https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

INFLOW-AI v2.1: A Machine Learning Framework for Predicting Out-of-Sample Extreme Seasonal Flood Extents

Jessica Rapson[1], Elisabeth Stephens[2,3], Ross I. Maidment[2], and Rogerio Bonifacio[4]

[1]Department of Statistics, University of Oxford, Oxford, United Kingdom
[2]Department of Meteorology, University of Reading, Reading, United Kingdom
[3]Red Cross Red Crescent Climate Centre, The Hague, Netherlands
[4]World Food Programme, Rome, Italy

Correspondence to: Ross Maidment (r.i.maidment@reading.ac.uk)

Abstract. Forecasting flood extent during extreme events remains a critical challenge for hydrological modelling, particularly in data-scarce and highly dynamic floodplain systems. Accurate and timely forecasts of these events are essential for effective disaster preparedness and response. Traditional physically based methods are often not well-suited for modelling complex hydrodynamic systems, as they depend on fixed structural parameterisations of surface water processes, groundwater interactions, and evapotranspiration that are difficult to calibrate and scale in catchments with highly heterogeneous vegetation, climatology, and terrain. Machine learning approaches, which can learn nonlinear relationships directly from data without explicit physical parameterisation, offer a promising alternative for modelling flooding in these regions.

We present INFLOW-AI v2.1, a machine learning framework for predicting extreme seasonal flood extent beyond what was observed in the training set. To enhance predictive accuracy for these out-of-sample extreme events, the framework employs a two-stage neural network architecture that combines (1) extreme-sensitive temporal thresholds with (2) dynamic spatial predictions. The first stage employs transformer-based models with multi-headed attention mechanisms to capture long– and short-term hydrometeorological patterns in total flood extent over the past 36 dekads. To enable more effective detection of extremes, this stage predicts the first difference of the seasonal anomaly in total flood extent, rather than the raw total flood extent. The second stage then dynamically models spatial flooding patterns using a ConvLSTM to predict local inundation probabilities at 1 km resolution, with the basin-scale inundation extent predicted by the first stage used to constrain the spatial predictions. The model generates forecasts with a lead time of up to six dekads (two months).

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

A case study was conducted over the Sudd wetland in South Sudan, one of the world's largest freshwater ecosystems which has experienced unprecedented catastrophic flooding beginning in June 2019, severely impacting Jonglei, Unity, and Upper Nile States. INFLOW-AI was tested on this catchment, demonstrating the two-stage model's ability to predict extreme out-of-sample post-2019 flooding with only exposure to pre-2019 data. INFLOW-AI has been deployed operationally since the 2024 flood season (August–November) on the Joint Analysis System Meeting Infrastructure Needs (JASMIN), providing real-time predictions to humanitarian organisations and informing flood preparedness in South Sudan.

1. Introduction

Flooding is one of the most significant hydrological phenomena, shaping landscapes, ecosystems, and human societies across the globe (Doocy et al., 2013). In many river basins, seasonal flooding is a natural and necessary process that sustains biodiversity, replenishes groundwater, and supports agriculture and fisheries. However, extreme flood events can have devastating consequences, disrupting livelihoods, damaging infrastructure, and leading to loss of life. Understanding and predicting floodplain inundation dynamics is therefore crucial for both ecological management and disaster preparedness.

Seasonal flooding along rivers and wetlands plays a critical role in maintaining ecological and socio-economic stability, influencing biodiversity (Rebelo, Senay, & McCartney, 2012), agricultural and fisheries productivity, and overall ecosystem dynamics (Reis et al., 2019). Additionally, these flood events contribute to global carbon and methane fluxes (Mitsch et al., 2010; Lunt et al., 2019; Dong et al., 2024), affect the timing and severity of disease outbreaks (Fonseca et al., 2022), and have even been linked to increased social conflict in vulnerable regions (Sosnowski et al., 2016). While seasonal inundation can be beneficial in sustaining livelihoods and ecosystems, extreme flooding events often overwhelm the coping capacity of communities (Pinho, Marengo, & Smith, 2015). In such cases, prolonged submersion of croplands can lead to food insecurity and economic hardship, with floodwaters persisting for months in some regions, such as the Lower Indus Basin in Pakistan (Atif et al., 2021). In the Sudd wetlands of South Sudan, the impact of extreme rainfall in the upstream catchment can have long-term consequences, with floodwaters lingering for years (Hardy, Palmer, & Oakes, 2023). Given these wide-ranging effects, the ability to forecast the timing and magnitude of seasonal flooding is essential for disaster preparedness, ecosystem management, and resource allocation (Pinho et al., 2015). This is especially true for major river basins such as the Amazon, Indus, Nile, and Ganges-Meghna-Brahmaputra, where nearly a billion people depend on the seasonal flow of water for their livelihoods.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

The predictability of seasonal flooding in large transboundary river basins is well established in hydrological literature. Major climate variability patterns, including the El Niño Southern Oscillation (Ward et al., 2014) and the Indian Ocean Dipole (Ficchì et al., 2021), are known to influence not only the magnitude but also the timing of peak river flows (Ficchì and Stephens, 2019). Advances in remote sensing technology now allow for continuous satellite monitoring of large upstream lakes (Hou et al., 2024), while established water management policies regulate the outflows from these reservoirs (Hardy et al., 2023). Furthermore, satellite-based monitoring techniques make it possible to track the onset and duration of seasonal flooding in near real-time (Papa et al., 2023).

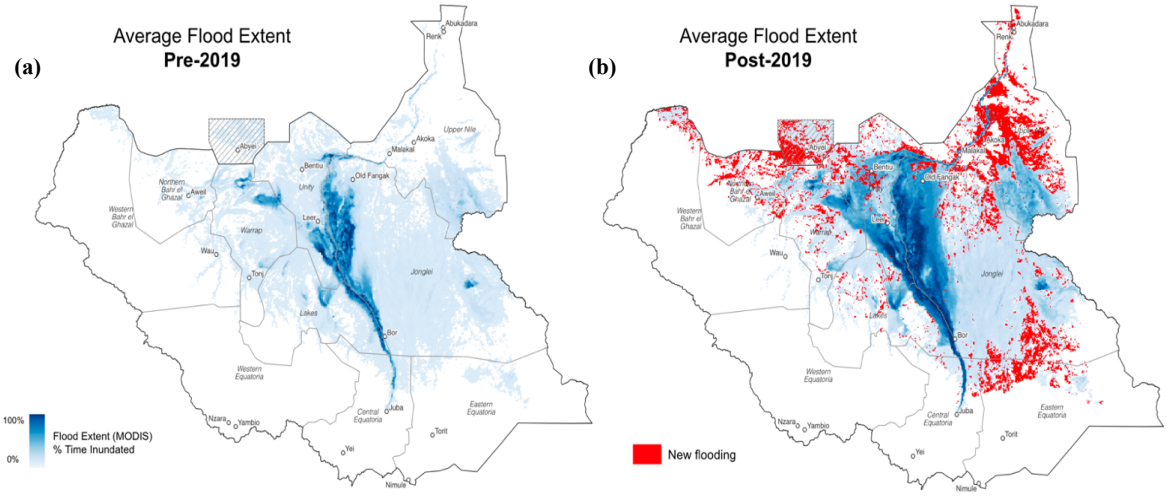

Figure 1. Maps of South Sudan showing **(a)** average Sudd wetlands extends from 2002-2019 and **(b)** from 2019-2024 based on flood extents detected by the MODIS satellite mission, with key population centres labelled. New areas of the country exclusively affected by post-2019 flooding, with no record of inundation from 2002-2019, are highlighted in red.

In 2024 alone, over 40 million people globally were displaced by water-related disasters (Global Water Monitor, 2024). South Sudan is particularly vulnerable, with floods severely impacting food security, livelihoods, and infrastructure. The humanitarian consequences include displacement, loss of agricultural productivity, and increased risks of waterborne diseases. As climate change drives more frequent and severe flood events, robust early warning systems are urgently needed to support anticipatory action and disaster preparedness. Predicting floods in the Sudd Wetlands of South Sudan with adequate lead time should be possible, given that water from Lake Victoria takes at least one month to reach the area (Sene, 2000). Satellite observations of lake levels and pre-existing flood extents therefore offer a foundation for reliable early warnings (Koriche & Rientjes, 2016). However, the development of physically based hydrological models to include the physical processes and data assimilation required to incorporate these datasets remains a complex challenge, highlighting the need for alternative approaches for real-time flood forecasting.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Traditional physically based flood forecasting models have significant limitations in wetlands like the Sudd, struggling with surface water-groundwater interactions, spatially variable evapotranspiration, and dynamic wetland storage (Zhang et al., 2025; Alfieri et al., 2024; Bukhari & Brown, 2021). Under shifting climatic regimes, such models can struggle to capture nonstationary hydrological responses that diverge from the conditions used for model calibration, creating systematic errors that conventional physically based structures cannot easily correct (Slater et al., 202). These uncertainties, along with computational constraints and challenges in data assimilation, reduce their operational effectiveness (Özdoğan-Sarıkoç & Dadaser-Celik, 2024). Notably, physically based hydrological models struggled to simulate the unprecedented flooding experienced in the Sudd wetlands since 2019 (Zhang et al., 2025). The flood extent in South Sudan peaked in 2022 and has remained persistently high since then, suggesting a shift in the regional flood regime. This shift is characterised by the emergence of inundation in previously unaffected areas and prolonged flooding in regions that historically experienced only seasonal inundation (Figure 1). With this change in flood extent, up to a million people have been affected by the flooding every year since 2019 (UNICEF South Sudan, n.d.). Forecasting this flood extent is thus vital for supporting community preparedness. Data-driven approaches, such as machine learning (ML), offer a promising solution (Özdoğan-Sarıkoç & Dadaser-Celik, 2024). Advances in ML coupled with high-resolution satellite data and improved inundation monitoring, enable data-driven models that can learn complex hydrological patterns. These data-driven models can more easily integrate numerous diverse datasets without presuming numerical structure in their interactions, including satellite observations, climate indices, hydrological measurements, and local subbasin properties to generate more accurate and timely flood predictions.

This study presents INFLOW-AI (Improved Anticipation of Floods on the White Nile using Artificial Intelligence) v2.1, a ML-based flood forecasting model for predicting flood extent. Leveraging state-of-the-art deep learning architectures, the INFLOW-AI v2.1 predictions have already enabled early warnings to inform humanitarian anticipatory and response efforts in South Sudan. Operationalised on the JASMIN supercomputing cluster, it delivered real-time flood predictions during the 2024 and 2025 flood seasons, providing real-time information to the South Sudan Red Cross, Médecins Sans Frontières (MSF), and the United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

2. Background

2.1 Machine learning-based flood forecasting

The application of artificial intelligence (AI) in flood forecasting has gained increasing attention as a means of improving prediction accuracy, particularly for extreme flood events that exceed historical observations (Slater et al., 2023). Machine learning (ML)

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

approaches offer the potential to learn nonlinear relationships from diverse datasets where physical hydrological effects are unknown or difficult to parameterise, enhancing predictive skill in both gauged and ungauged regions. ML models can also ingest near-real-time data without the need for traditional data assimilation, offering speed and flexibility in operational contexts (Frame et al., 2022; Slater et al., 2023).

One of the most well-known large-scale ML flood forecasting systems is Google's AI Flood model (Nearing et al., 2023), which aims to predict extreme fluvial floods in ungauged basins. It is benchmarked against the Global Flood Awareness System (GloFAS), a physics-based global flood forecasting system, and demonstrates higher predictive skill at a five-day lead time than GloFAS forecasts at comparable spatial scales, including near-real-time (zero-day) products. Using an encoder-decoder LSTM architecture and forced with meteorological and geophysical data, the model has achieved notable success but still exhibits variability in its ability to predict rare or severe events, a challenge that remains common to many ML models (Frame et al., 2022; Slater et al., 2023).

Beyond purely ML models, there is a growing movement toward hybrid hydroclimatic forecasting systems that combine dynamical (physically based) and data-driven methods (Slater et al., 2023). These hybrid systems aim to leverage the strengths of both model types: the physical consistency and interpretability of numerical models, and the data-driven pattern recognition and bias correction capabilities of ML. Hybrid structures typically fall into three broad categories: statistical-dynamical hybrids, where ML models are driven by outputs from numerical weather or climate models; serial hybrids, which sequence physics-based and ML models (e.g., post-processing streamflow forecasts); and coupled hybrids, which run both models in parallel or integrate ML directly into physical model components.

As highlighted by Slater et al. (2023), hybrid approaches are increasingly recognised as a promising direction for improving flood prediction skill across timescales, from hourly forecasts driven by Numerical Weather Prediction (NWP) models to seasonal or dekadal outlooks informed by climate ensembles. Their operational advantages include reduced computational cost, the ability to combine multiple sources of predictability (e.g., weather forecasts and El Niño-Southern Oscillation indices), and improved local scale forecast reliability. Importantly, hybrid models also offer a pathway toward physically informed ML, where ML models are constrained or guided by physical understanding, reducing risks of overfitting or implausible outputs.

Several recent studies highlight innovations in ML model design for flood prediction. Yeditha et al. (2020) developed a wavelet-based ML model for flood-prone regions in India, demonstrating that maximal overlap discrete wavelet transforms (MODWT) can improve feature selection and prediction skill. Pasche & Engelke (2022) proposed an extreme quantile regression network (EQRN)

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

that integrates extreme value theory with neural networks, estimating conditional Generalized Pareto Distribution (GPD) parameters for more stable extreme quantile prediction in Switzerland.

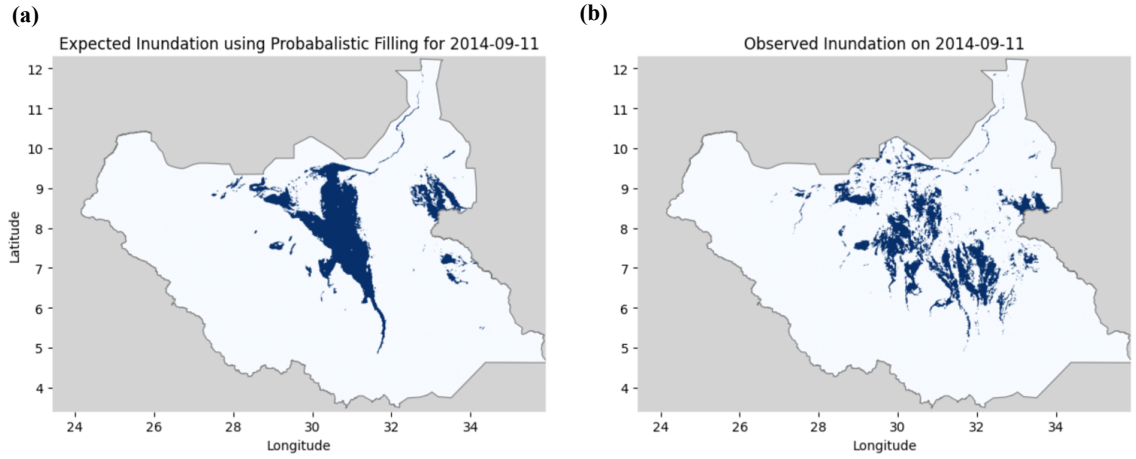

Figure 2. Diagrams of Sudd wetlands inundation extent on September 11th, 2014 (using data from the MODIS satellite mission). **(a)** Theoretical inundation using constant fill methodology, where each 1km² region is filled based on historic inundation records (i.e. pixels with the highest frequency of historical inundation are predicted to flood first until the total area reported as inundated is reached). **(b)** Actual map of the Sudd wetlands for the target date. The large number of violations of the constant-fill model indicate that this model is not appropriate for modelling variable wetlands extent in the White Nile basin.

In the South Sudanese context, El-Mahdy et al. (2024) developed a feedforward neural network to predict urban flooding across several cities between 2010 and 2019. Their study represents one of the few applications of ML-based flood prediction in the region and highlights the growing interest in data-driven approaches in climate-vulnerable and data-scarce settings. However, the model is limited in spatial and temporal scope, producing binary predictions at the city scale and excluding post-2019 flood events, many of which were among the most extreme on record. The study also does not explicitly address uncertainty quantification or potential temporal leakage, which may affect the generalisability and robustness of the reported results.

2.2 Case study: Flooding in the White Nile basin

A significant portion of South Sudan's flooding occurs in the Sudd wetlands, one of Africa's largest wetland ecosystems. This region experiences seasonal inundation, with nearly half of the incoming river water lost between Juba (upstream) and Malakal (downstream) due to evapotranspiration and water retention within the wetland system (Petersen and Fohrer, 2010). The extent of permanent wetland coverage is influenced by long-term variations in outflow from upstream lakes, particularly Lake Victoria, as well as regional precipitation and evaporative losses. Flooding in the Sudd wetlands begins when the seasonal floodplains become saturated, requiring approximately 350 mm of recharge, most of which originates from river overflow rather than direct precipitation. Once inundated,

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

floodwaters recede extremely slowly due to limited outflows, with evapotranspiration occurring at an average rate of approximately 5 mm per day during the inundation period. This hydrological behaviour makes forecasting particularly difficult, as water movement is driven by a complex interplay of inflows, groundwater storage, and surface retention. Purely empirical methods that predict inundation extent using historical inundation patterns are insufficient to accurately represent the hydrological complexity of the region (Figure 2).

The White Nile basin presents additional forecasting challenges due to the potential for unprecedented extreme flood events. Operational statistical and semi-empirical flood extent models, which rely heavily on historical trends and observed data, often struggle to capture these out-of-sample extremes. For instance, the Famine Early Warning Systems Network (FEWS NET) employs a modelling approach based on LASSO regression applied to lagged streamflow, runoff, and precipitation variables to forecast inundation extent. The spatial distribution of flooding is then predicted by statically filling 375 m × 375 m pixels in descending order of historical inundation probability, based on a flood frequency map derived from monthly VIIRS flood observations (USGS, 2024). While effective for typical seasonal events, this approach relies on expected flood patterns and does not allow flood probabilities to evolve spatially over time in response to changing hydrological conditions. That is, it assumes a static probability of flooding for each pixel, regardless of changing local conditions. For hydrologically complex regions such as the Sudd wetlands, a more robust approach that can both predict out-of-sample extreme events and dynamically represent spatial flood evolution is required.

2.3 Flood forecasting with INFLOW-AI

Flood forecasting in the White Nile basin requires a modelling approach capable of representing complex hydrological and climatic processes that govern floodplain inundation in the region. Existing operational models have limited skill in predicting extreme out-of-sample flood events, constraining their usefulness for anticipatory action. To address these challenges, INFLOW-AI was developed as a two-stage neural network framework designed to improve flood extent forecasting, with a particular focus on extreme events.

The first stage of INFLOW-AI employs transformer-based models with multi-headed attention mechanisms, as introduced by Vaswani et al. (2017), to capture both long- and short-term hydrometeorological dependencies. The model ingests 36 dekads of historical climate and hydrological inputs to generate basin-scale forecasts of total flood extent. Rather than predicting raw flood extent directly, this stage predicts first-differenced seasonal anomalies of flood extent. This transformation removes the dominant seasonal cycle and reduces sensitivity to scale effects, improving the model's ability to identify anomalous deviations associated

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

with extreme flooding. Full details of the model architecture, preprocessing, and training procedure are provided in the Methods section.

The second stage of INFLOW-AI uses a ConvLSTM architecture (Shi et al., 2015) to generate spatially explicit inundation probability estimates at a 1 km spatial resolution. This stage conditions local-scale spatial predictions on the basin-scale flood extent forecast produced by the first stage, enabling the model to dynamically represent spatial flood evolution rather than relying on static inundation patterns. By explicitly modelling spatial dependencies, the ConvLSTM stage supports more realistic representation of localised flooding dynamics across the Sudd wetlands and surrounding floodplains.

INFLOW-AI operates in near real-time, ingesting live hydrometeorological inputs and issuing forecasts on a dekadal basis. The system is deployed on JASMIN, a high-performance computing facility managed by the Science & Technology Facilities Council (STFC) for the Natural Environment Research Council (NERC). Owing to its computational efficiency, the full forecasting workflow, including data retrieval, preprocessing, and inference, typically completes within ten minutes. Model training can be performed on a single A100 GPU, facilitating use by research and operational groups without access to large-scale computational infrastructure.

The model produces updated forecasts every dekad and predicts Sudd wetland inundation extent up to six dekads in advance, corresponding to an approximate two-month lead time. This forecast horizon is intended to support anticipatory planning by humanitarian organisations and local authorities in flood-affected regions. Model performance is explicitly evaluated on extreme out-of-sample flood events, allowing assessment of robustness under conditions that differ from those observed during training.

INFLOW-AI is built entirely using open-source software and publicly available datasets, supporting transparency and reproducibility. The framework allows flexible geographic subsetting, enabling forecasts to be generated for specific administrative regions, river sub-basins, or areas of operational interest. The model has been operationally deployed and used to generate live flood forecasts during the 2024 and 2025 flood seasons (August–November). By combining temporally focused anomaly-based prediction with spatially explicit inundation modelling, INFLOW-AI provides a data-driven framework for forecasting flood extent in highly dynamic floodplain systems such as the White Nile basin.

3. Data

3.1 Inundation

The target variable for this study is the extent of inundation, predicted temporally in the first stage of the model and spatially in the

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

second (Figure 3). To generate accurate flood extent predictions, we utilise the Moderate Resolution Imaging Spectroradiometer (MODIS) Flood Masks dataset provided by the World Food Programme (WFP). This dataset offers dekadal (approximately 10-day) binary inundation masks for the White Nile basin, derived from MODIS instruments aboard NASA's Aqua and Terra satellites that are processed to capture the magnitude of the diurnal temperature cycle; specifically, the difference between daily minimum and maximum temperatures. The spatial resolution of these masks is 1 km, and the data is updated at the end of each dekad, requiring no further forecasting. The dataset is freely available from the WFP website (World Food Programme, 2024).

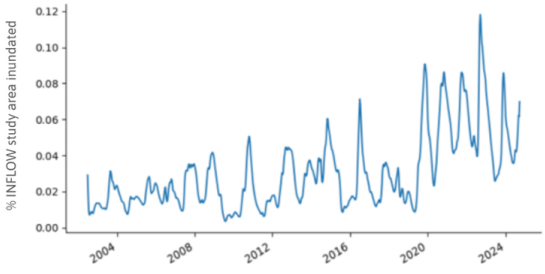

Figure 3. Sudd wetland extent as a percentage of the INFLOW study area from 2002-2025, per MODIS-derived satellite data provided by the WFP. Extreme increases in the inundation extent post-2019 are evident.

MODIS is particularly well suited for this study due to its long and consistent time series, covering 2002 to the present, which supports robust flood prediction based on extensive historical observations. Although alternative satellite products such as the Visible Infrared Imaging Radiometer Suite (VIIRS) offer higher spatial resolution, MODIS provides a more stable and homogeneous multi-decadal record for the White Nile basin. The WFP MODIS flood product used here relies on diurnal temperature range-based detection, which performs well in the Sudd wetlands where water driven thermal damping helps differentiate inundated from non-inundated surfaces. This consistency and suitability for densely vegetated wetland environments make MODIS a strong choice for mapping inundation in the region.

The dekadal data provided by the WFP MODIS Flood Masks dataset is organised into three 10-day periods within each calendar month. The first dekad of the month refers to the days from the 1st to the 10th, the second dekad covers the 11th to the 20th, and the third dekad encompasses the remainder of the month (21st to the end of the month). Notably, the last dekad of each month may contain a varying number of days depending on the length of the month. This temporal organization ensures that flood events are captured in a granular and consistent manner, aiding in accurate temporal predictions.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

To process the MODIS Flood Masks data and ensure it is suitable for flood forecasting, the data is pre-processed using the Whittaker smoother, a fast and efficient smoothing and interpolation algorithm designed for time-series data. Introduced by Eilers, Pesendorfer, and Bonifacio (2017), the Whittaker smoother optimises the amount of smoothing using a variation of the L-curve, known as the V-curve. This technique allows the model to handle missing data points effectively and correct for noise in the data.

The objective function of the Whittaker smoother is:

$$S = \sum_i w_i (y_i - z_i)^2 + \lambda \sum_i (\Delta^2 z_i)^2. \quad (1)$$

where the first term represents the weighted sum of squares of residuals, capturing the "distance" between observed and smoothed values. The second term penalises roughness by summing the squares of second-order differences of the smoothed series z where:

$$\begin{aligned} \Delta^2 z_i = \Delta(\Delta z_i) &= (z_i - z_{i-1}) - (z_i - z_{i-2}) \\ &= z_i - 2z_{i-1} - z_{i-2}. \end{aligned} \quad (2)$$

The parameter λ determines the amount of smoothing, larger values of λ result in smoother curves. Optimal smoothing is achieved by minimising the Euclidean distance between:

$$\psi = \log \sum_i w_i (y_i - \hat{z}_i)^2, \qquad \phi = \log \sum_i (\Delta_i^z)^2. \quad (3)$$

over a finite grid of λ.

The Whittaker smoother is well-suited for processing earth observation data, as it is computationally fast due to the use of sparse matrices and can handle large datasets with ease. Additionally, the application of expectile smoothing corrects for gaps in the data caused by cloud cover, further improving the quality of the inundation estimates.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

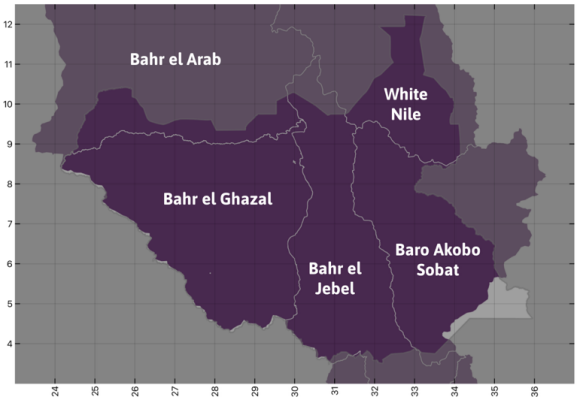

Figure 4. Major catchment boundaries for the White Nile basin of South Sudan, composing the INFLOW study area.

3.2 Catchment boundaries

The INFLOW Catchment Boundaries dataset was developed as part of the INFLOW project, using HydroSHEDS v1 data to delineate hydrologically relevant catchment areas for flood anticipation in the White Nile basin (Figure 4; Lehner, Verdin, & Jarvis, 2008). Catchment boundaries were defined using HydroSHEDS Level 4-6 sub-basins and cross-referenced against national hydrological planning frameworks to ensure consistency with management-relevant units. Specifically, Uganda's Water Management Zones (Ministry of Water and Environment, 2009) and South Sudan's Irrigation Development Master Plan (Ministry of Electricity, Dams, Irrigation & Water Resources, 2015) were used to validate and refine sub-catchment definitions, particularly in regions where sub-national boundaries influence rainfall aggregation and flood response assessment. This dataset provides a static, vector-based representation of catchment boundaries, meaning it does not change over time and does not require periodic updates. By defining these catchment areas, the dataset allows for a structured analysis of regional flood dynamics and hydrological processes that contribute to inundation.

3.3 Satellite altimetry

Satellite altimetry provides essential data for tracking water level variations in lakes, reservoirs, rivers, and wetlands, contributing to improved flood forecasting (Albright, 2011). Radar altimeters, operating at Ku or Ka band frequencies, measure surface height by emitting microwave pulses and recording the two-way travel time. The footprint size varies based on surface roughness, ranging from approximately 200 meters over calm water to several kilometres over rougher open water. Because each satellite follows a

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

repeat orbit, it revisits nearly the same location within approximately 1 km after a fixed period, typically between 10 to 35 days depending on the satellite's orbit, enabling the construction of time series datasets to monitor surface height changes. These data are widely validated, with mission continuity ensured until at least 2030.

Laser altimeters, which operate at green or infrared wavelengths, complement radar altimeters by enabling off-pointing toward specific targets of interest. These instruments are primarily used or monitoring ice sheets, mapping vegetation, and studying atmospheric properties. Notable laser altimetry missions include ICESat-1, ICESat-2, and GEDI, with ICESat missions providing high-latitude elevation data and GEDI focusing on mid- to low-latitude vegetation structure from the International Space Station.

For flood forecasting in the White Nile basin, three key datasets provide critical water level measurements. NASA's Lake Victoria Height Variation dataset delivers monthly water level estimates referenced to the EGM2008 vertical datum, with a forecasting requirement of up to three months (NASA Goddard Space Flight Center, 2025). Theia Hydroweb's Lake Albert Water Level dataset updates monthly with a one-month lag and allows forecasting up to six months ahead (Theia, 2025). NASA's Lake Kyoga Height Variation dataset provides dekadal water level estimates, updated at the end of each dekad, without requiring forecasts (NASA Goddard Space Flight Center, 2025). These datasets offer crucial insights into hydrological trends, supporting anticipatory action and flood risk management in the region.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Data Name	Source	Spatial Resolution	Temporal Resolution	Update Frequency
MODIS Flood Masks	World Food Programme (WFP)	1 km	Dekadal	End of each dekad
INFLOW Catchment Boundaries	INFLOW Project	Vector-based	Static	No updates
Lake Victoria Height Variation	NASA	Not applicable	Approximately weekly	One-week lag
Lake Albert Water Level	Theia Hydroweb	Not applicable	Approximately weekly	One-week lag
Lake Kyoga Height Variation	NASA	Not applicable	Approximately weekly	One-week lag
TAMSAT v3.1 Lake Victoria Rainfall	TAMSAT	Not applicable	Daily	1st, 6th, 11th, 16th, 21st, 26th of each month
CHIRPS v3 Lake Victoria Rainfall	Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)	Not applicable	Daily	1st, 6th, 11th, 16th, 21st, 26th of each month
TAMSAT v3.1 Rainfall Estimates	TAMSAT	~4km	Daily	Two-day lag
TAMSAT v3.1 Soil Moisture	TAMSAT	~25 km	Daily	Seven-day lag
Oceanic Niño Index (ONI)	Climate Prediction Center (CPC)	Not applicable	Monthly	One-month lag
Niño Regions SST & SOI	Various	Not applicable	Monthly	One-month lag
Madden-Julian Oscillation (MJO) Indices	Climate Prediction Center (CPC)	Not applicable	Every 5 days	Five-day lag
Indian Ocean Dipole Mode Index (DMI)	Climate Prediction Center (CPC)	Not applicable	Monthly	One-month lag
HydroATLAS Basin Attributes (v1.0)	HydroATLAS	Vector-based	Static	No updates
South Sudan Digital Elevation Model (DEM)	Shuttle Radar Topography Mission (SRTM)	~100m	Static	No updates

Table 1. Summary of data sources used in INFLOW-AI, along with spatio-temporal resolution.

3.4 Rainfall

Rainfall estimates for the White Nile basin used in this study are sourced from two complementary satellite-based datasets: the Tropical Applications of Meteorology using SATellite data and ground-based observations (TAMSAT) v3.1 program and the Climate Hazards Group InfraRed Precipitation with Station (CHRIPS) v3 data (University of Reading, 2025a). These datasets provide independent perspectives on rainfall, which is a critical driver of downstream inundation, particularly in a region where

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

satellite-based precipitation estimates can vary significantly.

TAMSAT provides high-resolution rainfall and soil moisture data across Africa (Maidment et al., 2014; Tarnavsky et al., 2014; Maidment et al., 2017). The TAMSAT v3.1 gridded rainfall dataset offers spatiotemporal rainfall estimates from 1983-present over the INFLOW study area (consisting of the Bahr el Arab, Bahr el Ghazal, Bahr el Jebel, Baro Akobbo Sobat, and White Nile catchments) at a resolution of 0.0375° (approximately 4 km), updated every five days with a two-day lag.

The TAMSAT Lake Victoria Rainfall dataset provides daily rainfall estimates subset and spatially average to Lake Victoria with a two-day update delay. To improve the robustness of rainfall inputs into the model, this dataset is supplemented with CHIRPS rainfall estimates over Lake Victoria. CHIRPS version 2.0 provides quasi-global (50°S-50°N) rainfall estimates from 1981-present at a spatial resolution of 0.05° (Funk et al., 2015). While the final version of the product is updated with a latency of approximately 6 weeks, the preliminary version is updated typically within 7 days (for most regions of Africa, the difference between the final and preliminary versions of CHIRPS are very small, including the White Nile Basin). Within INFLOW-AI, the preliminary version is added onto the final CHIRPS version to extend the rainfall record up to the near-present.

TAMSAT and CHIRPS are alike in that that they utilise thermal-infrared satellite imagery to identify the cold cloud tops of convective rain-bearing cloud systems and incorporate rain gauge information to bias correct the satellite rainfall estimates. However, CHIRPS employs an additional step to directly merge contemporaneous rain gauge measurements with the satellite estimates to improve skill. Despite limited studies of their skill over the region of interest, both TAMSAT and CHIRPS demonstrate skill in reliably estimating rainfall amounts over the While Nile Basin (Dinku et al., 2018). Incorporating both sources helps mitigate uncertainty and capture a more complete picture of upstream rainfall, which plays a critical role in determining flood dynamics in the White Nile Basin. Data from both TAMSAT and CHRIPS was used in the temporal model, as separate features.

3.5 Soil moisture

The TAMSAT v3.1 soil moisture dataset offers both historical records and forecasts of soil moisture over the top 1m of soil, providing crucial insights into surface hydrological conditions (University of Reading, 2025b). With a spatial resolution of 0.25°, it captures large-scale variations in soil moisture that influence flooding dynamics in the White Nile basin. The dataset is updated daily with a seven-day lag, ensuring near-real-time monitoring of soil saturation levels. This information is essential for improving flood forecasting models, as soil moisture directly affects runoff generation and the likelihood of inundation.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

3.6 Climate teleconnection indices

To analyse broader climate patterns that influence flooding in the White Nile basin, several climate teleconnection indices are utilised. The Oceanic Niño Index (ONI), provided by the Climate Prediction Center (CPC), tracks sea surface temperature anomalies in the Niño 3.4 region with a three-month running mean (CPC, 2025a). This index is updated monthly, with a one-month lag and a forecast horizon of up to four months. Additionally, the Niño Regions Sea Surface Temperatures (SST) and Southern Oscillation Index (SOI) offer monthly data on sea surface temperature anomalies and pressure differences, respectively, with updates at the end of each month and a three-month forecast horizon (CPC, 2025b; CPC, 2025c). The Madden-Julian Oscillation (MJO) Pentad Indices monitor an eastward-moving atmospheric wave and are updated every five days with a five-day lag, extending to a forecast horizon of two intervals (CPC, 2025d).

Lastly, the Indian Ocean Dipole Mode Index (DMI) measures the sea surface temperature gradient in the Indian Ocean, updated monthly with a one-month lag and a forecast horizon of five months (CPC, 2025e). These indices provide valuable insights into the climatic conditions that influence rainfall and river flow over East Africa, contributing to more accurate flood predictions (Pohl & Camberlin, 2007; Kebacho, 2022).

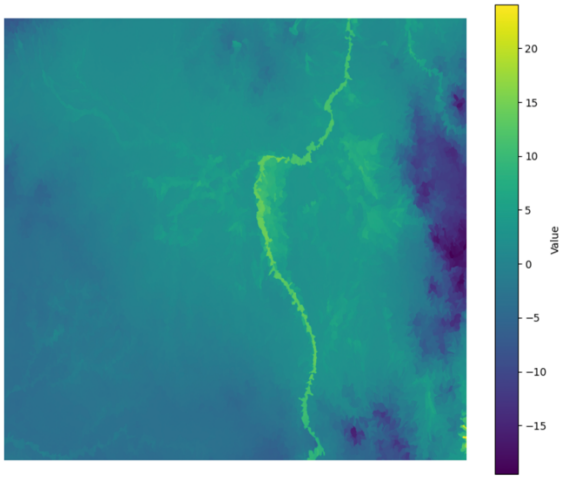

Figure 5. First principal component of HydroATLAS basins features for the INFLOW study area, combining information about elevation, soil type, vegetation, and average temperature.

3.7 Basin attributes

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

The HydroATLAS Basin Attributes (version 1.0) dataset provides a suite of static, sub-basin-level characteristics such as elevation, land cover, soil type, lithology, vegetation, and climatic variables that are relevant to hydrological and ecological modelling (Linke et al., 2019). This vector-based dataset has no temporal resolution and does not require regular updates. It integrates data from a range of global geospatial sources to offer harmonized hydro-environmental attributes for river basins at multiple spatial scales.

Key data sources underlying HydroATLAS data used in the INFLOW-AI model include:

- Elevation and hydrological network topology from HydroSHEDS (Lehner & Grill, 2013) and WaterGAP v2.2 runoff models (Döll, Kaspar, & Lehner, 2003),
- Soil type from SoilGrids1km (Hengl et al., 2014) and the Global Lithological Map (GLiM; Hartmann & Moosdorf, 2012),
- Land cover from Global Land Cover in 2000 (GLC2000; Bartholomé & Belward, 2005) and EarthStat (Ramankutty & Foley, 1999),
- Climate variables such as mean temperature and potential evapotranspiration from WorldClim v1.4 and Global Potential Evapotranspiration (Global-PET) v1 (Hijmans et al., 2005),
- Wetland and lake distribution from Global Lakes and Wetlands Database (GLWD; Lehner & Döll, 2004), HydroLAKES (Messager et al., 2016), and Global Inundation Extent from Multi-Satellites (GIEMS-D15; Fluet-Chouinard et al., 2015),
- Population and socioeconomic data from Gridded Population of the World (GPW) v4 (Center for Integrated Earth System Information Network at Columbia University, 2016) and gross domestic product (GDP) purchasing power parity (PPP) v2 (Kummu et al., 2018),
- Ecoregions and biodiversity zones from Freshwater Ecoregions of the World (FEOW; Abell et al., 2008), Terrestrial Ecoregions of the World (TEOW; Dinerstein et al., 2017), and Global Environmental Stratification (GEnS; Metzger et al., 2013),
- And groundwater and topography from the Global Groundwater Map (Fan et al., 2013) and EarthEnv-DEM90 (Robinson et al., 2014).

HydroATLAS Basin Attributes, including factors such as soil type, vegetation cover, and proximity to water sources, were examined for their correlation with local flood patterns at the sub-basin level. Pre-testing revealed that many of these attributes exhibited strong linear relationships with observed inundation, underscoring their relevance for flood forecasting. Based on this analysis, a subset of 70 HydroATLAS features with high correlation to inundation extent was selected for use in the INFLOW-AI framework (Appendix

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

A).

To further enhance computational efficiency and eliminate redundancy in the feature set, Principal Component Analysis (PCA) was applied to these sub-basin characteristics (Jolliffe, 2002). This transformation reorients the data into a new coordinate system, where the first axis, known as the first principal component, captures the maximum possible variance. The second principal component accounts for the next highest variance, and this pattern continues for subsequent components, ensuring that the most significant patterns in the data are retained while reducing dimensionality. This is done by finding the first unit weight vector $\mathbf{w}_{(1)}$ to satisfy:

$$\mathbf{w}_{(1)} = \text{argmax}_{\|\mathbf{w}\|=1} \left\{ \sum_i (v_1)_{(i)}^2 \right\} = \text{argmax} \left\{ \frac{\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \right\}, \quad (4)$$

where $\mathbf{X}$ is a $n \times p$ data matrix with the sample means of each column shifted to zero and individual variables $v_1, v_2, \ldots, v_l$ with $l < p$. Subsequent components are determined by subtracting the previous component from $\mathbf{X}$ and finding the new unit weight vector that extracts the maximum variable variance from this new matrix $\widehat{\mathbf{X}}$:

$$\mathbf{w}_{(k)} = \text{argmax} \left\{ \frac{\mathbf{w}^T \widehat{\mathbf{X}}^T \widehat{\mathbf{X}} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \right\}, \quad (5)$$

with

$$\widehat{\mathbf{X}}_k = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X} \mathbf{w}_{(s)} \mathbf{w}_{(s)}^T. \quad (6)$$

Through this dimensionality reduction technique, the dataset was refined to 16 principal components, each representing key hydrological and geographical patterns within the catchment areas. An example of these principal components is shown in Figure 5. Of these, only the first (which explained 28.8% of the variation in the chosen HydroATLAS features for the region) was ultimately used in the dynamic spatial modelling model. The contribution of each HydroATLAS variable to this first principal component (PC) is shown in Appendix A.

By structuring the dataset in this way, the INFLOW Catchment Boundaries dataset provides a compact yet information-rich

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

representation of spatial variables influencing flood occurrence. This ensures that the flood prediction model can efficiently process hydrologically significant features while minimising noise and redundancy in the input data.

3.8 Elevation

The South Sudan Digital Elevation Model (DEM) provides a high-resolution representation of the region's topography, essential for hydrological modelling and flood prediction (de Ferranti, 2012). Derived from data collected during the 2000 Shuttle Radar Topography Mission (SRTM), this DEM captures bare-earth elevation at a spatial resolution of 100 meters. Unlike dynamic datasets that require frequent updates, the DEM remains static, serving as a foundational layer for assessing drainage patterns, catchment boundaries, and floodplain characteristics. Elevation influences water flow and accumulation, making it a critical input for hydrodynamic simulations and inundation forecasting in the White Nile basin.

4. Methods

The following Methods subsections present the full technical specification of INFLOW-AI. Sections 4.1-4.4 cover data processing and imputation, the seasonally differenced dekadal-anomaly transformation, the temporal transformer and thresholding strategy, and the ConvLSTM dynamic spatial model, respectively; readers seeking implementation detail or reproducibility information should consult those subsections and the accompanying figures.

A good flood forecasting model must retain adaptive skill, i.e. the capacity to generalise to new climate and hydrological regimes with statistics that are not present in the training data (Beucler et al., 2024). This challenge is especially pronounced in South Sudan, where a steep change in flooding was observed after 2019, and models trained only on prior years may struggle to extrapolate into this new, more extreme regime. To explicitly test this capability, INFLOW-AI is trained only on data prior to July 21st, 2018. This temporal cutoff removes all post-2019 extreme events from the training dataset, allowing us to rigorously evaluate model performance on truly unseen extremes. Additionally, this approach ensures that the model produces an ex-ante forecast, using only inputs and variables available at the time of prediction, with no information leaked from future events or outcomes.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

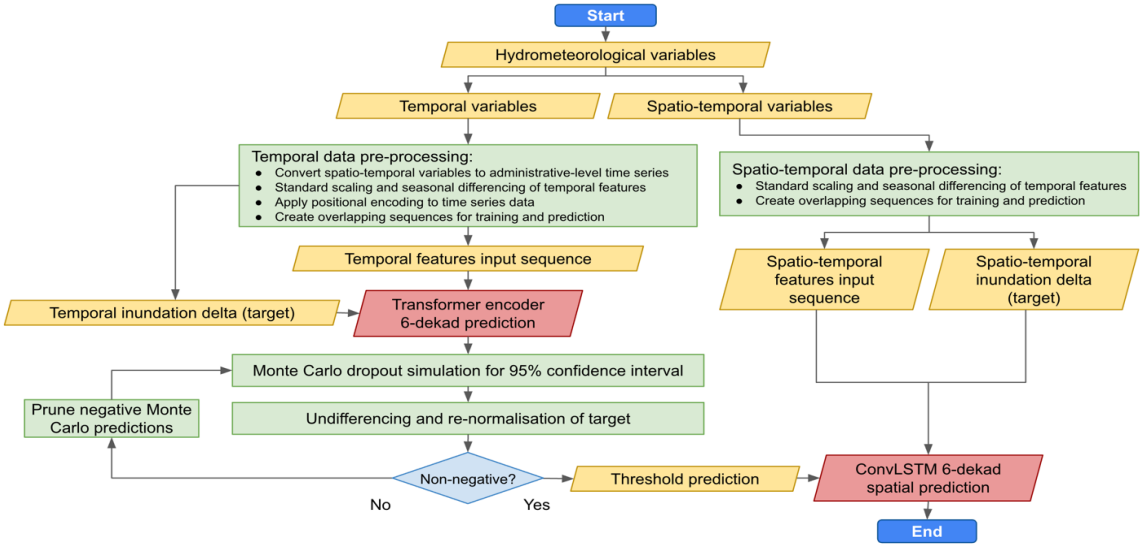

Figure 6. Overview of two-stage model and data processing pipeline for INFLOW-AI v2.1.

To evaluate model performance under these constraints, we formally define OSEVs using a standard deviation-based threshold. Let $\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{test}}$ represent the full dataset, where $y_i \in \mathbb{R}$ is the scalar response variable representing the percent inundation of the INFLOW study area. The mean and standard deviation of the training dataset $\mathcal{D}_{\text{train}}$ are given by:

$$\mu_{\text{train}} = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} y_i \tag{7}$$

$$\sigma_{\text{train}} = \sqrt{\frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} (y_i - \mu_{\text{train}})^2}. \tag{8}$$

Then, for a given threshold $k > 0$ in standard deviations (in this study, $k = 5$), we define an OSEV as any test observation $(x_j, y_i) \in \mathcal{D}_{\text{test}}$ satisfying the conditions $|y_j - \mu_{\text{train}}| > k\sigma_{\text{train}}$ and $\{y_i : (x_i, y_i) \in \mathcal{D}_{\text{train}}\}$. That is, an observation that exceeds the training distribution's extreme threshold and is only present in the testing data. Using this definition with $k = 5$, we identify 35 OSEV instances in the test set following the onset of extreme flooding after 2019. These provide a benchmark for model performance under the most difficult conditions.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Predicting such extremes is a known challenge for machine learning models, particularly high-variance models like neural networks. These models often fit training data well but exhibit poor generalization when exposed to rare or extreme out-of-distribution events (Blum & Reynaud-Bouret, 2020). Successfully predicting OSEVs, within a small margin of error and at the correct lead time, is therefore a strong indicator of a model's generalization and extrapolation capabilities.

To address this challenge, INFLOW-AI employs a two-stage architecture designed to separate the modelling of flood magnitude and spatial extent (Figure 6). The first stage is a temporal model that uses temporal-only information to predict the total percent inundation across the study area. This model is chosen for its interpretability and compatibility with operational anticipatory action frameworks. Crucially, it is also optimised to identify and predict OSEVs, forming the backbone of the system's ability to anticipate extreme events.

The second stage is a spatio-temporal model that generates pixel-level flood predictions across the region. It uses the temporal model's prediction as a thresholding mechanism, i.e. the total inundation predicted by the temporal model is used to determine the number of pixels that should be classified as inundated. The spatio-temporal model itself outputs probabilities for each pixel, and these probabilities are ranked to match the thresholded inundation value, ensuring coherence between regional-scale flood predictions and local flood patterns.

This two-stage approach enhances prediction accuracy by allowing the temporal model to specialise in forecasting extreme aggregate behavior, while the spatio-temporal model fine-tunes the spatial distribution. By explicitly decoupling these tasks, INFLOW-AI is better able to generalise from limited historical data and make skillful predictions during novel and severe flooding regimes.

4.1 Data processing

All input data were standardised to a common temporal resolution using dekads, in alignment with the MODIS flood mask data, which forms the basis of the target variable. As the input variables originated from heterogeneous sources with varying temporal frequencies (Table 1), preprocessing was required to ensure temporal alignment. For data sources that reported multiple values within a single dekad, values were averaged to produce a single representative value for that interval. In contrast, for sources updated less frequently than once per dekad, the most recent available value preceding each dekad was used. This forward-fill approach mimics a real-time forecasting environment by ensuring no future data is used to fill past intervals, thereby preserving the integrity of ex-ante prediction scenarios.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

For the temporal model, three-dimensional satellite data arrays (e.g., latitude × longitude × time) were aggregated into two-dimensional temporal sequences by spatially averaging over the area of interest, such that each feature represents a mean value aggregation of the original gridded observations. This spatial aggregation was conducted at three administrative levels: (1) the national boundaries of South Sudan, (2) its ten constituent states, and (3) the contested Abyei Area along the Sudan-South Sudan border. These compressed temporal arrays allowed the temporal model to learn flooding dynamics over the entire region based purely on temporal fluctuations of predictor variables.

Minimal data was missing from the data sources used. No spatio-temporal data was missing and less than 1% of temporal data was missing. To address missing temporal data, linear interpolation was applied using surrounding non-missing values. Given a missing observation $y(t_i)$ with t_{i^-} and t_{i^+} denoting the nearest previous and subsequent time steps with observed data, the value was imputed as:

$$y(t_i) = y_0 + \frac{(t_i - t_{i^-})}{t_{i^+} - t_{i^-}}(y_{i^+} - y_{i^-}), \tag{9}$$

where $t_{i^-} < t_i < t_{i^+}$ and y_{i^-}, y_{i^+} are the observed values temporally surrounding the missing data.

Once missing values were imputed, all predictor variables were standardised using *z*-score normalisation:

$$z_i = \frac{y_i - \mu_{\text{train}}}{\sigma_{\text{train}}}, \tag{10}$$

where μ_{train} and σ_{train} are the mean and standard deviation computed exclusively over the training set. This strict separation was essential to prevent data leakage from the test set into the training process. Specifically, standardising over the full dataset *including test values* would implicitly reveal information about future extreme values (OSEV), thus artificially boosting model awareness of events it should be oblivious to. By using only training data for normalisation, the model remains blind to the distributional characteristics of post-2019 extremes.

For real-time deployment, where future values may be missing due to delays in data availability, univariate OLS regression is used to extrapolate missing future values based on the past six dekads of available data. This was found to be the most effective imputation method for short-term forecasting, as it adapts to recent trends in temporal predictors and supports the continuity required for operational use. This approach was necessary to maintain consistent inputs across prediction cycles, as live data sources often update

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

asynchronously and with varying latencies.

4.2 Seasonal differencing

A central contribution of this study is improving the model's ability to predict extreme out-of-sample values (OSEV), which remain a persistent challenge for machine learning models due to their reliance on data distributions in the training set. When extreme events are absent or underrepresented in the training data, as is the case with the 2019 and post-2020 flooding regimes in South Sudan, models often fail to capture the complex, nonlinear dynamics that lead to such events. This is especially problematic for high-capacity models such as neural networks, which tend to overfit the training set and generalise poorly to regimes with substantially different statistical properties.

To address this, we adopt a seasonally differenced transformation, inspired by methods from classical time series analysis. While traditional forecasting models such as ARIMA routinely perform transformations such as first differencing to induce stationarity, machine learning models rarely apply such preprocessing (Hewamalage, Bergmeir, & Bandara, 2021). This is because many ML architectures can learn implicit temporal trends, making explicit transformations appear redundant.

However, this assumption breaks down in the presence of extreme values, where training data may be fundamentally unrepresentative of future conditions. In these contexts, detrending and deseasonalising the target variable offer underexplored benefits for extreme value detection.

First differencing, a standard transformation in ARIMA modelling (Box, Jenkins, & Reinsel, 2008), is used to convert a non-stationary time series y_t into a stationary series of deltas Δy_t defined as:

$$\Delta y_t = y_t - y_{t-1}. \quad (11)$$

In the context of inundation prediction, this transformation has two primary advantages. First, it reduces the variance of the target values by focusing on changes rather than absolute levels, which are often autocorrelated and smoother across time. Second, it reduces the effective number of OSEV, as large absolute extremes in the undifferenced series may correspond to relatively moderate changes between consecutive time steps in the differenced series. Thus, the model is trained to predict changes in inundation, not absolute levels, and can more feasibly generalise from moderate changes in the training set to large values in the test set.

To further reduce seasonal effects, we standardise the target variable seasonally, applying z-score scaling across dekads using only

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

training data. Specifically, for each time step t, let $\mathcal{D}_t$ denote all observations in the training set that fall on the same dekad of the year as t. The dekadal mean $\mu_{\mathcal{D}_t}$ and standard deviation $\sigma_{\mathcal{D}_t}$ are calculated from $\mathcal{D}_t$ and the dekadal anomaly is computed as:

$$\tilde{y}_t = \frac{y_t - \mu_{\mathcal{D}_t}}{\sigma_{\mathcal{D}_t}}. \quad (12)$$

Applying the first-difference operator to this scaled series yields the dekadal anomaly differenced transformation:

$$\Delta\tilde{y}_t = \tilde{y}_t - \tilde{y}_{t-1}, \quad (13)$$

where $\tilde{y}_t$ and $\tilde{y}_{t-1}$, are standardised using their respective dekad-based seasonal statistics as in (11). This transformation removes both long-term trends and cyclical seasonal variation from the target, enabling the model to focus exclusively on anomalous deviations in inundation dynamics. This is particularly valuable when training data are limited in duration or range. Without such transformations, models are at high risk of regressing to the mean of the training distribution, which may differ substantially from future distributions where OSEV are more prevalent. By learning to predict changes in standardised inundation levels rather than absolute magnitudes, the model gains robustness to statistical regime shifts, such as those caused by the persistent high-flooding periods post-2019. Moreover, these two transformations, seasonal scaling and first differencing, serve a joint purpose: they reduce the effective number of OSEV by simplifying the structure of the target variable, while preserving enough information to reconstruct the original series if needed (Figure 7 and Figure 8). This makes it possible for the model to learn generalisable patterns of change even when extreme values in their original form are absent from the training data.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

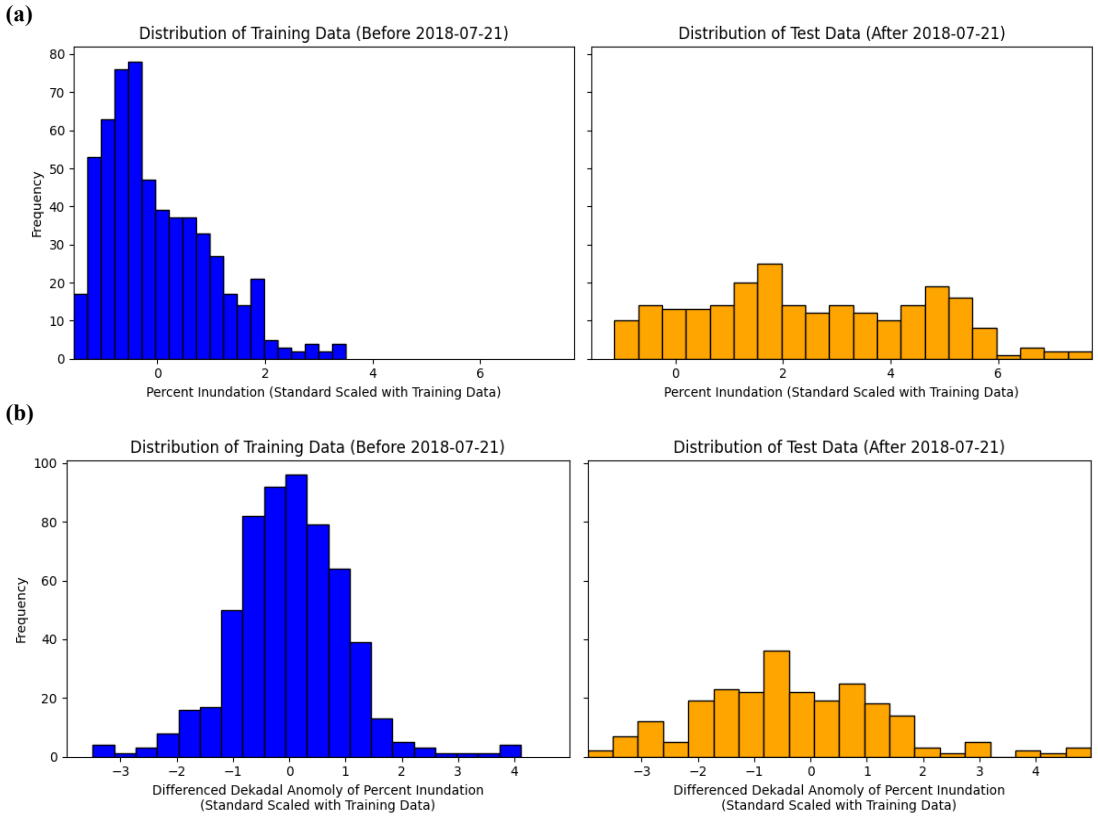

Figure 7. Distribution of training and test sets for the target variable (percent inundation of the INFLOW study area). **(a)** The distribution of the test data (right) does not resemble the training data (left) and contains 35 OSEV that exceed 5 standard deviations from the training data. **(b)** The distribution of the test data (right) when transformed using seasonal deltas more closely resembles that of the training data and contains no OSEV that exceed 5 standard deviations from the training data.

While this methodology is applied to dekads in the present study, it is not inherently limited to dekadal groupings. The same principles can be extended to any meaningful categorical structure, such as regional divisions or other domain-informed groupings, allowing this approach to generalise across spatial or thematic contexts where targeted transformations can enhance learning and reduce spurious variability.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

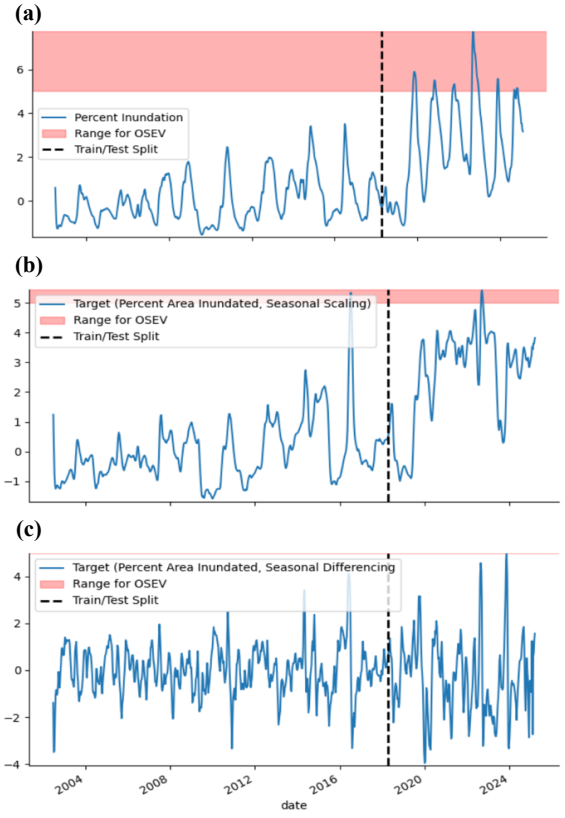

Figure 8. Transforming the target variable using seasonal deltas reduces the effective number of OSEV. **(a)** The unmodified standard scaled target data, with 35 OSEV (0.0% of target values in the 95th percentile for the training set, and 16.8% of varget values in the 95th percentile for the test set). **(b)** Seasonally scaled target data, with only 3 OSEV (0.9% of target values in the 95th percentile for the training set, and 14.9% of varget values in the 95th percentile for the test set). **(c)** Seasonally differenced target data, with 0 OSEV (3.3% of target values in the 95th percentile for the training set, and 9.1% of varget values in the 95th percentile for the test set).

4.3 Temporal dynamic thresholding: Transformer model

The temporal thresholding model in this study serves a dual purpose: it translates spatial probabilistic outputs from the dynamic spatial model into a single interpretable temporal signal, used to determine the number of flood-affected pixels to be filled, and it provides a simplified time series output that is more interpretable and actionable for humanitarian and governmental organisations operating in South Sudan. To achieve this, we implemented a transformer-based deep learning model, chosen for its capacity to learn long-range dependencies in sequential data via its multi-headed self-attention mechanism. This approach draws on the seminal

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

architecture introduced by Vaswani et al. (2017), in which each token in the sequence attends to all others through a scaled dot-product attention mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \qquad (14)$$

where Q, K, and V represent the query, key, and value matrices derived from the input embeddings, and d_k is the dimensionality of the key vectors. This mechanism allows the model to simultaneously learn relationships across multiple temporal scales. For instance, it can identify links between a strongly positive Indian Ocean Dipole phase several months prior, anomalously high rainfall over Lake Victoria weeks earlier, and elevated inundation over the White Nile basin in the present.

The final temporal model uses an encoder-only transformer architecture, which was selected over encoder-decoder variants after extensive testing due to its superior performance on fixed-length forecasting tasks and its robustness when training data are limited. The model predicts inundation levels for six dekads (roughly two months) into the future in a single forward pass, avoiding error accumulation inherent in recursive predictions. Training data consisted of overlapping sequences from before July 21st, 2018, with each sequence spanning the prior 36 dekads (approximately one year) thereby capturing both short- and long-term hydrometeorological drivers of flooding. This temporal split leaves all the OSEV arising from the unprecedented post-2019 flooding in South Sudan in the test set, meaning that the model will be trained exclusively on "typical" inundation cases without any exposure to extreme values. All target values were transformed using the differenced dekadal anomaly approach (Equation 12), which includes both first differencing and seasonal standardisation by dekad to remove trend and cyclicality. All predictor values were calculated as dekadal anomalies, without differencing (Equation 11). To reduce dimensionality, the predictor matrix was transformed using PCA, and the principal components explaining a total of 95% of the variance were selected (Equations 3-5).

To encode temporal information, we applied sinusoidal positional encodings, following Vaswani et al., where the positional encoding matrix $PE \in \mathbb{R}^{l \times d_{model}}$ is defined as:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{1000^{2i/d_{\text{model}}}}\right) \qquad (15)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{1000^{2i/d_{\text{model}}}}\right), \qquad (16)$$

for a given position $pos \in [0, l]$ (where l is the number of positional lookback sequences, equal to 36 in this case) and dimension

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

$i \in [0, d_{\text{model}}]$. These encodings are added to the input embeddings to provide the model with a sense of temporal order.

The network architecture consists of a transformer encoder with 8 attention heads, a feed-forward dimension of 128, and a dropout rate of 0.1. After encoding, the output is flattened and passed through two dense layers of sizes 100 and 50 with ReLU activations and dropout of 0.1, followed by a final dense output layer predicting the next six dekads. The model was trained over 500 epochs using the Adam optimiser with a learning rate of 0.0001, a batch size of 32, and early stopping based on validation loss with 10% of the training data (random split) used for validation. A total of 93 epochs were completed on an A100 GPU via Google Colab to accommodate the large spatio-temporal dataset. Once the data has been pre-processed, however, the model can be retrained on CPU hardware in under ten minutes.

We implemented a custom loss function, L_t, tailored to the forecasting objectives of this task. This function combines standard mean squared error (MSE) with penalties on both the directionality and the magnitude of cumulative predictions:

$$L_t = \underbrace{\text{MSE} \cdot \text{SignLoss}}_{\text{Directional accuracy}} + 0.1 \cdot \underbrace{\text{SumLoss}}_{\text{Total accuracy}}, \tag{17}$$

where

$$\text{SignLoss} = \mathbb{E}\left[\begin{cases} 20, & \text{if } \text{sign}(y_{\text{true}}) \neq \text{sign}(y_{\text{pred}}) \\ 1, & \text{otherwise} \end{cases}\right], \tag{18}$$

$$\text{SumLoss} = \mathbb{E}\left[\left(\sum_t y_{\text{true}}^t - \sum_t y_{\text{pred}}^t\right)\right], \tag{19}$$

and

$$\text{MSE} = \frac{1}{t}\sum_t \left(y_{\text{true}}^t - y_{\text{pred}}^t\right)^2 \tag{20}$$

To estimate uncertainty in forecasts, we used Monte Carlo dropout with 1000 forward passes during inference. Dropout layers were left active at test time, sampling from the model's approximate posterior to derive prediction distributions. The predictive mean μ_{M} and standard deviation σ_{M} for each time step were calculated across these samples. Confidence intervals were computed using the standard z-score formulation:

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

$$\mathrm{CI}_{\mathrm{lower}} = \mu_{\mathrm{M}} - z\sigma_{\mathrm{M}},\ \mathrm{CI}_{\mathrm{upper}} = \mu_{\mathrm{M}} + z\sigma_{\mathrm{M}}, \tag{21}$$

where $z = \Phi^{-1}(1 - \frac{1-\alpha}{2})$ and $\alpha = 0.95$ corresponds to a 95% confidence interval, giving $z \approx 1.96$.

The final output predictions were transformed back to the original inundation space by reversing the differenced dekadal anomaly transformation. For each prediction $\hat{y}_t$, this involved rescaling using the saved dekadal mean $\mu_{\mathcal{D}_t}$and standard deviation $\sigma_{\mathcal{D}_t}$ from the training set, and then cumulatively summing the predicted deltas over the forecast horizon:

$$\hat{y}_t = \Delta\tilde{y}_t \cdot \sigma_{\mathcal{D}_t} + \mu_{\mathcal{D}_t} + \hat{y}^{\mathrm{unscaled}}{}_{t-1} \tag{22}$$

This recursive reconstruction ensured that predictions aligned with the original target distribution, while preserving the advantages of delta transformation in reducing the frequency and magnitude of OSEV during training.

4.4 Dynamic spatial modelling: ConvLSTM model

The dynamic spatial modelling model in INFLOW-AI is designed to translate regional predictions of flooding extent into high-resolution, pixel-wise inundation maps (Figure 7). While the temporal thresholding transformer model focuses on detecting extreme flooding events and estimating total inundated area, including OSEV, the ConvLSTM (Convolutional Long Short-Term Memory) spatial model uses this estimate as a threshold to fill a probabilistic spatial prediction (Shi et al., 2015). Rather than targeting the prediction of extreme values themselves, the spatial model prioritises accurate, dynamic predictions of flood extent at a local scale.

Specifically, the ConvLSTM model predicts a binary inundation target derived from MODIS satellite data. The output consists of gridded values over the White Nile basin, where each 1 km² grid cell is assigned a probability of inundation ranging from 0 to 1. These probabilities are then thresholded based on the output of the temporal model: i.e. pixels are ranked in descending order of predicted inundation probability and sequentially filled until the total predicted inundated area is matched, producing a binary flood mask. This dynamic thresholding approach ensures consistency between the temporal and spatial models and allows the system to adapt to extreme events without overfitting to rare occurrences in training data.

ConvLSTM networks were chosen for this task due to their capacity to handle spatiotemporal data effectively. LSTMs are a class of recurrent neural networks (RNNs) that address the vanishing gradient problem by introducing gating mechanisms that regulate the

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

flow of information through time. ConvLSTM, proposed by Shi et al. (2015), extends the traditional LSTM architecture by replacing fully connected layers with convolutional layers in both the input-to-state and state-to-state transitions, enabling the network to retain spatial structure while learning temporal dependencies. This architecture has shown promise in precipitation nowcasting and spatiotemporal sequence forecasting.

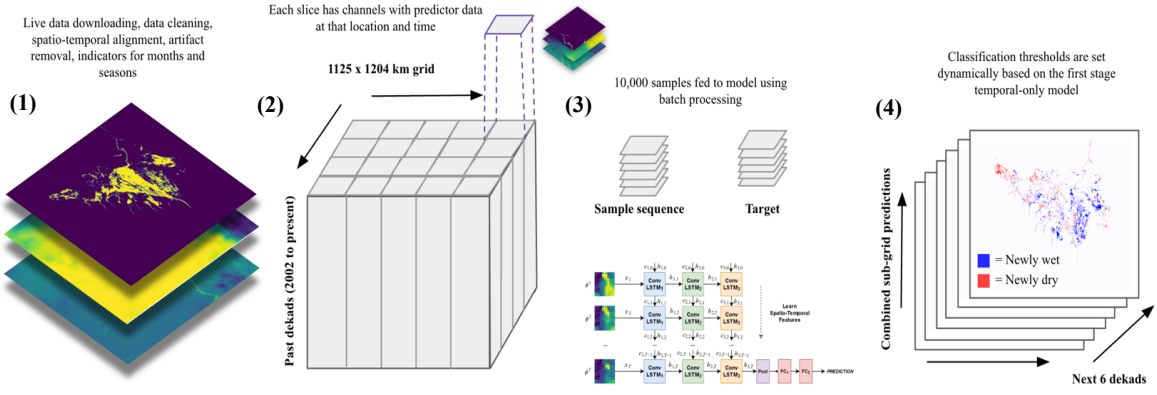

Figure 9. Workflow for dynamic spatial modelling stage of INFLOW-AI. (1) Spatio-temporal and spatial data is processed and combined into 4D array with channels as predictors. (2) 4D array is split into 64x64 overlapping sub-grids. (3) Model is trained on sequences of 6 dekads (2 months) to predict a sequence of 6 dekads using a NVIDIA A100 Tensor Core GPU. Focal loss is used to dynamically weight imbalanced data so that predictions for newly flooded and newly dry pixels are given more attention by the model than the regions with no change that dominate the spatial dataset. (4) Predictions at the sub-grid level are cropped to remove prediction irregularities at the image edges along the overlapping border and are recombined into a single map of the study area, using temporal thresholding from the first stage temporal model to determine the size of the inundated region to fill.

A ConvLSTM model determines the future state of a given grid cell using the inputs and past states of its local neighbours. This is achieved by using a convolutional operator in the state-to-state and input-to-state transitions, replacing the matrix multiplication used in LSTM models. For a ConvLSTM cell index i; where X_t is the input data at time t, $\mathcal{H}_t$ is the hidden state at time t, W_{xi} is a convolution filter connecting the input X_t to the input gate i_t, $\circledast$ is a convolution operation, $\odot$ is the Hadamard product, and σ is a sigmoid activation function. The input gate:

$$i_t = \sigma(W_{xi} \circledast X_t + W_{hi} \circledast \mathcal{H}_{t-1} + W_{ci} \odot \mathcal{C}_{t-1}) \quad (23)$$

decides how much of the new input should influence the current memory, while the forget gate:

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

$$f_t = \sigma(W_{xf} \circledast X_t + W_{hf} \circledast \mathcal{H}_{t-1} + W_{cf} \odot \mathcal{C}_{t-1}) \quad (24)$$

controls what information from the previous cell state $\mathcal{C}_{t-1}$ should be discarded (thereby avoiding the vanishing and exploding gradient problems present in other RNNs). The candidate cell state:

$$\mathcal{C}_t = f_t \odot \mathcal{C}_{t-1} + i_t \odot \tanh\,(W_{xc} \circledast X_t + W_{hc} \circledast \mathcal{H}_{t-1} \quad (25)$$

proposes new information to be added and is updated by combining this candidate with retained past memory from the previous cell state $\mathcal{C}_{t-1}$. Finally, the output gate:

$$o_t = \sigma(W_{xo} \circledast X_t + W_{ho} \circledast \mathcal{H}_{t-1} + W_{co} \odot \mathcal{C}_t) \quad (26)$$

filters the updated cell state $\mathcal{C}_t$ to produce the hidden state:

$$\mathcal{H}_t = o_t \odot \tanh\,(\mathcal{C}_t) \quad (27)$$

which is passed onto the next time step and used for prediction.

The INFLOW-AI spatial model uses a 2D ConvLSTM to forecast local inundation for the next six dekads, drawing on spatiotemporal input data from the past six dekads for local rainfall, local soil moisture, and local inundation. Additionally, static spatial predictors for elevation and sub-basin attributes, with the dimensionality of the sub-basin attributes reduced using PCA to retain only the first principal component (Figure 5). These datasets were processed into 4D arrays (time × height × width × channels), with each array representing one input sample.

To increase the effective size of the training set and enable the model to generalise across different hydrological regimes, the spatial input arrays were divided into overlapping 64×64 sub-grids. Each training sample thus corresponded to a spatiotemporal slice of six dekads over a 64×64-pixel region. The model was trained only on pre-July 21, 2018 data, while post-2019 data, which includes almost all the OSEV, was reserved for testing.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

The model architecture consisted of three ConvLSTM2D layers with 64, 128, and 128 filters, each using 3×3 kernels, a stride of 1×1, and 'same' padding. The final ConvLSTM2D layer produced the full six-dekad forecast sequence. Batch normalization and dropout layers were applied for regularization. Two subsequent Conv2D layers with 3×3 and 1×1 kernels, 'same' padding, and sigmoid activation refined spatial features and generated the single-channel probabilistic prediction map.

To handle extreme class imbalance in the inundation data, where most pixels are non-flooded, a focal loss function was used:

$$FL(p_t) = -\alpha_t \cdot (1 - p_t)^{\gamma} \cdot \log(p_t), \tag{28}$$

where the model's predicted probability for the correct class $p_t = y_{\text{true}} \cdot y_{\text{pred}} + (1 - y_{\text{true}})(1 - y_{\text{pred}})$ and $\alpha_t = y_{\text{true}} \cdot \alpha + (1 - y_{\text{true}})(1 - \alpha)$. A focussing parameter of $\gamma = 3.0$ was chosen to increase the model's attention to difficult-to-classify pixels and $\alpha = 0.95$ was used to assign greater weight to the minority (i.e. inundated) class.

Training was conducted using an NVIDIA A100 Tensor Core GPU for up to 100 epochs, with early stopping applied when the validation loss did not improve for 5 consecutive epochs. The model typically converged well before the maximum epoch limit, and training required approximately three hours to complete. At inference, sub-grid predictions were stitched back into full-domain inundation maps. Edge artifacts from overlapping regions were cropped along the 8-pixel overlapping border prior to recombination. The final binary inundation mask was generated by applying the threshold derived from the temporal model's predicted flood extent to each pixel (i, j):

$$\text{FloodMask}(i,j) = \begin{cases} 1, & \text{if } (i,j) \in \text{top } N \text{ of } P(i,j) \\ 0, & \text{otherwise} \end{cases}, \tag{29}$$

where $P(i,j)$ is the predicted probability of inundation from the dynamic spatial modelling and N is the number of pixels corresponding to the total predicted flood area by the first stage temporal thresholding.

For deployment, both the temporal and spatial models were retrained using all available historical data, including previously held-out OSEV, to maximise learning and performance in real-time operations.

5. Model Selection and Hyperparameter Tuning

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

5.1 Model selection

To identify the most effective approach for temporal thresholding in the first stage of INFLOW-AI, a broad comparison of model architectures was conducted. Each model used the same set of predictors and a lookback period of 36 dekads, and all were evaluated on both the full test set and on OSEV. Models were tested in increasing order of complexity: a baseline persistence model, a linear interpolation model, linear regression, lasso regression, random forest, feed-forward neural networks (FFNNs) with no temporal structure, and finally transformer models. All models were tested using both the raw target values and the differenced dekadal anomalies introduced in Equation (12).

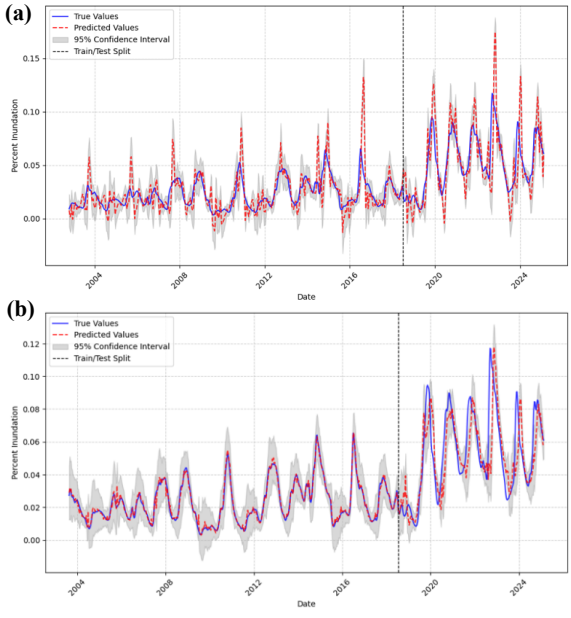

Figure 10. **(a)** Predictions from the persistence model $\Delta\tilde{y}_t$ predicting the differenced dekadal anomaly on the training and test set for the sixth dekad prediction, illustrating high volatility of predictions that contribute to it not being an appropriate model. **(b)** Predictions from the transformer model $\Delta\tilde{y}_t$ predicting the differenced dekadal anomalay on the training and test set for the sixth dekad, with less volatlit, but still high ability to detect OSEV.

The persistence model served as a baseline, projecting the most recent inundation extent forward. While simplistic, this model is essential for benchmarking; if a candidate model fails to outperform persistence, it cannot be justified for operational forecasting. Confidence intervals for this baseline were constructed using the standard deviation of the training set, following the method in (20). The linear interpolation model was built by extrapolating the linear trend from the past six dekads and forcing it through the most

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

recent data point. The lasso model's L1 regularization parameter was tuned to its optimal value through cross-validation.

Table 2 summarizes model performance across the six-dekad prediction window. The transformer model predicting the differenced dekadal anomaly ($\Delta\tilde{y}_t$) emerged as the top-performing model overall. It achieved the lowest mean absolute error (MAE) and mean squared error (MSE) on both the general test set and OSEV-specific metrics. Compared to the persistence model forecasting raw inundation (y), the transformer model ($\Delta\tilde{y}_t$) reduced test MAE by 50%, test MSE by 78%, OSEV MAE by 19%, and OSEV MSE by 41%. Even when compared to the persistence model applied to the same differenced target ($\Delta\tilde{y}_t$), the transformer still outperformed it by 26% for test MAE, 59% for test MSE, and 23% for OSEV MSE.

Model	Test		OSEV	
	MAE	MSE	MAE	MSE
Persistence ($\Delta\tilde{y}_t$)	0.0147	0.0448	**0.0262**	0.1271
Persistence (y)	0.0176	0.0501	0.0318	0.1462
Linear interpolation ($\Delta\tilde{y}_t$)	0.0274	0.1451	0.0482	0.4089
Linear interpolation (y)	0.0196	0.0835	0.0414	0.2727
Linear regression ($\Delta\tilde{y}_t$)	0.0260	0.1123	0.0350	0.1737
Linear regression (y)	0.0258	0.1113	0.0351	0.1758
Lasso ($\Delta\tilde{y}_t$)	0.0127	0.0325	0.0282	0.1114
Lasso (y)	0.0127	0.0324	0.0285	0.1138
Random forest ($\Delta\tilde{y}_t$)	0.0135	0.0362	0.0293	0.1180
Random forest (y)	0.0134	0.0359	0.0267	0.1199
FFNN ($\Delta\tilde{y}_t$)	0.0147	0.0447	0.0338	0.1602
FFNN (y)	0.0146	0.0444	0.0342	0.1638
Transformer ($\Delta\tilde{y}_t$)	**0.0117**	**0.0282**	0.0267	**0.1034**
Transformer (y)	0.0545	0.3637	0.0934	0.8819

Table 2. Mean absolute error (MAE) and mean squared error (MSE) for different model configurations predicting the percentage of inundated area over the White Nile basin six dekads in the future, with best model performance for each column bolded. $\Delta\tilde{y}_t$ represents predictions on targets that have been trasformed using differenced dekadal anomolies and y represents predictions on raw targets (percentage of inundated area). Predictions on $\Delta\tilde{y}_t$ are first reconverted back to raw inundation values using (21) before error is calculated. All MSE values are multiplied by 100 to convert raw values from percentages for ease of presentation.

One exception was a slight edge in MAE for OSEV at the sixth dekad held by the persistence model ($\Delta\tilde{y}_t$); however, this came at a cost of extremely high volatility. Because the model naively projected the most recent differenced anomaly across the next six dekads,

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

its undifferenced and de-seasonalised predictions frequently overshot actual changes in inundation extent, resulting in disproportionately high MSE and unstable forecasts, especially for extreme events (Figure 10). These outcomes confirm that despite this occasional short-term advantage, the persistence model's oversensitivity to recent extremes renders it unsuitable for robust forecasting.

The next most competitive model after the transformer was lasso regression, which also benefited from using differenced dekadal anomalies but still fell significantly short of the transformer's performance. Across both raw and transformed targets, the transformer demonstrated clear dominance.

These results validate the selection of a transformer-based architecture using seasonally differenced dekadal anomalies. Not only does this configuration outperform simpler alternatives, but it also justifies the added complexity by substantially improving forecast accuracy, especially in predicting OSEV; a key goal of the INFLOW-AI framework.

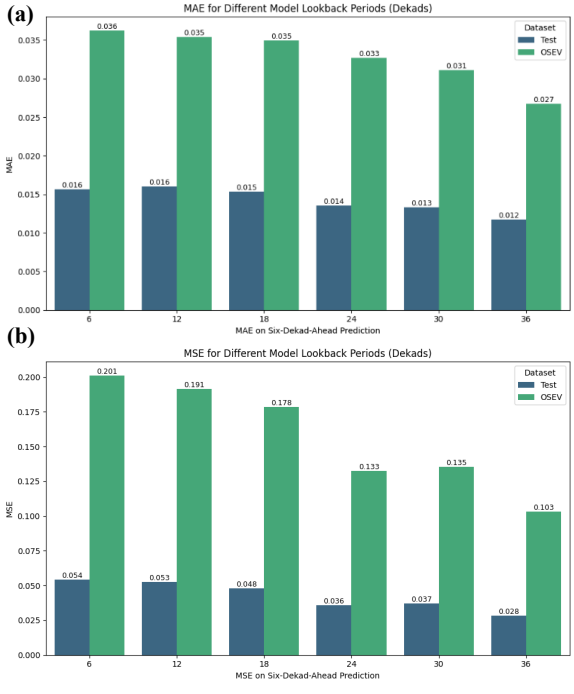

Figure 11. **(a)** MAE and **(b)** MSE (multiplied by 100) for different lookback periods of the transformer model $\Delta\tilde{y}_t$ predicting the differenced dekadal anomaly for the sixth dekad prediction.

5.3 Hyperparameter tuning

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Hyperparameter tuning for the INFLOW-AI framework was conducted separately for both the temporal thresholding model and the dynamic spatial modelling model. For the temporal thresholding model, training times were relatively short, enabling a comprehensive grid search over a range of hyperparameters. The search focused on optimising the learning rate, number of attention heads in the multi-headed self-attention mechanism, dimensionality of the feed-forward layers, and the number of neurons in the fully connected layers. Performance was assessed using the validation set, and hyperparameter choices were selected based on improvements in MAE and MSE for the post-2019 test set and OSEV events.

In contrast, the dynamic spatial modelling model required significantly longer training times. As a result, hyperparameter tuning was restricted to a narrower set of variables. Small-scale experiments were conducted on a reduced training set of 1,000 samples over five epochs to explore the effects of kernel size and dropout rate on performance. These limited tests helped prevent overfitting while ensuring a reasonable degree of regularization in the final model.

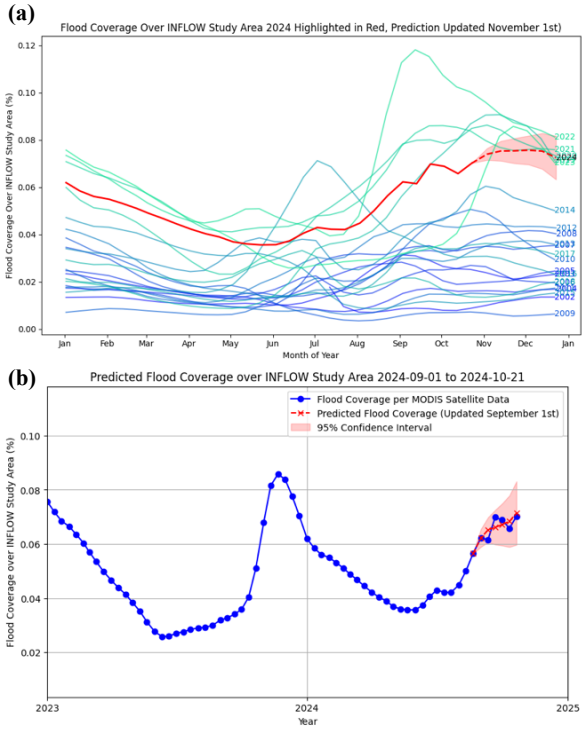

Figure 12. **(a)** Real-time predictions made using the deployed INFLOW-AI model during the 2024 flood season with 95% confidence interval , compared to historic inundation extents for the same time period (MODIS satellite data). Forecast produced on November 1st, 2024 and distributed to stakeholders at technical meetings. **(b)** Retrospective performance analysis of INFLOW-AI performance during the real-time prediction period.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Across both models, one of the most influential hyperparameters was the lookback period; the temporal window of past data used to inform predictions. For the transformer-based temporal model, longer lookback periods consistently led to better performance. Multiple configurations were tested, with lookbacks ranging up to a maximum of 36 dekads (one calendar year). As shown in Figure 11, performance improvements were observed on both the full test set and OSEV subset as the lookback period increased. This suggests that the model benefits from having access to a full year of hydrometeorological history, which likely contains seasonal and antecedent signals relevant for flood prediction. Consequently, the maximum tested lookback period of 36 dekads was adopted in the final transformer model configuration. To limit model complexity and maximise training data availability, no additional dekads were used.

Region (Capital)	MAE	MSE
Central Equatoria (Juba)	**0.0099**	**0.0224**
Eastern Equatoria (Torit)	**0.0073**	**0.0151**
Jonglei (Bor)	0.0712	0.9679
Lakes (Rumbek)	0.0290	0.1583
Northern Bahr el Ghazal (Aweil)	0.0159	0.0965
Unity (Bentiu)	0.0574	0.6958
Upper Nile (Malakal)	0.0381	0.2837
Warrap (Kuajok)	0.0428	0.4706
Western Bahr el Ghazal (Wau)	**0.0027**	**0.0034**
Western Equatoria (Yambio)	**0.0003**	**0.0000**
Abyei Area (Abyei)	0.0608	1.7940

Table 3. MAE and MSE (multiplied by 100) for different regions of South Sudan predicted by the region-specific transformer model $\Delta\tilde{y}_t$ predicting the differenced dekadal anomaly for the sixth dekad prediction. Capital states of regions provided forgeographical context. Scores better than the respective scores from the global South Sudan model are bolded.

Due to computational constraints, lookback period tuning was not performed as extensively for the ConvLSTM model, though a default 6-dekad input and output sequence length was maintained for consistency with the operational prediction window.

6. Performance

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

6.1 Temporal thresholding model performance

The temporal thresholding model was evaluated across multiple dimensions to assess its utility as a forecasting tool for flood extent prediction in South Sudan. Live model outputs were deployed during the 2024 and 2025 flood seasons and disseminated in real time to humanitarian organizations including the Red Cross, MSF, and OCHA. Forecasts, shared in the form of CSVs and visual time series graphs, were presented and explained through weekly technical briefings to key stakeholders. Figure 12(a) presents an example of a distributed real-time forecast compared to historical inundation patterns, while Figure 12(b) retrospectively compares predicted (red) and observed (blue) inundation extents for the 6-dekad prediction horizon. Notably, all forecasted values during the operational prediction period fell within the 95% confidence interval generated via Monte Carlo dropout, as in Gal and Ghahramani (2016), confirming the model's reliability under uncertainty.

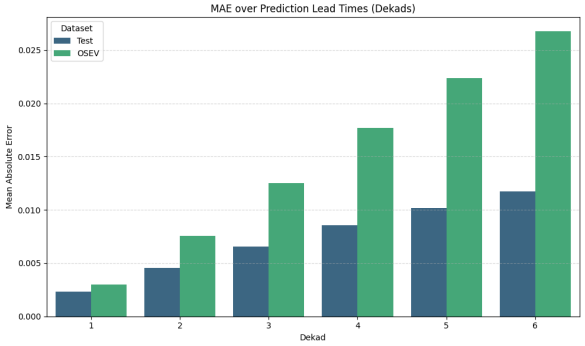

Figure 13. MAE for different lead times of the transformer model $\Delta\tilde{y}_t$ predicting the differenced dekadal anomaly for each dekad prediction.

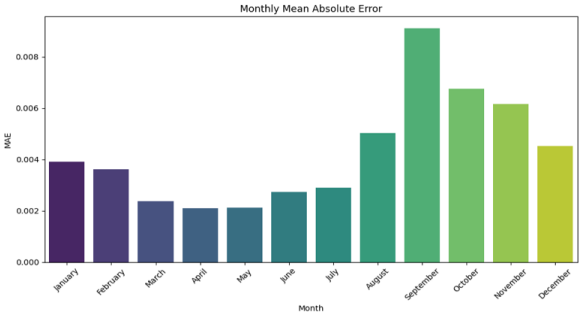

Figure 14. MAE for different months of the transformer model $\Delta\tilde{y}_t$ predicting the differenced dekadal anomaly for the sixth dekad prediction.

Performance metrics on the hold-out test set (post-July 21st, 2018) are reported in Table 2. The transformer model predicting the differenced dekadal anomaly achieved a mean absolute error (MAE) of 0.0117 and mean squared error (MSE) of 0.0282 at the 6-dekad lead time. This represents a 33.5% and 43.7% improvement, respectively, over the baseline persistence model predicting raw inundation extent. It also outperformed the next best model, the lasso, by 7.9% (MAE) and 13.0% (MSE). For OSEV predictions specifically, the transformer model achieved an MSE of 0.1034, which is 29.3% better than the persistence baseline and 7.2% better than the lasso model. While the persistence model predicting differenced anomalies marginally outperformed the transformer in MAE for OSEV on the sixth dekad, its high volatility caused by propagating recent anomaly spikes resulted in much higher MSE and frequent overestimations (Figure 10).

Temporal error dynamics were consistent with expectations for a forecast model. As shown in Figure 13, prediction accuracy declined

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

with longer lead times. The model had near-zero error for one-dekad-ahead forecasts (MAE = 0.0030, MSE = 0.0022), but errors increased at the 6-dekad horizon (MAE = 0.0267, MSE = 0.0010). Forecasts beyond the sixth dekad proved unreliable and were excluded from further use.

Model performance by month (Figure 14) revealed seasonal variability in accuracy. Prediction errors peaked during the flood-intensive month of September, reflecting high inundation variability, but even then, MAE remained under 1%. In contrast, accuracy was highest from March to July, when inundation variance was minimal.

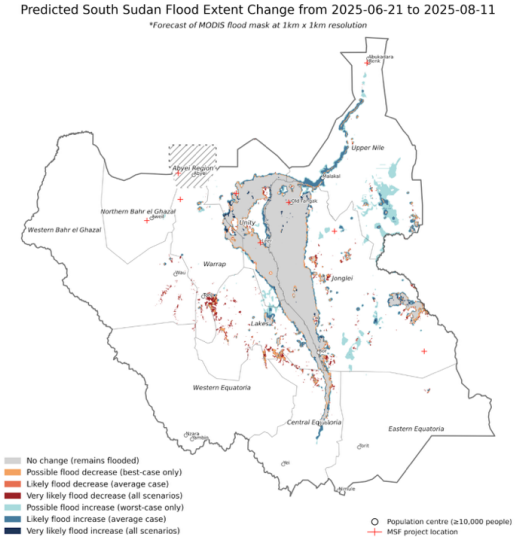

Figure 15. Example spatial prediction at the 1km x 1km resolution made 6 dekads in advance for the 2025 flood season. "Best-case scenario" is calculated by filling predictions from the dynamic spatial modelling stage in descending order of predicted inundation probably until the lower bound of the 95% confidence interval from the temporal thresholding stage is reached. "Worst-case scenario" is calculated by filling predictions from the dynamic spatial modelling stage in descending order of predicted inundation probably until the upper bound of the 95% confidence interval from the temporal thresholding stage is reached. "Average-case scenario" is calculated by filling predictions from the dynamic spatial modelling stage in descending order of predicted inundation probably until mean prediction from the temporal thresholding stage is reached. Major population centres and Médecins Sans Frontières (MSF) locations are shown on the deployed map to aid in operational decsion-making.

Spatially, performance varied by administrative region (Table 3). The model was most accurate in higher elevation areas with infrequent flooding. However, forecasts over individual states were generally less accurate than those over the entire White Nile catchment. This is likely due to misalignment between political boundaries and hydrological patterns, reducing the model's ability

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

to detect coherent spatiotemporal trends within arbitrary administrative divisions. These results suggest that regional-level temporal forecasts should be deprioritized in favor of outputs from the second-stage dynamic spatial modelling model for finer spatial detail, as was ultimately done for the deployed model.

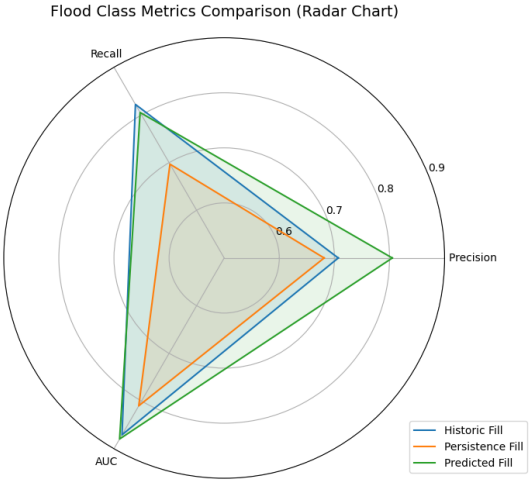

Figure 16. Comparison of key classification metrics for the ConvLSTM model (predicted fill) with the historic average and the persistence model.

Finally, to validate the superiority of the differenced anomaly modelling approach, a Wilcoxon signed-rank test was conducted on paired model configurations in Table 2 (e.g., transformer vs. transformer, raw vs. differenced target). The test showed statistically significant improvement ($W = 0.0$, $p = 0.03125$) for models using the differenced dekadal anomaly when predicting OSEV. Interestingly, this advantage did not extend to the full test set, indicating that differencing the target is uniquely beneficial for capturing extreme events.

6.2 Dynamic spatial modelling model performance

To ensure that the dynamic spatial modelling model (Figure 15) was functioning as a dynamic spatio-temporal predictor rather than a static risk map, a key evaluation focused on whether it captured variation in flood risk across time for a given location. A truly effective forecasting model should generate different predictions for different time periods, even within the same sub-grid region. To verify this, the spatial variance in ConvLSTM model predictions across time for identical sub-grid regions was calculated using the same out-of-sample test set employed for evaluating the temporal thresholding model (beginning on July 21st, 2018). If the model were simply outputting static probabilities based on historical inundation frequency, the variance in predicted values across time

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

would approach zero. However, the average spatial variance in 6-dekad-ahead predictions was 0.1406, very close to the true spatial variance at the 6th dekad (0.1463), indicating the model effectively captures meaningful temporal variation at local spatial scales.

Metric	Historic Fill	Persistence Fill	Predicted Fill
Precision (Flood)	0.71	0.68	**0.95**
Recall (Flood)	**0.82**	0.70	0.81
F1-Score (Flood)	0.76	0.69	**0.81**
Accuracy	0.90	0.88	**0.93**
Macro Avg. F1	0.85	0.81	**0.88**
Weighted Avg. F1	0.90	0.88	**0.93**
Precision (Total)	0.71	0.68	**0.81**
Recall (Total)	**0.82**	0.70	0.81
AUC	0.87	0.81	**0.88**

Table 4. Classification metrics for the INFLOW-AI dynamic spatial modelling models and the ConvLSTM model predictions. The best score for each metric is bolded.

Two baseline models were used to provide context for performance evaluation. The first is a historic fill model, which ranks pixels based on their historical inundation frequency from the training set and fills them sequentially until the forecasted total flood extent is reached, essentially a static reconstruction of past flood likelihoods (Figure 2). The second baseline is a persistence fill model, which projects the most recent observed inundation extent from the training data forward to the 6-dekad prediction horizon. Although technically dynamic, this model is incapable of forecasting increases or reductions in flood extent and serves as a critical benchmark; outperforming it is essential to justify the predictive value of the spatial model.

Given that most areas in the forecast domain are not inundated during any given dekad, naïvely high accuracy scores can be misleading. To address this, a wide array of classification metrics were used to isolate performance on flooded pixels. These included: precision (the fraction of predicted flooded pixels that were correct), recall (the fraction of actual flooded pixels correctly predicted), F1-score (the harmonic mean of precision and recall), and AUC (Area Under the Receiver Operating Characteristic Curve), which quantifies the likelihood that the model will rank a randomly selected flooded pixel higher than a randomly selected unflooded pixel. The ROC curve itself is created by plotting the true positive rate against the false positive rate across various thresholds; the AUC summarizes this curve as a single scalar, where 1.0 indicates perfect ranking ability.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

As shown in Table 4 and Figure 16, the ConvLSTM-based dynamic spatial modelling model outperformed or matched both baselines across all key metrics. Notably, the model demonstrated substantially higher precision in detecting flooded pixels, indicating it predicted very few false positives compared to the baseline approaches. This reflects a strong capacity for spatially precise flood detection, which is critical for localized decision-making.

Regionally, the spatial model provided significantly higher resolution than the temporal thresholding model, confirming the value of the two-stage forecasting system. While the temporal model captures broader-scale trends and extreme events, the spatial model excels at interpolating localized flood dynamics. As expected, the greatest errors occurred along the boundaries of floodplains; transitional zones where inundation presence fluctuates rapidly (Figure 17). Nevertheless, even in these challenging zones, the model maintained strong predictive accuracy at a 1km spatial resolution, demonstrating its capability for fine-scale forecasting. Compared to the temporal model's performance at the state level, the dynamic spatial model is clearly more suitable for local-scale flood prediction tasks.

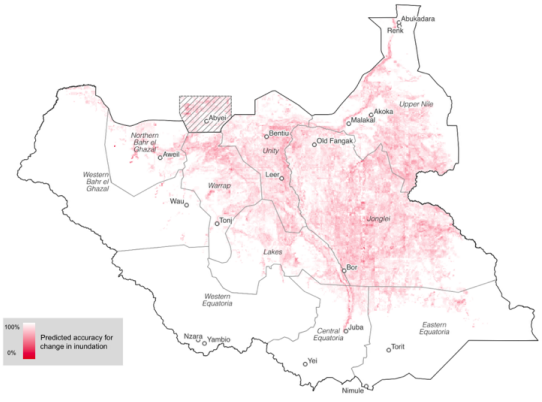

Figure 17. Accuracy for change detection for the out-of-sample test set (i.e. the average accuracy for correctly detecting either an increase or decrease in the inundation extent for each 1km x 1km region.

7. Discussion

7.1 Findings

The results of this study demonstrate the utility of two core advancements in the development of the INFLOW-AI model: (1) the use of differenced seasonal anomalies as a predictive target, and (2) a two-stage model architecture combining temporal thresholding and dynamic spatial modelling. INFLOW-AI's performance in predicting out-of-sample extreme values (OSEV) was consistently

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

strong and outperformed all alternative model configurations tested, confirming the effectiveness of both design choices.

Differenced seasonal anomalies enabled the model to better focus on seasonal deviations in the target variable without needing to explicitly learn seasonal patterns. This formulation shifts the prediction target from raw values to changes, which simplifies the learning task, particularly for time series where extreme values may appear in the test set but are not present in the training data. The cumulative differenced predictions across lead times can be summed and converted back to raw values, allowing the model to reproduce the original target variable while avoiding the need to extrapolate from a limited training distribution. This is particularly impressive given that the model was not provided any information about the existence of post-2019 extreme values during training.

The two-stage model architecture further enhanced performance by separating the modelling of temporal and spatial dynamics. The first-stage temporal thresholding model focuses specifically on detecting the timing and magnitude of OSEVs, while the second-stage dynamic spatial model uses this threshold to guide the spatial extent of inundation. This structure allows the temporal model to specialize in detecting shifts in flood magnitude, and the spatial model to focus on predicting local dynamic changes in inundation probability. The result is a highly effective division of labour between detecting national-scale extremes and resolving localized spatial risk, enabling more accurate and interpretable predictions. This approach may be beneficial for forecasting other spatio-temporal hydrometeorological phenomena that require sensitivity to OSEVs, including rainfall, cyclones, and heatwaves.

In terms of forecast skill, model performance was unsurprisingly better for shorter lead times and declined with increasing forecast horizon. Experimental tests that expanded the lead time of the temporal thresholding model beyond the six dekads used in the final model produced prediction intervals that were larger than the standard deviation of the data. This rendered those extended forecasts not particularly useful and reinforced the decision to adopt a 6-dekad lead time for operational deployment.

The first-stage dynamic thresholding model proved highly effective at predicting OSEVs in the test set despite not having been exposed to any similar inundation conditions during training. Meanwhile, the second-stage dynamic spatial modelling model outperformed both a persistence-based baseline and a constant-fill baseline based on historic inundation frequencies. The ability of the spatial model to capture local variance in flood risk responsive to local spatio-temporal variables was likely supported by the decision to divide the study area into overlapping 64×64 km sub-grids for training. This approach multiplied the number of training samples per time period by over 330 and exposed the model to a broader variety of local geographic and climatic conditions. It also improved the computational efficiency of the ConvLSTM architecture by reducing the number of neurons required due to the smaller input dimensions. In contrast, training on the entire 1125×1204 km study area as a single input tended to result in relatively static

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

predictions for each region, as the model defaulted to learning coarse regional averages instead of local spatial relationships.

During deployment, the temporal model has proven most useful for communicating high-level, country-wide forecasts. However, efforts are ongoing to make the spatial predictions more accessible to humanitarian agencies, given that localized spatial information is critical for operational decision-making. This includes decisions about where to deploy food aid, increase dyke heights around populated areas, or anticipate where roads might become impassable due to flooding. Future iterations of INFLOW-AI are expected to expand the use of spatial predictions to better support these applications.

Interpretability analysis using Shapley values (Lundberg & Lee, 2017; Appendix B) indicated that the model's predictions aligned well with established hydrological knowledge of the region. This suggests that the model was detecting meaningful correlates of inundation, rather than overfitting or relying on persistence patterns derived from recent inundation events.

Finally, in terms of cost-effectiveness, INFLOW-AI offers significant benefits. For aid agencies, the marginal cost-effectiveness of a new forecasting model must be weighed against the next best alternative. INFLOW-AI's configuration significantly outperformed other tested models and enabled both OSEV detection and high-resolution dynamic spatial modelling. Since the model is already deployed on JASMIN and has very low ongoing training or inference costs, it represents a cost-effective solution for improving the timeliness and accuracy of early warning systems and early action decision-making for flood responses in South Sudan.

7.2 Limitations

While INFLOW-AI demonstrates strong predictive performance, several limitations remain that constrain both its interpretability and operational reliability. Fundamentally, the model is clearly lacking a complete set of input data required to make perfect predictions. This limitation is particularly salient in the context of post-2020 flooding in South Sudan, where variability in inundation extent may be substantially influenced by human factors not captured in the available datasets. It may be possible that planned dam releases upstream in Uganda played a significant role in driving flood dynamics during this period. This interpretation is supported by documented instances where release schedules from the Nalubaale and Kiira dams altered downstream flow independently of preceding rainfall, effectively decoupling the expected positive correlation between rainfall and inundation and sometimes even reversing it (The EastAfrican, 2024; see Appendix C), with several of these deviations plausibly attributable to unobserved upstream water management activities. The spatial data is also severely limited by the accuracy of the MODIS flood masks, which may not be perfectly accurate given the high density of vegetation in the swamp areas of the White Nile Basin (Lin et al., 2019).

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

In addition to gaps in explanatory coverage, INFLOW-AI is subject to the well-documented problem of technical debt in machine learning systems, particularly as it relates to data dependencies. As noted by Sculley et al. (2015), maintaining machine learning systems in real-world deployments often incurs substantial long-term costs, with data dependencies proving especially burdensome. INFLOW-AI's performance depends on the availability of multiple upstream datasets with consistent formatting and methodological stability over time. These dependencies are more difficult to detect and untangle than code dependencies, and failures can significantly impair performance. For instance, the MODIS satellite product used to generate the binary flood masks that serve as the model's training target is scheduled for retirement in the coming years. This will necessitate model retraining on a new satellite product, introducing both technical complexity and the potential for performance degradation.

Further limitations were identified in how the model's outputs align with end-user needs. During deployment, humanitarian agencies operating in South Sudan frequently requested forecasts of predicted water levels or flood depth, which was not provided by the model. Because INFLOW-AI was trained using MODIS-derived binary flood masks, it provides no estimates of water depth or volume, which can be crucial for planning decisions such as dyke construction, evacuation protocols, or transport logistics. The spatial resolution of the MODIS flood masks also posed challenges, as they are relatively coarse compared to the high-resolution (e.g., 1 km × 1 km) flood maps that humanitarian actors often seek. When selecting a training target, the MODIS flood masks were chosen due to their relatively higher accuracy and consistency compared to alternative products, but this choice involved a trade-off between resolution and data quality that continues to affect the model's practical utility.

7.3 Recommendations for future work

Given the promising results and operational deployment of INFLOW-AI, several avenues for future work are recommended to enhance the model's utility, robustness, and relevance for humanitarian decision-making. First, efforts should be made to increase the practical applicability of the model's outputs by integrating its predicted flood inundation maps with auxiliary datasets such as infrastructure maps (e.g. hospitals, schools, roads) from OpenStreetMap (OSM), as well as gridded population datasets. Such integration would allow responders to more easily assess the potential impacts of predicted floods on critical infrastructure and population centres. However, population data should be used cautiously in contexts such as South Sudan, where large numbers of refugees, internally displaced persons (IDPs), and semi-permanent settlements, coupled with logistical difficulties in collecting survey data, limit the accuracy and completeness of available population datasets.

Second, the model's value to humanitarian agencies would be significantly enhanced by associating flood predictions with estimated

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

water depth. As discussed in the limitations, INFLOW-AI is currently trained on a binary flood mask derived from MODIS satellite data, which provides no information on floodwater depth or volume. Future work could explore the possibility of linking MODIS-derived inundation data with local physical water level measurements or Surface Water and Ocean Topography (SWOT) satellite data, thereby enabling the derivation of statistical relationships between predicted inundation and actual water depth. This would address a frequently cited information gap among humanitarian actors and would directly support operational decision-making tasks such as determining dyke height requirements or planning evacuation routes.

Additionally, INFLOW-AI's current live deployment in South Sudan presents a strong foundation for expanding its operational integration and usage. The model is already in use for real-time early warning and has been incorporated into a Simplified Early Action Protocol (sEAP) to enable the South Sudan Red Cross to trigger financial support from the International Federation of Red Cross and Red Crescent Societies' (IFRC) Disaster Response Emergency Fund (DREF). Future work should focus on embedding the model more systematically into broader humanitarian coordination systems. For instance, INFLOW-AI should be integrated into the World Food Programme's (WFP) dashboards and forecasting toolkits, and its outputs used to generate regular bulletins and situation reports. Doing so would help ensure that predictive analytics directly inform early action decision-making at scale.

8. Conclusions

This study presents INFLOW-AI v2.1, a machine learning framework applied to predicting out-of-sample extreme values (OSEV) in South Sudan seasonal flood extents. It demonstrates the utility of two core advancements of this model: (1) the use of differenced seasonal anomalies as a predictive target, and (2) a two-stage model architecture combining temporal thresholding and dynamic spatial modelling. These architectural decisions enabled the model to outperform alternative model configurations evaluated in this study when predicting OSEV, confirming the effectiveness of the model architecture. The integration of INFLOW-AI into operational systems demonstrate its potential for informing early action decisions for flood preparedness in South Sudan. Future work should focus on enhancing model interpretability, linking predicted inundation extent to water depth, and incorporating exposure data for more comprehensive risk assessments. These findings highlight the potential for machine learning approaches to address forecasting challenges in data-scarce, hydrologically complex regions and suggest that similar architectures may be transferable to other flood-prone systems.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Appendix A: HydroTLAS attributes used in model

Variable	Description	Source	Citation	PC1 Loadings
tmp_dc_syr	Average annual temperature in sub-basin, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al, 2005	0.2319
tmp_dc_s11	Average monthly temperature in sub-basin for November, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.2311
tmp_dc_s03	Average monthly temperature in sub-basin for March, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.2276
tmp_dc_s10	Average monthly temperature in sub-basin for October, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.2237
tmp_dc_s08	Average monthly temperature in sub-basin for August, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.2232
tmp_dc_s09	Average monthly temperature in sub-basin for September, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.2227
tmp_dc_smn	Annual minimum temperature in sub-basin, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.2204
tmp_dc_s04	Average monthly temperature in sub-basin for April, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al, 2005	0.2086
tmp_dc_s12	Average monthly temperature in sub-basin for December, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.1989
tmp_dc_s02	Average monthly temperature in sub-basin for February, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.1959
tmp_dc_s01	Average monthly temperature in sub-basin for January, in degrees Celsius (x10)	WorldClim v1.4	Hijmans et al., 2005	0.1722
inu_pc_slt	Percent of sub-basin covered by the inundation extent long-term maximum	GIEMS-D15	Fluet-Chouinardetal, 2015	0.1136
wet_pc_sg1	Percent of sub-basin that is well-defined wetlands	GLWD	Lehner & Döll, 2004	0.1121
wet_pc_sg2	Percent of sub-basin that is well-defined wetlands or uncertain/complex wetland mosaics	GLWD	Lehner & Döll, 2004	0.1076
wet_pc_s04	Percent of sub-basin that is freshwater marsh or floodplain	GLWD	Lehner & Döll, 2004	0.1058
wet_pc_ug1	Percent of total watershed upstream of sub-basin pour point that is well-defined wetlands	GLWD	Lehner & Döll, 2004	0.0964

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

inu_pc_ult	Percent of total watershed upstream of sub-basin pour point that is covered by the inundation extent long-term maximum	GIEMS-D15	Fluet-Chouinardetal, 2015	0.0948
inu_pc_smx	Percent of sub-basin that is covered by the inundation extent annual maximum	GIEMS-D15	Fluet-Chouinardetal, 2015	0.092
pop_ct_usu	Population count in total watershed upstream of sub-basin pour point, in thousands	GPW v4	CIESIN, 2016	0.0915
gdp_ud_usu	Gross Domestic Product (GDP) per capita adjusted for Purchasing Power Parity (PPP) for total watershed upstream of sub-basin pour point, in US dollars	GDP PPP v2	Kummuetal, 2018	0.0915
lkv_mc_usu	Total lake volume in total watershed upstream of sub-basin pour point, in million cubic metres	HydroLAKES	Messager et al., 2016	0.0899
clz_cl_smj_17	Binary indicator for whether spatial majority of sub-basin is an extremely hot and xeric climate zone	GEnS	Metzger et al. 2013	0.0889
lka_pc_use	Percent total watershed upstream of sub-basin pour point that is lake	HydroLAKES	Messager et al., 2016	0.0887
wet_pc_u01	Percent off total watershed upstream of sub-basin pour point that is lake	GLWD	Lehner & Döll, 2004	0.0882
wet_pc_ug2	Percent of total watershed upstream of sub-basin pour point that is well-defined wetlands or uncertain/complex wetland mosaics	GIEMS-D15	Fluet-Chouinardetal, 2015	0.0864
glc_pc_u20	Percent of total watershed upstream of sub-basin pour point that is water bodies (natural and artificial)	GLC2000	Bartholomé & Belward, 2005	0.0855
riv_tc_usu	River volume in total watershed upstream of sub-basin pour point, in thousand cubic metres	HydroSHEDS and WaterGap v2.2	Lehner & Grill, 2013	0.0849
wet_pc_u04	Percent of total watershed upstream of sub-basin pour point that is freshwater marsh or floodplain	GLWD	Lehner & Döll, 2004	0.0848
ria_ha_usu	River area in total watershed upstream of sub-basin pout point, in hectares	HydroSHEDS and WaterGap v2.2	Lehner & Grill, 2013	0.0848
inu_pc_umx	Percent of total watershed upstream of sub-basin pour point that is covered by the inundation extent annual maximum	GIEMS-D15	Fluet-Chouinardetal, 2015	0.0843
glc_pc_u02	Percent of total watershed upstream of sub-basin pour point that is broadleaved, deciduous, and closed tree cover	GLC2000	Bartholomé & Belward, 2005	0.0836
dis_m3_pmn	Annual minimum natural discharge at sub-basin pour point, in cubic metres/second	WaterGAP v2.2	Döll et al., 2003	0.0815

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

dis_m3_pyr	Annual average natural discharge at sub-basin pour point, in cubic metres/second	WaterGAP v2.2	Döll et al., 2003	0.0812
cls_cl_smj_123	Binary indicator for whether spatial majority of sub-basin is an extremely hot and xeric climate strata (Q4)	GEnS	Metzger et al. 2013	0.0784
inu_pc_smn	Percent of sub-basin that is covered by the inundation extent annual minimum	GIEMS-D15	Fluet-Chouinardetal, 2015	0.0761
inu_pc_umn	Percent of total watershed upstream of sub-basin pour point that is covered by the inundation extent annual minimum	GIEMS-D15	Fluet-Chouinardetal, 2015	0.0737
riv_tc_ssu	River volume in sub-basin, in thousand cubic metres	HydroSHEDS and WaterGap v2.2	Lehner & Grill, 2013	0.0669
fec_cl_smj_522	Binary indicator for whether spatial majority of sub-basin is the Upper Nile freshwater ecoregion	FEOW	Abell et al., 2008	0.0621
glc_pc_u15	Percent of total watershed upstream of sub-basin pour point that is regularly flooded shrub or herbaceous cover	GLC2000	Bartholomé & Belward, 2005	0.0571
lka_pc_sse	Percent total sub-basin that is lake	HydroLAKES	Messager et al., 2016	0.0532
ria_ha_ssu	River area in sub-basin, in hectares	HydroSHEDS and WaterGap v2.2	Lehner & Grill, 2013	0.0528
wet_pc_s03	Percent of total watershed upstream of sub-basin pour point that is river	GLWD	Lehner & Döll, 2004	0.0481
fmh_cl_smj_10	Binary indicator for whether spatial majority of sub-basin is tropical and subtropical floodplain rivers and wetlands	FEOW	Abell et al., 2008	0.0474
wet_pc_s01	Percent of total watershed upstream of sub-basin pour point that is lake	GLWD	Lehner & Döll, 2004	0.0433
tbi_cl_smj_9	Binary indicator for whether majority of sub-basin is flooded grasslands and savannas	TEOW	Dinerstein et al., 2017	0.0397
tec_cl_smj_74	Binary indicator for whether spatial majority of sub-basin is the Sudd flooded grasslands terrestrial ecoregion	TEOW	Dinerstein et al., 2017	0.0397
wet_cl_smj_4	Binary indicator for whether spatial majority of sub-basin is freshwater marsh or floodplain	GLWD	Lehner & Döll, 2004	0.0339
DIST_MAIN	Distance from polygon outlet to the most downstream sink along the river network, in kilometres	HydroBASINS	Lehner & Grill, 2013	0.0315
DIST_SINK	Distance from polygon outlet to the next downstream sink along the river network, in kilometres	HydroBASINS	Lehner & Grill, 2013	0.0314

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

tec_cl_smj_53	Binary indicator for whether spatial majority of sub-basin is the Sahelian Acacia savanna terrestrial ecoregion	TEOW	Dinerstein et al., 2017	0.0263
glc_cl_smj_15	Binary indicator for whether spatial majority of sub-basin is regularly flooded shrub and/or herbaceous cover	GLC2000	Bartholomé & Belward, 2005	0.0128
cls_cl_smj_122	Binary indicator for whether spatial majority of sub-basin is an extremely hot and xeric climate strata (Q3)	GEnS	Metzger et al. 2013	0.0115
pnv_cl_smj_9	Binary indicator for whether spatial majority of sub-basin is savanna	EarthStat	Ramankutty & Foley, 1999	0.0091
glc_cl_smj_3	Binary indicator for whether spatial majority of sub-basin is tree cover, broadleaved, deciduous, or open	GLC2000	Bartholomé & Belward, 2005	-0.0142
tbi_cl_smj_7	Binary indicator for whether spatial majority of sub-basin is tropical and subtropical grasslands, savannas and shrublands	TEOW	Dinerstein et al., 2017	-0.0171
tec_cl_smj_43	Binary indicator for whether spatial majority of sub-basin is the East Sudanian savanna terrestrial ecoregion	TEOW	Dinerstein et al., 2017	-0.0197
glc_cl_smj_9	Binary indicator for whether spatial majority of sub-basin is regularly flooded shrub and/or herbaceous cover	GLC2000	Bartholomé & Belward, 2005	-0.0252
tec_cl_smj_52	Binary indicator for whether spatial majority of sub-basin is the Northern Congolian Forest-Savanna terrestrial ecoregion	TEOW	Dinerstein et al., 2017	-0.0256
clz_cl_smj_18	Binary indicator for whether spatial majority of sub-basin is an extremely hot and moist climate zone	GEnS	Metzger et al. 2013	-0.0555
lit_cl_smj_8	Binary indicator for whether spatial majority of sub-basin metamorphic rocks	GLiM	Hartmann & Moosdorf, 2012	-0.0561
soc_th_sav	Average organic carbon content in soil of sub-basin, in tonnes/hectare	SoilGrids1km	Hengl et al., 2014	-0.0698
soc_th_uav	Average organic carbon content in soil of total watershed upstream of sub-basin pour point, in tonnes/hectare	SoilGrids1km	Hengl et al., 2014	-0.0965
pre_mm_s02	Average monthly precipitation in sub-basin for February, in millimetres	WorldClim v1.4	Hijmans et al., 2005	-0.1379
cmi_ix_s02	Average monthly Climate Moisture Index (CMI) in sub-basin, in index values (x100)	WorldClim v1.4 and Global-PET v1	Hijmans et al., 2005	-0.1429
gwt_cm_sav	Groundwater table depth for sub-basin, in centimetres	Global Groundwater Map	Fan et al., 2013	-0.1653
slp_dg_sav	Average terrain slope in sub-basin, in degrees (x10)	EarthEnv-DEM90	Robinson et al. 2014	-0.1835

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

ele_mt_smx	Maximum elevation in sub-basin, in metres above sea level	EarthEnv-DEM90	Robinson et al., 2014	-0.1998
ele_mt_smn	Minimum elevation in sub-basin, in metres above sea level	EarthEnv-DEM90	Robinson et al., 2014	-0.2125
ele_mt_sav	Average elevation in sub-basin, in metres above sea level	EarthEnv-DEM90	Robinson et al., 2014	-0.2197

Table A1 HydroATLAS variables used to construct the first principal component (PC) for the INFLOW-AI dynamic spatial modelling model with their PCA loadings indicating the degree to which each variable contributes to the PC used in the model. Mathematically, $PC_1 = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ where a_i are the PC loadings (shown in the table) and X_i is the original value of the corresponding variable. Positive loadings increase the value of the component while negative loadings decrease the value of the component. This process removes redundancy (multicollinearity) between highly similar variables. The table is sorted in decending order of descending postive contributions to PC1.

Appendix B: Explainability

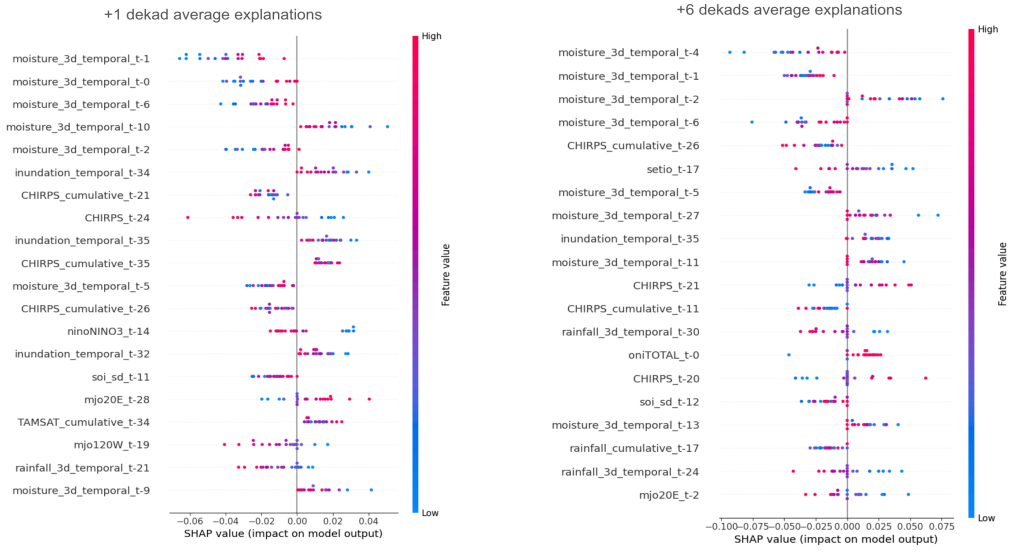

Figure B1. Variable importance at each temporal lag was analyzed using SHAP (Shapley Additive Explanations; Lundberg & Lee, 2017), which interpret machine learning predictions by attributing each feature a contribution value to the output. SHAP values were computed with the model-agnostic KernelExplainer, which estimates feature contributions by perturbing input features and fitting a locally weighted linear model around each prediction.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

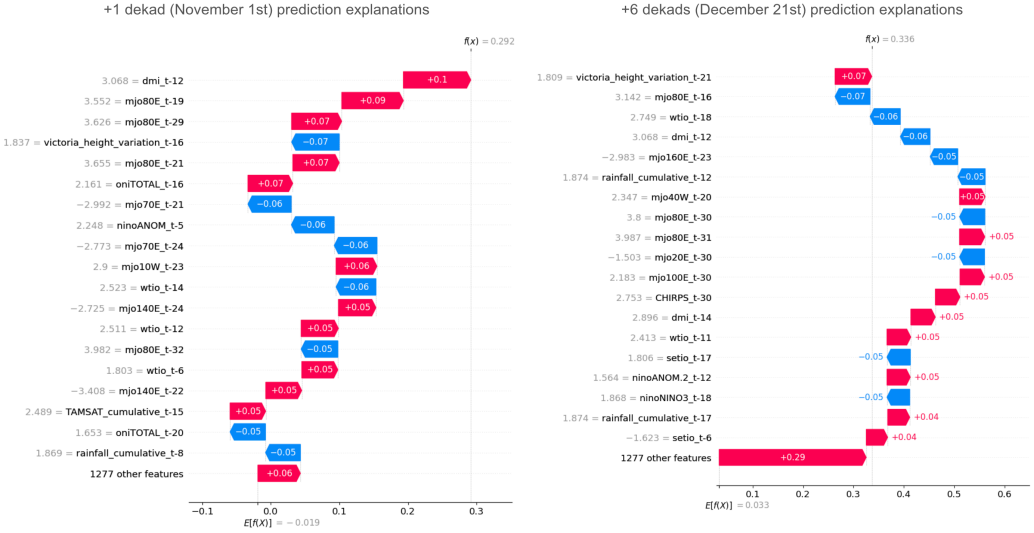

Figure B2. SHAP values indicating the contribution of each variable towards specific predictions made by the deployed model during the 2024 flood season. Summing the individual SHAP values produces the final prediction given by the model.

Appendix C: Water management confounds

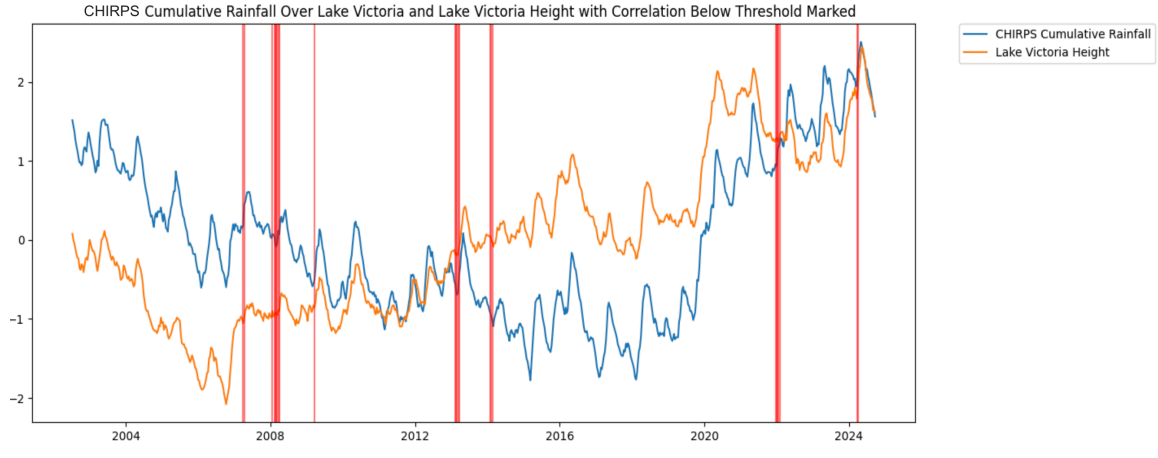

Figure C1. Periods of negative linear correlations between CHIRPS rainfall over Lake Victoria and Lake Victoria Height over a five-dekad period. Indicates potential dam releases from Lake Victoria that may make purely hydrometerological prediction in the area challenging.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

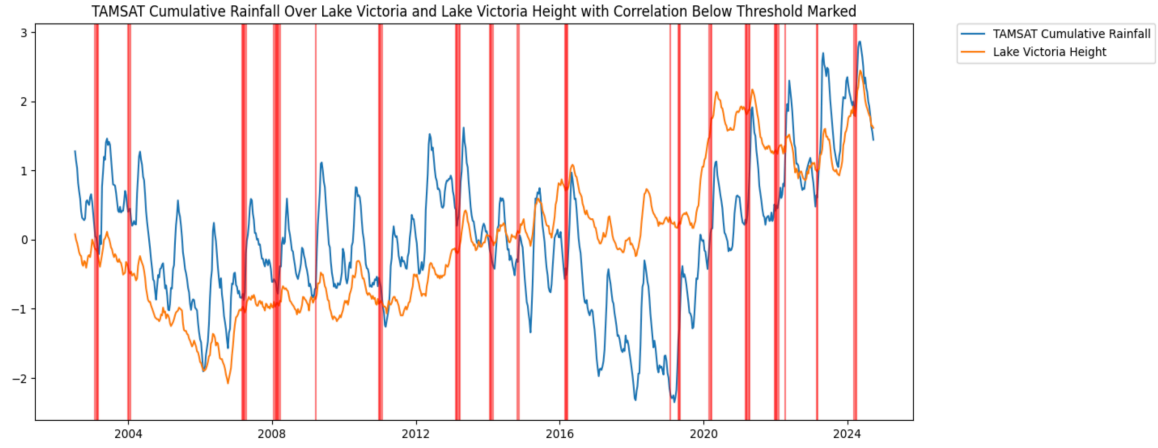

Figure C2. Periods of negative linear correlations between TAMSAT rainfall over Lake Victoria and Lake Victoria Height over a five-dekad period. Indicates potential dam releases from Lake Victoria that may make purely hydrometerological prediction in the area challenging, though less definitive periods are identified compared with the CHRIPS data.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Code and data availability: The released version of INFLOW-AI v2.1 is available on Zenodo under the MIT licence. The exact version of the model used to produce the results presented in this paper is archived on Zenodo under DOI 10.5281/zenodo.18258570 and is distributed under the MIT licence (Rapson et al., 2026a). The release contains all scripts, modules, and live data processing files required to reproduce the results shown in this paper. Detailed instructions on setting up the environment, downloading the data, and running the code are provided in the README file.

Live model predictions from the deployed system are available on JASMIN at https://gws-access.jasmin.ac.uk/public/tamsat/INFLOW/flood-inundation-predictions_inflow-ai-v2.1/. These include raw predictions and plots showing the next six-dekad prediction made by INFLOW-AI. As of August 2025, predictions are updated approximately dekadally.

All figures used in the Results section can be reproduced using the data and Python notebook archived on Zenodo under DOI 10.5281/zenodo.18261701 and distributed under the Creative Commons Attribution 4.0 International licence (Rapson et al., 2026b). This archive contains the exact code and data used to generate the plots presented in this paper.

Author contribution: Jessica Rapson was responsible for conceptualisation, methodology, model development, formal analysis, and visualisation, and led the writing of the original manuscript. Elisabeth Stephens contributed to the investigation through a hydrological literature review and participated in manuscript review and editing. Ross I. Maidment contributed data curation by providing descriptions of the TAMSAT dataset and reviewed data descriptions for other data sources, and participated in manuscript review and editing. Rogerio Bonifacio contributed software development and data curation through the provision of code and methodological descriptions for processing MODIS data. All authors reviewed and approved the final manuscript.

Financial support: This work was conducted under the INFLOW project, which was funded through the Climate Adaptation and Resilience (CLARE) programme, with support from the UK Foreign, Commonwealth & Development Office (FCDO) and the International Development Research Centre (IDRC), Canada, and additionally supported by the Natural Environment Research Council (NERC) via NCAS NC-International funding (NC/X006263/1) and a Knowledge Exchange Fellowship (Drought Risk Information for the African Finance Sector (DRIAFS);

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

NE/Y005058/1).

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

References

Abell, R., Thieme, M.L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., ... & Wikramanayake, E. (2008). Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience, 58*(5), 403-414.

Albright, R., Kaiser, L., Madsen, S., Policelli, F., & Jasinski, M. (2011). *Measuring reservoir heights via satellite altimetry products for global flood modeling*. In *Forty Years of Earth Observation at NASA*.

Alfieri, L., Libertino, A., Campo, L., Dottori, F., Gabellani, S., Ghizzoni, T., ... & Massabò, M. (2024). Impact-based flood forecasting in the Greater Horn of Africa. *Natural Hazards and Earth System Sciences, 24*(1), 199-224.

Atif, S., Umar, M. & Ullah, F. Investigating the flood damages in Lower Indus Basin since 2000: Spatiotemporal analyses of the major flood events. *Nat Hazards 108*, 2357–2383 (2021). https://doi.org/10.1007/s11069-021-04783-w

Bartholomé, E., Belward, A.S. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing, 26*(9), 1959-1977.

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., Yu, S., Rasp, S., Ahmed, F., O'Gorman, P. A., Neelin, J. D., Lutsko, N. J., & Pritchard, M. (2024). Climate-invariant machine learning. *Science Advances, 10*(6), eadj7250. https://doi.org/10.1126/sciadv.adj7250

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bukhari, H., & Brown, C. A. (2022). A comparative review of decision support tools routinely used by selected transboundary River Basin Organisations. *African Journal of Aquatic Science, 47*(3), 318-337.

Caldwell, J. (2022, March 10). Flood risk for South Sudan's 2022 rainy season. The Centre for Humanitarian Data. https://centre.humdata.org/flood-risks-for-south-sudans-2022-rainy-season/

Center for International Earth Science Information Network at Columbia University (CIESIN). (2016). *Gridded Population of the World, Version 4 (GPWv4)*: Population Count. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4X63JVC. Accessed 23 May 2017.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Climate Prediction Centre (CPC). (2025a). *Cold & Warm Episodes by Season*. Retrieved August 7, 2025 from https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

Climate Prediction Centre (CPC). (2025b). *Seasonal climate forecast from* CFSv2. Retrieved August 7, 2025 from https://www.cpc.ncep.noaa.gov/products/CFSv2/CFSv2seasonal.shtml

Climate Prediction Centre (CPC). (2025c). *The Southern Oscillation Index*. Retrieved August 7, 2025 from https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensocycle/soi.shtml

Climate Prediction Centre (CPC). (2025d). *Madden / Julian Oscilliation.* Retrieved August 7, 2025 from https://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/mjo.shtml

Climate Prediction Centre (CPC). (2025e). *Positive and Negative IOD by Month.* Retrieved August 7, 2025 from https://www.cpc.ncep.noaa.gov/products/international/ocean_monitoring/IODMI/DMI_month.html

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., & Zsoter, E. (2016). Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences, 20*(9), 3549–3560. https://doi.org/10.5194/hess-20-3549-2016

de Ferranti, J. (2022) Digital elevation data. *Viewfinder Panoramas*. Accessed August 7, 2025 from https://viewfinderpanoramas.org/dem3.html

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., ... & Hansen, M. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience, 67*(6), 534-545. doi:10.1093/biosci/bix014.

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., & Ceccato, P. (2018). Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quarterly Journal of the Royal Meteorological Society*, 144, 292-312. https://doi.org/10.1002/qj.3244.

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology, 270*, 105-134.

Dong, B., Peng, S., Liu, G., Pu, T., Gerlein-Safdi, C., Prigent, C., & Lin, X. (2024). Underestimation of methane emissions from the Sudd wetland: Unraveling the impact of wetland extent dynamics. *Geophysical Research*

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Letters, 51(16), e2024GL110690.

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The human impact of floods: a historical review of events 1980-2009 and systematic literature review. *PLoS currents, 5*, ecurrents-dis. https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

Eilers, P. H., Pesendorfer, V., & Bonifacio, R. (2017, June). Automatic smoothing of remote sensing data. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)* (pp. 1-3). IEEE.

El-Mahdy, M. E.-S., Mousa, F. A., Morsy, F. I., Kamel, A. F., & El-Tantawi, A. (2024). Flood classification and prediction in South Sudan using artificial intelligence models under a changing climate. Alexandria Engineering Journal, 97, 127–141.

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science, 339*(6122), 940-943.

Ficchì, A., Cloke, H. L., Neves, C., Woolnough, S., Coughlan de Perez, E., Zsótér, E., Pinto, I., Meque, A., & Stephens, E. (2021). Beyond El Niño: Unsung climate modes drive African floods. *Weather and Climate Extremes, 33*, 100345. https://doi.org/10.1016/j.wace.2021.100345.

Ficchì, A. & Stephens, L. (2019). Climate variability alters flood timing across Africa. *Geophysical Research Letters*, *46*(15), 8809-8819.

Fluet-Chouinard, E., Lehner, B., Rebelo, L. M., Papa, F., & Hamilton, S. K. (2015). Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. *Remote Sensing of Environment, 158*, 348-361.

Fonseca F, Worfarth-Couto B, Santos A, Marinho R, Martinez J-M, Filizola N. Hydrological Scenarios and Malaria Incidence in the Amazonian Context. *Water*. 2022; 14(8):1283. https://doi.org/10.3390/w14081283.

Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L. M., Gupta, H. V., & Nearing, G. S. (2022). Deep learning rainfall–runoff predictions of extreme events. *Hydrology and Earth System Sciences, 26*(13), 3377–3392. https://doi.org/10.5194/hess-26-3377-2022.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific*

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Data, 2(1), 1-21. doi: 10.1038/sdata.2015.66.

Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning* (pp. 1050-1059). PMLR.

Global Water Monitor. (2024). *2024 Summary Report*, Retrieved from https://www.globalwater.online/globalwater/report/index.html#:~:text=Water%2Drelated%20disasters%20caused%20major,of%20casualties%20and%20economic%20damage.

Hardy, A., Palmer, P. I., & Oakes, G. (2023). Satellite data reveal how Sudd wetland dynamics are linked with globally-significant methane emissions. *Environmental Research Letters*, *18*(7), 074044.

Harrigan, S., Zsoter, E., Cloke, H., Salamon, P., & Prudhomme, C. (2020). "Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System." *Hydrology and Earth System Sciences Discussions, 2020*, 1–22.

Hartmann, J., Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems, 13*, Q12004.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J., Walsh, M., Gonzalez, M.R. (2014). SoilGrids1km—global soil information based on automated mapping. *PLoS ONE, 9*(8), e105992. doi:10.1371/journal.pone.0105992.

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting, 37*(1), 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*(15), 1965-1978.

Hou, J., Van Dijk, A. I. J. M., Renzullo, L. J., and Larraondo, P. R.: GloLakes: water storage dynamics for 27 000 lakes globally from 1984 to present derived from satellite altimetry and optical imaging, *Earth Syst. Sci. Data, 16*, 201–218, https://doi.org/10.5194/essd-16-201-2024.

INFLOW-AI Team. (2025). INFLOW-AI (v2.1) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.18258570.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science* (pp. 1094-1096). Springer, Berlin, Heidelberg.

Kebacho, L. L. (2022). Large-scale circulations associated with recent interannual variability of the short rains over East Africa. *Meteorology and Atmospheric Physics, 134*(1), 10. https://doi.org/10.1007/s00703-021-00846-6.

Kummu, M., Taka, M., Guillaume, J.H.A. (2018) Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015. *Scientific Data, 5*, 180004. https://doi.org/10.1038/sdata.2018.4

Lehner, B., Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology, 296*(1), 1-22.

Lehner, B., Grill G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes, 27*(15), 2171-2186. doi: 10.1002/hyp.9740.

Lehner, B., Verdin, K., Jarvis, A. (2008): New global hydrography derived from spaceborne elevation data. *Eos, Transactions, 89*(10): 93-94. Data available at https://www.hydrosheds.org.

Lin, L., Di, L., Tang, J., Yu, E., Zhang, C., Rahman, M. S., ... & Kang, L. (2019). Improvement and validation of NASA/MODIS NRT global flood mapping. *Remote Sensing, 11*(2), 205.

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data 6: 283. DOI: 10.1038/s41597-0190300-6.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30.*

Lunt, M. F., Palmer, P. I., Feng, L., Taylor, C. M., Boesch, H., & Parker, R. J. (2019). An increase in methane emissions from tropical Africa between 2010 and 2016 inferred from satellite data. *Atmospheric Chemistry and Physics, 19*(23), 14721-14740.

Koriche, S. A., & Rientjes, T. H. M. (2016). Application of satellite products and hydrological modelling for flood early warning. *Physics and Chemistry of the Earth, Parts A/B/C, 93*, 12–23. https://doi.org/10.1016/j.pce.2016.03.007

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

967 Maidment, R. I., Grimes, D., Allan, R. P., Tarnavsky, E., Stringer, M., Hewison, T., ... & Black, E. (2014). The 30
968 year TAMSAT African rainfall climatology and time series (TARCAT) data set. *Journal of Geophysical*
969 *Research: Atmospheres, 119*(18), 10-619. https://doi.org/10.1002/2014JD021927.

970 Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., ... & Alcántara, E. M. U. (2017). A
971 new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Scientific Data, 4*(1), 1-
972 19. doi: 10.1038/sdata.2017.63.

973 Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O. (2016). Estimating the volume and age of water stored
974 in global lakes using a geo-statistical approach. *Nature Communications, 7*, 13603. doi: 10.1038/ncomms13603

975 Mitsch, W. J., Nahlik, A., Wolski, P., Bernal, B., Zhang, L., & Ramberg, L. (2010). Tropical wetlands: Seasonal
976 hydrologic pulsing, carbon sequestration, and methane emissions. *Wetlands Ecology and Management, 18*(5),
977 573–586. https://doi.org/10.1007/s11273-009-9164-4

978 Ministry of Electricity, Dams, Irrigation & Water Resources (MEDIWR). (2015). *Irrigation Development Master*
979 *Plan (IDMP) in the Republic of South Sudan*.

980 Ministry of Water and Environment (MWE). (2009). Water Management Zones (WMZ) | Ministry of Water and
981 Environment. https://www.mwe.go.ug/library/water-management-zones-wmz

982 Metzger, M.J., Bunce, R.G., Jongman, R.H., Sayre, R., Trabucco, A., Zomer, R. (2013). A high-resolution bioclimate
983 map of the world: a unifying framework for global biodiversity research and monitoring. *Global Ecology and*
984 *Biogeography, 22*(5), 630-638.

985 NASA Goddard Space Flight Center. (2025). *Blue Ice: Global Water Monitor (timeseries data)*. NASA.
986 https://blueice.gsfc.nasa.gov/gwm/timeseries

987 Nile Basin Initiative. (2024, July 8). *Nile Basin seasonal hydrological outlook: June – September 2024* (Issue No.
988 NBI002). Nile Basin Initiative. https://nilebasin.org/content/nile-river-basin-seasonal-hydrological-outlook-
989 june-september-2024

990 Özdoğan-Sarıkoç, G., & Dadaser-Celik, F. (2024). Physically based vs. data-driven models for streamflow and
991 reservoir volume prediction at a data-scarce semi-arid basin. *Environmental Science and Pollution Research,*
992 *31*(27), 39098-39119.

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Papa, F., Crétaux, J.-F., Grippa, M., Robert, E., Trigg, M., Tshimanga, R. M., Kitambo, B., Paris, A., Carr, A., Fleischmann, A. S., de Fleury, M., Gbetkom, P. G., Calmettes, B., & Calmant, S. (2023). Water resources in Africa under global change: Monitoring surface waters from space. Surveys in Geophysics, 44(1), 43–93. https://doi.org/10.1007/s10712-022-09700-9Petersen, G. and Fohrer, N., 2010. Flooding and drying mechanisms of the seasonal Sudd flood plains along the Bahr el Jebel in southern Sudan. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55*(1), pp.4-16. https://doi.org/10.1080/02626660903525278

Pasche, O. C., & Engelke, S. (2024). Neural networks for extreme quantile regression with an application to forecasting of flood risk (arXiv:2208.07590). *arXiv*. https://arxiv.org/abs/2208.07590

Petersen, G. & Fohrer, N. (2010) Flooding and drying mechanisms of the seasonal Sudd flood plains along the Bahr el Jebel in southern Sudan. *Hydrol. Sci. J. 55*(1), 4–16.

Pinho, P.F., Marengo, J.A. & Smith, M.S. Complex socio-ecological dynamics driven by extreme events in the Amazon. *Reg Environ Change 15*, 643–655 (2015). https://doi.org/10.1007/s10113-014-0659-z.

Pohl, B., & Camberlin, P. (2006). Influence of the Madden–Julian oscillation on East African rainfall: II. March–May season extremes and interannual variability. Quarterly Journal of the Royal Meteorological Society: *A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography, 132*(621), 2541-2558. https://doi.org/10.1256/qj.05.223.

Ramankutty, N., Foley, J.A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles, 13*(4), 997-1027.

Rapson, J., Stephens, E., Maidment, R. I., and Bonifacio, R. (2026). *INFLOW-AI v2.1*. Zenodo. https://doi.org/10.5281/zenodo.18258570.

Rapson, J., Stephens, E., Maidment, R. I., and Bonifacio, R. (2026). *INFLOW-AI v2.1 Dataset and Code for Paper Figures*. Zenodo. https://doi.org/10.5281/zenodo.18261701.

Rebelo, L., G. B. Senay, & M. P. McCartney. (2012). Flood Pulsing in the Sudd Wetland: Analysis of Seasonal Variations in Inundation and Evaporation in South Sudan. *Earth Interact., 16*, 1–19, https://doi.org/10.1175/2011EI382.1

Reis, V., Hermoso, V., Hamilton, S. K., Bunn, S. E., Fluet-Chouinard, E., Venables, B., & Linke, S. (2019).

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Characterizing seasonal dynamics of Amazonian wetlands for conservation and decision making. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*(7), 1073-1082.

Robinson, N., Regetz, J., Guralnick, R.P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. ISPRS *Journal of Photogrammetry and Remote Sensing, 87*, 57-67. doi: 10.1016/j.isprsjprs.2013.11.002.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., ... & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems, 28*.

Sene, K. J. (2000). Theoretical estimates for the influence of Lake Victoria on flows in the upper White Nile. *Hydrological Sciences Journal, 45*(1), 125–145. https://doi.org/10.1080/02626660009492310

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems, 28*.

Sosnowski, A., Ghoneim, E., Burke, J. J., Hines, E., & Halls, J. (2016). Remote regions, remote data: A spatial investigation of precipitation, dynamic land covers, and conflict in the Sudd wetland of South Sudan. *Applied Geography*, *69*, 51-64.

Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., & Zappa, M. (2023). Hybrid forecasting: Blending climate predictions with AI models. H*ydrology and Earth System Sciences, 27*(9), 1865–1889. https://doi.org/10.5194/hess-27-1865-2023

University of Reading. (2025a). *TAMSAT – Tropical Applications of Meteorology using SATellite and ground-based observations*. University of Reading. https://research.reading.ac.uk/tamsat/

University of Reading. (2025b). *TAMSAT Soil Moisture*. https://research.reading.ac.uk/tamsat/soil-moisture/

The EastAfrican. (2020, March 9). *Uganda drains excess dam water into the Nile*. Retrieved August 7, 2025 from https://www.theeastafrican.co.ke/tea/news/east-africa/uganda-drains-excess-dam-water-into-the-nile-1438230

Theia. (2025). Hydroweb: *Operational altimetry water levels for rivers and lakes* [Data set]. Theia. https://www.theia-land.fr/en/blog/product/water-levels-of-rivers-and-lakes-hydroweb/

https://doi.org/10.5194/egusphere-2026-66
Preprint. Discussion started: 27 February 2026

Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R. P., Stringer, M., ... & Kayitakire, F. (2014). Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. *Journal of Applied Meteorology and Climatology, 53*(12), 2805-2822. https://doi.org/10.1175/JAMC-D-14-0016.1.

United Nations High Commissioner for Refugees (UNHCR). (2022, March 29). *UNHCR warns of dire impact from floods in South Sudan as new wet season looms*. https://www.unhcr.org/uk/news/briefing-notes/unhcr-warns-dire-impact-floods-south-sudan-new-wet-season-looms

United Nations Office for the Coordination of Humanitarian Affairs (OCHA). (2024, October 25). *South Sudan: Floods snapshot (as of 25 October 2024)*. https://www.unocha.org/publications/report/south-sudan/south-sudan-floods-snapshot-25-october-2024

U.S. Geological Survey. (2024, July 15). *Flood_Modelling_v2.docx* [Data file]. Famine Early Warning Systems Network (FEWS NET). https://edcftp.cr.usgs.gov/project/FEWSNET/spervez/SouthSudan_flooding/Flood_Modelling_v2.docx

UNICEF South Sudan. (n.d.). *Climate change and flooding*. UNICEF. Retrieved January 16, 2026, from https://www.unicef.org/southsudan/what-we-do/climate-change-and-flooding

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Ward, P. J., Jongman, B., Kummu, M., Dettinger, M. D., Sperna Weiland, F. C., & Winsemius, H. C. (2014). Strong influence of El Niño Southern Oscillation on flood risk around the world. *Proceedings of the National Academy of Sciences*, *111*(44), 15659-15664.

Yeditha, P. K., Kasi, V., Rathinasamy, M., & Agarwal, A. (2020). Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 30*(6), 063115. https://doi.org/10.1063/5.0008195

Zhang, T., Dong, Z., Wang, W., Qian, Z., Zhang, J., Lu, P., ... & Li, H. (2025). Flood risk assessment in data-scarce South Sudan using a flood modeling framework. *Journal of Hydrology: Regional Studies, 61*, 102743.