



P-ISSN: 2706-7483
E-ISSN: 2706-7491
NAAS Rating (2026): 4.5
IJGGE 2026; 8(8): 138-145
www.geojournal.net
Received: 26-06-2026
Accepted: 30-07-2026

Edward Namusasi Masibayi
Department of Disaster
Management, International
Peace Support Training
Centre, National Defence
University–Kenya, Nairobi,
Kenya

An integrated geospatial–hydrometeorological framework for developing HEC-HMS input datasets in data-constrained river basins: Application to the Tana River Basin, Kenya

Edward Namusasi Masibayi

DOI: <https://www.doi.org/10.22271/27067483.2026.v8.i8b.856>

Abstract

Reliable HEC-HMS modelling depends on spatially consistent, quality-controlled environmental inputs, yet input-data development is often treated as routine preprocessing. This study developed an integrated geospatial–hydrometeorological workflow for preparing model-ready HEC-HMS input datasets in data-constrained basins, using Kenya’s Tana River Basin as a case study. A 30 m SRTM DEM, 10 m ESA WorldCover land-use/land-cover data, Hydrologic Soil Groups, daily rainfall from 33 stations, CHIRPS precipitation, and observed discharge were quality assessed and harmonised to a common spatial framework. Terrain conditioning generated 23 computational sub-basins. Land-cover and soil data were integrated at 30 m resolution to derive a distributed Curve Number (CN) surface and area-weighted sub-basin CN values of 63–83. Of 1,023 expected rainfall station-days, 50 (4.89%) were missing and infilled using CHIRPS estimates. Daily rainfall was then interpolated using inverse distance weighting (IDW; $p = 2$), evaluated by Leave-One-Out Cross-Validation (LOOCV), and aggregated to sub-basin Mean Areal Precipitation (MAP). Interpolation errors were lowest in the upper and central catchments (MAE 0.25–0.45 mm; RMSE 0.35–0.60 mm) and higher in the lower and eastern basin (MAE 0.60–1.10 mm; RMSE up to 1.50 mm). The resulting terrain, runoff-parameter, meteorological, and hydrometric datasets form a traceable model-ready database. The contribution lies in an auditable, transferable input-development architecture that links quality assurance, spatial harmonisation, distributed parameter derivation, rainfall validation, and hydrological aggregation before model implementation.

Keywords: HEC-HMS, geospatial hydrology, curve number, rainfall interpolation, mean areal precipitation, Tana River Basin

Introduction

Flood risk is shaped by climate variability and change, land-use transformation, population growth, and increasing occupation of flood-prone areas, while weaknesses in hydrometeorological monitoring continue to constrain risk reduction in many developing regions (IPCC, 2022; UNDRR, 2015; WMO, 2021). Reliable hydrological information is therefore central to flood forecasting, disaster-risk reduction, climate adaptation, and water-resources planning.

Hydrological models provide a scientific basis for representing rainfall–runoff processes and supporting flood forecasting and watershed management (Beven, 2012; Chow *et al.*, 1988; Maidment, 1993) ^[2, 4, 14]. HEC-HMS is widely used because it supports precipitation–runoff transformation, loss methods, routing, continuous simulation, and GIS-based watershed representation (Feldman, 2000; USACE, 2023, 2024). Model reliability, however, depends not only on model structure and calibration but also on the quality, resolution, spatial alignment, and preprocessing of terrain, land-use/land-cover (LULC), soil, rainfall, and discharge inputs. Errors introduced before simulation can propagate through watershed delineation, runoff parameterisation, meteorological forcing, and subsequent flood-risk products (Beven, 2012) ^[2].

The problem is particularly important in data-constrained basins, where datasets from different institutions and remote-sensing products often differ in coordinate systems, spatial resolution, extent, classification, temporal coverage, and quality-control procedures.

Corresponding Author:
Edward Namusasi Masibayi
Department of Disaster
Management, International
Peace Support Training
Centre, National Defence
University–Kenya, Nairobi,
Kenya

Rainfall presents an additional challenge because gauges are discrete and frequently unevenly distributed; interpolation is therefore necessary, but uncertainty increases where observational support is weak (Goovaerts, 2000; Li & Heap, 2014; Ly *et al.*, 2013) [8, 12, 13]. Semi-distributed modelling also requires representative sub-basin parameters even though land cover, soils, topography, and rainfall are spatially heterogeneous. A transparent approach is therefore to derive variables at an appropriate spatial resolution, evaluate their quality, and only then aggregate them to model computational units.

This study develops and demonstrates an integrated geospatial hydrometeorological framework for producing HEC-HMS input datasets for the Tana River Basin, Kenya. The objectives are to quality-assess and harmonise heterogeneous environmental data; derive terrain, runoff-parameter, and rainfall datasets compatible with semi-distributed modelling; evaluate interpolated rainfall using Leave-One-Out Cross-Validation (LOOCV); and formulate transferable principles for input-data development. The methodological contribution lies in integrating established procedures into a single auditable workflow rather than proposing a new terrain-processing or rainfall-interpolation

algorithm. The study is deliberately limited to input-dataset development and validation rather than HEC-HMS calibration or simulation-performance evaluation.

2. Materials and Methods

2.1 Study Area and Data

The framework was applied to the Tana River Basin, Kenya, which covers approximately 126,000 km² and extends from the humid highlands of Mount Kenya and the Aberdare Range through semi-arid lowlands to the Indian Ocean. The basin exhibits pronounced gradients in elevation, rainfall, land cover, soils, and hydrological response, while the rainfall-station network is unevenly distributed. Six datasets were used: a ~30 m SRTM DEM (Farr *et al.*, 2007) [5]; 10 m ESA WorldCover 2021 v200 LULC (Zanaga *et al.*, 2022) [27]; Kenya Soil Survey information classified into Hydrologic Soil Groups (HSG A–D); daily rainfall from 33 Kenya Meteorological Department stations for 1–31 March 2024; CHIRPS daily rainfall for gap filling (Funk *et al.*, 2015) [7]; and Water Resources Authority discharge records retained for subsequent model evaluation. All spatial datasets were referenced to WGS 84 / UTM Zone 37S (EPSG:32737).

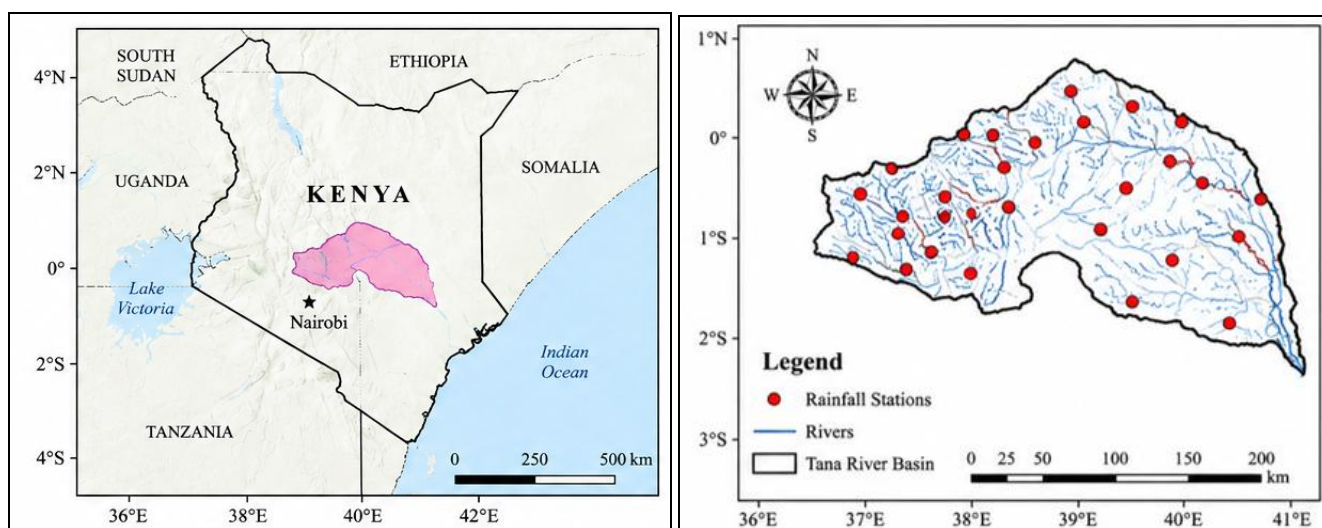


Fig 1: Location of the Tana River Basin, Kenya, showing the watershed boundary, major river network, and rainfall stations.

2.2 Integrated Workflow, Quality Control and Spatial Harmonisation

Dataset development followed eight linked stages: acquisition, quality assessment, spatial harmonisation, terrain conditioning, parameter derivation, rainfall interpolation, LOOCV, and hydrological aggregation/HEC-HMS integration (Figure 2). The DEM was examined for completeness, elevation continuity, voids, and artificial depressions; LULC and HSG data were checked for classification completeness and spatial integrity; rainfall records were checked for completeness, physical plausibility, temporal consistency, coordinate accuracy, and spatial coherence; and discharge records were checked for completeness, continuity, anomalous observations, and hydrological plausibility.

Rainfall-record completeness and the proportion of missing station-days were quantified as:

$$C = \frac{n_{\text{obs}}}{n_{\text{exp}}} \times 100, \quad M = \frac{n_{\text{miss}}}{n_{\text{exp}}} \times 100$$

where C is data completeness (%), M is the missing-data percentage (%), n_{obs} is the number of observed station-days, n_{miss} is the number of missing station-days, and n_{exp} is the expected number of station-days. For 33 stations observed over 31 days, $n_{\text{exp}} = 1,023$; therefore $C = 95.11\%$ (973/1,023) and $M = 4.89\%$ (50/1,023).

Missing station-days were filled using CHIRPS values extracted at the affected station coordinates, while available gauge measurements were retained unchanged (Funk *et al.*, 2015). Spatial layers were projected to EPSG:32737 and aligned to the 30 m DEM reference grid using a common raster origin, cell size, extent, and grid alignment; categorical LULC and HSG data were resampled by nearest neighbour and clipped to the basin boundary. This harmonisation reduced positional inconsistencies before raster overlay, parameter derivation, rainfall interpolation, and zonal aggregation.

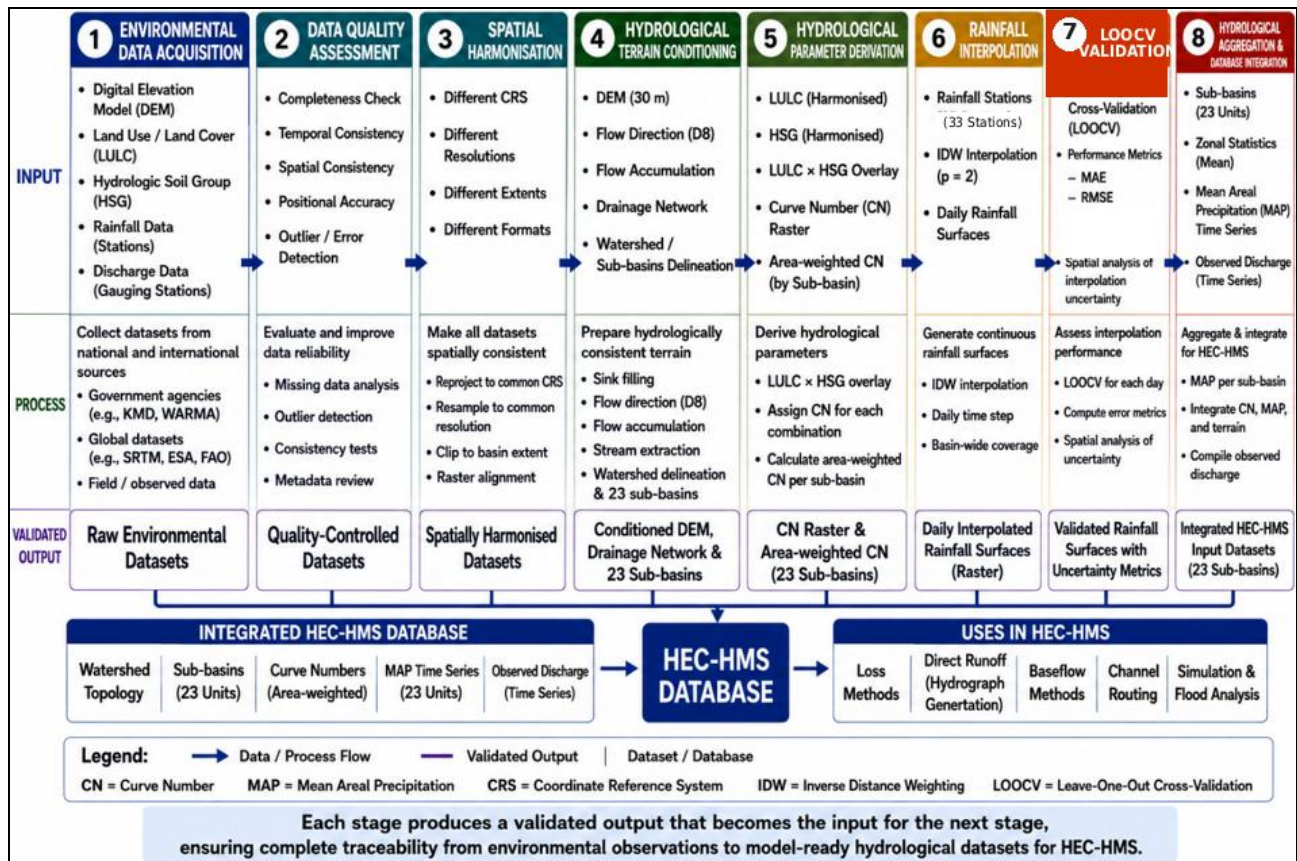


Fig 2: Integrated geospatial-hydrometeorological workflow for developing quality-controlled HEC-HMS inputs.

2.3 Terrain Conditioning and Curve Number Development

The 30 m DEM was hydrologically conditioned in ArcGIS Pro 3.4 using sink filling, D8 flow direction, flow accumulation, stream extraction, and watershed delineation, following established raster-based drainage-analysis principles (O'Callaghan & Mark, 1984; Jenson & Domingue, 1988) [38, 10]. Flow accumulation represents the number of upstream raster cells draining through a cell and was used to define the drainage network. The SRTM terrain source provides near-global elevation data suitable for basin-scale topographic analysis, while its known resolution constraints were considered during interpretation (Farr *et al.*, 2007; Mukherjee *et al.*, 2013) [5, 17].

The 25,000-cell stream-definition threshold was converted to contributing drainage area using:

$$A_t = N_t(\Delta x \Delta y) = 25,000(30 \text{ m} \times 30 \text{ m}) = 22.5 \text{ km}^2$$

where A_t is the threshold contributing area, N_t is the flow-accumulation threshold in cells, and Δx and Δy are the raster cell dimensions. With $30 \text{ m} \times 30 \text{ m}$ cells, the threshold corresponds to approximately 22.5 km^2 . This criterion suppresses minor drainage pathways while retaining the principal basin-scale drainage structure. The conditioned terrain and extracted drainage network were then used to delineate the watershed and 23 computational sub-basins.

Runoff-loss parameters were developed with the SCS Curve Number (CN) method under average antecedent moisture conditions (AMC II) (USDA-SCS, 1986; USDA-NRCS, 2004a, 2004b; USACE, 2023). Harmonised LULC and HSG rasters were overlaid at 30 m resolution, and each LULC–

HSG combination was assigned a CN from the adopted lookup matrix. Higher CN values represent lower infiltration capacity and greater direct-runoff potential, whereas lower CN values represent greater infiltration and retention. The corresponding potential maximum retention, S , and initial abstraction, I_a , were calculated using the standard SCS relationships.

Potential maximum retention was computed from CN as:

$$S = \frac{25400}{CN} - 254 \quad (1)$$

where S is potential maximum retention (mm) and CN is the dimensionless Curve Number.

Initial abstraction was represented by the conventional SCS relationship:

$$I_a = 0.2S \quad (2)$$

where I_a is initial abstraction (mm).

The CN-derived retention and abstraction parameters enter the standard SCS-CN precipitation-excess relationship as:

$$P_e = \begin{cases} \frac{(P - I_a)^2}{P - I_a + S}, & P > I_a \\ 0, & P \leq I_a \end{cases} \quad (3)$$

where P_e is direct runoff or precipitation excess (mm), P is accumulated precipitation (mm), I_a is initial abstraction (mm), and S is potential maximum retention (mm). This standard SCS-CN rainfall-runoff formulation links the derived CN parameter to precipitation excess used in event

modelling (USDA-NRCS, 2004b; USDA-SCS, 1986). The equation is included to show how the derived CN parameters are used in subsequent HEC-HMS rainfall–runoff modelling; this study does not evaluate the resulting simulated discharge.

For CN assignment, the harmonised land-cover data were represented by the eight hydrological LULC classes reported in Table 1 and combined with HSG classes A–D. The HSG sequence represents progressively decreasing infiltration capacity and increasing runoff potential from Group A to Group D, consistent with the NRCS hydrologic soil-cover framework (USDA-NRCS, 2004a). The AMC II matrix below was then used to reclassify the LULC–HSG combinations; only combinations occurring within the basin contributed to the distributed CN surface.

Table 1: LULC–Hydrologic Soil Group Curve Number lookup matrix used for distributed CN derivation.

Hydrological LULC class	HSG A	HSG B	HSG C	HSG D
Forest	25	55	70	77
Grassland / Rangeland	50	69	79	84
Agricultural land / Cropland	67	78	85	89
Shrubland	55	72	81	86
Wetland	30	58	71	78
Built-up / Settlement	81	88	91	93
Bare land / Rock / Sand	78	86	91	93
Water bodies	98	98	98	98

Note: HSG = Hydrologic Soil Group; CN values represent AMC II conditions.

Representative CN values for the 23 sub-basins were obtained by area-weighted aggregation of the distributed raster:

$$CN_j = \frac{\sum_{i=1}^n CN_i A_i}{\sum_{i=1}^n A_i} \quad (4)$$

where CN_j is the representative CN for sub-basin j , CN_i is the CN of spatial unit i , A_i is its area, and n is the number of contributing units. This implements the principle of deriving spatially before aggregating hydrologically.

2.4 Rainfall Interpolation, Validation and Mean Areal Precipitation

After quality control and CHIRPS gap filling, daily rainfall from the 33 stations was interpolated to continuous basin-wide surfaces using inverse distance weighting (IDW) with distance-decay exponent $p = 2$ (Shepard, 1968; Li & Heap, 2011, 2014; Ly *et al.*, 2013) [19, 11, 12, 13]. IDW is a deterministic interpolation method in which observations nearer to a prediction location receive greater weight than more distant observations. Station coordinates were maintained in the common projected reference system so that interpolation distances were computed consistently, and the same $p = 2$ parameterisation was applied to each daily surface. Rainfall at an unsampled location x was estimated using normalized distance weights:

$$w_i(x) = \frac{d_i(x)^{-p}}{\sum_{k=1}^n d_k(x)^{-p}}, \quad \hat{P}(x) = \sum_{i=1}^n w_i(x) P_i \quad (5)$$

where $w_i(x)$ is the normalized weight assigned to station i at

prediction location x , $d_i(x)$ is the corresponding distance, p is the distance-decay exponent ($p = 2$), P_i is observed station rainfall, and n is the number of contributing stations. The weights sum to one, so nearer stations exert greater influence while preserving a weighted-average estimate. The same parameterisation was applied consistently to all daily rainfall surfaces.

Interpolation performance was evaluated by LOOCV. Each rainfall station was omitted in turn, rainfall at the omitted location was estimated from the remaining stations, and the prediction was compared with the corresponding observation. Prediction error was quantified using mean absolute error (MAE) and root mean square error (RMSE) (Chai & Draxler, 2014) [3]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - \hat{P}_i| \quad (6)$$

and

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - \hat{P}_i)^2} \quad (7)$$

where P_i and $\hat{P}_i$ are observed and LOOCV-predicted rainfall values, respectively, and n is the number of validation observations. MAE summarises the average absolute prediction error, whereas RMSE gives greater weight to relatively large errors. Validation preceded hydrological aggregation so that interpolation performance was assessed before model forcing was generated.

Validated daily rainfall rasters were aggregated to each of the 23 sub-basins using zonal statistics. Mean Areal Precipitation (MAP) for sub-basin j at time t was represented as:

$$\bar{P}_{j,t} = \frac{1}{A_j} \iint P(x,y,t) dA \quad (8)$$

where $\bar{P}_{j,t}$ is sub-basin MAP, A_j is sub-basin area, and $P(x,y,t)$ is the interpolated rainfall field. Zonal aggregation was repeated for all 31 daily rainfall surfaces, generating 23 continuous MAP time series directly compatible with the semi-distributed HEC-HMS meteorological framework. For the common equal-area raster grid used in the analysis, the continuous area-average expression was implemented numerically as:

$$\bar{P}_{j,t} \approx \frac{\sum_{k=1}^{m_j} P_{k,t} A_k}{\sum_{k=1}^{m_j} A_k} = \frac{1}{m_j} \sum_{k=1}^{m_j} P_{k,t} \quad \text{for equal-area cells} \quad (9)$$

where $P_{k,t}$ is rainfall in raster cell k on day t , A_k is cell area, and m_j is the number of contributing cells in sub-basin j . Because the harmonised grid uses equal-area 30 m cells, the area-weighted expression reduces to the arithmetic mean of the rainfall-cell values within each sub-basin. This discrete form corresponds directly to the zonal-statistics implementation used to generate the 23 daily MAP time

series.

2.5 HEC-HMS Dataset Integration

The final database linked four components to the same 23-sub-basin framework for HEC-HMS Version 4.12: conditioned terrain and drainage topology for the Basin Model; area-weighted CN values for the SCS-CN loss method; daily MAP series for the Meteorological Model; and quality-controlled discharge records for subsequent calibration, validation, and performance assessment. The latter stage should use transparent quantitative and graphical model-evaluation criteria when undertaken (Moriassi *et al.*, 2007) [16]. This structure maintained traceability between the original environmental observations and the model-ready inputs. The present study evaluates the preparation of these

inputs rather than the resulting HEC-HMS simulations.

3. Results

3.1 Terrain and Spatial Harmonisation

Hydrological conditioning produced a connected drainage network and 23 computational sub-basins from the 30 m DEM. The principal outputs were the conditioned DEM, flow-direction and flow-accumulation rasters, drainage network, watershed boundary, sub-basin boundaries, junctions, and basin outlet (Figure 3). Spatial harmonisation placed the DEM, LULC, HSG, rainfall stations, and sub-basin boundaries in EPSG:32737 with a common 30 m analytical grid, providing pixel-level correspondence for subsequent overlay and zonal aggregation.

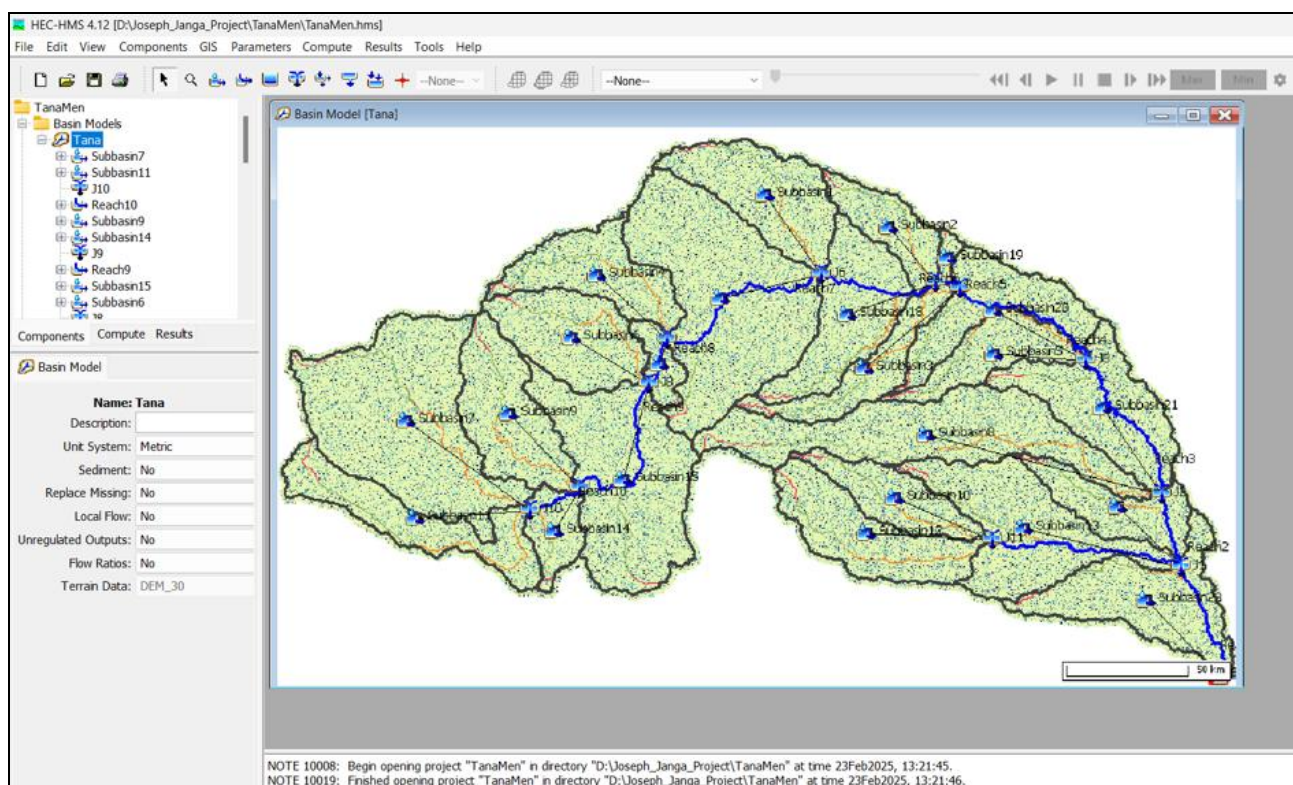


Fig 3: Hydrologically conditioned DEM, extracted drainage network, watershed boundary, and 23 computational sub-basins.

3.2 Curve Number Dataset

Overlay of the harmonised LULC and HSG datasets generated a distributed 30 m CN surface. For cartographic presentation, the raster was displayed using classes extending from approximately 40 to >90 (Figure 4); these are display intervals rather than exact minimum and maximum raster statistics. Lower CN values represented areas with greater infiltration and retention potential, whereas higher values represented greater direct-runoff potential. Area-weighted aggregation produced sub-basin

CN values ranging from 63 to 83. The minimum occurred in SB2 (CN = 63) and the maximum in SB20 (CN = 83), demonstrating substantial hydrological variability among computational units.

Deriving CN at pixel scale before sub-basin aggregation retained the proportional contribution of different land-cover and soil combinations within each hydrological unit. The resulting representative values therefore preserve more of the basin's environmental heterogeneity than direct assignment of a single unexamined sub-basin value.

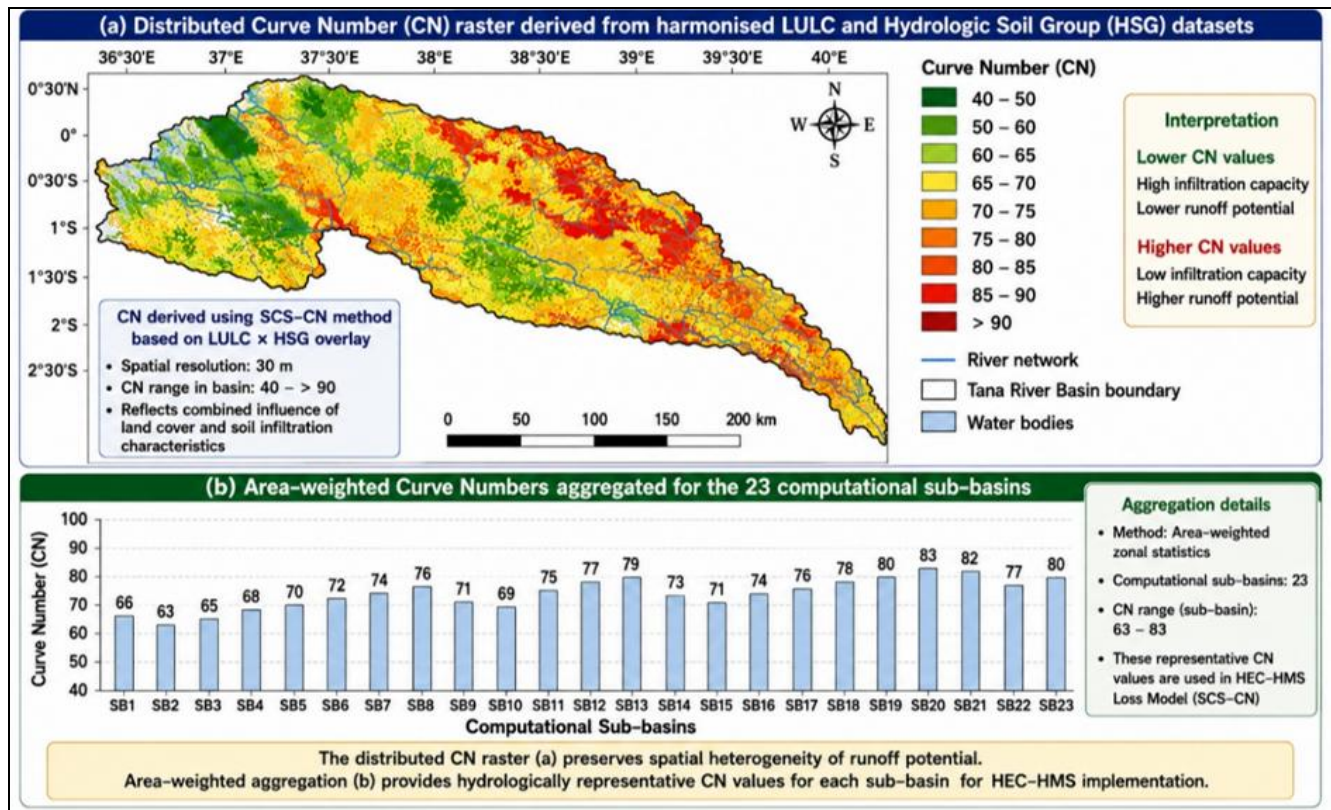


Fig 4: Distributed Curve Number surface and area-weighted aggregation to the 23 computational sub-basins.

3.3 Rainfall Interpolation and Validation

The rainfall network comprised 33 stations within and around the basin. Station density was higher in the upper and central catchments and lower in downstream and eastern areas. After quality control and gap filling, IDW produced 31 daily rainfall rasters for March 2024. Representative surfaces for 1, 15, and 29 March show

marked spatial and temporal heterogeneity, with broader rainfall gradients where gauges were more widely spaced (Figure 5). These maps represent interpolated estimates between stations rather than direct observations of rainfall at every location.

Daily application of IDW generated the complete rainfall-surface dataset used for subsequent MAP estimation.

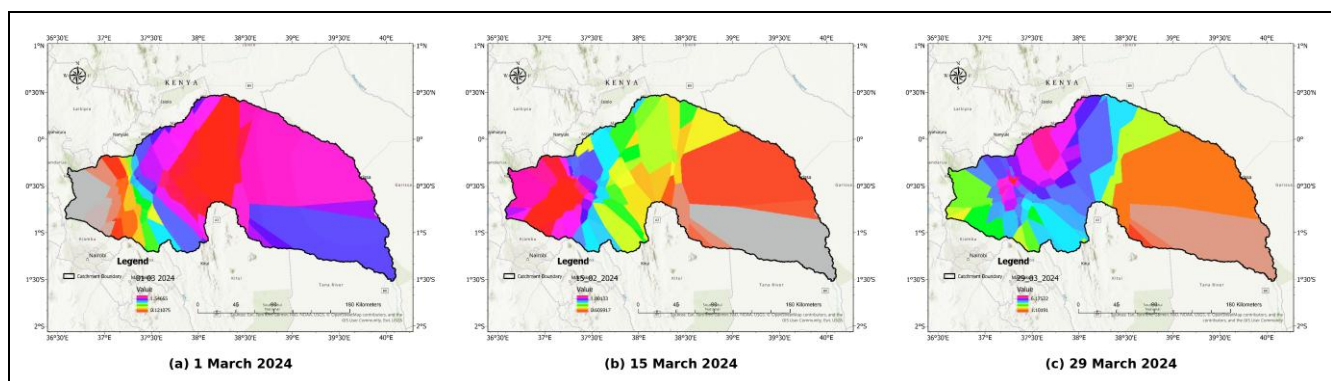


Fig 5: Representative IDW rainfall surfaces for 1, 15, and 29 March 2024.

LOOCV revealed a clear spatial pattern in interpolation performance. Upper and central catchments had the lowest station-specific errors, with MAE of 0.25–0.45 mm and RMSE of 0.35–0.60 mm. In the lower and eastern basin, where station spacing was greater, MAE increased to 0.60–1.10 mm and RMSE reached approximately 1.50 mm. RMSE was generally higher than MAE, indicating the

influence of relatively large individual prediction errors. Table 2 reports minimum–maximum ranges of station-specific error statistics within each basin region; the values are not basin-wide average errors. The results identify the lower and eastern basin as the principal zone of rainfall-interpolation uncertainty.

Table 2: Regional ranges of station-specific LOOCV error statistics for IDW rainfall interpolation.

Basin region	Station density	Station-specific MAE range (mm)	Station-specific RMSE range (mm)	Interpretation
Upper and central catchments	High	0.25–0.45	0.35–0.60	Lowest prediction errors
Lower and eastern basin	Low	0.60–1.10	Up to 1.50	Higher prediction errors
Entire Tana River Basin	Variable	Spatially variable	Spatially variable	Accuracy varied with station distribution

3.4 Mean Areal Precipitation and Integrated HEC-HMS Inputs

Zonal aggregation of the 31 validated rainfall surfaces generated 23 continuous daily MAP series. At the beginning of March, sub-basin rainfall was relatively low and spatially uniform, averaging about 1.55 mm. By mid-March, representative sub-basin MAP values ranged from approximately 0.606 to 1.801 mm. The greatest contrast among the selected dates occurred on 29 March, when MAP ranged from approximately 1.102 to 6.175 mm, confirming that a single basin-average rainfall value would have obscured important differences in meteorological forcing.

The final model-ready database consisted of: (i) conditioned terrain, drainage topology, and 23 sub-basins; (ii) a distributed CN raster and area-weighted sub-basin CN values; (iii) 31 interpolated rainfall surfaces and 23 MAP time series; and (iv) quality-controlled discharge observations. All components were spatially and temporally referenced to the same computational framework, providing a traceable basis for subsequent HEC-HMS parameterisation and rainfall–runoff simulation.

4. Discussion

The results demonstrate that HEC-HMS input preparation is not merely a preliminary GIS task but a scientific workflow in its own right. In data constrained basins, datasets commonly originate from different institutions and products and carry distinct spatial, temporal, and quality characteristics. Treating acquisition, quality assessment, harmonisation, parameter derivation, meteorological interpolation, validation, and aggregation as linked quality-assurance stages improves traceability and reduces the likelihood that unrecognised input errors are transferred into model calibration and flood-risk products (Beven, 2012; Maidment, 1993)^[2, 14]. The contribution therefore lies in the architecture and sequencing of established methods, not in claiming a new terrain, Curve Number, or interpolation algorithm.

A central methodological principle is to derive spatially and aggregate hydrologically. LULC and HSG were first combined at 30 m resolution to represent within-sub-basin variability before CN values were area-weighted to HEC-HMS computational units. Rainfall followed the same logic: observations were quality controlled and completed, interpolated as continuous fields, evaluated by LOOCV, and only then aggregated to MAP. This input-oriented workflow complements terrain-focused preprocessing tools such as Arc Hydro and HEC-GeoHMS and differs from model-specific HRU construction in SWAT by emphasising a transparent chain for preparing terrain, runoff-parameter, meteorological, and hydrometric inputs (Maidment, 2002; Arnold *et al.*, 2012)^[15, 1].

The rainfall results show why uncertainty should be evaluated before hydrological simulation. Interpolation errors were lowest where station density was greatest and increased in the sparsely monitored lower and eastern basin, consistent with the sensitivity of spatial interpolation to

sampling configuration and method choice (Goovaerts, 2000; Li & Heap, 2011, 2014; Ly *et al.*, 2013)^[8, 11, 12, 13]. CHIRPS gap filling improved temporal completeness but introduced a separate source of uncertainty because the replacement values were gridded estimates rather than direct gauge measurements (Funk *et al.*, 2015)^[7]. LOOCV evaluates IDW prediction under the available station configuration; it does not independently validate the CHIRPS-filled observations. Consequently, interpolation uncertainty and infilling uncertainty should be treated as distinct components when interpreting MAP forcing and any subsequent runoff or flood estimates.

The framework is transferable because it integrates established methods rather than depending on a basin-specific algorithm. Its principal limitations are the one-month demonstration period, uneven rainfall-station distribution, the assumptions of IDW, the absence of a separate validation of CHIRPS infilling, and the use of a 30 m DEM that may not resolve fine-scale drainage features in low-relief terrain (Goovaerts, 2000; Mukherjee *et al.*, 2013)^[8, 17]. Future applications should compare geostatistical and elevation-informed rainfall methods, validate gap-filling products independently where observations permit, test longer periods and contrasting basins, incorporate higher-resolution terrain or additional satellite/radar precipitation products, and quantify uncertainty propagation from environmental observations through hydrological simulation.

5. Conclusions

An integrated geospatial–hydrometeorological workflow was developed for producing quality-controlled HEC-HMS input datasets in the Tana River Basin. The procedure generated a conditioned terrain model with 23 computational sub-basins, harmonised LULC and HSG datasets, distributed and area-weighted Curve Numbers, 31 daily rainfall surfaces from 33 stations, 23 daily MAP series, and quality-controlled discharge records. These products form a coherent, traceable database for subsequent semi-distributed HEC-HMS modelling.

The principal methodological contribution is the explicit integration of quality assurance, spatial harmonisation, distributed parameter derivation, rainfall cross-validation, and hydrological aggregation within one input-development architecture. Deriving variables spatially before aggregating them to model units preserves environmental heterogeneity while producing model-compatible inputs. The framework provides a transparent basis for hydrological modelling and flood-risk assessment in data-constrained basins, while future work should evaluate alternative rainfall interpolation and gap-filling methods, longer periods, higher-resolution geospatial data, and cumulative uncertainty through the full modelling chain.

6. Declarations

Author contributions

Edward Namusasi Masibayi conceived and designed the

study, acquired and processed the data, performed the analyses, interpreted the results, prepared the figures and tables, and wrote, reviewed, and approved the manuscript.

Funding

No specific funding was received from public, commercial, or not-for-profit agencies.

Conflict of interest

The author declares no competing interests.

Data availability

Processed datasets are available from the corresponding author on reasonable request, subject to the data-sharing policies of the originating institutions.

Code availability

Geospatial processing was implemented in ArcGIS Pro; no custom software was developed specifically for this study.

Ethics/consent

Ethics approval, consent to participate, and consent for publication: Not applicable.

Acknowledgements

The author acknowledges the Kenya Meteorological Department, Water Resources Authority, and agencies responsible for the DEM, land-use, and soil datasets used in this study.

References

- Arnold JG, Kiniry JR, Srinivasan R, Williams JR, Haney EB, Neitsch SL. Soil and water assessment tool input/output documentation version 2012. Texas: Texas Water Resources Institute; 2012.
- Beven KJ. Rainfall-runoff modelling: the primer. 2nd ed. Chichester: John Wiley & Sons; 2012.
- Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014;7(3):1247–1250. doi:10.5194/gmd-7-1247-2014.
- Chow VT, Maidment DR, Mays LW. Applied hydrology. New York: McGraw-Hill; 1988.
- Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, *et al.* The Shuttle Radar Topography Mission. *Reviews of Geophysics*. 2007;45(2). doi:10.1029/2005RG000183.
- Feldman AD. Hydrologic Modeling System HEC-HMS technical reference manual. Davis (CA): U.S. Army Corps of Engineers, Hydrologic Engineering Center; 2000.
- Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, *et al.* The climate hazards infrared precipitation with stations a new environmental record for monitoring extremes. *Scientific Data*. 2015;2:150066. doi:10.1038/sdata.2015.66.
- Goovaerts P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*. 2000;228(1–2):113–129. doi:10.1016/S0022-1694(00)00144-X.
- Intergovernmental Panel on Climate Change. Climate change 2022: impacts, adaptation and vulnerability. Cambridge: Cambridge University Press; 2022.
- Jenson SK, Domingue JO. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering & Remote Sensing*. 1988;54(11):1593–1600.
- Li J, Heap AD. A review of comparative studies of spatial interpolation methods in environmental sciences. Canberra: Geoscience Australia; 2011.
- Li J, Heap AD. Spatial interpolation methods applied in the environmental sciences: a review. *Environmental Modelling & Software*. 2014;53:173–189. doi:10.1016/j.envsoft.2013.12.008.
- Ly S, Charles C, Degré A. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modelling at watershed scale: a review. *Biotechnology, Agronomy, Society and Environment*. 2013;17(2):392–406.
- Maidment DR, editor. Handbook of hydrology. New York: McGraw-Hill; 1993.
- Maidment DR. Arc Hydro: GIS for water resources. Redlands (CA): ESRI Press; 2002.
- Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*. 2007;50(3):885–900. doi:10.13031/2013.23153.
- Mukherjee S, Joshi PK, Mukherjee S, Ghosh A, Garg RD, Mukhopadhyay A. Evaluation of vertical accuracy of open-source digital elevation models. *International Journal of Applied Earth Observation and Geoinformation*. 2013;21:205–217.
- O'Callaghan JF, Mark DM. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*. 1984;28(3):323–344. doi:10.1016/S0734-189X(84)80011-0.
- Shepard D. A two-dimensional interpolation function for irregularly spaced data. *Proceedings of the 1968 ACM National Conference*. 1968:517–524.
- United Nations Office for Disaster Risk Reduction. Sendai framework for disaster risk reduction 2015–2030. Geneva: United Nations Office for Disaster Risk Reduction; 2015.
- U.S. Army Corps of Engineers. Hydrologic Modeling System HEC-HMS technical reference manual. Version 4.12. Davis (CA): Hydrologic Engineering Center; 2023.
- U.S. Army Corps of Engineers. Hydrologic Modeling System HEC-HMS applications guide. Davis (CA): Hydrologic Engineering Center; 2024.
- USDA Natural Resources Conservation Service. National engineering handbook, Part 630 hydrology, Chapter 9: hydrologic soil-cover complexes. Washington (DC): U.S. Department of Agriculture; 2004.
- USDA Natural Resources Conservation Service. National engineering handbook, Part 630 hydrology, Chapter 10: estimation of direct runoff from storm rainfall. Washington (DC): U.S. Department of Agriculture; 2004.
- USDA Soil Conservation Service. Urban hydrology for small watersheds (Technical Release 55). Washington (DC): U.S. Department of Agriculture; 1986.
- World Meteorological Organization. State of the global climate 2021. Geneva: World Meteorological Organization; 2021.
- Zanaga D, Van De Kerchove R, Daems D, De Keersmaecker W, Brockmann C, Kirches G, *et al.* ESA WorldCover 10 m 2021 v200 [data set]. Zenodo; 2022. doi:10.5281/zenodo.7254221.