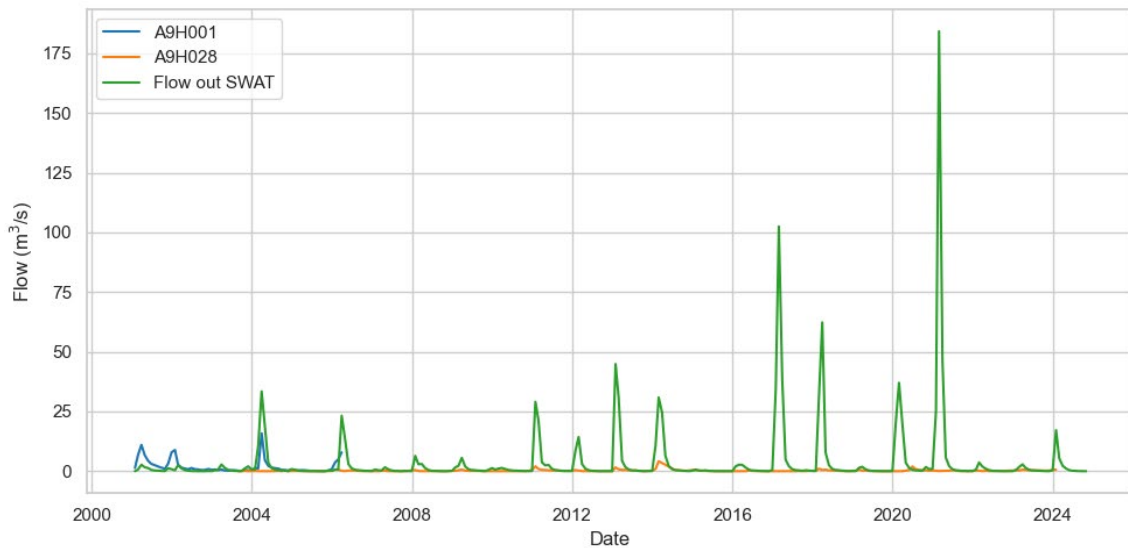
Query 5: Extract monthly aggregated **discharge station data** from the **verified_discharges** table for a given **swat_channel** and excluding data with quality codes 255 and 151

```
SELECT vs.station_name, AVG(vd.discharge_m3s) as "avg_discharge_m3s", last_day(vd.date) as "date"
FROM eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.swat_channel = 'cha0586' and vd.quality_code not in (255, 151)
group by vs.station_name, last_day(vd.date)
```

	RPC station_name	123 avg_discharge_m3s	📅 date
1	A9H001	1.346129032	2001-01-31
2	A9H001	6.768928571	2001-02-28
3	A9H001	11.022590645	2001-03-31
4	A9H001	6.732	2001-04-30
5	A9H001	4.35483871	2001-05-31
6	A9H001	3.028	2001-06-30
7	A9H001	2.51483871	2001-07-31
8	A9H001	1.919354839	2001-08-31
9	A9H001	1.457666667	2001-09-30
10	A9H001	0.987741935	2001-10-31
11	A9H001	3.2	2001-11-30
12	A9H001	7.911612903	2001-12-31
13	A9H001	8.867741935	2002-01-31
14	A9H001	2.698928571	2002-02-28
15	A9H001	1.474193548	2002-03-31
16	A9H001	1.121666667	2002-04-30
17	A9H001	0.776774194	2002-05-31
18	A9H001	1.349	2002-06-30
19	A9H001	0.844516129	2002-07-31
20	A9H001	0.720322581	2002-08-31
21	A9H001	0.469666667	2002-09-30
22	A9H001	0.512590645	2002-10-31



between the results of **Query 5** and SWAT flow out data for the same channel:

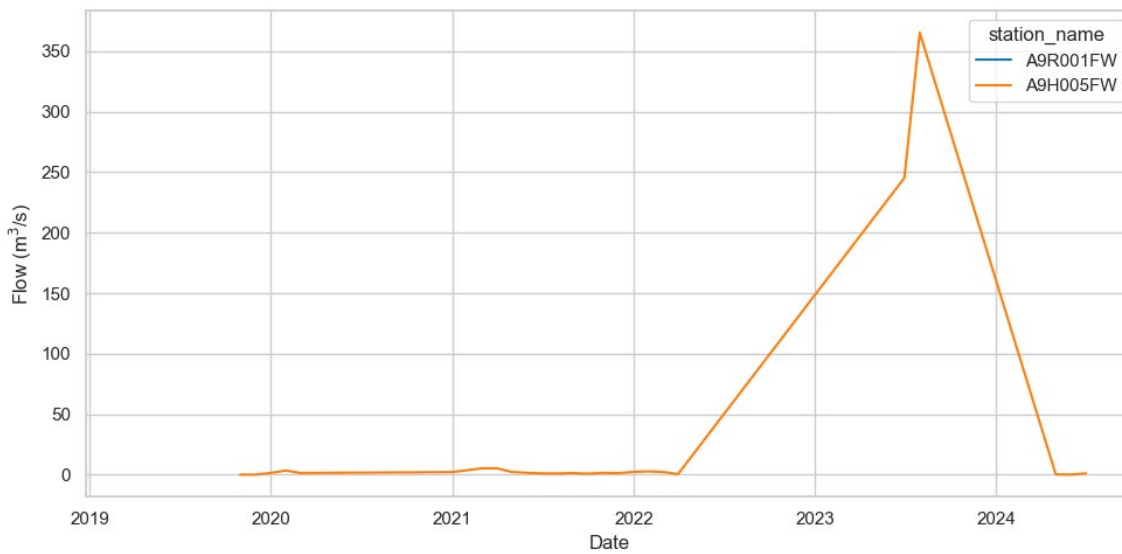


All unverified discharges linked to a SWAT channel

Query 6: Extract monthly aggregated **discharge station data** from the **unverified_discharges** table for a given **swat_channel**

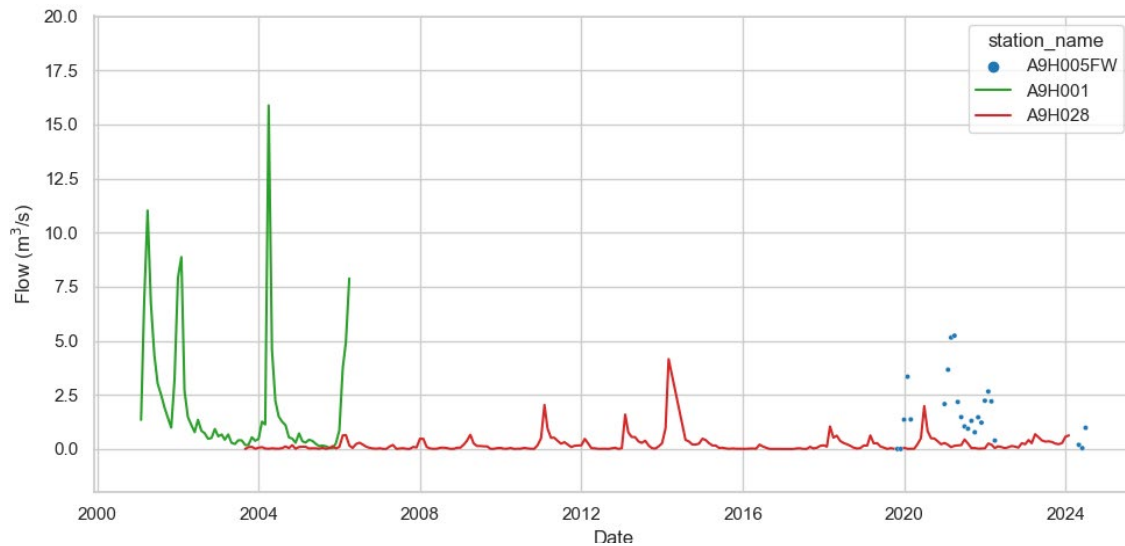
```
SELECT us.station_name, AVG(ud.discharge_m3s) as "avg_discharge_m3s", last_day(ud.date) as "date"
FROM eflow_iwmi.unverified_discharges ud
inner join unverified_stations us ON us.id = ud.station_id
where us.swat_channel = 'cha0586'
group by us.station_name, last_day(ud.date)
```

	ABC station_name	123 avg_discharge_m3s	date
1	A9R001FW	76.7843172367	2019-03-31
2	A9H005FW	0	2019-10-31
3	A9H005FW	0.0001232666	2019-11-30
4	A9H005FW	1.3559644367	2019-12-31
5	A9H005FW	3.3397342047	2020-01-31
6	A9H005FW	1.3684602564	2020-02-29
7	A9H005FW	2.0789368056	2020-12-31
8	A9H005FW	3.6648137711	2021-01-31
9	A9H005FW	5.1490152886	2021-02-28
10	A9H005FW	5.2368516129	2021-03-31
11	A9H005FW	2.1758876282	2021-04-30
12	A9H005FW	1.4766349462	2021-05-31
13	A9H005FW	1.0431995375	2021-06-30
14	A9H005FW	0.9406499074	2021-07-31
15	A9H005FW	1.3007456409	2021-08-31
16	A9H005FW	0.7738875	2021-09-30
17	A9H005FW	1.4641410372	2021-10-31
18	A9H005FW	1.2185590383	2021-11-30
19	A9H005FW	2.2313134783	2021-12-31
20	A9H005FW	2.6582749114	2022-01-31



Comparison of

the results in Queries 5 and 6:

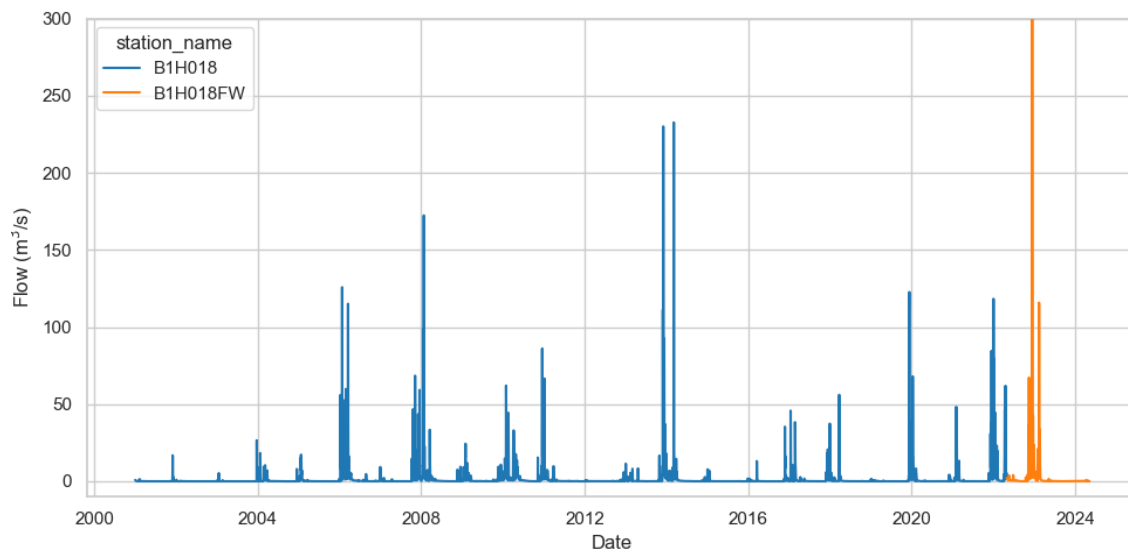


Combining verified and unverified station discharge data

Query 7: Extract **discharge station data** from the **verified_discharges** and **unverified_discharges** tables for a given verified station_name and corresponding unverified station name. Verified data is extracted until the last available date and unverified data is extracted for dates beyond that date.

```
(
-- Select all records from the verified table
SELECT vs.swat_channel, vs.station_name ,vd.discharge_m3s, vd.date
FROM eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.station_name = 'B1H018' and vd.quality_code not in (255, 151)
)
UNION ALL
(
-- Select records from the unverified table that are later than the last date in the verified table
SELECT us.swat_channel, us.station_name ,ud.discharge_m3s, ud.date
FROM eflow_iwmi.unverified_discharges ud
inner join unverified_stations us ON us.id = ud.station_id
inner join verified_stations vs ON us.verified_station = vs.id
WHERE vs.station_name = 'B1H018' and ud.date > (SELECT MAX(vd.date) FROM
eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.station_name = 'B1H018'and vd.quality_code not in (255, 151))
)
ORDER BY date
```

	ABC swat_channel	ABC station_name	123 discharge_m3s	date
1	cha1409	B1H018	0.86	2001-01-01
2	cha1409	B1H018	0.76	2001-01-02
3	cha1409	B1H018	0.66	2001-01-03
4	cha1409	B1H018	0.51	2001-01-04
5	cha1409	B1H018	0.39	2001-01-05
6	cha1409	B1H018	0.31	2001-01-06
7	cha1409	B1H018	0.25	2001-01-07
8	cha1409	B1H018	0.2	2001-01-08
9	cha1409	B1H018	0.14	2001-01-09
10	cha1409	B1H018	0.12	2001-01-10
11	cha1409	B1H018	0.1	2001-01-11
12	cha1409	B1H018	0.07	2001-01-12
13	cha1409	B1H018	0.06	2001-01-13
14	cha1409	B1H018	0.05	2001-01-14
15	cha1409	B1H018	0.05	2001-01-15
16	cha1409	B1H018	0.07	2001-01-16
17	cha1409	B1H018	0.06	2001-01-17
18	cha1409	B1H018	0.05	2001-01-18



Aggregating verified and unverified station discharge data

Query 8: Same as Query 7 but with aggregation of data by **month** (to compare with swat model data)

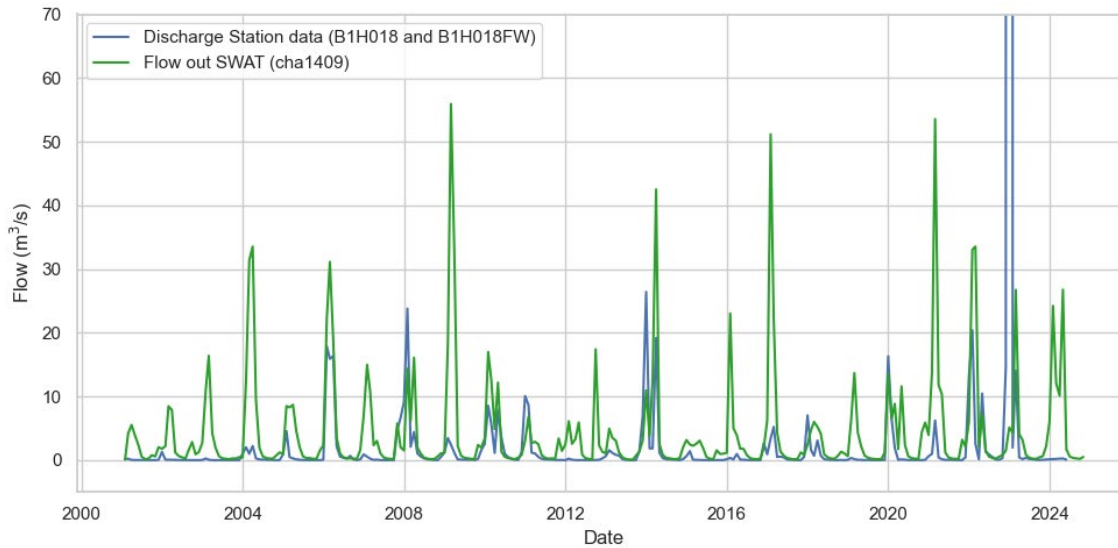
```

SELECT LAST_DAY(date) AS last_date, AVG(discharge_m3s) AS avg_discharge_m3s
FROM
(
(
-- Select all records from the verified table
SELECT vs.station_name ,vd.discharge_m3s, vd.date
FROM eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.station_name = 'B1H018' and vd.quality_code not in (255, 151)
)
)
UNION ALL

```

```
(
-- Select records from the unverified table that are later than the last date in the verified table
SELECT us.station_name ,ud.discharge_m3s, ud.date
FROM eflow_iwmi.unverified_discharges ud
inner join unverified_stations us ON us.id = ud.station_id
inner join verified_stations vs ON us.verified_station = vs.id
WHERE vs.station_name = 'B1H018' and ud.date > (SELECT MAX(vd.date) FROM
eflow_iwmi.verified_discharges vd
inner join verified_stations vs ON vs.id = vd.station_id
where vs.station_name = 'B1H018'and vd.quality_code not in (255, 151))
)
)
AS combined_data
GROUP BY last_date
```

	last_date	123 avg_discharge_m3s
1	2001-01-31	0.1680645161
2	2001-02-28	0.2146428571
3	2001-03-31	0.0358064516
4	2001-04-30	0.0166666667
5	2001-05-31	0.0206451613
6	2001-06-30	0.0176666667
7	2001-07-31	0.0151612903
8	2001-08-31	0.0032258065
9	2001-09-30	0
10	2001-10-31	0
11	2001-11-30	0.0383333333
12	2001-12-31	1.2990322581
13	2002-01-31	0.0719354839
14	2002-02-28	0.0575
15	2002-03-31	0.0441935484
16	2002-04-30	0.015
17	2002-05-31	0.0048387097
18	2002-06-30	0.01
19	2002-07-31	0.0019354839
20	2002-08-31	0
21	2002-09-30	0.0016666667
22	2002-10-31	0
23	2002-11-30	0
24	2002-12-31	0.0216129032
25	2003-01-31	0.2606451613



Reservoirs

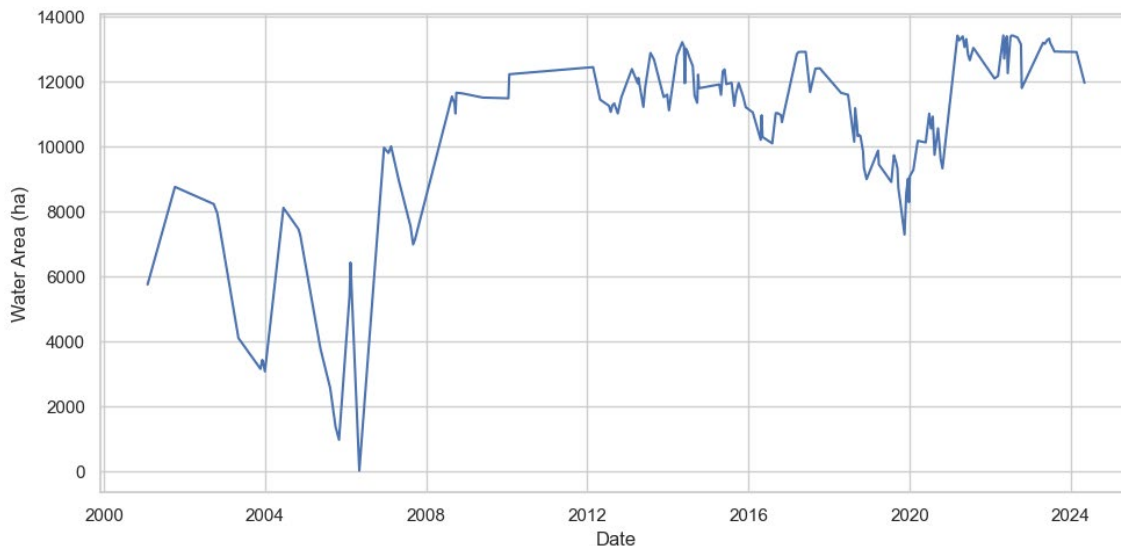
This chapter details SQL examples of the tables to the eflow_iwmi MySQL database, specifically reservoir information (Digital Earth Africa DEA) for the Limpopo River Basin.

Water area time series

Query 1: Extract **water area time series** from the **waterbodies_data** table given a reservoir name and filtering by percent_invalid

```
SELECT r.name, r.waterbody_id, wbd.date, wbd.water_area_ha
FROM eflow_iwmi.waterbodies_data wbd
inner join waterbodies w on w.id = wbd.waterbody_id
inner join reservoirs r on wbd.waterbody_id = r.waterbody_id
where r.name = 'Massingir Dam' and wbd.percent_invalid < 10
```

	abc name	122 waterbody_id	date	123 water_area_ha
1	Massingir Dam	1,088	2001-01-30	5,742.81
2	Massingir Dam	1,088	2001-10-05	8,745.12
3	Massingir Dam	1,088	2002-09-22	8,216.82
4	Massingir Dam	1,088	2002-10-24	7,934.58
5	Massingir Dam	1,088	2003-05-04	4,095
6	Massingir Dam	1,088	2003-11-20	3,147.39
7	Massingir Dam	1,088	2003-12-06	3,423.78
8	Massingir Dam	1,088	2003-12-14	3,349.98
9	Massingir Dam	1,088	2003-12-30	3,069.63
10	Massingir Dam	1,088	2004-06-15	8,102.25
11	Massingir Dam	1,088	2004-10-29	7,446.69
12	Massingir Dam	1,088	2004-11-14	7,269.93
13	Massingir Dam	1,088	2005-05-17	3,762.99
14	Massingir Dam	1,088	2005-08-13	2,555.55
15	Massingir Dam	1,088	2005-09-30	1,378.98
16	Massingir Dam	1,088	2005-11-01	967.59
17	Massingir Dam	1,088	2006-02-05	5,421.78
18	Massingir Dam	1,088	2006-02-13	6,417.36
19	Massingir Dam	1,088	2006-05-04	17.55
20	Massingir Dam	1,088	2006-12-14	9,952.11
21	Massingir Dam	1,088	2007-01-23	9,788.76
22	Massingir Dam	1,088	2007-02-16	9,992.07
23	Massingir Dam	1,088	2007-04-29	8,917.83
24	Massingir Dam	1,088	2007-08-11	7,538.04
25	Massingir Dam	1,088	2007-09-04	6,975.18
26	Massingir Dam	1,088	2007-09-20	7,116.48
27	Massingir Dam	1,088	2008-08-21	11,526.12

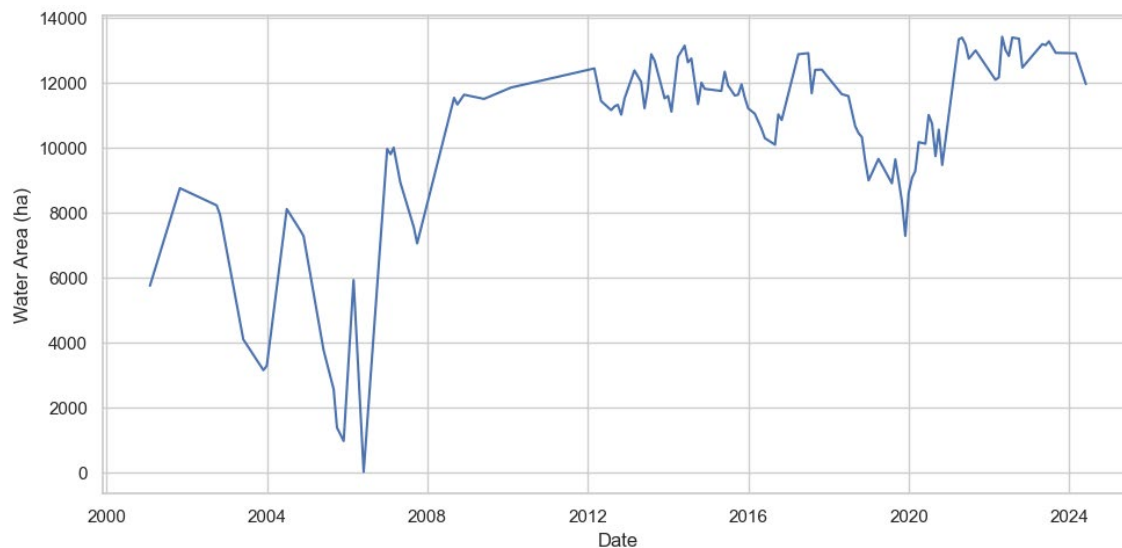


Aggregated water area time series

Query 2: Extract monthly aggregated **water area time series** from the **waterbodies_data** table given a reservoir name

```
SELECT r.name, r.waterbody_id, last_day(wbd.date) as date, AVG(wbd.water_area_ha ) as avg_water_area_ha
FROM eflow_iwmi.waterbodies_data wbd
inner join waterbodies w on w.id = wbd.waterbody_id
inner join reservoirs r on wbd.waterbody_id = r.waterbody_id
where r.name = 'Massingir Dam' and wbd.percent_invalid < 10
group by r.name, r.waterbody_id, last_day(wbd.date)
```

	ABC name	waterbody_id	date	avg_water_area_ha
1	Massingir Dam	1,088	2001-01-31	5,742.81
2	Massingir Dam	1,088	2001-10-31	8,745.12
3	Massingir Dam	1,088	2002-09-30	8,216.82
4	Massingir Dam	1,088	2002-10-31	7,934.58
5	Massingir Dam	1,088	2003-05-31	4,095
6	Massingir Dam	1,088	2003-11-30	3,147.39
7	Massingir Dam	1,088	2003-12-31	3,281.13
8	Massingir Dam	1,088	2004-06-30	8,102.25
9	Massingir Dam	1,088	2004-10-31	7,446.69
10	Massingir Dam	1,088	2004-11-30	7,269.93
11	Massingir Dam	1,088	2005-05-31	3,762.99
12	Massingir Dam	1,088	2005-08-31	2,555.55
13	Massingir Dam	1,088	2005-09-30	1,378.98
14	Massingir Dam	1,088	2005-11-30	967.59
15	Massingir Dam	1,088	2006-02-28	5,919.57
16	Massingir Dam	1,088	2006-05-31	17.55
17	Massingir Dam	1,088	2006-12-31	9,952.11
18	Massingir Dam	1,088	2007-01-31	9,788.76
19	Massingir Dam	1,088	2007-02-28	9,992.07
20	Massingir Dam	1,088	2007-04-30	8,917.83
21	Massingir Dam	1,088	2007-08-31	7,538.04
22	Massingir Dam	1,088	2007-09-30	7,045.83
23	Massingir Dam	1,088	2008-08-31	11,526.12
24	Massingir Dam	1,088	2008-09-30	11,317.83

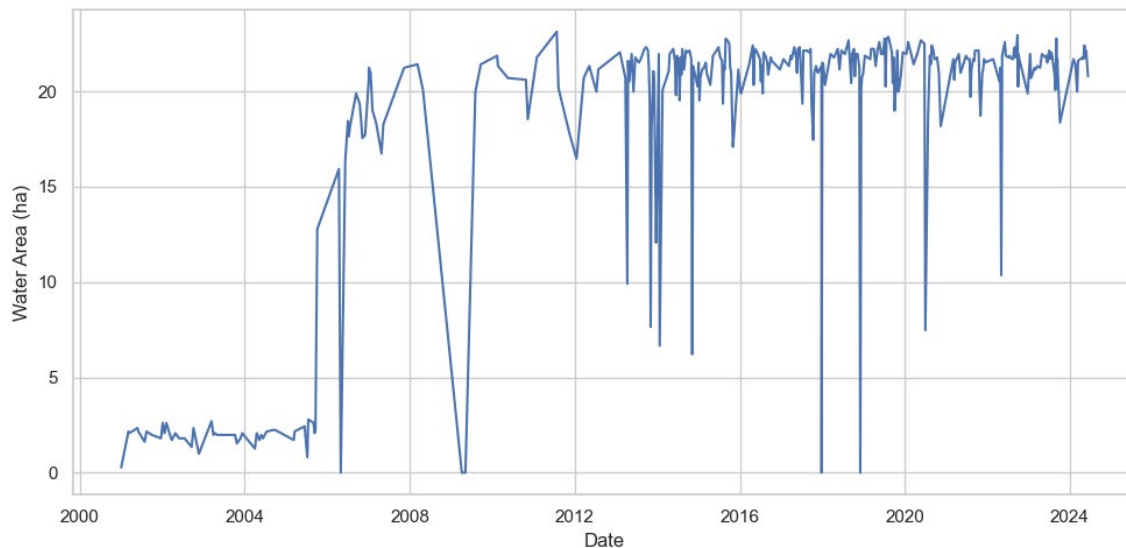


Water area time series given DEA waterbody ID

Query 3: Extract water area time series from the waterbodies_data table given a waterbody original_id

```
SELECT w.original_id, w.id, wbd.date, wbd.water_area_ha FROM eflow_iwmi.waterbodies_data wbd
inner join waterbodies w on w.id = wbd.waterbody_id
where w.original_id = 143369 and wbd.percent_invalid < 10
```

	original_id	id	date	water_area_ha
1	143,369	1	2001-01-02	0.27
2	143,369	1	2001-03-07	2.16
3	143,369	1	2001-03-15	2.07
4	143,369	1	2001-05-26	2.34
5	143,369	1	2001-06-03	2.16
6	143,369	1	2001-07-29	1.62
7	143,369	1	2001-08-14	2.16
8	143,369	1	2001-10-01	1.98
9	143,369	1	2001-12-20	1.8
10	143,369	1	2002-01-05	2.61
11	143,369	1	2002-01-21	2.07
12	143,369	1	2002-02-06	2.61
13	143,369	1	2002-03-10	1.98
14	143,369	1	2002-03-26	1.71
15	143,369	1	2002-04-27	2.07
16	143,369	1	2002-05-29	1.8
17	143,369	1	2002-06-30	1.8
18	143,369	1	2002-07-16	1.8
19	143,369	1	2002-09-18	1.35
20	143,369	1	2002-10-04	2.34
21	143,369	1	2002-11-21	0.99
22	143,369	1	2003-03-13	2.7
23	143,369	1	2003-03-29	1.98
24	143,369	1	2003-04-14	2.07
25	143,369	1	2003-04-30	1.98

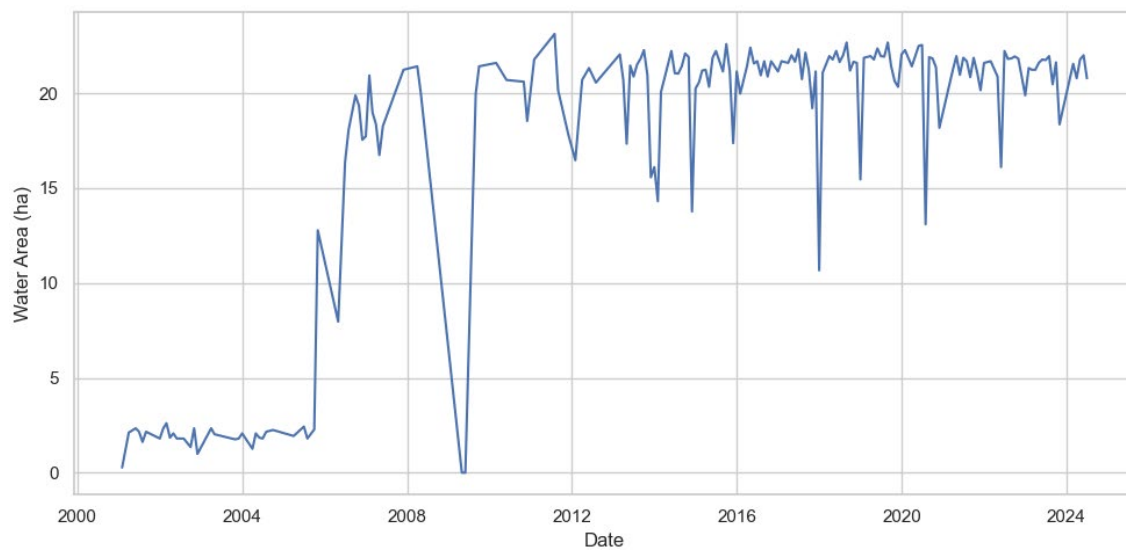


Aggregated water area time series given DEA waterbody ID

Query 4: Extract monthly aggregated **water area time series** from the **waterbodies_data** table given a waterbody **original_id**

```
SELECT w.original_id, w.id, last_day(wbd.date) as date, AVG(wbd.water_area_ha ) as avg_water_area_ha FROM
eflow_iwmi.waterbodies_data wbd
inner join waterbodies w on w.id = wbd.waterbody_id
where w.original_id = 143369 and wbd.percent_invalid < 10
group by w.original_id, w.id, last_day(wbd.date)
```

	123 original_id	123 id	123 date	123 avg_water_area_ha
1	143,369	1	2001-01-31	0.27
2	143,369	1	2001-03-31	2.115
3	143,369	1	2001-05-31	2.34
4	143,369	1	2001-06-30	2.16
5	143,369	1	2001-07-31	1.62
6	143,369	1	2001-08-31	2.16
7	143,369	1	2001-10-31	1.98
8	143,369	1	2001-12-31	1.8
9	143,369	1	2002-01-31	2.34
10	143,369	1	2002-02-28	2.61
11	143,369	1	2002-03-31	1.845
12	143,369	1	2002-04-30	2.07
13	143,369	1	2002-05-31	1.8
14	143,369	1	2002-06-30	1.8
15	143,369	1	2002-07-31	1.8
16	143,369	1	2002-09-30	1.35
17	143,369	1	2002-10-31	2.34
18	143,369	1	2002-11-30	0.99
19	143,369	1	2003-03-31	2.34
20	143,369	1	2003-04-30	2.025
21	143,369	1	2003-10-31	1.755
22	143,369	1	2003-11-30	1.8



Aggregated water area time series by channel

Query 5: Extract monthly aggregated **total water area time series** from the **waterbodies_data** table given a **swat_channel**

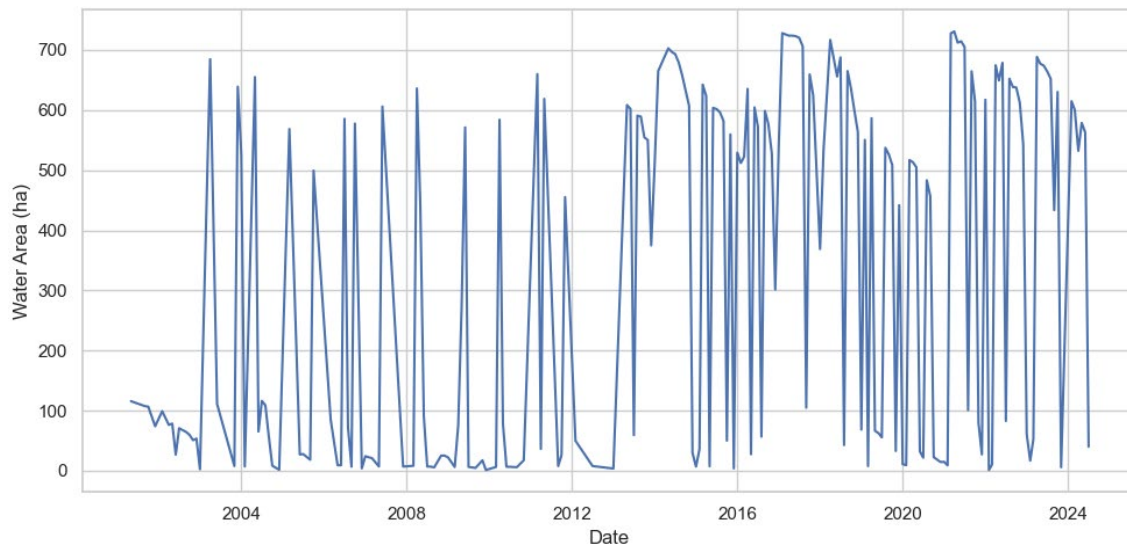
```

SELECT aggregated_month.swat_channel, aggregated_month.last_date as date,
sum(aggreated_month.avg_water_area_ha) as total_water_area_ha
FROM (
SELECT w.swat_channel, w.id, last_day(wbd.date) as last_date, AVG(wbd.water_area_ha ) as
avg_water_area_ha
FROM eflow_iwmi.waterbodies_data wbd
inner join waterbodies w on w.id = wbd.waterbody_id
where wbd.percent_invalid < 10
group by w.swat_channel, w.id, last_day(wbd.date)
) as aggregated_month
where aggregated_month.swat_channel = 'cha0075'
group by aggregated_month.swat_channel, aggregated_month.last_date

```

order by aggregated_month.last_date

	swat_channel	date	total_water_area_ha
1	cha0075	2001-04-30	115.83
2	cha0075	2001-08-31	107.55
3	cha0075	2001-09-30	106.74
4	cha0075	2001-11-30	74.07
5	cha0075	2002-01-31	99.09
6	cha0075	2002-03-31	76.275
7	cha0075	2002-04-30	78.57
8	cha0075	2002-05-31	26.91
9	cha0075	2002-06-30	70.83
10	cha0075	2002-08-31	64.53
11	cha0075	2002-09-30	59.67
12	cha0075	2002-10-31	50.85
13	cha0075	2002-11-30	53.55
14	cha0075	2002-12-31	2.61
15	cha0075	2003-03-31	684.72
16	cha0075	2003-05-31	111.24
17	cha0075	2003-10-31	7.92
18	cha0075	2003-11-30	639.03
19	cha0075	2003-12-31	526.095
20	cha0075	2004-01-31	7.11
21	cha0075	2004-04-30	655.02
22	cha0075	2004-05-31	65.34
23	cha0075	2004-06-30	116.37



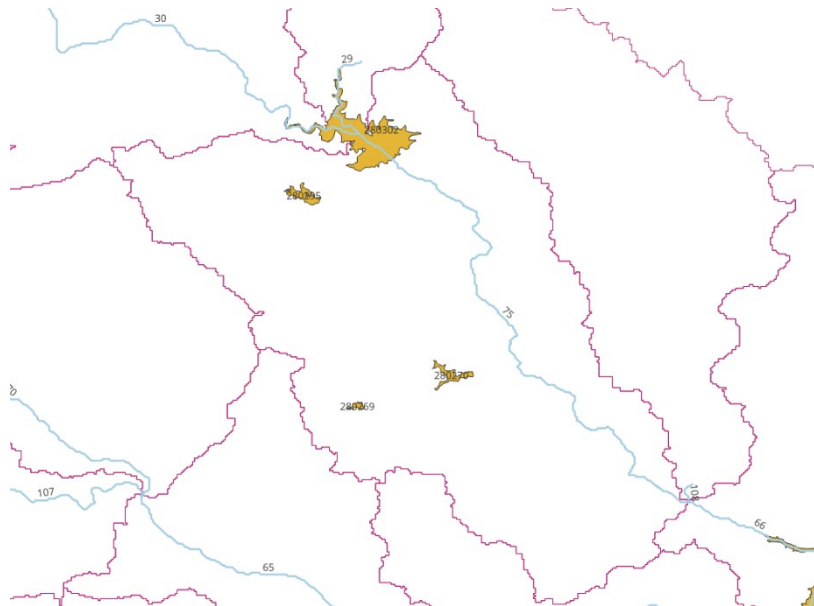
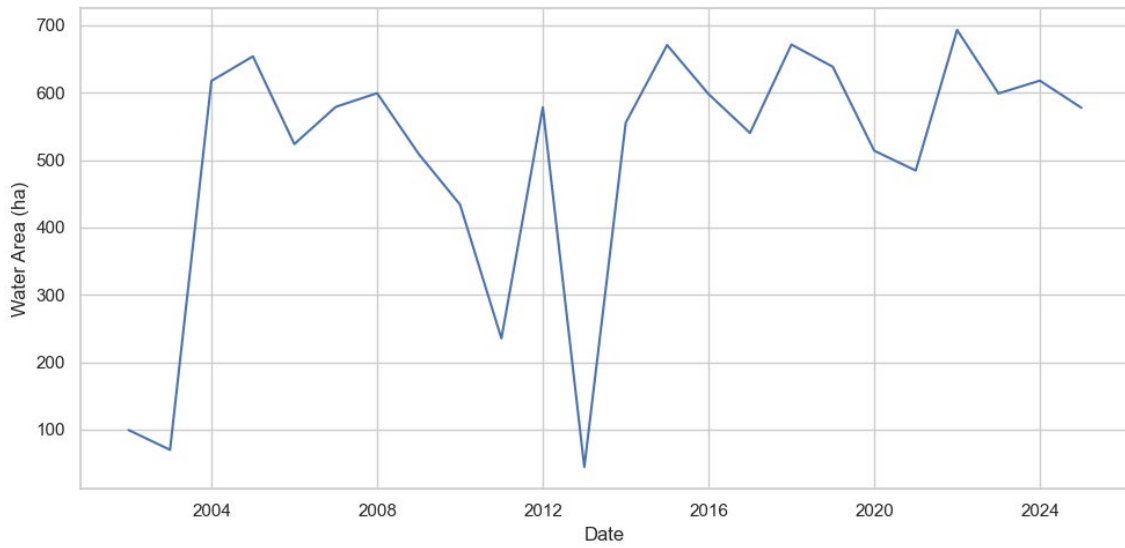
Yearly aggregated water area time series by channel

Query 6: Extract yearly aggregated **total water area time series** from the **waterbodies_data** table given a **swat_channel**

```
SELECT aggregated_month.swat_channel, aggregated_month.last_day_of_year as date,
sum(aggregated_month.avg_water_area_ha) as total_water_area_ha
FROM (
```

```
SELECT w.swat_channel, w.id, DATE_SUB(MAKEDATE(YEAR(wbd.date) + 1, 1), INTERVAL 1 DAY) AS
last_day_of_year, AVG(wbd.water_area_ha ) as avg_water_area_ha
FROM eflow_iwmi.waterbodies_data wbd
inner join waterbodies w on w.id = wbd.waterbody_id
where wbd.percent_invalid < 10
group by w.swat_channel, w.id, last_day_of_year
) as aggregated_month
where aggregated_month.swat_channel = 'cha0075'
group by aggregated_month.swat_channel, aggregated_month.last_day_of_year
order by aggregated_month.last_day_of_year
```

	swat_channel	date	total_water_area_ha
1	cha0075	2001-12-31	99.405
2	cha0075	2002-12-31	70.0741258741
3	cha0075	2003-12-31	617.34
4	cha0075	2004-12-31	653.886
5	cha0075	2005-12-31	523.692
6	cha0075	2006-12-31	578.8875
7	cha0075	2007-12-31	599.2155
8	cha0075	2008-12-31	509.2634805195
9	cha0075	2009-12-31	433.92825
10	cha0075	2010-12-31	235.4325
11	cha0075	2011-12-31	578.1394285714
12	cha0075	2012-12-31	44.514
13	cha0075	2013-12-31	555.2284090909
14	cha0075	2014-12-31	670.6585227273
15	cha0075	2015-12-31	597.9253846154
16	cha0075	2016-12-31	540.1370535714
17	cha0075	2017-12-31	671.3328202386
18	cha0075	2018-12-31	638.42
19	cha0075	2019-12-31	513.8806561086
20	cha0075	2020-12-31	484.4328571429
21	cha0075	2021-12-31	693.0303506494
22	cha0075	2022-12-31	598.8511363636
23	cha0075	2023-12-31	617.7959337731
24	cha0075	2024-12-31	577.5226785714



Citation

Chambel-Leitão, P.; Santos, F.; Barreiros, D.; Santos, H.; Silva, P.; Madushanka, T.; Matheswaran, K.; Muthuwatta, L.; Vickneswaran, K.; Retief, H.; Dickens, C.; Garcia Andarcia, M. 2024. *Operational SWAT+ model: advancing seasonal forecasting in the Limpopo River Basin*. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Initiative on Digital Innovation. 97p.

Disclaimer

This publication has been prepared as an output of the [CGIAR Initiative on Digital Innovation](#), which researches pathways to accelerate the transformation towards sustainable and inclusive agrifood systems by generating research-based evidence and innovative digital solutions. This publication has not been independently peer reviewed. Responsibility for editing, proofreading, and layout, opinions expressed, and any possible errors lies with the authors and not the institutions involved. The boundaries and names shown and the designations used on maps do not imply official endorsement or acceptance by IWMI, CGIAR, our partner institutions, or donors. In line with principles defined in the [CGIAR Open and FAIR Data Assets Policy](#), this publication is available under a [CC BY 4.0](#) license. © The copyright of this publication is held by [IWMI](#). We thank all funders who supported this research through their contributions to the [CGIAR Trust Fund](#).