

# COMPARISON OF MACHINE LEARNING AND RAINFALL-RUNOFF MODELS FOR RUNOFF PREDICTION IN URBAN DRAINAGE SYSTEMS

(Working Paper)

Eng. Jordan Mafumbo, Eng. Noah Mesfin, Eng. Dr. Seith Mugume

February 2026

## Abstract

Urban pluvial flooding remains a persistent challenge in Kampala, Uganda, particularly in the Lubigi catchment—a crucial catchment affected by rapid urbanization, impervious surface expansion, wetland degradation, and intense rainfall events. This study evaluates and compares a Random Forest Regression Machine Learning (ML) model against two widely used physical-based rainfall-runoff models: HEC-HMS (Hydrologic Engineering Center's Hydrologic Modeling System) and SWMM (Storm Water Management Model).

The analysis draws on historical daily rainfall data (1981–2022) from the Uganda National Meteorological Authority (UNMA), Digital Elevation Models (DEM), land use, and soil data from the Kampala Capital City Authority (KCCA). The HEC-HMS model was configured with the SCS Curve Number loss, unit hydrograph transform, and Muskingum routing; The SWMM model incorporated subcatchment hydrology, impervious/pervious parameters, and detailed conduit hydraulics; the ML model leveraged rainfall-derived features (for instance depth, antecedent conditions) and calibrated runoff proxies.

The models were assessed on an 80% calibration/ 20% validation split using comprehensive performance metrics: Root Mean Square Error (RMSE) for absolute error, Nash-Sutcliffe Efficiency (NSE) for overall hydrograph fit, Kling-Gupta Efficiency (KGE) for decomposition of correlation, bias, and variability, and peak flow error (%PE) for flood peak accuracy.

The results indicate SWMM's superiority in urban hydraulic contexts (highest NSE/KGE, lowest %PE), owing to its explicit routing of channels and structures. The Random Forest ML model delivered competitive performance with markedly lower computational cost and faster execution, making it suitable for real-time applications. HEC-HMS provided solid hydrologic simulation but was less effective for detailed urban drainage dynamics.

These findings highlight ML's potential as an efficient, data-untaxing alternative or complement to traditional models in resource-constrained urban settings like Kampala, while underscoring the value of physical-based approaches for hydraulic precision.

## 1. Introduction

Kampala's rapid urban growth has amplified flood risks in catchments such as Lubigi, where horrible drainage infrastructure, solid waste/sediment deposition, and short-duration intense rainfall lead to frequent flooding. Physical-based models like HEC-HMS and SWMM are the usually-relied-upon tools for rainfall-runoff and hydraulic simulation, yet they demand substantial data and calibration effort. Machine learning, particularly ensemble methods like Random Forest, offers promise for pattern-based prediction with reduced input requirements.

### **Main objective:**

Compare the performance of a Random Forest ML model with HEC-HMS and SWMM for runoff prediction in the Lubigi catchment.

### **Specific objectives:**

Configure and calibrate the three models using available meteorological, topographic, and land-use data.

Evaluate model performance using RMSE, NSE, KGE, and %PE across calibration and validation periods.

Discuss implications for urban flood forecasting and management in data-scarce environments.

## 2. Literature Review

Current urban drainage modeling usually employs SWMM for integrated hydrologic-hydraulic processes—for instance, conduit flow, backwater effects—and HEC-HMS for event-focused rainfall-runoff simulation—for example, SCS-CN and unit hydrograph methods. Studies in Kampala and similar African cities emphasize SWMM's advantages in capturing urban-specific features, while HEC-HMS suits broader catchment hydrology but may overlook detailed channel dynamics.

Machine learning approaches, including Random Forest, have demonstrated strong performance in runoff prediction where gauged data are limited, often achieving high NSE and KGE values. However, ML models can lack physical interpretability. Single-metric assessments—for instance, RMSE alone—are inadequate for flood applications; multi-metric evaluations incorporating KGE (for bias and variability) and peak error metrics are essential, as recommended in hydrological guidelines.

Recent work on Lubigi highlights the role of urban growth, sediment/waste blockage, and pluvial extremes in flood risk, with modeling efforts focusing on HEC-RAS, SWMM, or integrated approaches.

### 3. Methodology

**Data sources:** Daily rainfall series (1981–2022, UNMA); DEM, land use/soil layers (KCCA). Runoff references derived from model calibration, literature proxies (for example, water levels), and synthetic generation.

**HEC-HMS configuration:** SCS-CN loss method for infiltration, SCS unit hydrograph for transform, Muskingum for reach routing; subbasins delineated from DEM.

**SWMM configuration:** Subcatchments with percent imperviousness, curve numbers, conduit network geometry, hydraulic structures (for example, inlets, junctions); continuous and design-storm simulations.

**Random Forest ML:** Input features: rainfall depth, antecedent precipitation index, seasonal indicators; target: proxy runoff; preprocessing (normalization, outlier handling); 5-fold time-series cross-validation; hyperparameter tuning (number of trees, depth, etc.).

#### Performance metrics:

RMSE : Quantifies average error magnitude ( $m^3/s$ ; lower is better).

NSE: Assesses model fit relative to mean observed (1 = perfect, 0 = no better than mean).

KGE: Balances correlation, bias ratio, and variability ratio (1 = ideal).

%PE: Peak flow relative error  $|(simulated\ peak - reference\ peak)/reference\ peak| \times 100$  (lower better for flood peaks).

**Data split:** 80% calibration (training/tuning), 20% independent validation.

### 4. Results and Discussion

All models reproduced observed runoff patterns reasonably, but divergence appeared in urban-specific metrics.

**Table 1:** Model Performance Comparison (Validation Set, Focus on 2022 Events)

Model	RMSE ( $m^3/s$ )	NSE	KGE	%PE	Key Strengths / Limitations
HEC-HMS	1.45	0.72	0.78	$\pm 18\%$	Solid event hydrology; weaker on urban routing/timing
SWMM	1.12	0.85	0.88	$\pm 9\%$	Best hydraulic detail, peak accuracy, volume capture
Random Forest ML	1.28	0.81	0.84	$\pm 12\%$	Excellent efficiency, low bias; rapid for real-time use

SWMM consistently outperformed due to its explicit representation of drainage conduits and hydraulic processes, yielding the tightest hydrograph matches and lowest peak errors—vital for flood warning in Lubigi. Random Forest provided strong overall fit (high NSE/KGE) with minimal computational overhead, outperforming HEC-HMS in efficiency and often closely rivaling SWMM in non-peak metrics. HEC-HMS was reliable for hydrologic response but showed timing/volume shifts in complex urban channels.

Hydrograph visualizations—for instance, selected 2022 events—confirmed SWMM's alignment with reference peaks, ML's smooth low-bias predictions, and HEC-HMS's minor offsets.

**Key limitation:** Reliance on daily rainfall data and proxy runoff (due to sparse gauging); sub-hourly resolution would better capture urban flash-flood dynamics.

## **5. Conclusion and Recommendations**

SWMM emerges as the most suitable for detailed urban drainage and flood modeling in the Lubigi catchment, excelling in hydraulic precision. Random Forest ML proves a compelling, efficient alternative—especially in data-scarce settings—for quick predictions, scenario analysis, and integration into early-warning systems. HEC-HMS remains valuable for simpler hydrologic tasks.

### **Recommendations:**

- Install sub-daily rainfall and flow/water-level gauges in Lubigi for robust observed-data validation.
- Develop hybrid models (for example, ML-enhanced parameters in SWMM or HEC-HMS).
- Incorporate sediment/waste deposition effects (from recent studies) into future simulations.
- Apply models to climate projections and urban planning under KCCA frameworks for long-term resilience.

### **References (Selected key sources; full list available if needed)**

- Moriasi et al. (guidelines on NSE, KGE, and hydrological evaluation).
- EPA SWMM documentation (Rossman et al.).
- Relevant Lubigi studies (for instance, integrated flood modeling, HEC-RAS applications in Kampala).