

ENFORCING CATCHMENT WATER BALANCE IMPROVES THE RELIABILITY OF NEURAL STREAMFLOW PREDICTION IN TROPICAL CATCHMENTS

Zafar Aqthaf¹* and Lalith Rajapakse¹

¹ Department of Civil Engineering, University of Moratuwa, Katubedda, Sri Lanka

*Corresponding author: aqthafahz.21@uom.lk

Abstract

Reliable streamflow prediction underpins flood early warning and water-resource planning, yet remains difficult in tropical catchments where monsoonal extremes and sparse monitoring limit both process-based and data-driven models. Data-driven models post high skill scores, but by minimising only statistical error they can produce hydrographs that silently violate the catchment water balance, generating runoff unsupported by rainfall. This study asks whether physical consistency, not accuracy alone, should rank such models. A compact neural network constrained to honour a lumped water-balance relation was tested on two contrasting Sri Lankan basins, a perennial wet-zone basin and a regulated intermittent dry-zone basin, against a process-based model and four deep-learning baselines. In the wet-zone basin the constrained model was both the most accurate and the most mass-conserving. In the dry-zone basin the highest Nash–Sutcliffe efficiency belonged to a model overpredicting runoff volume by about twenty percent, whereas the constrained model held volumetric bias near six percent with the best Kling–Gupta efficiency. Conventional efficiency scores can thus mask implausible predictions; embedding mass conservation is an auditable reliability safeguard for operational hydrology.

Keywords: Hydrological modelling, Physics-informed machine learning, Mass conservation, Kling–Gupta efficiency

1. Introduction

Floods and droughts are among the costliest natural hazards, and dependable streamflow prediction underpins flood early warning, reservoir operation, and irrigation scheduling. Prediction is hardest in tropical monsoon settings of intense seasonal rainfall and sparse gauging. Process-based models demand dense data and laborious calibration rarely available in developing regions, whereas data-driven models learn non-linear rainfall-runoff behaviour directly from records and now dominate the literature [1]. Their strength is also their weakness: minimising only statistical error, they carry no guarantee of physical consistency and can overfit dominant regimes, degrading beyond the training distribution [2]. A model may reproduce a hydrograph's shape while generating water no rainfall supports, a failure conventional skill scores do not reveal.

Physics-informed neural networks instead constrain learning with domain knowledge, embedding governing equations directly in the training loss so predictions stay physically admissible [3]. The decisive question for operational deployment is not which model attains the highest efficiency score, but which respects conservation of mass. Nash–Sutcliffe efficiency, the default ranking metric, overweights high-flow fit and can hide systematic volume error, exactly what decomposed and volumetric metrics were designed to expose [4]. This study therefore reframes evaluation around physical consistency, testing a water-balance-constrained network on two hydro-climatically opposite Sri Lankan basins.

2. Materials and Methods

Two basins spanning Sri Lanka's hydro-climatic extremes were used. The Kalu Ganga (wet zone; ~2,719 km²; 4,000-5,000 mm yr⁻¹; south-west monsoon; 2001-2010) is perennial and flood-prone; the Maduru Oya (dry zone; ~1,487 km²; ~2,158 mm yr⁻¹; north-east monsoon; 2006-2015) is intermittent and regulated by an irrigation-tank cascade. Daily precipitation, temperature, evaporation, and discharge were quality-controlled, gap-filled, and reduced to a common 19-feature set of antecedent-rainfall and cyclical-seasonality predictors, split chronologically 70/30 into calibration and validation. Treating the catchment as a single control volume, continuity with a linear-reservoir closure gives a first-order discharge relation [5]:

$$dQ/dt = \alpha_p \bar{P} - \alpha_e E - \beta Q + \delta$$

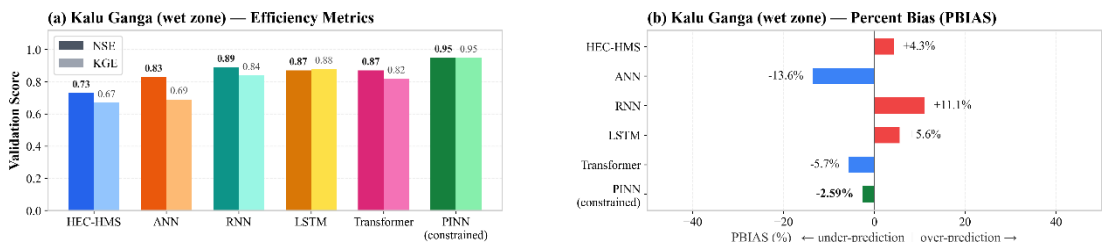
where the precipitation, evaporation, and recession coefficients are constrained non-negative to stay physically admissible. The residual was evaluated using daily finite differences, with learned coefficients representing effective basin responses rather than directly measurable parameters. The training loss of a compact feed-forward network (~11,000 parameters, with a linear skip connection) summed the streamflow prediction error and this water-balance residual, the residual weighted by a penalty coefficient ramped from near-zero over a warm-up so the network first fit the hydrograph before conservation was enforced. The process-based HEC-HMS was calibrated over the same periods and splits as the deep-learning baselines (ANN, RNN, LSTM, Transformer). Skill used Nash–Sutcliffe (NSE) and Kling–Gupta (KGE) efficiencies, the volumetric KGE component, and percent bias (PBIAS), the last two chosen to expose water-balance error that NSE can mask [4].

3. Results and Discussion

Wet zone, where accuracy and consistency agree. In the perennial Kalu Ganga the constrained network was best on every criterion (Figure 1), reaching validation NSE and KGE of 0.95 at only -2.6% volumetric bias; enforcing mass conservation cost nothing.

Regulated dry zone, where they diverge. In the tank-regulated Maduru Oya the LSTM ranks first by NSE (0.68) yet overpredicts runoff volume by ~20% (PBIAS = +20%, $\beta = 1.20$), manufacturing water unsupported by rainfall; HEC-HMS underpredicts by 37%. The constrained network ranks third by NSE (0.60) but gives the lowest absolute bias (+6%) and highest KGE (0.72), so the NSE leaderboard is physically misleading: the apparently best model is the least credible for water allocation.

A transferable reliability safeguard. The same architecture generalised across both regimes without basin-specific tuning, so the water-balance term supplies transferable regularisation. Its limitation is instructive: in the dry-zone basin it attenuated flash peaks to ~64% of observed magnitude while keeping recession and volume consistent, because the single-reservoir closure cannot capture threshold tank-spilling (Figure 2), an honest, interpretable failure rather than a hidden one.



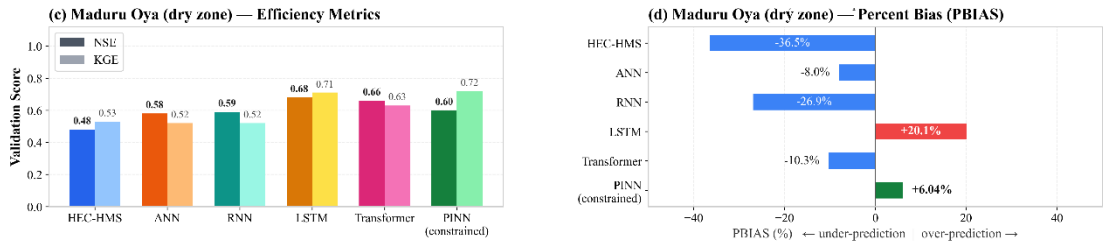


Figure 1: Validation NSE, KGE, and PBIAS for all models in both basins.

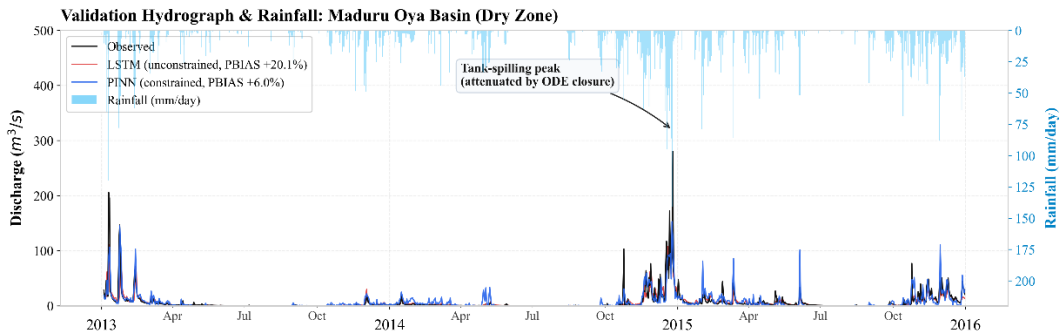


Figure 2: Observed vs. predicted daily discharge for the dry-zone validation period.

4. Conclusions

Reframing evaluation around physical consistency materially improves neural streamflow prediction in tropical basins: where accuracy and consistency agree the constrained model was best on all metrics; where they diverge it preserved the mass conservation that the highest-efficiency model silently violated. Mass conservation should therefore be a first-class evaluation criterion, reported through volumetric and decomposed metrics, a light, auditable safeguard for operational water management in data-scarce regions. The dry-zone behaviour further suggests that the physical constraint need not use the same response structure in every basin. Perennial flow can be represented by a smooth storage-response prior, whereas intermittent or strongly regulated systems may benefit from regime-aware response functions for dry-period persistence and rapid event activation. As the framework uses contemporaneous observed forcing, its immediate role is continuous simulation and planning support rather than lead-time flood forecasting.

References

- [1] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 11, pp. 6005–6022, 2018.
- [2] K. H. T. Chamika and R. L. H. L. Rajapakse, "Artificial intelligence techniques in basin hydrology and their applicability to Sri Lankan river basins," in *Proc. 2025 Moratuwa Eng. Res. Conf. (MERCon)*, 2025, pp. 1–6.
- [3] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: a deep learning framework for forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.
- [4] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling," *J. Hydrol.*, vol. 377, no. 1–2, pp. 80–91, 2009.
- [5] K. Beven, *Rainfall-Runoff Modelling: The Primer*, 2nd ed. Chichester, U.K.: Wiley, 2012.

ENFORCING THE WATER BALANCE FOR RELIABLE NEURAL STREAMFLOW PREDICTION IN TROPICAL CATCHMENTS

Aim To test whether embedding the catchment water balance makes neural streamflow prediction physically reliable, not just accurate, across a wet and a dry tropical basin

