Task: Forecast the next 96 days of daily streamflow from 96 days of history.
Context: Four past-only weather covariates; max target–covariate |r| = 0.281; distribution shift = 0.116.
Objective: Minimize four-fold expanding-window CV MSE (seed 42); Evaluate the selected checkpoint on held-out test.

Data Profiling:

Persistent dynamics
lag-1 ACF = 0.920, indicates persistent runoff, consistent with storage/recession dynamics.

Distribution shift
drift score = 0.116, flags level/scale variation, so window magnitude may encode wet/dry states.

Delayed covariate signal
Max |r| rises from 0.281 to 0.439 at lag 24, |r| = 0.439 at lag 24 days suggesting delayed or shared seasonal effects.

Literature Extraction:

Normalization Retrieve from: *Non-Stationary Transformer*
Mechanism: window z-score + output rescaling.
Effect (expected): Stable inputs; better generalization.
Applicability: Shifting mean/scale (drift=0.116).
Limitation: Useful level/scale cues may be lost;

Variable interaction Retrieve from: *iTransformer*
Mechanism: Channel Mixing.
Effect (expected): Covariate context can reduce error.
Applicability: Useful lagged coupling (|r|=0.439).

Hypothesis
Normalized Mamba can provide a compact carrier for persistent runoff dynamics.

Architecture / Change
instance_norm + independent + 2-layer Mamba + linear head

Experimental Results
CV MSE = 0.701954. Slight improvement over previous SSM.

Reflection
Reusing Mamba allowed a fair SSM comparison.

Hypothesis
Instance normalization should help under shift. Test by removing it.

Architecture / Change
instance_norm → none
All other settings fixed.

Experimental Results
CV MSE = 0.596769. Lower CV spread supports a consistent gain.

Reflection
We expected normalization to help. The local test changed that view. Update the test results on **E1**.

Hypothesis
Channel mixing may help on the no-norm carrier.

Architecture / Change
independent → feature_mlp
All other settings fixed.

Experimental Results
CV MSE = 0.588419. 3/4 folds improve. Gain is inconsistent.

Reflection
Re-test mixing after removing normalization. Keep the small gain as an uncertain result.

Champion Model

5 channel variables
→ No preprocess
→ Channel mixing
→ 2-layer Mamba
→ Linear head
→ Output

Test MSE: 0.981216
Test MAE: 0.544346

Champion on held-out test

Stored as reusable asset.

Accumulated Memory

Knowledge

[K1] Shift → window norm → stable inputs.
Limit: useful level/scale cues may be lost.

[K2] Useful covariates → mix → lower error.
Limit: attention and lag alignment untested.

Experience

[E1] When: ETTh1 linear; PJM N-BEATS
Effect: +norm → MSE +1.2% / +3.2%.
Scope: MA paths; L/H=96/96 &168/24.

[E2] When: MOPEX raw+norm; TCN/SSM
Effect: +mix → MSE −16.9% / −0.6%.
Scope: L=H=96, seed42; carrier-specific.

Model Assets

[A1] Reused Mamba-1 module
Use: ×2 layers; width64 / state32.
Swap SSM backbone; scaffold fixed.
Performed well on persistent hourly load forecasting.