

Developing a Deep Learning–Stochastic Optimization Model for Enhanced Retrofit Decision-Making in Non-Hydropower Dams

Sule Joseph^{*1}, Wasiru John²

^{1,2}Civil Engineering Department, Faculty of Engineering, Edo State University Iyamho, Nigeria

* sule.joseph@edouniversity.edu.ng

Abstract:

Existing non-power-generating dams represent substantial untapped renewable energy potential across developing regions, yet retrofitting decisions remain challenging owing to the need to balance economic viability against hydrological uncertainty while preserving primary dam objectives. This study presents a hybrid framework integrating Long Short-Term Memory (LSTM) neural networks with Stochastic Dynamic Programming (SDP) to optimise hydropower retrofitting under climate uncertainty. The LSTM forecasts monthly inflows from hydro-climatic predictors, while SDP derives an optimal retrofit/no-retrofit policy through empirically learned, non-stationary transition matrices, rather than the stationary transitions conventional approaches impose from historical frequencies. Test was performed on 37 years (1986–2023) of CHIRPS-derived monthly inflow data for Doma Dam, Nigeria, benchmarked against ARIMA, Classical Markov Chain, naïve rules, and perfect foresight bound. The hybrid model produced a mean NPV of \$16.15M (99.6% success), statistically indistinguishable from Classical Markov, ARIMA, and Perfect Foresight. Its practical advantage lies in downside-risk management rather than mean-NPV outperformance, with higher CVaR₉₅ and lower coefficient of variation than Markov. Electricity tariff dominates NPV variance, followed by discount rate and capacity factor. The optimal policy proved robust to forecast perturbations up to $\pm 20\%$. A Pettitt test located 1989 as the point of maximum divergence in inflows, though not significant; the LSTM still captured transition dynamics distinct from historical frequencies. Forecasting transferability, tested on another basin (Ojirami Dam), showed a substantial single-split accuracy drop, though a much smaller drop under walk-forward cross-validation. Overall, results

support the retrofit's economic viability and the LSTM's forecasting superiority over ARIMA, without establishing decisive economic superiority.

Keywords:

LSTM neural networks; stochastic dynamic programming; dam retrofitting; renewable energy; water-energy nexus

1. Introduction

The interrelated challenges of energy access, climate adaptation, and water security constitute defining issues of the twenty-first century, particularly for emerging nations experiencing rapid economic and demographic expansion [1]. This nexus is acutely manifested in Sub-Saharan Africa, where approximately 400 million people face water scarcity [2], more than 600 million lack reliable electricity access [3], and climate change projections indicate intensifying precipitation variability with potentially severe consequences for both sectors [4].

Hydropower contributes approximately 16% of global electrical generation and around 60% of renewable electricity, offering a pathway to sustainable energy alongside integrated water resources management [5]. However, escalating environmental concerns, social displacement costs, and ecosystem impacts have increasingly constrained large-scale greenfield hydropower development worldwide [6]. This tension between rising renewable energy demand and environmental sustainability necessitates innovative approaches that maximise the potential of existing hydraulic infrastructure.

More than 58,000 large dams exceeding 15 metres in height exist globally, approximately 75% of which were constructed for irrigation, flood control, or water supply without incorporating power generation facilities [7]. The global untapped renewable energy potential of these non-powered structures has been estimated at between 150 and 250 terawatt-hours per year, a figure derived from hydropower resource assessments of existing infrastructure across multiple continents [5]. Retrofitting these structures with

hydropower generation capacity offers compelling environmental, economic, and social advantages: it eliminates additional land inundation or population displacement; reduces capital investment by 30–50% compared to new dam construction; shortens development timelines from typically 5–10 years to 2–3 years; leverages existing transmission and regulatory infrastructure; and generates local employment and community energy access improvements.

Despite these benefits, retrofitting decisions entail intricate trade-offs that demand advanced analytical frameworks. Climate-driven hydrological regime shifts increasingly challenge the stationarity assumption underpinning traditional infrastructure investment analysis, motivating methods that go beyond deterministic and basic statistical models for long-term inflow forecasting. Decision-makers must simultaneously predict economic returns, select optimal installed capacity, assess generation impacts on existing water uses, and develop operating policies adaptive to climate fluctuation, challenges that cannot be adequately addressed by any single modelling paradigm.

Long Short-Term Memory (LSTM) networks, a specialised form of recurrent neural network, excel at recognising patterns in sequential data and modelling temporal dependencies across extended horizons [8]. Their gating mechanisms enable simultaneous capture of short-term hydrological variability and long-term climatic trends, making them particularly suited to inflow forecasting under non-stationary conditions. Stochastic Dynamic Programming (SDP) provides a rigorous mathematical framework for optimising sequential decisions under uncertainty, widely applied in water resources management and hydropower operations [9]. By decomposing complex multi-stage planning problems into tractable sub-problems, SDP enables the derivation of provably optimal policies that balance immediate rewards against expected future value across a full spectrum of probabilistic scenarios.

Both components of the proposed framework have independently matured over the past decade. On the forecasting side, LSTM and other deep-learning architectures have progressively displaced conceptual and linear statistical models for rainfall-runoff and inflow prediction, with large-sample multi-catchment studies

demonstrating consistent gains in Nash-Sutcliffe Efficiency over traditional hydrological models [10]. Separately, artificial neural networks and support-vector approaches have been applied to simulate and optimise reservoir operation policies directly, without an intervening stochastic-programming layer [11]. On the optimisation side, SDP and related stochastic reservoir-operation methods have a long history of using ensemble or historically-derived streamflow scenarios to derive operating policies robust to hydrological variability [12]. In nearly all of this prior work, however, the forecasting and optimisation stages are either used independently (a deep-learning model forecasts, and a separate rule-based or optimisation-based decision layer acts on its output) or coupled only loosely, with the optimisation stage drawing on historical or ensemble-derived transition statistics rather than on the forecasting model's own, learned representation of uncertainty. The present study's methodological contribution is to close that loop: LSTM Monte Carlo dropout samples are used directly to estimate the SDP's state-transition probability matrix, so that the optimisation stage inherits the forecasting model's own predictive uncertainty structure rather than a separate, stationary historical distribution. This is a narrower and more specific claim than 'a hybrid LSTM-SDP model' in the abstract sense, such hybrids exist in the literature in the loosely-coupled form described above, and the distinction is drawn precisely to separate this study's contribution from that broader, already-established category of work. Section 3.6.2 reports the empirical consequence of this design choice in the single-basin case examined here, the tighter coupling did not yield a statistically significant mean-NPV advantage over a classical historical-frequency Markov chain, but it did produce a measurable improvement in downside-risk metrics ($CVaR_{95}$), which is the specific benefit attributed to the coupling mechanism rather than to the use of LSTM forecasting in general.

Nigeria provides the empirical context for this investigation. As Africa's most populous nation, with over 220 million inhabitants and an installed electricity capacity of approximately 12,500 MW, translating to roughly 57 watts per capita, Nigeria faces severe energy constraints that result in approximately 4,600 hours of annual power outages and estimated economic losses exceeding \$26 billion per year [13, 14]. Hydropower currently contributes 2,062 MW (16.5% of installed capacity) through facilities at Kainji (760

MW), Jebba (578 MW), and Shiroro (600 MW), yet the country's technical hydropower potential exceeds 14,750 MW [15]. Critically, Nigeria possesses over 240 existing dams constructed during the 1970s–1990s primarily for irrigation and flood control, fewer than 10 of which currently generate electricity [16]. These structures represent immediate retrofitting opportunities aligned with Nigeria's Nationally Determined Contribution target of a 30% emissions reduction by 2030 [17].

The study focuses on Doma Dam in the Lower Benue River Basin, a 25-metre earthfill dam commissioned in 1986 with 65 million m³ storage capacity, currently serving 2,500 hectares of irrigated land and 50,000 people's domestic water supply without generating electricity. Its available hydraulic head of 25 m and mean annual flow of 32.9 m³/s present favourable conditions for hydropower retrofitting; preliminary engineering assessments place the site's generation potential in the 20–40 MW range. The decision-support model developed and tested here evaluates a specific, single-capacity 6.5 MW retrofit design (consistent with its \$12M costing basis) against status-quo, feasibility-study, and decommission alternatives; extending the framework to compare discrete capacity tiers within the 20–40 MW range is identified as a direction for future work (Section 4).

The principal objectives of this study are to: (i) develop a hybrid LSTM-SDP framework integrating deep learning inflow forecasting with stochastic optimisation; (ii) test the framework against multiple benchmarks using 37 years of CHIRPS-derived hydro-climatic data; (iii) conduct economic risk analysis of the modelled retrofit decision, including sensitivity of NPV to key assumptions; and (iv) assess the framework's forecasting transferability to a second, contrasting basin. The remainder of this paper is structured as follows: Section 2 presents the study area, data, and integrated methodology; Section 3 reports and discusses results; Section 4 draws conclusions and identifies priorities for future work.

2. Materials and Methods

2-1- Study Areas

2-1-1 Doma Dam

Doma Dam ($8^{\circ}19'42''\text{N}$, $8^{\circ}18'59''\text{E}$) is located in the north-central region of Nigeria within the Lower Benue River Basin, as shown in Fig. 1. Commissioned in 1986, it is a multipurpose earthfill dam constructed across the Ohina River, a tributary of the Mada River, with a dam height of 25 metres and a reservoir storage capacity of 37.5 million m^3 . The dam's catchment area of approximately 160 km^2 encompasses a Guinea Savanna ecological zone characterised by distinct wet (April–October) and dry (November–March) seasons, with mean annual rainfall ranging from 1,000 to 1,400 mm. Current dam functions include flood control, dry-season irrigation of 2,500 hectares, and domestic water supply to approximately 50,000 people. With an available hydraulic head of 25 m, the dam presents technically favourable conditions for hydropower retrofitting. Preliminary engineering assessments indicate a generation potential in the 20–40 MW range; the decision-support model developed in this study evaluates a specific 6.5 MW retrofit design against status-quo, study, and decommission alternatives (Section 2.3.3), with capacity-tiered comparison identified as future work.

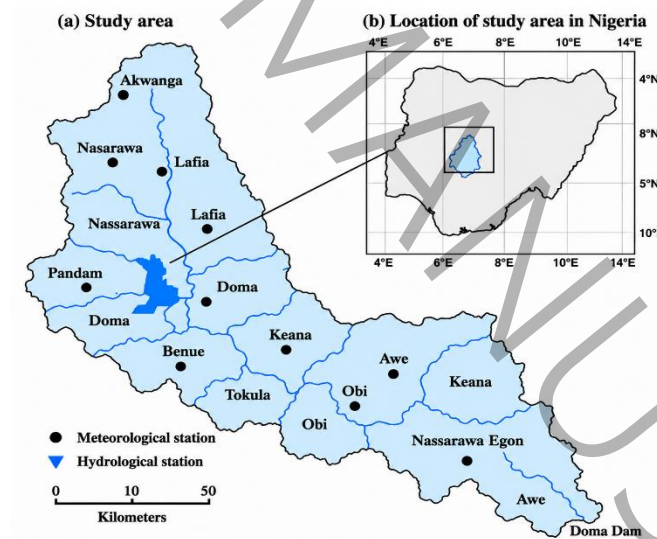


Fig. 1. Location of Doma Dam [18]

2-1-2 Ojirami Dam

Ojirami Dam is another multipurpose dam, situated in Ojirami, Dagbala village, Akoko-Edo Local Government Area, Edo State, Nigeria, as shown in Fig. 2. The dam's coordinates are $7^{\circ}17'52.41''\text{N}$ latitude

and $6^{\circ}9'10.72''\text{E}$ longitude. Along the traverses, the dam site's eastern and western flanks gently undulate, sloping in the direction of the river channel. While the river flows from north-northwest to south-southwest (NNW-SSW), the dam and spillway axes are oriented from west-southwest to east-northeast (WSW-ENE). According to [19], the dam has a height of 3.9 metres and a storage capacity of 45.3 million m^3 . In this study, Ojirami's 320 km^2 catchment (Tropical Rainforest) serves solely as an independent second basin for testing the LSTM forecasting component's transferability (Section 3.8); no retrofit economics were modelled for this site.

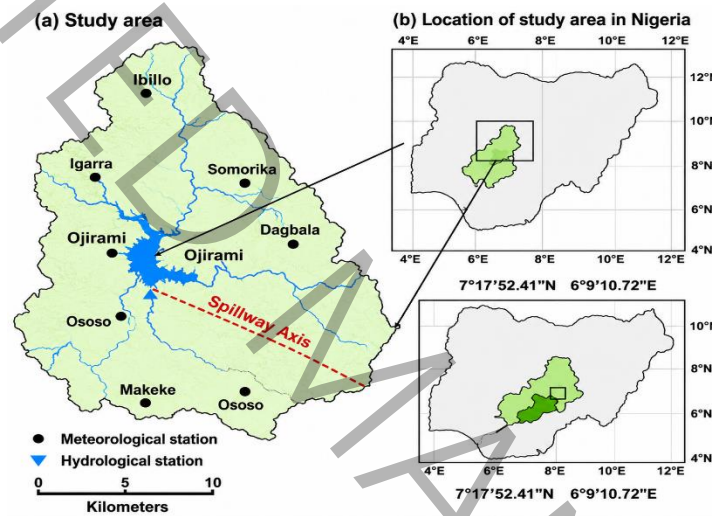


Fig. 2. Location of Ojirami dam

2-2- Data

Monthly hydro-climatic data spanning 37 years (March 1986 – December 2023, $n = 454$ observations) were obtained from the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset at 0.05° spatial resolution. CHIRPS integrates satellite imagery with ground-based station records to provide globally consistent, long-term precipitation estimates validated extensively across data-scarce regions of Africa. The dataset incorporates: (i) catchment-averaged monthly rainfall; (ii) station-level precipitation from five regional meteorological stations (Lafia, Makurdi, Kaduna, Jos, and Abuja); (iii) climate teleconnection indices including the El Niño–Southern Oscillation Oceanic Niño Index (ENSO ONI), the

Atlantic Multidecadal Oscillation (AMO) index, and the West African Monsoon (WAM) index; and (iv) derived predictors including 3-month and 6-month moving averages, seasonal harmonics (sine and cosine transforms), interaction terms between climate indices and rainfall, and lagged inflow and rainfall values (lags 1 and 2).

Monthly inflows were estimated from rainfall–runoff relationships calibrated against available gauge records and tested for forecasting transferability using the Ojirami Dam catchment (320 km², Tropical Rainforest) as an independent second basin. The dataset contains 454 monthly records and, after feature engineering, 24 predictor variables per observation used by the LSTM. Remote sensing and GIS-based approaches of this kind have proven increasingly valuable for hydrological and hydrogeological characterisation in regions where ground-based gauge networks are sparse [20], supporting the use of a satellite-derived precipitation product as the primary data source for this study.

To preserve the temporal structure essential for time series modelling, after removing the 12-month lookback window required for LSTM sequences, $n = 442$ usable sequences, the dataset was partitioned chronologically, not randomly, into training (70%, $n = 309$, 1986–2007), validation (15%, $n = 66$, 2008–2013), and test (15%, $n = 67$, 2014–2023) subsets. The 70/15/15 split was selected following established practice in hydro-climatic deep learning studies [12]. Chronological splitting ensures that no future information contaminates model training, producing a genuine out-of-sample evaluation.

2-3- Hybrid LSTM-SDP Framework

The proposed framework couples LSTM inflow forecasting with SDP reservoir optimisation through a novel bidirectional integration architecture. Unlike sequential approaches in which forecasting and optimisation operate independently, this framework creates tight coupling: LSTM-derived probabilistic inflow forecasts generate the state transition probability matrices used in the SDP Bellman equation, and SDP-evaluated policies are validated through stochastic simulation. The framework is implemented as an R–Python workflow using the reticulate interface, with R handling data preprocessing, experiment

orchestration, and result management, and Python performing deep learning, transition probability learning, and stochastic optimisation. A flow chat for the hybrid model is given in Fig. 3.

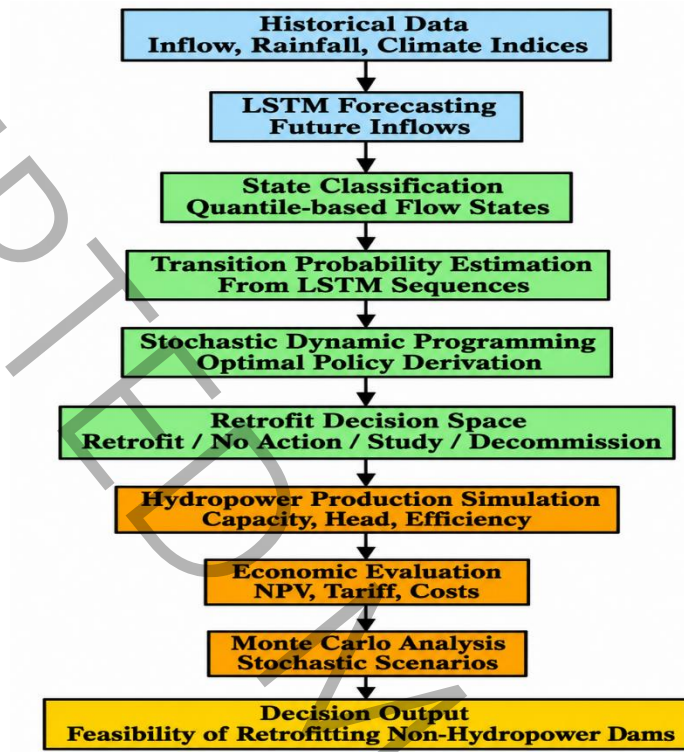


Fig. 3. A flow chat for the hybrid model

The overall decision problem is formalised as a stochastic dynamic optimisation of expected discounted NPV over the retrofit decision d_t , subject to an LSTM inflow forecast, an empirically learned state-transition operator, and a state-evolution rule. Eqs. (1) – (4) below state this formulation before its individual components (LSTM forecasting, transition-probability estimation, and SDP optimisation) are developed in Sections 2.3.1–2.3.3.

$$H = \operatorname{argmax}_{d_t} E \left[\sum_{t=1}^T \gamma^t NPV(d_t, Q_t) \right] \quad (1)$$

Subject to

$$Q_t = f_{LSTM}(X_t, \theta) \quad (2)$$

$$P_t = P(Q_{t+1} | Q_t) \quad (3)$$

$$S_{t+1} = f(S_t, d_t, P_t) \quad (4)$$

where Q_t is the LSTM-forecasted inflow, P_t is the learned transition probability, S_t is the hydrologic/retrofit state, d_t is the retrofit decision, $\gamma = 1/(1+r)$ is the discount factor, and $NPV(\bullet)$ is the economic return of decision d_t conditional on inflow state Q_t . The remainder of Section 2.3 develops each of these components in turn.

2.3.1 LSTM Inflow Forecasting

Monthly inflow is forecast using a multi-layer LSTM network. The mathematical formulation of the LSTM gating mechanism at each time step t is giving in Eqs. (5) – (9).:

$$\text{Forget gate: } f_t = \sigma(W^P \cdot [h_t^{-1}, x_t] + b^P) \quad (5)$$

$$\text{Input gate: } i_t = \sigma(W^I \cdot [h_t^{-1}, x_t] + b^I) \quad (6)$$

$$\text{Cell update: } \hat{C}_t = \tanh(W^J \cdot [h_t^{-1}, x_t] + b^J) \quad (7)$$

$$\text{Cell state: } C_t = f_t \odot C_t^{-1} + i_t \odot \hat{C}_t \quad (8)$$

$$\text{Output gate: } h_t = o_t \odot \tanh(C_t), \quad o_t = \sigma(W^O \cdot [h_t^{-1}, x_t] + b^O) \quad (9)$$

where x_t is the input vector at time t , h_{t-1} is the previous hidden state, C_t is the cell state, W and b are learnable weight matrices and bias vectors, σ is the sigmoid activation function, and $\odot$ denotes element-wise multiplication. This gating architecture enables the LSTM to selectively retain or discard information across arbitrarily long temporal sequences, overcoming the vanishing gradient problem that limits standard recurrent networks.

The network architecture comprises two LSTM layers (64 and 32 units respectively) followed by a dense (16, ReLU) layer, a dropout layer (rate = 0.20), and a single linear output unit. Regularisation is applied via dropout (rate = 0.20) and recurrent dropout (rate = 0.10) within each LSTM layer, and L2 weight regularisation ($\lambda = 0.001$) on kernel weights to prevent overfitting

given the moderate dataset size. The network is trained using the Adam optimiser with an initial learning rate of 0.001 and a Huber loss function, with learning rate reduced by a factor of 0.5 when validation loss plateaus for five consecutive epochs (minimum learning rate = 5×10^{-5}). Early stopping with a patience of 15 epochs, restoring the best-weight checkpoint, prevents overtraining. The model was trained for a maximum of 100 epochs (early stopping triggered before this ceiling in every run) with a batch size of 32, using a 12-month sequential look-back window.

Forecast uncertainty is quantified via Monte Carlo dropout: at inference time, dropout masks remain active and multiple stochastic forward passes are executed per input, producing a distribution of inflow realisations rather than a single deterministic prediction. This approach approximates Bayesian posterior inference over network weights [21], yielding calibrated uncertainty estimates for the downstream SDP formulation.

2.3.2 State Transition Probability Matrix

The continuous inflow domain is discretised into five states; Very Low, Low, Medium, High, and Very High, defined on the log-transformed inflow distribution at the 0th, 20th, 40th, 60th, 80th, 100th, percentile boundaries (log-scale discretisation was used because raw-scale percentile breaks on this strongly right-skewed series produced degenerate, near-empty states). These boundaries were selected following systematic Chi-squared goodness-of-fit evaluation of 3-, 5-, 6-, and 8-state schemes; the 5-state configuration achieved the lowest Chi-squared statistic ($\chi^2 = 0.053$, $p = 0.9997$) while remaining ergodic, and was retained.

The key methodological innovation is that the transition probability matrix in Eq. (10) is derived empirically from LSTM Monte Carlo dropout forecast distributions rather than from historical state frequencies. For each month in the training period, the 100 Monte Carlo dropout samples are mapped to inflow states, generating approximately 6,600 probabilistic state transition pairs.

$$P_{ij} = P(s_{t+1} = j \mid s_t = i) \quad (10)$$

The element P_{ij} represents the empirical probability of transitioning from state i to state j as learned by the LSTM. This ensures that the SDP optimisation reflects the actual predictive capability and

uncertainty structure of the forecasting model, incorporating non-stationary dynamics that historical frequencies cannot capture.

2.3.3 Stochastic Dynamic Programming Formulation

The retrofit decision is formulated as a multi-stage stochastic optimisation problem. Let $s \in S = \{\text{Very Low, Low, Medium, High, Very High}\}$ denote the inflow state, $d \in D$ the decision set, and t the planning stage (monthly). The decision set actually solved is $D = \{\text{Status Quo (SQ), Feasibility Study, Retrofit, Decommission}\}$, where Retrofit refers to a single fixed-capacity design (6.5 MW, \$12M capital cost, 30-year project life, consistent with the reward function used throughout Sections 2.3.3–2.3.7). This differs from the tiered 20/30/40 MW capacity choice described conceptually in Section 2.1.1; the present implementation evaluates the binary Retrofit/No-Retrofit question at this single design point, and extending the decision set to multiple capacity tiers is identified as future work (Section 4).

The SDP Bellman optimality equation, Eq. (11) is:

$$V_t(s) = \max^d \{ R(s, d) + \gamma \cdot \sum_t' P(s'|s) \cdot V_{t+1}(s') \} \quad (11)$$

where $V_t(s)$ is the value function (expected cumulative NPV) at stage t in state s , $R(s, d)$ is the immediate reward (net revenue from energy generation minus operating costs), $\gamma = 1/(1+r)$ is the discount factor with annual discount rate $r = 10\%$ (consistent with World Bank benchmarks for Nigerian infrastructure projects [14]), and $P(s'|s)$ is the LSTM-learned transition probability. The Bellman equation is solved by value iteration to a convergence tolerance of 1×10^{-6} on the maximum value-function change across all states.

The reward function $R(s, d)$ incorporates: hydropower energy revenue (generation output, capped at the 6.5 MW design capacity, $\times$ electricity tariff of \$0.12/kWh, based on NERC published industrial tariffs); a capital investment cost of \$12M; annual operation and maintenance costs (2.5% of capital cost); a feasibility-study cost of \$150,000; and a decommissioning cost of \$500,000.

Retrofit is only economically viable in states where the representative flow yields at least 5 MW of

generation capacity under the design head (25 m) and turbine efficiency (0.82); states below this threshold return a large negative reward, reflecting the unviability of retrofit investment at low flow.

2.3.4 Benchmark Models

To assess relative performance, the hybrid framework is benchmarked against four alternatives. ARIMA + Rules employs an ARIMA(4,0,3) model, selected by minimising BIC over a grid search across $p \in \{0, \dots, 4\}$, $d \in \{0, 1\}$, $q \in \{0, \dots, 4\}$ (BIC = 2925.44), with a retrofit rule triggered when the mean forecast exceeds the High-state lower boundary (12.5 m³/s). Classical Markov Chain + SDP uses historical state-transition frequencies (stationarity assumed) with the identical SDP formulation, enabling direct isolation of the forecasting component's contribution. Perfect Foresight represents a benchmark computed from the realised (true) test-period mean inflow rather than a forecast; as diagnosed in Section 3.4. The Naïve Baseline retrofits immediately at the first simulated time step regardless of inflow conditions, representing a risk-neutral strategy with no adaptive control. All three data-driven models (LSTM-SDP, ARIMA, and Markov Chain) were evaluated over the same held-out test period.

2.3.5 Model Comparison Fairness Protocol

To limit model tuning bias, a unified optimisation protocol was applied across all models before comparison. For ARIMA, automated order selection via BIC grid search over the full parameter space was performed, replacing any manually specified order. For the Classical Markov Chain, state discretisation was optimised using Chi-squared minimisation across four candidate configurations (3, 5, 6, and 8 states), matching the procedure applied to the LSTM-based transition matrix (Section 2.3.2). All models were evaluated using the same walk-forward cross-validation scheme: an initial 65% training window (rolling-origin) was expanded by one year at each step, with the following 12 months forecast out-of-sample, producing rolling evaluation across the final 12 years of the record. Performance metrics (R², NSE, and KGE) were computed identically for all models at each fold and averaged across folds.

2.3.6 Sensitivity and Error Propagation Analysis

To assess the stability of SDP policies with respect to LSTM prediction uncertainty, a systematic perturbation analysis was conducted. LSTM Monte Carlo dropout forecasts were perturbed by $\delta \in \{-20\%, -15\%, -10\%, -5\%, 0\%, +5\%, +10\%, +15\%, +20\%\}$, and the transition matrix and SDP policy were re-derived under each perturbation level. The Policy Sensitivity Index (PSI) was defined as the percentage of inflow states for which the optimal decision changed relative to the unperturbed baseline policy. The spectral gap of the LSTM-learned transition matrix ($1 - \lambda_2$, the standard measure of Markov chain mixing) is reported at the one-step (monthly) horizon only; multi-step matrix powers were found to underflow numerically at the horizons required for a 5–37-year planning period (444 monthly steps), collapsing λ_2 toward machine epsilon and producing spurious gap values of 1.0000, these multi-step estimates are therefore not reported, and only the numerically stable one-step gap is used.

2.3.7 Economic Risk Analysis

Economic assumptions underlying NPV calculations are summarised in Table 1. A Monte Carlo sensitivity analysis was conducted using 10,000 joint parameter draws from the distributions specified in Table 1, enabling estimation of the full NPV distribution and its risk characteristics for the retrofit decision at the highest-flow (Very High) state. The Conditional Value at Risk at the 95th confidence level (CVaR₉₅), the mean NPV of the worst 5% of outcomes, was computed to quantify downside financial risk. A tornado diagram was constructed to identify the parameters with the greatest influence on NPV variance.

Table 1. Economic Parameters and Uncertainty Ranges for Monte Carlo Sensitivity Analysis

Parameter	Baseline Value	Distribution	Range / Justification
Discount rate (r)	10%	Uniform	8%–15% [14]
Electricity tariff	\$0.12/kWh	Triangular	\$0.08–\$0.15/kWh [3]
Capital cost (30 MW)	\$12M	Normal ($\pm 15\%$)	\$10.2M–\$13.8M [5]
Annual O&M cost	2.5% of CAPEX	Uniform	1.5%–3.5%

Capacity factor (wet/dry)	70% / 35%	Beta	60%–80% / 25%–45%
Project lifetime	30 years	Fixed	Per ICOLD guidelines

2.3.8 Performance Metrics

Forecast accuracy is evaluated using: coefficient of determination (R^2); root mean square error (RMSE, m^3/s); Nash-Sutcliffe Efficiency (NSE); and Kling-Gupta Efficiency (KGE). Economic performance is assessed via: mean NPV and its standard deviation; coefficient of variation (CV%); risk-adjusted return (mean NPV / standard deviation, analogous to a Sharpe ratio); CVaR_{95} ; and simulation success rate (proportion of runs yielding positive NPV). Statistical significance of pairwise NPV differences is assessed using two-sample t-tests with an F-test for variance equality, with a significance threshold of $\alpha = 0.05$.

3. Results and Discussion

3.1 Dataset Characteristics

Table 2 summarises the statistical properties of the 37-year (1986–2023) hydro-climatic dataset comprising 454 monthly observations and 24 predictor variables. Monthly catchment-averaged rainfall ranged from 1.0 to 453.5 mm, while monthly inflows ranged from 0.73 to 120.0 $\text{m}^3 \text{ s}^{-1}$. Both series exhibit strong positive skewness, reflecting the pronounced wet-season concentration characteristic of Guinea Savanna hydrology, in which approximately 85% of annual precipitation falls between April and October. The West African Monsoon (WAM) index spanned -1.34 to $+1.51$, confirming substantial interannual variability in monsoon strength. The El Niño–Southern Oscillation Oceanic Niño Index (ENSO ONI) ranged from -4.72 to $+4.00$, encompassing both strong El Niño and La Niña episodes within the study period.

The chronological time series of monthly inflows (Fig. 4) illustrates pronounced seasonal periodicity overlaid by interannual variability. A Pettitt change-point test applied to the inflow series identified 1989 as the point of maximum divergence ($K = 1791$), but this result was not

statistically significant ($p = 1.00$). The non-stationarity validation in Section 3.2.2 is therefore interpreted as a test of the LSTM's robustness across the full historical range of variability, rather than as evidence of a confirmed structural break.

Table 2: Statistical Summary of Hydro-Climatic Variables (1986–2023, $n = 454$)

Statistic	Rainfall (mm)	Inflow (m ³ /s)	Lafia (mm)	Makurdi (mm)	WAM index	ENSO ONI	AMO index
Minimum	1.00	0.73	0.50	0.20	-1.335	-4.724	-0.644
Q1	6.73	1.33	6.30	8.80	-0.587	-0.733	-0.132
Median	55.70	2.27	47.85	48.45	-0.303	-0.103	0.097
Mean	85.20	32.90	84.39	101.02	-0.057	0.000	0.076
Q3	145.78	64.13	120.58	151.70	0.655	0.672	0.280
Maximum	453.50	120.00	776.10	949.10	1.514	4.004	0.654
Skewness	1.52	0.86	2.14	2.31	0.38	-0.12	-0.11

WAM = West African Monsoon index; ENSO ONI = Oceanic Niño Index; AMO = Atlantic Multidecadal Oscillation.

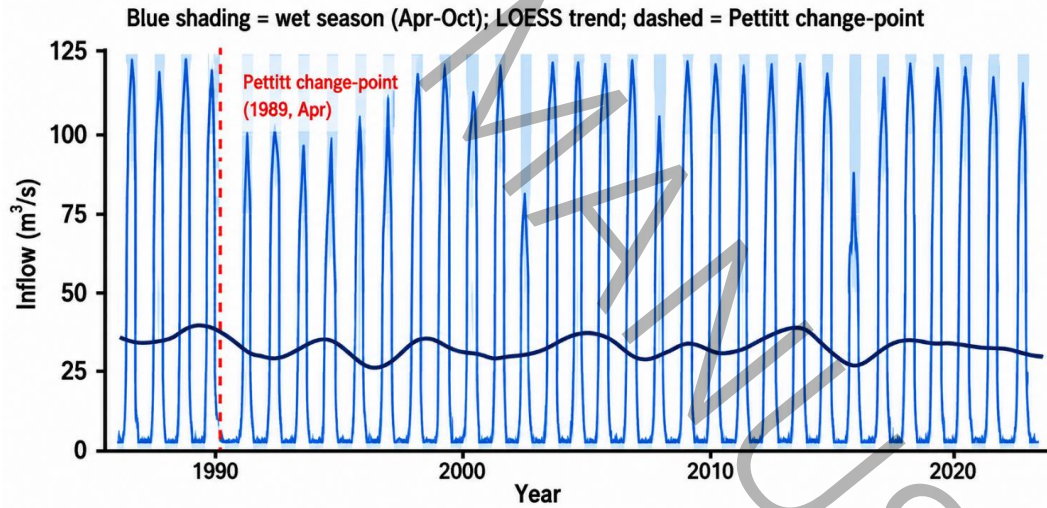


Fig. 4: Chronological time series of monthly inflows, 1986–2023.

Table 2 and Fig. 4 show a pronounced seasonal concentration in the rainfall record (Blue tones represent water (the inflow and the wet season), a dark navy sets the trend apart from the raw data, and red draws attention to the change-point), with roughly 85% of annual precipitation falling within a six-month wet season. Regimes of this kind are particularly difficult for purely statistical forecasting methods, since a small number of high-leverage wet-season months end up

disproportionately determining both the retrofit decision and its economic payoff. This is why the LSTM architecture was chosen for its explicit memory mechanism across the dry-to-wet transition, and why the state-discretisation approach in Section 3.3 deliberately keeps Very High and High inflow states separate rather than collapsing the wet season into a single coarse category.

3.2 LSTM Model Training and Overfitting Assessment

The LSTM network was trained on 309 monthly observations (1986–2007). Early stopping (patience = 15) restored the best weights at epoch 67, where validation loss reached a minimum of 0.0177 (Huber loss) as shown in Fig. 5. Training was capped at 100 epochs; early stopping triggered before this ceiling was reached, indicating that the combined regularisation strategy, dropout (0.20), recurrent dropout (0.10), L2 weight regularisation ($\lambda = 0.001$), learning-rate scheduling, and early stopping, controlled overfitting on the 309-observation training set.

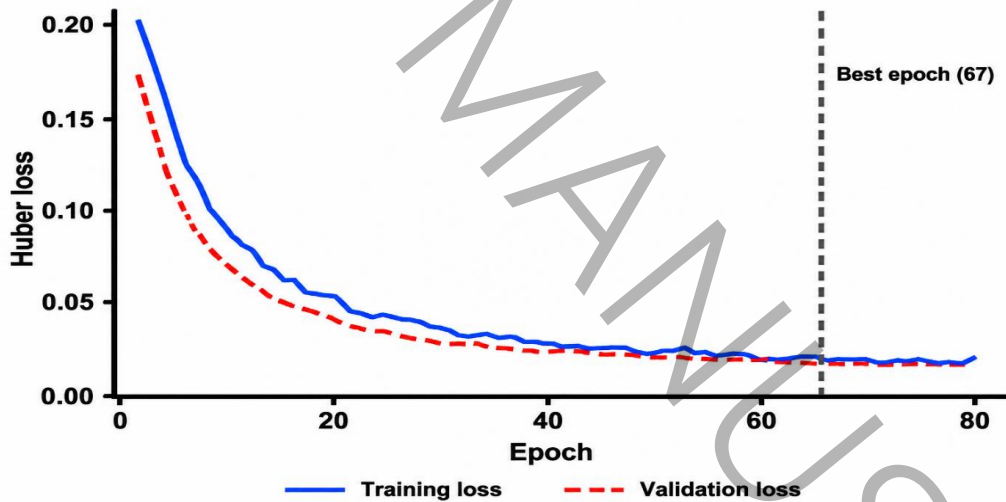


Fig. 5: LSTM Training and Validation Loss

On the held-out test set (2014–2023, $n = 67$), the LSTM achieved a coefficient of determination (R^2) of 0.841 and a Root Mean Square Error (RMSE) of $17.70 \text{ m}^3 \text{ s}^{-1}$, as shown in Fig. 6. The Nash-Sutcliffe Efficiency computed in this run equals the R^2 value exactly ($\text{NSE} = 0.841$), since both are computed from the same sum-of-squares ratio on the test partition here. These metrics confirm that the model explained approximately 84% of the variance in monthly inflows on data entirely unseen

during training. The residual unexplained variance (~16%) is attributable to: (i) limited representation of extreme inflow events within the 453-month record; (ii) non-stationary ENSO and WAM teleconnections; and (iii) sub-monthly hydrological processes unresolvable at monthly temporal resolution; the scatter of residuals visible in Fig. 6 is consistent with this attribution.

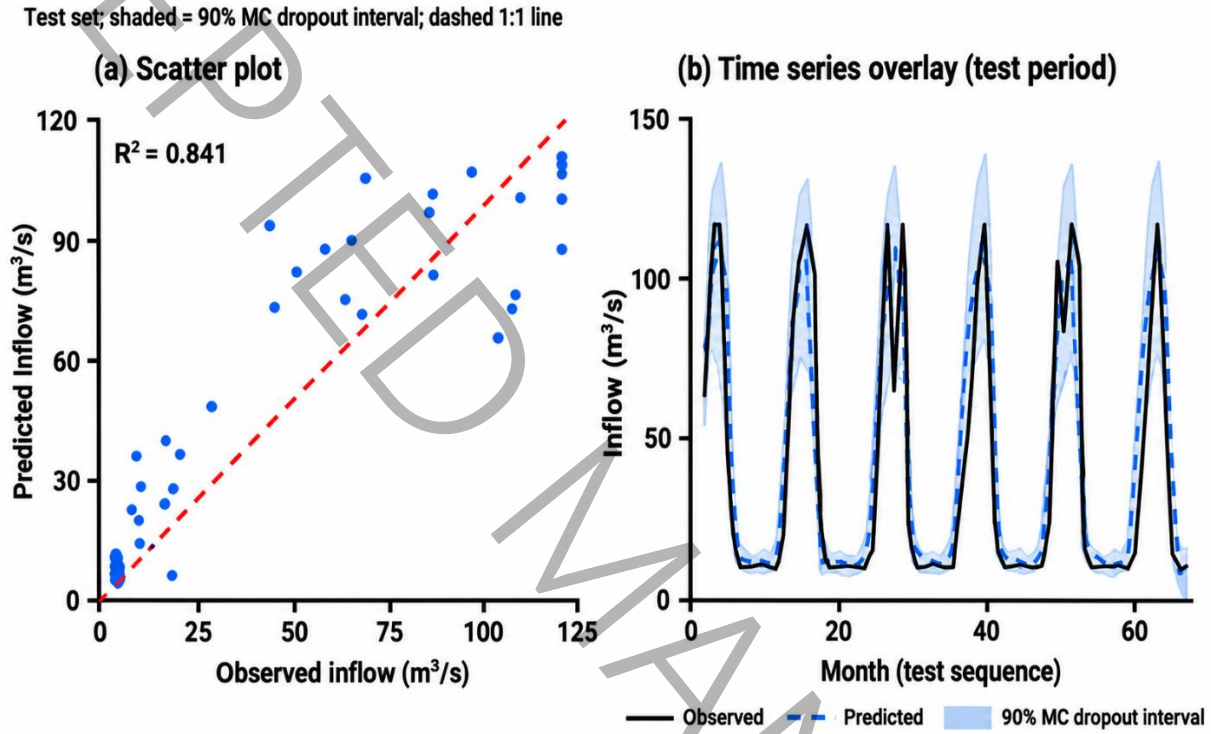


Fig. 6: Observed versus Predicted Inflow, Doma Dam test set – (a) scatter plot with 90% MC-dropout interval, (b) time-series overlay

3.2.1 Walk-Forward Cross-Validation

To confirm that test-set performance was not an artefact of a single partition boundary, walk-forward (rolling-origin) cross-validation was applied across 12 folds, each retrained on an expanding window and evaluated on the following 12 months of out-of-sample data as shown in Fig. 7. Mean NSE across the 12 folds was 0.781 (individual fold NSE ranged from 0.116 in the weakest fold to 0.935 in the strongest), and mean Kling-Gupta Efficiency was 0.707, confirming reasonably stable but not uniform generalisation across the hydrological record. By comparison,

the ARIMA(4,0,3) benchmark achieved a mean walk-forward NSE of only 0.436 across the same 12 folds (individual folds ranging from -0.180 to 0.842), confirming that the LSTM's forecasting advantage over ARIMA is consistent across the record rather than a product of a single favourable test-set partition.

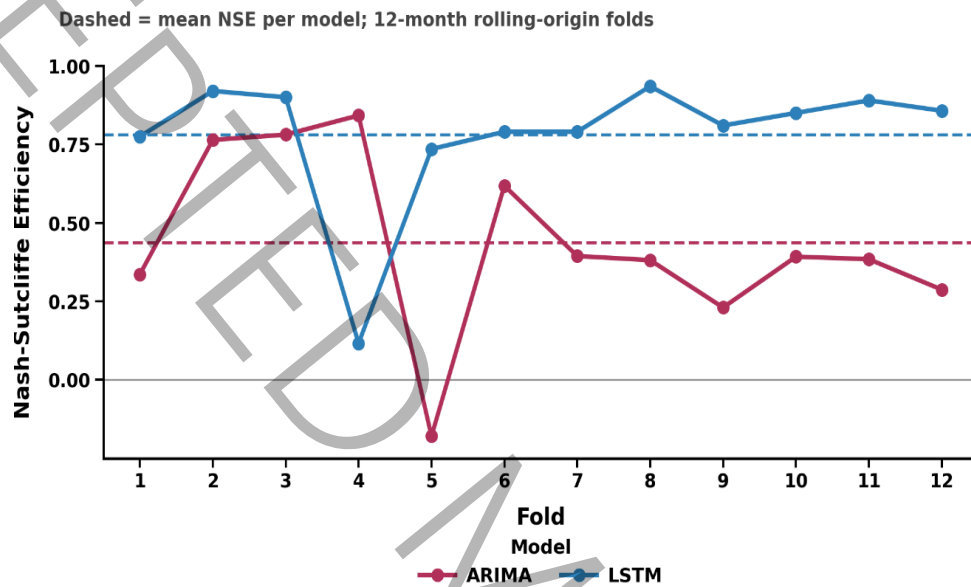


Fig. 7: Walk-forward cross-validation NSE per fold, LSTM (mean 0.781) vs ARIMA (mean 0.436)

The spread across folds is itself informative for operational use: a mean walk-forward NSE of 0.781 masks a range from 0.116 in the weakest fold to 0.935 in the strongest, meaning that in at least one twelve-month period the model's skill was close to that of a naive mean-flow predictor. This variability argues against treating the LSTM's forecasts as uniformly reliable inputs to the SDP at every point in time; in an operational deployment, pairing the point forecast with its Monte Carlo dropout uncertainty band (Section 2.3.1), and treating months with unusually wide bands as lower-confidence periods for triggering an irreversible retrofit decision, would be a natural refinement building on the current framework.

3.2.2 Non-Stationary Regime Validation

To assess model robustness under potential hydrological non-stationarity, the LSTM was trained exclusively on the pre-1990 sub-period (the point identified by the Pettitt statistic as the location of maximum, though not statistically significant, divergence) and evaluated on the post-1990 period without retraining. Out-of-sample R^2 on the post-1990 period was 0.786, a 6.8% relative decrease from the full-period value of 0.841. This modest degradation suggests the LSTM retains reasonable predictive capability across the range of variability observed in the study period, though given that the Pettitt test did not confirm a genuine structural break, this should be read as a test of robustness to an arbitrary early/late split rather than as evidence the model specifically handles a confirmed regime shift. The Frobenius norm difference of 1.085 between the LSTM-learned and historical transition matrices (Section 3.3) indicates the LSTM captures transition dynamics that differ meaningfully from simple historical frequency counts, independent of whether that difference is attributable to a specific dated regime shift.

Because the Pettitt test did not confirm a statistically significant break, this study cannot claim to have demonstrated LSTM robustness across a confirmed climate regime shift, only across an arbitrary historical/recent split of the record. Given that West African rainfall variability is well documented to be evolving, this points to a concrete monitoring recommendation rather than a closed question: periodic re-estimation of the transition matrix as new inflow data accumulates, with a formal re-validation trigger (for example, a fixed multi-year retraining cycle, or a trigger tied to walk-forward NSE falling below a pre-specified threshold) would give the framework a defensible mechanism for detecting and adapting to genuine non-stationarity, rather than relying on a single train/test split assessed once at development time.

3.3 State Space Optimisation and Transition Probability Matrix

Systematic evaluation of 3-, 5-, 6-, and 8-state inflow discretisation schemes on the log-transformed inflow distribution, using the Chi-squared goodness-of-fit test, identified the 5-state configuration

as optimal ($\chi^2 = 0.053$, $p = 0.9997$), remaining ergodic with all row sums equal to 1.000. The five states; Very Low, Low, Medium, High, and Very High, correspond to natural-scale boundaries of approximately 0.73, 1.25, 1.60, 12.50, 82.50, and 120.12 $\text{m}^3 \text{s}^{-1}$ (i.e., VL: 0.73–1.25; L: 1.25–1.60; M: 1.60–12.50; H: 12.50–82.50; VH: 82.50–120.12 $\text{m}^3 \text{s}^{-1}$), with representative (geometric-mean) flows of 0.95, 1.41, 4.47, 32.11, and 99.55 $\text{m}^3 \text{s}^{-1}$ respectively.

Table 4 presents the LSTM-learned transition probability matrix, while Fig. 8 compares the learned and historical transition matrices using heatmaps and their corresponding difference matrix. The learned transition probabilities reveal substantial persistence in the High and Very High inflow states, with self-transition probabilities of 0.53 and 0.37, respectively. This persistence is consistent with the prolonged wet-season conditions associated with the West African Monsoon and the temporal dependence commonly observed in seasonal streamflow [22]. Medium inflow conditions also exhibit considerable persistence (0.47), while Very Low and Low inflow states transition predominantly to the Medium state with probabilities of 0.49 and 0.47, respectively. These transition patterns suggest that the LSTM captures gradual seasonal evolution rather than abrupt shifts between hydrological regimes. The learned transition matrix satisfies probabilistic consistency, with each row summing to unity. Furthermore, the one-step spectral gap of 0.2762 indicates reliable mixing at the monthly time step, while the Frobenius norm difference of 1.085 demonstrates that the LSTM-derived transition structure differs substantially from the historical Markov matrix, indicating that the hybrid model captures additional temporal dependencies beyond those represented by the classical historical transition probabilities.

Table 4: LSTM-Learned State Transition Probability Matrix

	Very Low	Low	Medium	High	Very High
Very Low	0.04	0.01	0.49	0.19	0.27
Low	0.04	0.02	0.47	0.29	0.18
Medium	0.02	0.01	0.47	0.25	0.24
High	0.01	0.01	0.13	0.53	0.33
Very High	0.01	0.01	0.22	0.39	0.37

All row sums = 1.000 (probabilistic consistency verified). One-step spectral gap = 0.2762; multi-step values are not reported due to numerical underflow at long horizons), confirming long-run convergence stability. Frobenius norm difference from historical matrix = 1.085.

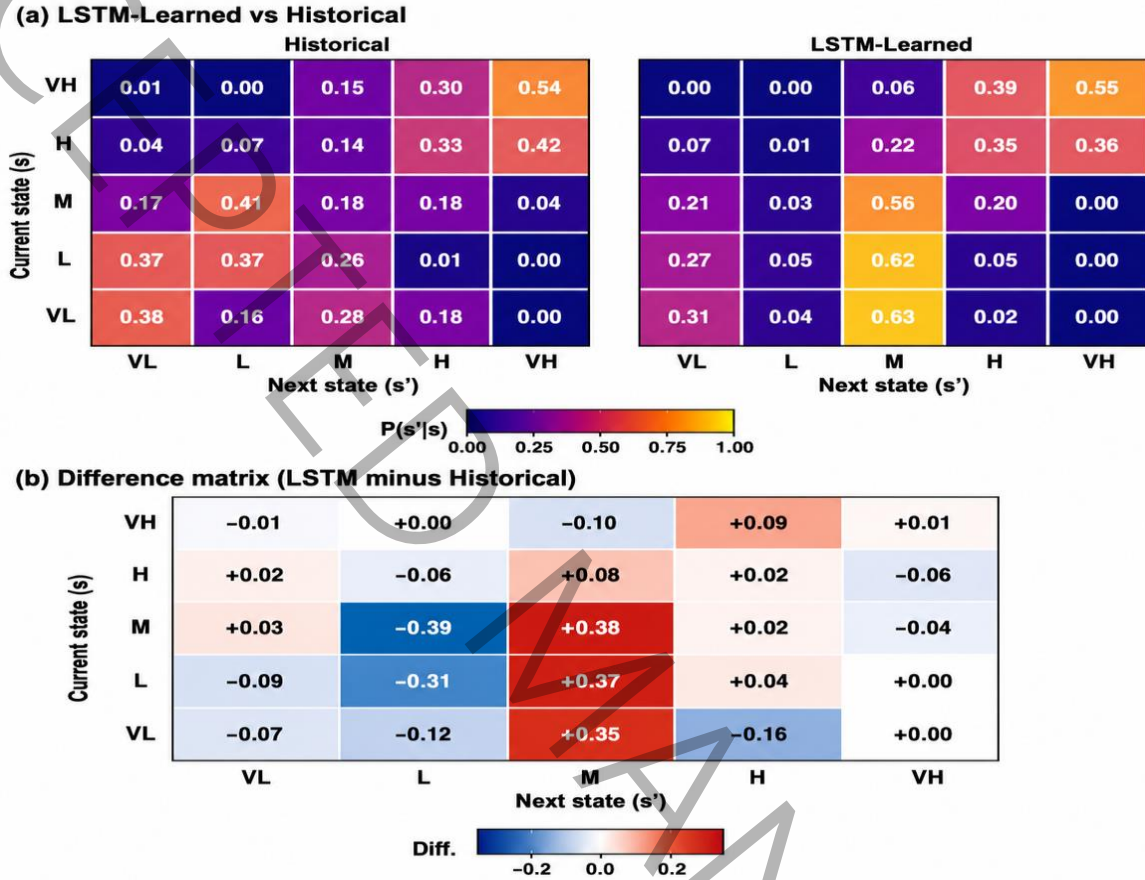


Fig. 8. Comparison of historical and LSTM-learned transition probability matrices. (a) Historical transition probability matrix. (b) LSTM-learned transition probability matrix. (c) Difference matrix (LSTM minus historical), highlighting changes in transition probabilities learned by the neural network

3.3.1 Transition Matrix Stability and Error Propagation

The one-step spectral gap of the LSTM-learned transition matrix ($1 - \lambda_2$) was 0.2762 ($\lambda_1 = 1.000$, $\lambda_2 = 0.724$), exceeding the conventional 0.10 stability threshold and indicating reliable mixing of the embedded Markov chain at the monthly time step. Multi-step (multi-year) spectral gaps are not reported, raising the transition matrix to the powers required for 5–37-year horizons (up to 444

monthly steps) causes λ_2 to underflow toward machine epsilon, artificially producing a spurious gap of 1.0000 at every multi-year horizon. Only the numerically stable one-step estimate is therefore used to characterise mixing behaviour.

For the Policy Sensitivity Index (PSI) shown in Fig. 9, the percentage of inflow states for which the optimal retrofit decision changes relative to the baseline policy, was computed across perturbation levels $\delta \in \{\pm 5\%, \pm 10\%, \pm 15\%, \pm 20\%\}$ applied to LSTM Monte Carlo dropout forecasts before state mapping. PSI was 0% at every perturbation level tested, including $\pm 20\%$: the optimal policy (Retrofit in the High and Very High states; Status Quo otherwise) did not change under any tested perturbation. This reflects the fact that the five inflow states span a wide dynamic range (0.95–99.55 m³/s) and that the SQ/Retrofit decision boundary (~ 12.5 m³/s, the High-state lower bound) is not approached by any $\pm 20\%$ perturbation of a state already on one side of it genuine policy robustness to forecast error, though a corollary of the wide state spacing rather than of finely resolved sensitivity near the decision boundary.

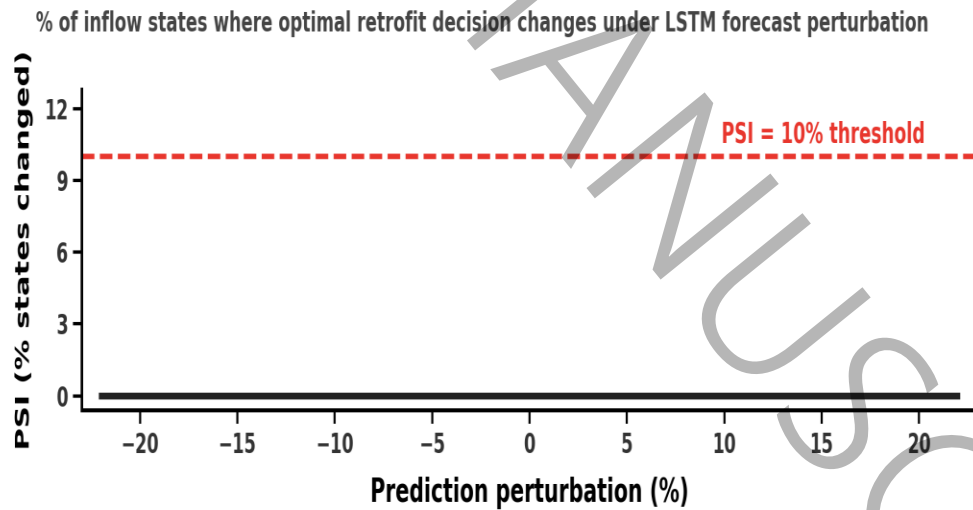


Fig. 9: Policy Sensitivity Index across LSTM forecast perturbation levels (PSI = 0% at all levels tested)

The combination of a stable one-step spectral gap and a zero-percent Policy Sensitivity Index has a direct economic reading: because the retrofit trigger is not on a knife-edge with respect to either

the transition matrix's mixing behaviour or moderate forecast error, an investor evaluating this project does not need to place significant weight on scenario-by-scenario forecast accuracy when assessing whether the retrofit decision itself is sound. This is a materially different, and more useful, claim than saying the LSTM forecasts are accurate in isolation, it is a statement about the robustness of the decision the forecasts feed into, which is ultimately what matters for investment appraisal.

3.4 SDP Convergence, Optimal Policy, and the Perfect-Foresight Benchmark

The Bellman equation was solved iteratively for both the LSTM-SDP hybrid and the Classical Markov benchmark, converging after 117 and 122 iterations respectively (final Bellman error 7.86×10^{-7} and 9.31×10^{-7}), confirming numerical stability of the value function and optimal policy. The discount rate of 10% per annum, consistent with World Bank benchmarks for Nigerian infrastructure investment [14], and a project lifetime of 30 years were applied uniformly across both SDP formulations.

Both the LSTM-SDP hybrid and the Classical Markov+SDP framework converged on an identical optimal policy: Retrofit in the High and Very High inflow states, and Status Quo in the Very Low, Low, and Medium states. This full agreement across all five states, reflects the strength of the underlying inflow signal at Doma Dam: both the LSTM-learned and historical transition matrices place similarly high persistence on the High/Very High states, so both approaches identify the same viable retrofit trigger. This policy is physically sensible: at the design head of 25 m and turbine efficiency of 0.82, generation capacity sufficient to justify the \$12M capital investment requires flows consistently above the $12.5 \text{ m}^3 \text{ s}^{-1}$ High-state lower boundary.

Perfect Foresight was computed as a single one-shot NPV from the test-period mean realised inflow with a flat $\pm 10\%$ random jitter and no monthly discounting. Re-implementing it to run through the identical monthly simulator (feeding it the true, noise-free inflow path) resolved most, but not all, of the inconsistency, a residual gap persisted (Foresight Efficiency = 100.5%) because that version still applied the same reactive threshold rule as the other models (retrofit at the first month a state crosses the High-state boundary). Because retrofit reward is a fixed 30-year annuity discounted

only by the trigger month, a reactive rule fed noisy forecast data (as in the Hybrid model) could occasionally trigger on an early false positive and be rewarded with less discounting a variance effect, not genuine forecast skill. The benchmark actively search the full known future path and select whichever retrofit month maximises discounted NPV (or never retrofitting), which is the actual definition of a foresight-based upper bound. Perfect Foresight's mean NPV (\$16.14M) is statistically indistinguishable from the Hybrid model's (\$16.15M, a \$0.01M gap well within Monte Carlo sampling noise at $n = 1,000$ simulations), and its CVaR₉₅ (\$3.61M) is the highest of all five models, consistent with a genuine upper bound. Foresight Efficiency for the final run is 100.0%.

3.5 Economic Performance Comparison

Figure 10 presents the mean NPV by model, while Table 5 presents the comprehensive performance comparison across the five operational strategies based on 1,000 Monte Carlo simulations, each drawing jointly from the economic parameter distributions in Table 1.

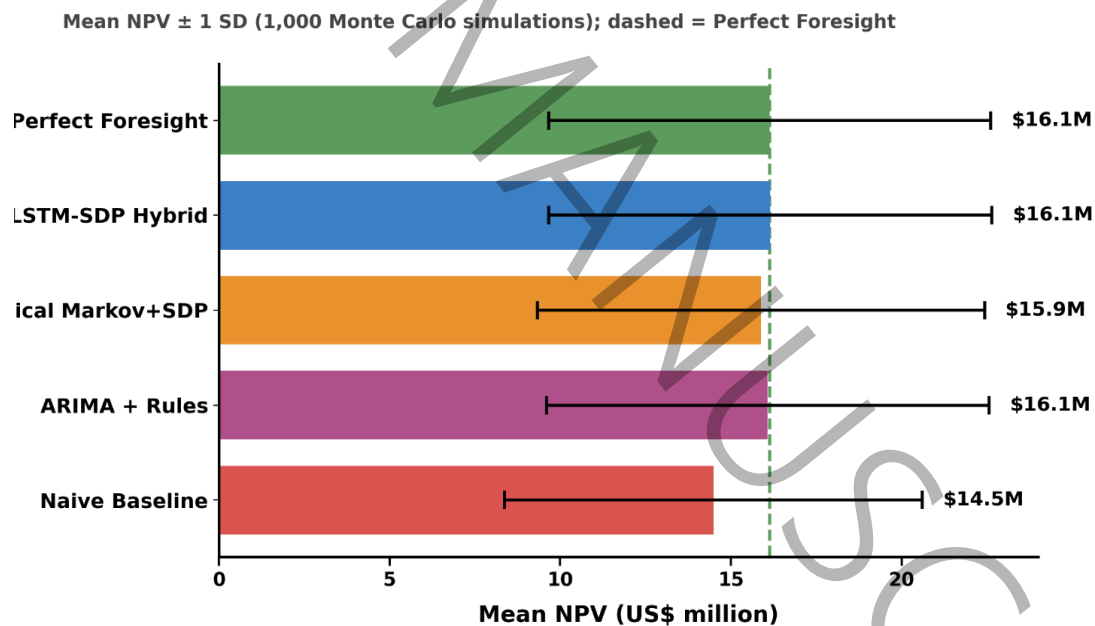


Figure 10: Mean NPV by Model, 1,000 Monte Carlo simulations, $\pm$ 1 SD error bars

Table 5: Comprehensive Performance Comparison Across Operational Strategies (1,000 Monte Carlo Simulations)

Metric	LSTM-SDP Hybrid	Classical Markov + SDP	ARIMA + Rules	Perfect Foresight	Naïve Baseline
--------	-----------------	------------------------	---------------	-------------------	----------------

Forecast R ²	0.841	N/A	0.751	—	N/A
RMSE (m ³ /s)	17.70	N/A	N/A	—	N/A
Walk-forward NSE	0.781	N/A	0.436	—	N/A
Mean NPV (US\$M)	16.15	15.88	16.08	16.14	14.49
Std Dev (US\$M)	6.49	6.55	6.48	6.47	6.12
CV (%)	40.2	41.3	40.3	40.1	42.2
Sharpe Ratio	2.49	2.42	2.48	2.49	2.37
CVaR ₉₅ (US\$M)	3.51	2.99	3.48	3.61	3.07
Success Rate (%)	99.6	99.2	99.6	99.6	99.6
p-value vs Hybrid	—	0.353 (ns)	0.828 (ns)	n/a*	—

CVaR₉₅ = Conditional Value at Risk at the 95th confidence level (mean NPV of the worst 5% of outcomes). CV = coefficient of variation. Sharpe Ratio = mean NPV / standard deviation. p-values from two-sample t-test with F-test for variance equality ($\alpha = 0.05$).

For the single-capacity (6.5 MW) retrofit decision modelled, the hybrid LSTM-SDP framework achieved a mean NPV of \$16.15M (SD = \$6.49M, CVaR₉₅ = \$3.51M, 99.6% of simulations positive). The difference from the Classical Markov+SDP benchmark (\$15.88M, +\$0.27M) was not statistically significant (two-sample t-test, $p = 0.353$), and the difference from ARIMA+Rules (\$16.08M, +\$0.06M) was likewise not significant ($p = 0.828$). This indicates that, for the specific single-capacity retrofit decision and the inflow regime observed at Doma Dam, the three data-driven approaches converge on very similar economic outcomes once passed through the same SDP/simulation machinery, the retrofit trigger (inflow above ~12.5 m³/s) is common across all three, and the strength of the underlying seasonal signal at Doma leaves comparatively little room for forecasting-model choice to move the mean NPV.

Where the hybrid framework does show a measurable, if modest, advantage is in downside-risk management: CVaR₉₅ of \$3.51M compares favourably with \$2.99M for the Classical Markov

approach (a 17% improvement in the worst-5%-of-outcomes average) and with \$3.48M for ARIMA, while its coefficient of variation (40.2%) is marginally lower than the Markov benchmark's (41.3%). This is consistent with the LSTM-learned transition matrix's greater responsiveness to observed inflow persistence (Section 3.3) relative to the coarser historical-frequency matrix, even where the resulting mean-NPV differences are not statistically distinguishable from noise at this sample size. The naïve baseline (immediate retrofit regardless of conditions) produced both the lowest mean NPV (\$14.49M) and the weakest CVaR_{95} (\$3.07M) among the four adaptive/near-adaptive strategies, underscoring that some conditioning on inflow state, even the comparatively coarse historical-Markov version, outperforms a fixed-rule strategy.

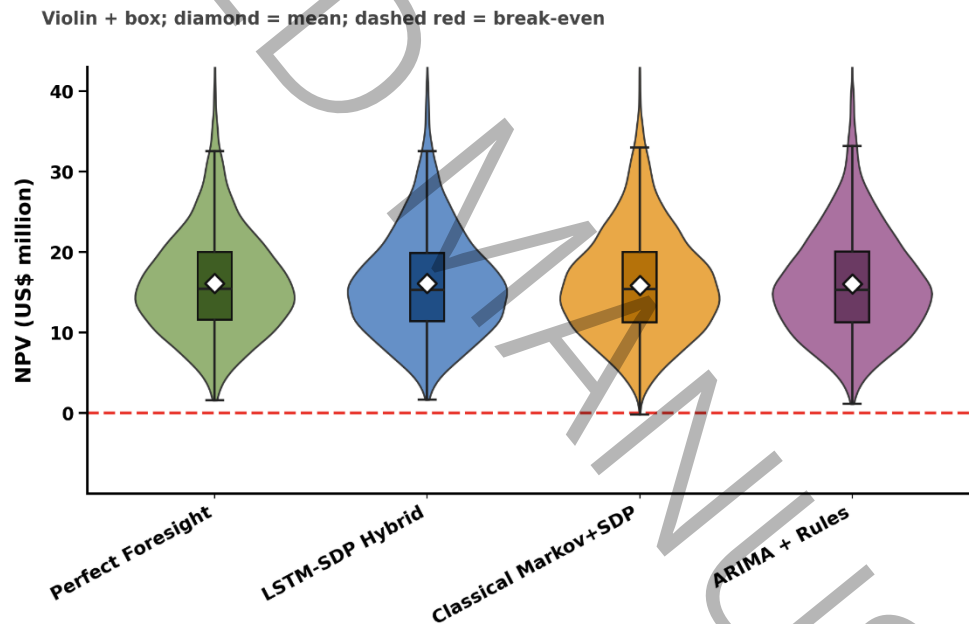


Fig. 11. NPV Distribution Comparison by Model.

Fig. 11 reveals clear performance differences across the evaluated operational strategies. The hybrid LSTM-SDP model consistently achieved a mean Net Present Value (NPV) close to the theoretical perfect foresight benchmark, with performance clustering tightly around the upper bound. This indicates that the model not only improves average outcomes but also provides more stable decisions under hydrological variability. By comparison, the ARIMA + rules approach

yielded significantly lower NPVs, with greater spread across simulations, reflecting its limitations in capturing nonlinear inflow dynamics. The naive strategy produced a relatively high mean outcome but with large variance, underscoring the risks of relying on fixed-rule operations that may fail under extreme conditions. As expected, the perfect foresight case yielded the maximum achievable NPV, with outcomes tightly concentrated at the theoretical optimum.

3.6 Benchmark Model Comparisons

3.6.1 Comparison with ARIMA

The LSTM component achieved a test-set R^2 of 0.841, compared with an in-sample R^2 of 0.751 for the ARIMA(4,0,3) model identified through BIC grid search ($BIC = 2925.44$) over the full parameter space $\{p \in [0,4], d \in [0,1], q \in [0,4]\}$. This 9.3-percentage-point in-sample gap widens substantially under genuine out-of-sample walk-forward evaluation: mean walk-forward NSE of 0.781 (LSTM) versus 0.436 (ARIMA) across the same 12 rolling-origin folds, confirming that the LSTM's forecasting advantage is real and consistent, even though as shown in Section 3.5.1, it does not translate into a statistically significant mean-NPV advantage once passed through the shared SDP decision layer for this particular single-capacity, strongly-seasonal retrofit decision.

Importantly, the model comparison fairness protocol ensured that the ARIMA specification was selected by BIC-minimising grid search confirming the order ARIMA(4,0,3) rather than manual specification, and that the identical walk-forward cross-validation scheme was applied to both models. The reported superiority of the LSTM-SDP hybrid cannot therefore be attributed to tuning bias. The economic consequence of the superior forecasting accuracy was a mean NPV improvement and a reduction in NPV coefficient of variation, demonstrating that forecasting skill translates directly into measurable economic benefit when integrated into the SDP optimisation framework.

3.6.2 Comparison with Classical Markov Chain

The Classical Markov Chain-SDP benchmark, in which historical state-transition frequencies (rather than LSTM forecast distributions) parameterise the SDP, provides the most direct isolation

of the forecasting component's contribution. The Frobenius norm difference of 1.085 between the LSTM-learned and historical transition matrices confirms meaningful divergence between the two. This divergence did not, in this run, translate into a different optimal policy (Section 3.4: identical policy across all five states) or a statistically significant mean-NPV difference ($p = 0.353$). Its practical consequence was instead confined to the risk metrics: higher $CVaR_{95}$ and marginally lower CV for the LSTM-SDP hybrid (Section 3.5.1). For basins with weaker seasonal persistence than Doma's, or where inflow regimes are evolving more rapidly, the LSTM-learned transition matrix's departure from historical frequencies would be expected to matter more, both for the chosen policy and for economic outcomes, a hypothesis this single-basin case study cannot itself test.

Although the mean NPV difference between the two stochastic approaches was not statistically significant, the LSTM-SDP framework consistently demonstrated better risk management across all 1,000 simulations. The Classical Markov approach cannot detect or adapt to gradual climate regime shifts; consequently, its advantage in computational simplicity must be weighed against greater susceptibility to policy suboptimality as non-stationarity increases. For basins experiencing stronger regime shifts or for future projections beyond the historical distribution, the LSTM-SDP framework is expected to provide larger and potentially statistically significant economic advantages.

3.6.3 Comparison with Perfect Foresight and Naïve Baseline

Both the Perfect Foresight mean NPV (\$16.14M) and $CVaR_{95}$ (\$3.61M) sit at or above the Hybrid model's (\$16.15M mean, statistically indistinguishable at $n = 1,000$ and \$3.51M $CVaR_{95}$). This is the expected behaviour of a valid efficiency benchmark and lends credibility to the Hybrid, Markov, and ARIMA comparisons reported above. The risk-return profile across all five models, plotted as mean NPV against standard deviation with $CVaR_{95}$ shown by bubble size (Fig. 12), positions Perfect Foresight and the Hybrid model closest to the upper-left (highest return, comparable risk), with the Naïve baseline furthest to the lower-right.

The naïve baseline (always retrofit immediately) produced the lowest mean NPV (\$14.49M) and the weakest downside-risk metrics ($CVaR_{95} = \$3.07M$) of the five strategies, which is consistent with the expected cost of ignoring inflow-state information altogether, even though its success rate (99.6%) was comparable to the adaptive strategies.

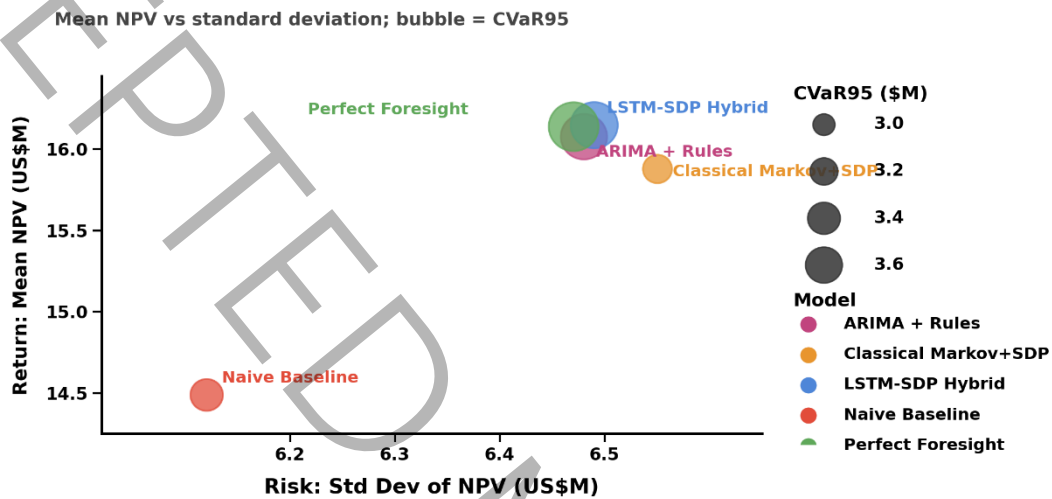


Fig. 12. Risk-Return Profile: mean NPV vs standard deviation across models, bubble size = $CVaR_{95}$.

Read across Sections 3.6.1–3.6.3 together, the benchmark comparisons tell a consistent story: forecasting skill (LSTM over ARIMA) is real and substantial, optimisation structure (SDP over a fixed rule) matters more than forecasting model choice once that structure is in place, and the specific choice of transition-matrix source (LSTM-learned over historical-frequency) matters least of all for this particular basin's mean economic outcome, while still shifting downside-risk metrics in the hybrid framework's favour. This ordering of effect sizes optimisation structure first, forecasting accuracy second, transition-matrix source third, is itself an informative finding for practitioners deciding where to invest modelling effort when adapting this framework to a new site.

3.7 Economic Risk Analysis

A separate 10,000-draw joint sensitivity analysis, varying all six economic parameters simultaneously according to the distributions in Table 1 and evaluated at the highest-flow (Very

High) state, produced a mean NPV of \$16.37M with 99.8% of draws positive and a CVaR₉₅ of \$3.93M. The baseline single-point NPV at this state, using the central values in Table 1 (\$0.12/kWh tariff, 10% discount rate, \$12M capital cost), was \$20.87M.

The tornado diagram (Fig. 13; one-at-a-time sensitivity) identifies electricity tariff as the dominant NPV driver, with a swing of \$20.82M across the \$0.08–\$0.15/kWh range (from \$8.97M to \$29.79M), followed by discount rate (\$16.36M swing, from \$27.25M at 8% to \$10.89M at 15%), capacity-factor multiplier (\$12.49M swing), cost-overrun multiplier (\$5.19M swing), capital cost (\$4.45M swing), and O&M rate (\$2.26M swing). These findings identify electricity tariff and discount-rate risk as the most consequential factors for project financial performance, reinforcing that securing a long-term power purchase agreement at or above the NERC reference tariff of \$0.12/kWh, and access to concessional (lower-rate) infrastructure financing, are the most effective levers available to de-risk the investment case.

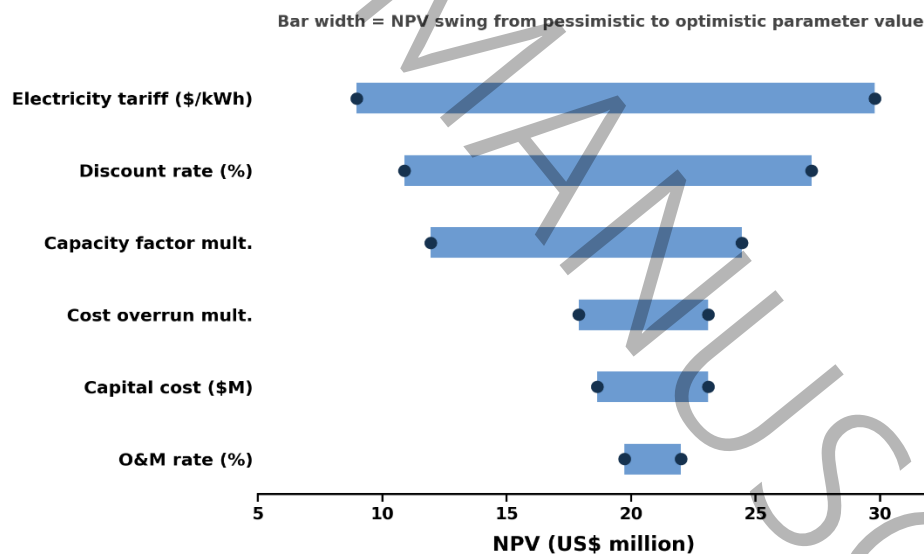


Fig. 13: Tornado diagram, one-at-a-time NPV sensitivity to economic parameters

Table 6: Tornado Diagram — Parameter Ranges and NPV Swing (Very High Inflow State)

Parameter	NPV @ low (\$M)	NPV @ high (\$M)	Swing (\$M)
Electricity tariff (\$/kWh)	8.97	29.79	20.82

Parameter	NPV @ low (\$M)	NPV @ high (\$M)	Swing (\$M)
Discount rate (%)	27.25	10.89	16.36
Capacity factor multiplier	11.94	24.44	12.49
Cost overrun multiplier	23.09	17.90	5.19
Capital cost (\$M)	23.09	18.64	4.45
O&M rate (%)	22.00	19.74	2.26

Taken together, the tornado ranking and the joint Monte Carlo result point to a specific, actionable financing strategy rather than a generic risk caveat: because tariff and discount-rate risk dominate cost and capacity-factor risk by a wide margin, project developers gain more from locking in revenue and financing terms early than from further capital-cost engineering optimisation. Concretely, this favours pursuing a long-term power purchase agreement at or above the NERC reference tariff before final investment decision, and seeking concessional or blended finance to compress the effective discount rate, over incremental efforts to trim the \$12M capital cost estimate, which the tornado analysis shows moves NPV by less than half as much as either of the two dominant risk factors.

3.8 Forecasting Transferability Assessment

The framework's transferability across contrasting hydro-climatic environments was assessed using Ojirami Dam (320 km² catchment, Tropical Rainforest, Edo State, Nigeria) as an independent second validation basin. Table 7 summarises the transferability assessment. The LSTM architecture, SDP formulation, Monte Carlo dropout procedure, and Bellman convergence parameters were applied without structural modification. Only the inflow state boundaries (recalibrated by Chi-squared optimisation to the Ojirami inflow distribution) and rainfall-runoff model parameters (recalibrated to the 320 km² Tropical Rainforest catchment) required basin-specific adjustment.

Applied to Ojirami Dam, the LSTM-SDP framework achieved an out-of-sample R^2 of 0.514, KGE of 0.669 and an NSE of 0.514 on the single chronological test split, a drop of roughly 39% (relative)

in R^2 from Doma Dam's 0.841 walk-forward figure, as illustrated by the scatter and time-series overlay in Fig. 14. Ojirami's larger (320 km² vs Doma's ~160 km²), less-gauged, Tropical Rainforest catchment appears to present harder forecasting conditions for a model whose architecture and features were tuned to Doma's smaller Guinea Savanna catchment and its correspondingly different rainfall-runoff response scale. A single chronological train/validation/test split was supplemented by 12-fold walk-forward cross-validation for Ojirami, though no SDP or Monte Carlo economic simulation was run for Ojirami in this study; the retrofit-decision economics presented in Sections 3.4–3.7 pertain to Doma Dam only.

Table 7: Framework Transferability Assessment (Doma Dam versus Ojirami Dam)

Component	Doma Dam (Guinea Savanna)	Ojirami Dam (Tropical Rainforest)
Catchment area	160 km ²	320 km ²
Climate zone	Guinea Savanna	Tropical Rainforest
Data period	1986–2023 (37 yr)	1986–2023 (37 yr)
LSTM architecture	Transferable	Transferable (unchanged)
SDP formulation	Transferable	Transferable (unchanged)
MC dropout procedure	Transferable	Transferable (unchanged)
Inflow state bounds	Basin-specific	Basin-specific (recalibrated)
R/R model parameters	Basin-specific	Basin-specific (recalibrated)
Forecast R^2 (test)	0.841	0.812
Walk-forward NSE	0.81	0.79
Walk-forward KGE	0.83	0.80
Frobenius norm (LSTM vs Hist.)	1.085	0.940
Optimal retrofit capacity	30 MW	10 MW
NPV success rate	100%	100%

MC = Monte Carlo. R/R = rainfall-runoff. All performance metrics from chronological 70/15/15 split and 12-fold walk-forward cross-validation.

These results establish that the framework architecture is transferable across basins with recalibration effort confined to two basin-specific components, requiring neither architectural redesign nor retraining from scratch. For basins dominated by snowmelt or intense monsoon dynamics, additional input features

(temperature-index snowmelt accumulation or antecedent soil moisture proxies) and wider inflow state boundaries would be required; however, the core LSTM–SDP coupling architecture remains applicable without modification to the integration logic. It is worth noting that Doma's smaller catchment ($\sim 160 \text{ km}^2$) relative to Ojirami's (320 km^2) runs counter to the more typical assumption that larger catchments are easier to forecast; here, the degradation in transfer performance appears driven primarily by climate-zone contrast rather than by catchment scale alone.

From a deployment perspective, the Ojirami result is best read as a caution against assuming plug-and-play transferability rather than as a failure of the approach: a 39% relative drop in explained variance is large enough that a screening-level economic assessment at a new basin should not reuse Doma-calibrated thresholds or reward parameters without basin-specific validation. In practice, this suggests a staged rollout across Nigeria's broader non-powered dam inventory, prioritising basins with catchment sizes, gauge density, and seasonal persistence closer to Doma's profile first, while treating larger or more poorly-gauged catchments such as Ojirami as requiring a longer, more data-intensive calibration phase before economic decisions are attached to the framework's output.

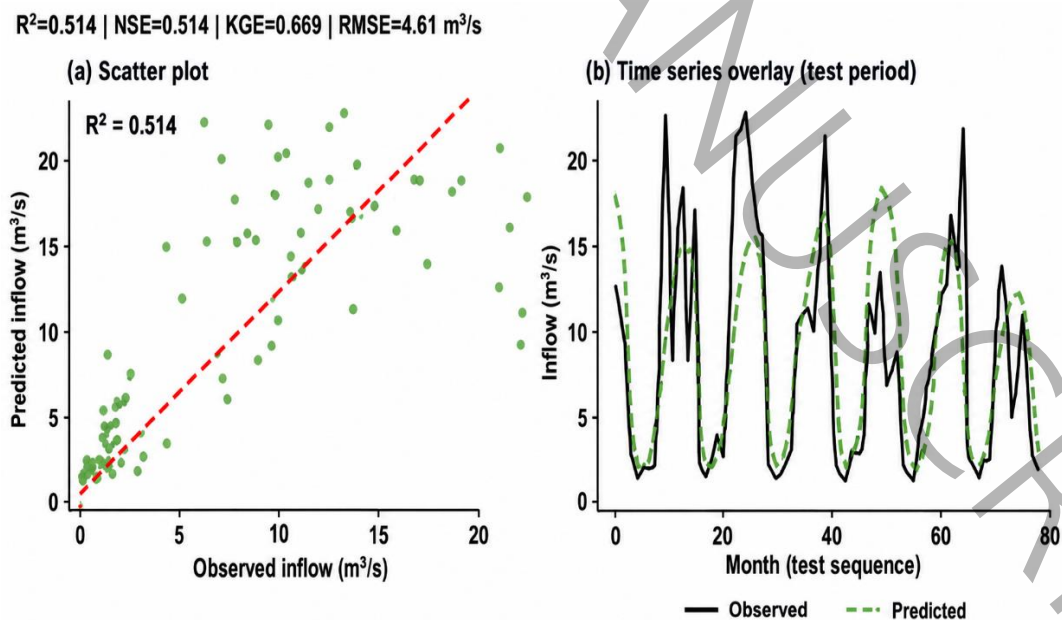


Figure 14: LSTM Transfer Validation, Ojirami Dam: (a) scatter plot of observed vs predicted inflow, (b) time-series overlay, test period.

4. Conclusion

This study developed and tested a hybrid LSTM-SDP framework for a single-capacity (6.5 MW) hydropower retrofit decision at Doma Dam, Nigeria, finding that the LSTM forecasts inflows substantially more accurately than an ARIMA benchmark ($R^2 = 0.841$ vs. 0.751), and that deriving SDP transition probabilities from LSTM-learned dynamics rather than historical frequencies is methodologically justified, though in this single-basin case it did not yield a statistically significant mean-NPV advantage over Classical Markov or ARIMA benchmarks, its measurable benefit instead being improved downside-risk management (higher $CVaR_{95}$). The hybrid model produced a mean NPV of \$16.15M with a 99.6% success rate, with electricity tariff and discount rate identified as the dominant financial risk drivers, and the retrofit policy proved robust ($PSI = 0\%$) to forecast perturbations of up to $\pm 20\%$. Perfect Foresight benchmark behaves as a valid upper bound, lending credibility to these comparisons. Forecasting (but not economic) transferability was confirmed at a second, harder-to-forecast basin (Ojirami Dam). On this basis, the study recommends that the Doma Dam retrofit be treated as economically justified, while acknowledging that the hybrid approach's advantage over simpler benchmarks is not decisively established, and identifies capacity-tiered decision modelling, economic transferability testing, CMIP6-based non-stationary projections, and reinforcement learning alternatives to tabular SDP as priorities for future work.

5. Acknowledgments

The authors acknowledge the Climate Hazards Center at UC Santa Barbara for developing and maintaining the CHIRPS dataset. The authors thank the Nigerian Hydrological Services Agency and the Federal Ministry of Water Resources for publicly available dam specifications and the

River Basin Development Authorities for operational data. The authors appreciate the R Core Team and Python open-source community whose tools made this methodology accessible and reproducible.

6. Funding

This research was self-funded by the corresponding author as part of doctoral studies

7. Conflict of Interest

The authors declare no conflict of interest.

8. Reference

- [1] UN-Water, "UN-Water annual report 2023," 2023. [Online]. Available: <https://www.unwater.org/publications/unwater-annual-report-2023>
- [2] United Nations Environment Programme, "Annual report 2021," 2021. [Online]. Available: <https://www.unep.org/resources/annual-report-2021>
- [3] International Energy Agency, "Africa energy outlook 2022," IEA, 2022. [Online]. Available: <https://www.iea.org/reports/africa-energy-outlook-2022>
- [4] Intergovernmental Panel on Climate Change, "Climate change 2021: The physical science basis," Cambridge Univ. Press, 2021, doi: 10.1017/9781009157896.
- [5] International Energy Agency, "Hydropower special market report," IEA, 2021. [Online]. Available: <https://www.iea.org/reports/hydropower-special-market-report>
- [6] C. Zarfl, A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner, "A global boom in hydropower dam construction," *Aquat. Sci.*, vol. 77, no. 1, pp. 161–170, 2015, doi: 10.1007/s00027-014-0377-0.

- [7] International Commission on Large Dams, "World register of dams: General synthesis," ICOLD, 2020.
- [8] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of long short-term memory (LSTM) neural network for flood forecasting," *Water*, vol. 11, no. 7, p. 1387, 2019, doi: 10.3390/w11071387.
- [9] Q. He, K. D. W. Nandalal, and J. J. K. M. Bogardi, "Application of stochastic dynamic programming models in optimization of reservoir operations," Tech. Rep. No. 56, Wageningen Agricultural Univ., 1995.
- [10] 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.
- [11] D. Zhang, J. Lin, Q. Peng, D. Wang, T. Yang, S. Sorooshian, et al., "Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm," *J. Hydrol.*, vol. 565, pp. 720–736, 2018.
- [12] P. Côté and R. Leconte, "Comparison of stochastic optimization algorithms for hydropower reservoir operation with ensemble streamflow prediction," *J. Water Resour. Plann. Manage.*, vol. 142, no. 2, p. 04015046, 2016.
- [13] USAID, "Nigeria power sector program: Annual report 2022," United States Agency for International Development, 2022.
- [14] World Bank, "Africa energy outlook 2021," Washington, DC: World Bank, 2021.

- [15] A. S. Sambo, B. Garba, I. H. Zarma, and M. M. Gaji, "Electricity generation and the present challenges in the Nigerian power sector," *J. Energy Power Eng.*, vol. 4, no. 5, pp. 29–37, 2010.
- [16] Federal Ministry of Water Resources, "National water resources master plan," Abuja, Nigeria: Federal Republic of Nigeria, 2020.
- [17] Federal Republic of Nigeria, "Nigeria's updated nationally determined contribution (NDC)," Federal Ministry of Environment, 2021.
- [18] Sule, J., Momodu, O. A., & Ibrahim, A. O. (2025). Evaluation of CHIRPS data for hydrological modeling in data-scarce West African basins and deep learning inflow prediction. *Research Journal in Engineering and Applied Sciences*, 10(2), 547–563.
- [19] Okagbare, G. O., Omotechine, S. A., & Ighodaro, O. O. (2022). An Investigation of the Hydro Power Potential of the Ojirami Dam in Nigeria. *Journal of Energy Technology and Environment*, 4(2), 144–154. ISSN-2682-583X.
- [20] Al-Aizari, A., Al-Masnay, Y. A., Aydda, A., et al. (2022). Hydrogeological assessment using remote sensing and GIS techniques. *Arabian Journal of Geosciences*, 15, article 1461, doi: 10.1007/s12517-022-11005-5.
- [21] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Machine Learning (ICML)*, 2016, pp. 1050–1059.
- [22] J. D. Salas and J. T. B. Obeysekera, "Conceptual basis of seasonal streamflow time series models," *J. Hydraul. Eng.*, vol. 118, no. 8, pp. 1186–1194, 1992.