Time Series Engression:

Modeling Conditional Distributions with Input-Dependent Noise and Temporal Structure

Yuri CROCI

University of Geneva, GSEM, Student in Master of Science in Statistics

14 November, 2025

Supervised by: Prof. Dr. Sebastian Engelke (GSEM)

Overview

1. Motivation
2. Literature Review:
   - Noise Structure Design
   - Engression Framework
3. Methodology:
   - Problem Formulation
   - Temporal extension
   - Heteroskedastic extension
   - Experimental Setup
4. Experiments:
   - Simulation Study
   - River Discharge Application
5. Conclusions

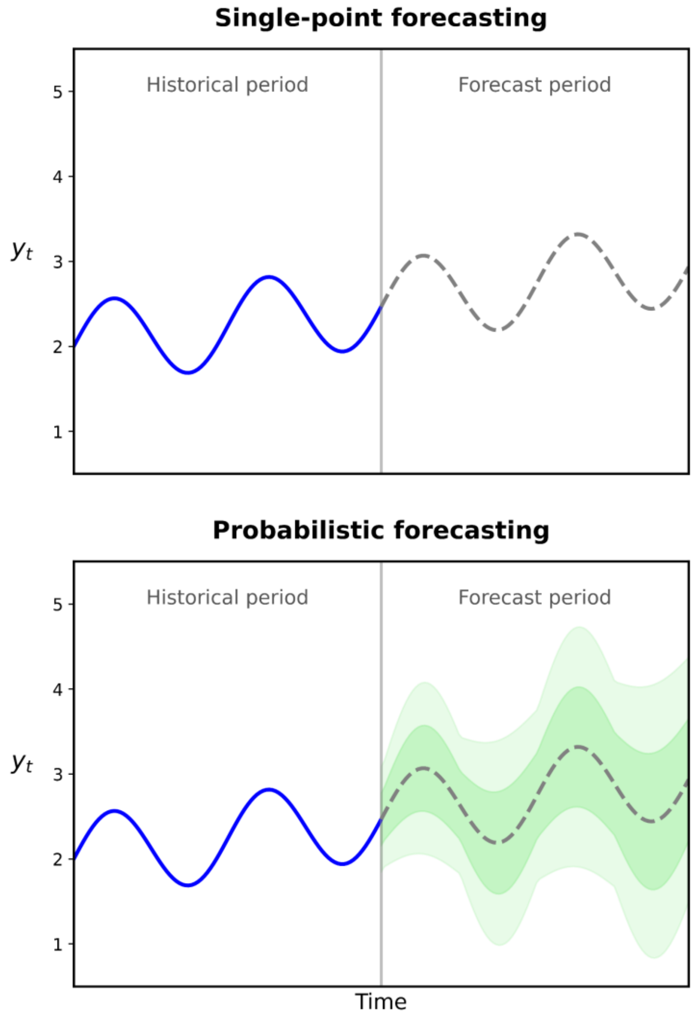
Motivation

The challenges:

- Temporal complexity of time series
- Limitation single-point forecasting
- Trade-off complexity models

What's needed:

- Conditional distribution for temporal forecasting
- Non-parametric flexibility
- Extrapolation capability



Literature Review: Noise Structure Design

Post-Additive Noise

$$Y = f(X) + \varepsilon, \quad \varepsilon \overset{\text{i.d.}}{\sim} \mathcal{D}$$

- Noise added **after** transformation
- Perturbs only the response variable
- No information of $f(\cdot)$ beyond support

Pre-Additive Noise

$$Y = g(X, \eta), \quad \eta \overset{\text{i.d.}}{\sim} \mathcal{D}$$

- Input-level noise perturbation
- Uncertainty propagates through g(·)
- Enables extrapolation beyond support

Literature Review: Engression

Model class:

Given covariates X and response Y, Engression models the conditional distribution $P(Y|X)$ through a general class of functions

$$\mathcal{M} = \{g(x,\eta) : g \in \mathcal{G}\}, \quad \eta \overset{\text{i.d.}}{\sim} \mathcal{N}(0,\sigma^2).$$

Goal: Find g such that

$$g(x,\eta) \sim P(y|x) \text{ for any } x,$$

by enabling Sampling-based inference from the estimated distribution.

Literature Review: Engression

How: Find optimal $g(\cdot)$ by minimizing the Energy Loss as proper scoring rule,

$$\tilde{g} \in \arg\min_{g \in \mathcal{M}} \mathbb{E}_{(X,Y)} \left[\boxed{\|Y - g(X,\eta)\|_2} - \frac{1}{2} \boxed{\|g(X,\eta) - g(X,\eta')\|_2} \right],$$

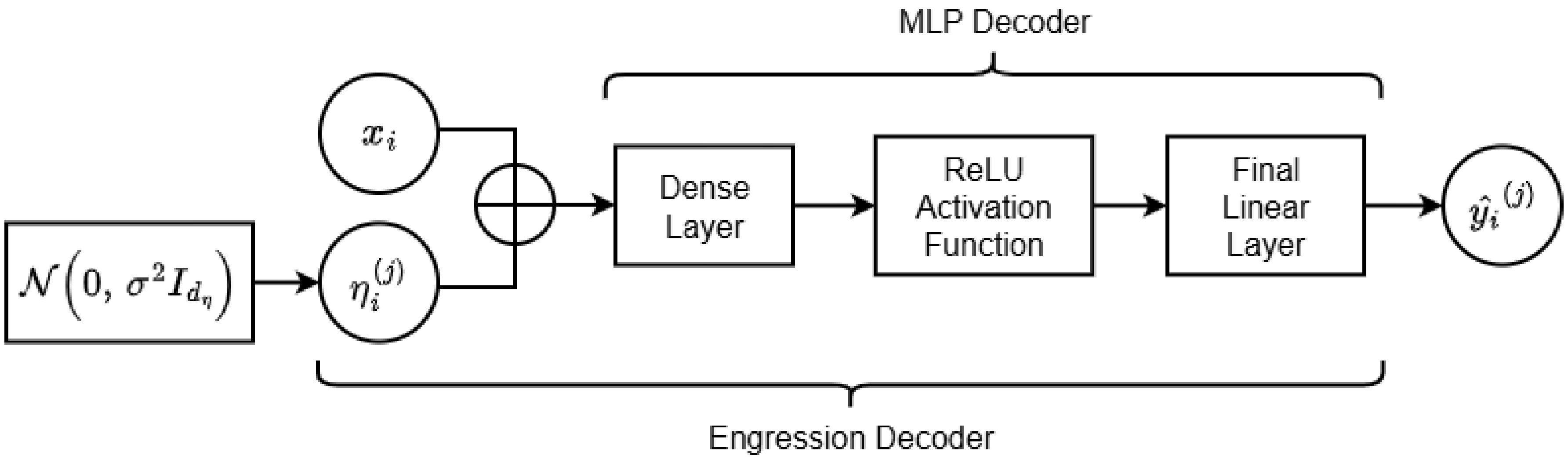
where η and η' are independent draws from $\mathcal{N}(0, \sigma^2)$.

Dual Nature:

- Accuracy: Pulls samples toward observations
- Dispersion: Spreads samples apart

Literature Review: Engression

Implementation:



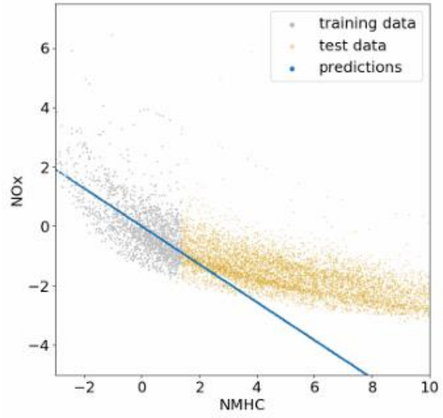
- Forward pass composed by m samples, where each $\eta_i^{(j)}$ produces one prediction $y_i^{(j)}$.
- Optimized by Adam optimizer under Energy Loss.

Extrapolation Capabilities:

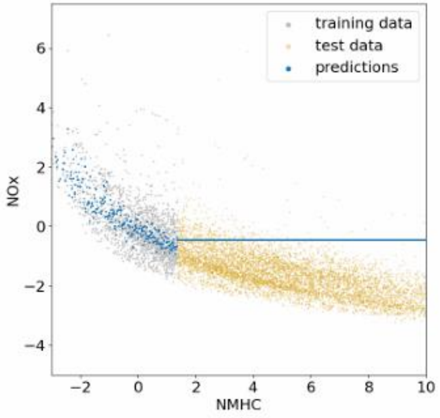
- Under local monotonic $g(x)$ near boundary, pre-additive structure, and bounded noise.

  Recover $g(x)$ between $[\, x_{train\_min} - \eta_{min}\, ;\, x_{train\_max} + \eta_{max}]$.

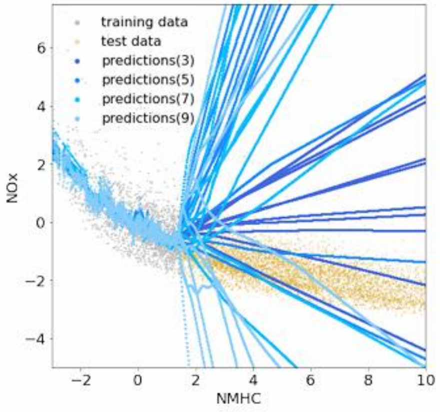
- Robust fallback to linear extrapolation in case of misspecification.



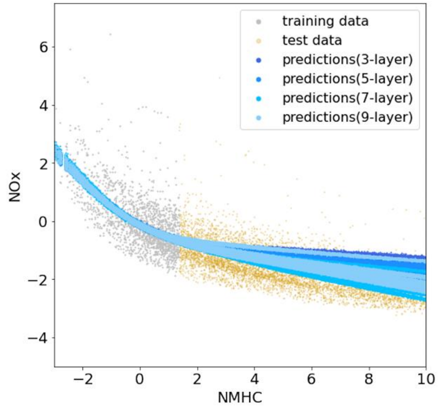
Liner regression



Random Forest



Neural network regression



Engression

Methodology: Problem Formulation

Adapting Engression to Time Series Forecasting:

- One-step-ahead probabilistic forecasting using only historical information

$$P\left(Y_t \mid \tilde{x}_t\right), \quad \text{where} \quad \tilde{x}_t = \{(x_\tau, y_\tau)\}_{\tau=t-s}^{t-1}.$$

- Lagged input space composed of past s observations of covariates x_t and response y_t.

Learning the Conditional Distribution:

- Find $\hat{g}$ such that $\hat{g}\left(\tilde{x}_t, \eta_t\right) \sim P\left(Y_t \mid \tilde{x}_t\right)$ by minimizing the empirical Energy Loss

$$\hat{g} \in \arg\min_{g \in \mathcal{M}} \hat{\mathcal{L}}(g).$$

Methodology: Temporal extension

Limitation of Original Architecture:

- Treatment of lagged features as independent inputs.
- Ignore sequential dependencies in temporal data.

Temporal Encoder with Gated Recurrent Unit (GRU):

- Process input sequence $\tilde{x}_t \in \mathbb{R}^{s\times(d_x+1)}$ through GRU encoder.
- Produce hidden states $H_t = \{h_{t-s+1}, \ldots, h_t\}$ via recurrent architecture.
- Able to capture sequential patterns in the historical window.
- GRU selected over LSTM for better encoder-decoder balance.

Methodology: Temporal extension

Limitations of Standard GRU Encoding:

- Final hidden state h_t creates information bottleneck in longer sequences.
- Shared recurrent structure fails to capture lag-specific contributions.

Static Attention Pooling:

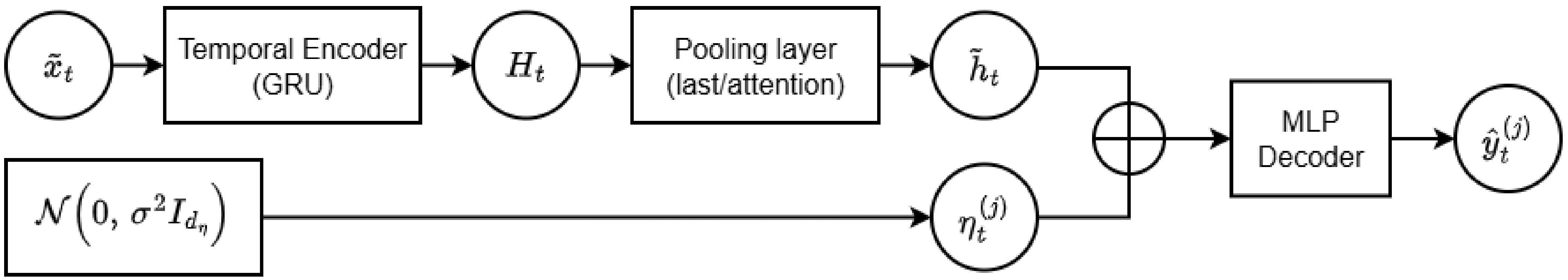
Weighted combination of all hidden states to preserve lag-specific information

$$\tilde{h}_t = \sum_{i=1}^{s} \alpha_i h_{t-s+i}, \quad \alpha_i = \frac{\exp(w_i)}{\sum_{k=1}^{s} \exp(w_k)} \ ,$$

where learnable weights w_i capture relative importance of each lag.

Methodology: Temporal extension

Sequential Engression architecture:



Methodology: Heteroskedastic extension

Limitation of Fixed Noise:

- Original framework injects homoskedastic noise $\eta \sim \mathcal{N}\left(0, \sigma^2 I_{d_\eta}\right)$.
- Fails to capture time-varying uncertainty in real-world time series.

Input-Dependent Noise Injection

- Learn conditional variance as function of input $\eta_t \sim \mathcal{N}\left(0, \Sigma_t\left(\tilde{h}_t\right)\right)$.
- Two parameterizations for varying complexity

$$\text{Scalar: } \Sigma_t = \tilde{\sigma}_t^2\left(\tilde{h}_t\right) I_{d_\eta}, \quad \text{Vectorized: } \Sigma_t = \operatorname{diag}\left(\tilde{\sigma}_t^2\left(\tilde{h}_t\right)\right).$$

Methodology: Heteroskedastic extension

Implementation:

- Linear projection of the pooled latent representation $\tilde{h}_t$ to conditional variance

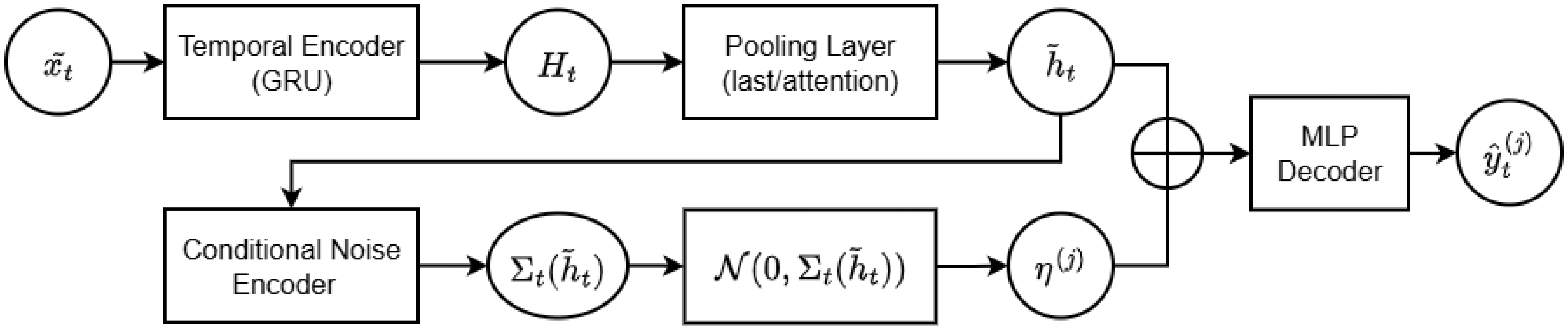
$$\tilde{\sigma}_t^2\left(\tilde{h}_t\right) = \mathrm{softplus}\left(W\tilde{h}_t + b\right),$$

where output space dimensions depend on representation.

- Softplus activation guarantees strictly positive variance and stable gradients.

Methodology: Heteroskedastic extension

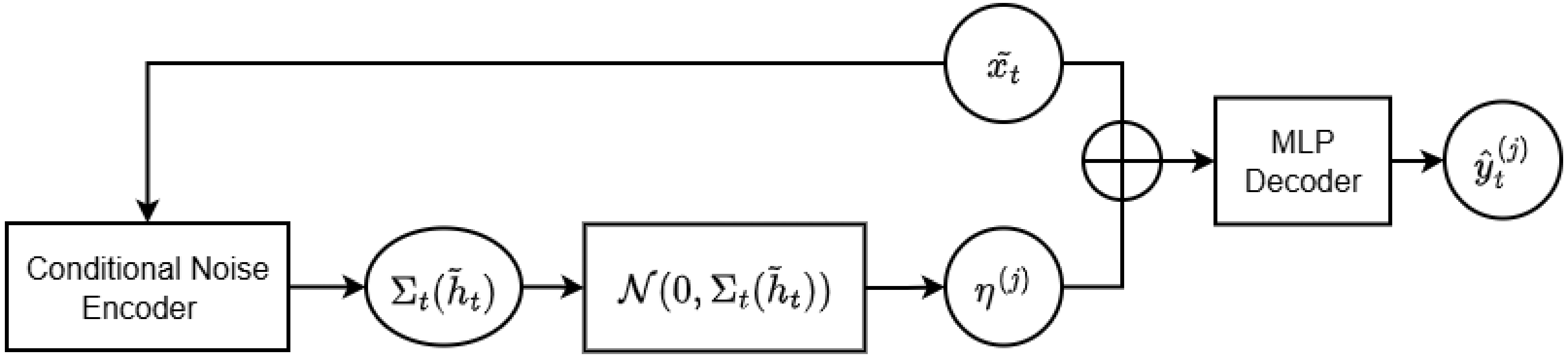
Heteroskedastic Sequential Engression architecture:



Methodology: Experimental Setup

Model Configurations evaluated:

- Deterministic baseline (MLP and Sequential MLP) trained to estimate $\mathbb{E}\left(Y_t \mid \tilde{x}_t\right)$.
- Distributional models (Engression, Heteroskedastic Engression, Sequential Engression, and Heteroskedastic Sequential Engression) trained to estimate $P(Y_t \mid \tilde{x}_t)$.



Heteroskedastic Engression Architecture

Methodology: Experimental Setup

Training & Validation:

- Chronological train-validation-test split to preserve temporal order.
- Standardization of features for training stability and equal noise perturbation.
- Grid search over hyperparameters validated on validation set using respective loss functions.
- Early stopping on validation loss to determine optimal number of epochs.
- Weight decay tuned as regularization across all models.

Methodology: Experimental Setup

Evaluation Metrics:

- RMSE: Mean prediction error
- Extreme RMSE: Performance on extreme events (>99.5% training quantile)
- Energy Loss: Joint measure of sharpness and calibration
- Coverage ($PICP_{80}$): Calibration of 80% prediction intervals
- $Sharpness_{80}$: Precision of prediction intervals
- Quantile RMSE: Accuracy over true quantiles(Q10, Q50, Q90)

Distributional models only

Robustness:

- All models trained and evaluated across 10 independent seeds

Experiments: Simulation Study

Process Design:

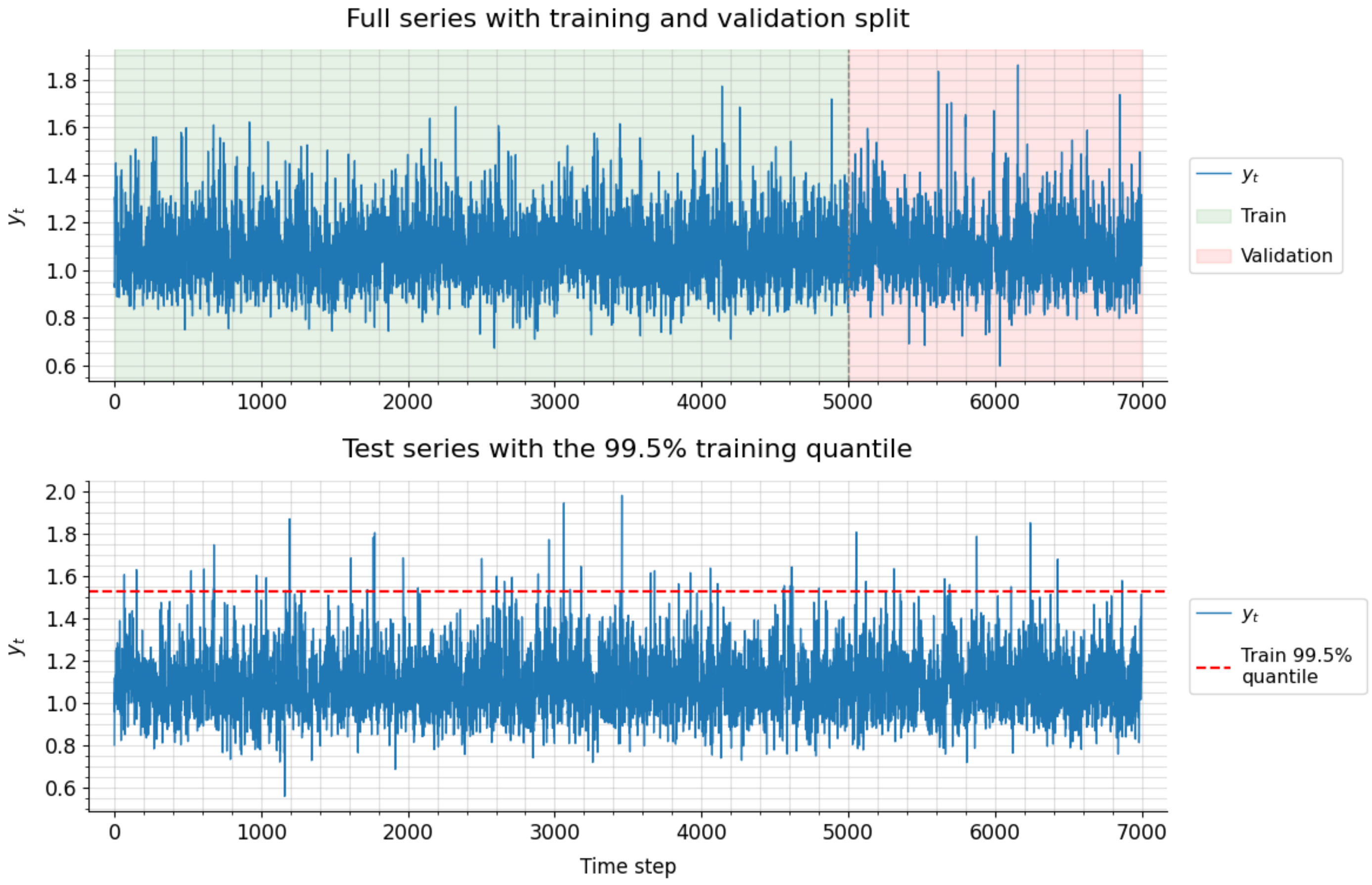
Synthetic time series (y_t, x_t) with three designed properties:

- Sequential dependencies: Latent state + lagged responses/covariates.
- Heteroskedastic variance: Time-varying volatility (GARCH-like).
- Pre-additive noise: Scaled exponential transformation applied to noise-perturbed signal to preserve Engression assumptions.

Closed-form quantiles enable evaluation against true distributions.

Experiments: Simulation Study

Experimental Setup:



Experiments: Simulation Study

Key Findings:

Mean Prediction Performance:

- Sequential architectures: small RMSE improvements ($\sim$0.5%).
- Engression-based vs deterministic: comparable mean, marginal extrapolation gains ($\sim$1-1.6%).

Distributional Quality:

- Sequential models achieve best Energy Loss and coverage.
- Heteroskedastic extension effective only when combined with temporal encoding.

Experiments: Simulation Study

Quantile Estimation:

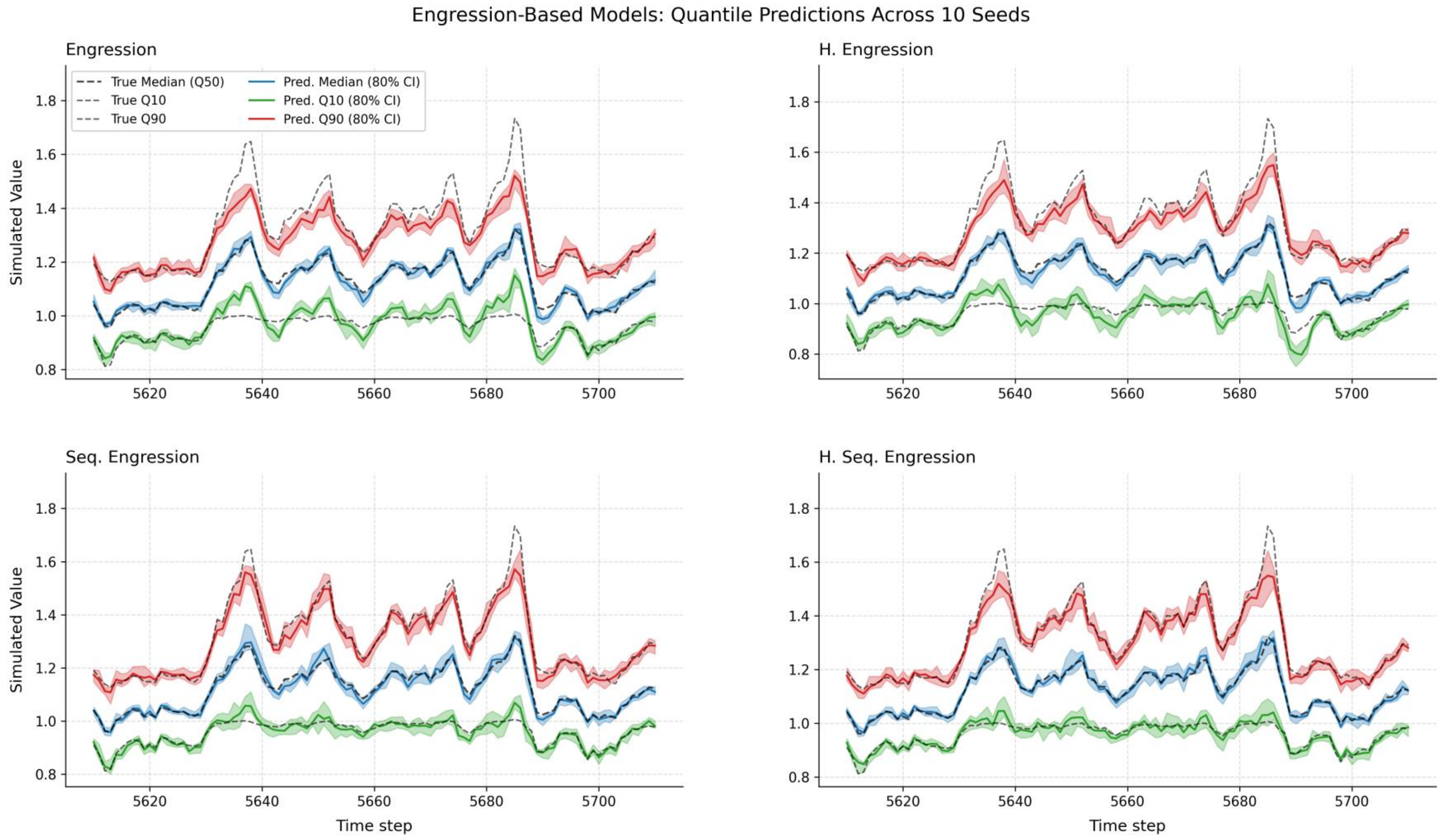
- Sequential models achieve major tail improvements: 21-32% better RMSE on quantiles (Q10/Q90) over baseline.

- Heteroskedastic Engression still provides gains in flexibility over baseline ($\sim$2-6% on tails).

Performance Trade-offs:

- 2-3× higher cross-seed variability in complex models.
- Increased computational cost with architectural extensions.

Experiments: Simulation Study



Experiments: River Discharge Application

Motivation:

Real-world application assessing model robustness in challenging conditions.

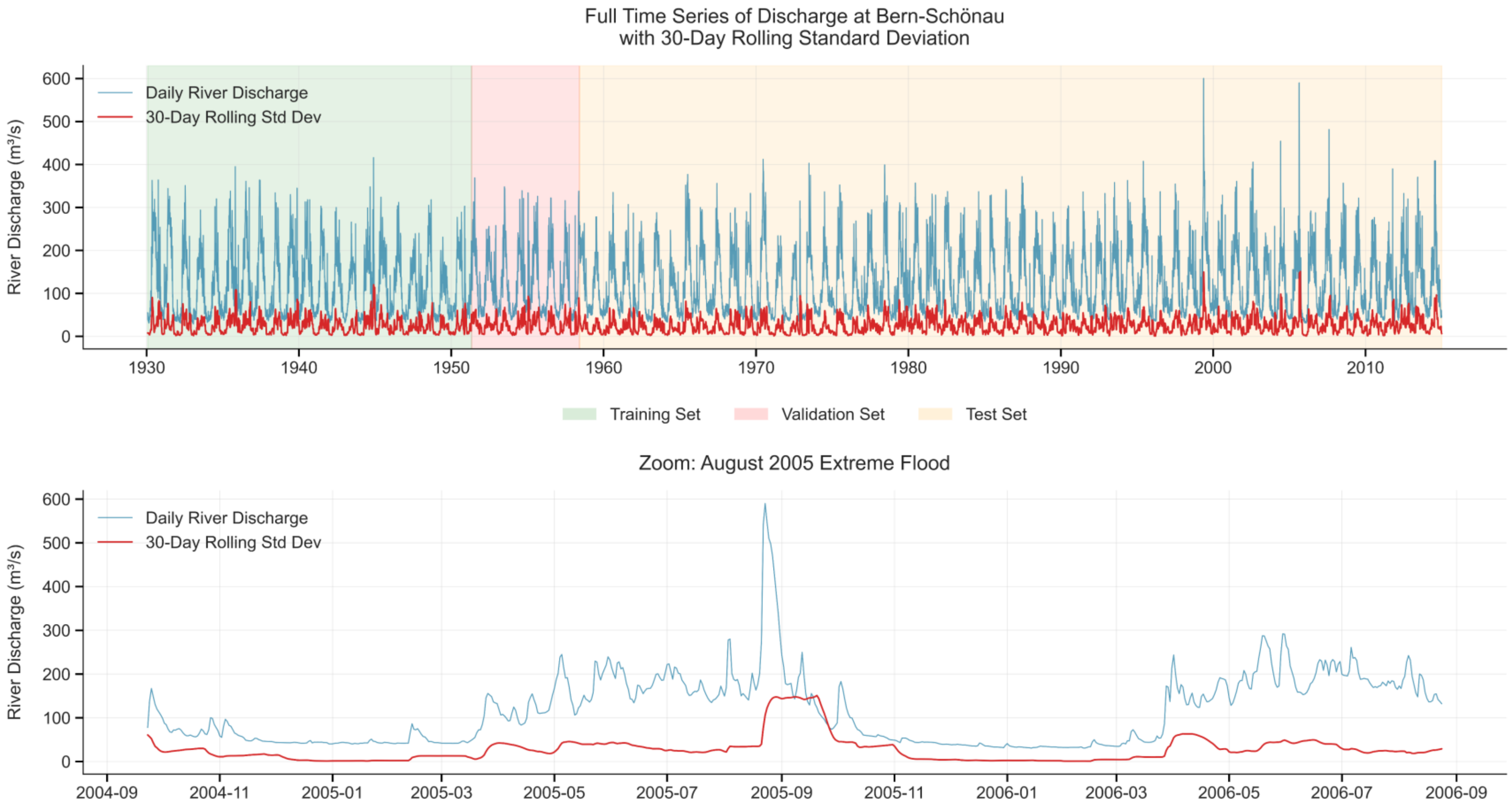
- Complex temporal dependencies from multiple physical processes.
- Non-stationary behavior from long-term climatic changes.
- Practical relevance for flood risk assessment in extreme event (August 2005)

Dataset: Swiss Aare river at Bern-Schönau(1930-2014).

- Response: Daily average discharge (m^3/s) at Bern-Schönau gauging station.
- Covariates: 7 upstream measurements (1 discharge from Gsteig + 6 precipitation stations).

Experiments: River Discharge Application

Experimental Setup:



Experiments: River Discharge Application

Key Findings:

Sequential Models Underperform in Mean Prediction:

- 2-3% worse RMSE, 44-65% worse extreme prediction than non-sequential variants.
- Daily aggregation weakens sequential signals.

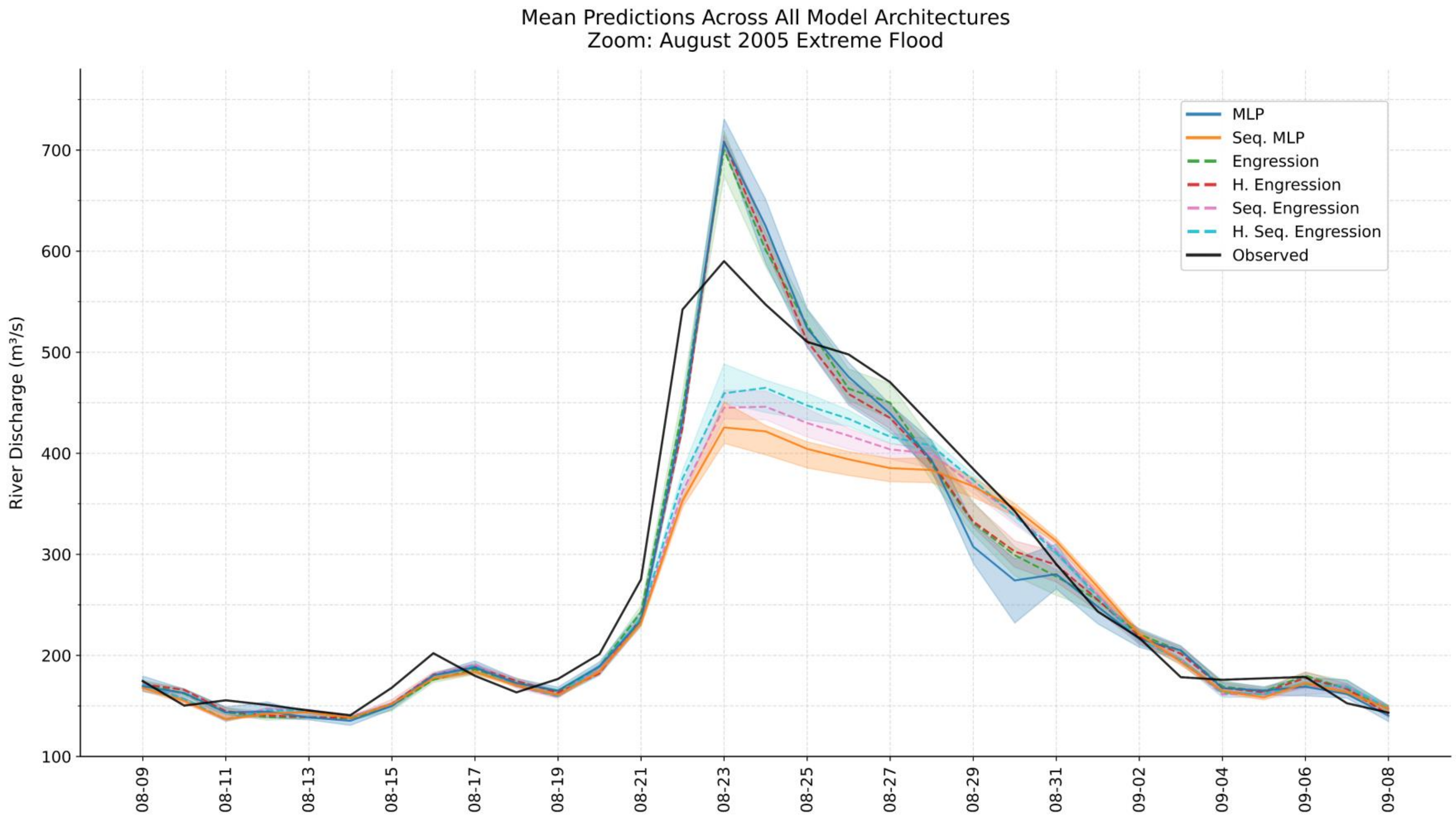
Engression Robustness Even Under Misspecification:

- 7-9% lower RMSE, 9-20% better extreme prediction vs deterministic models.
- Performance maintained despite suboptimal alignment with data structure.

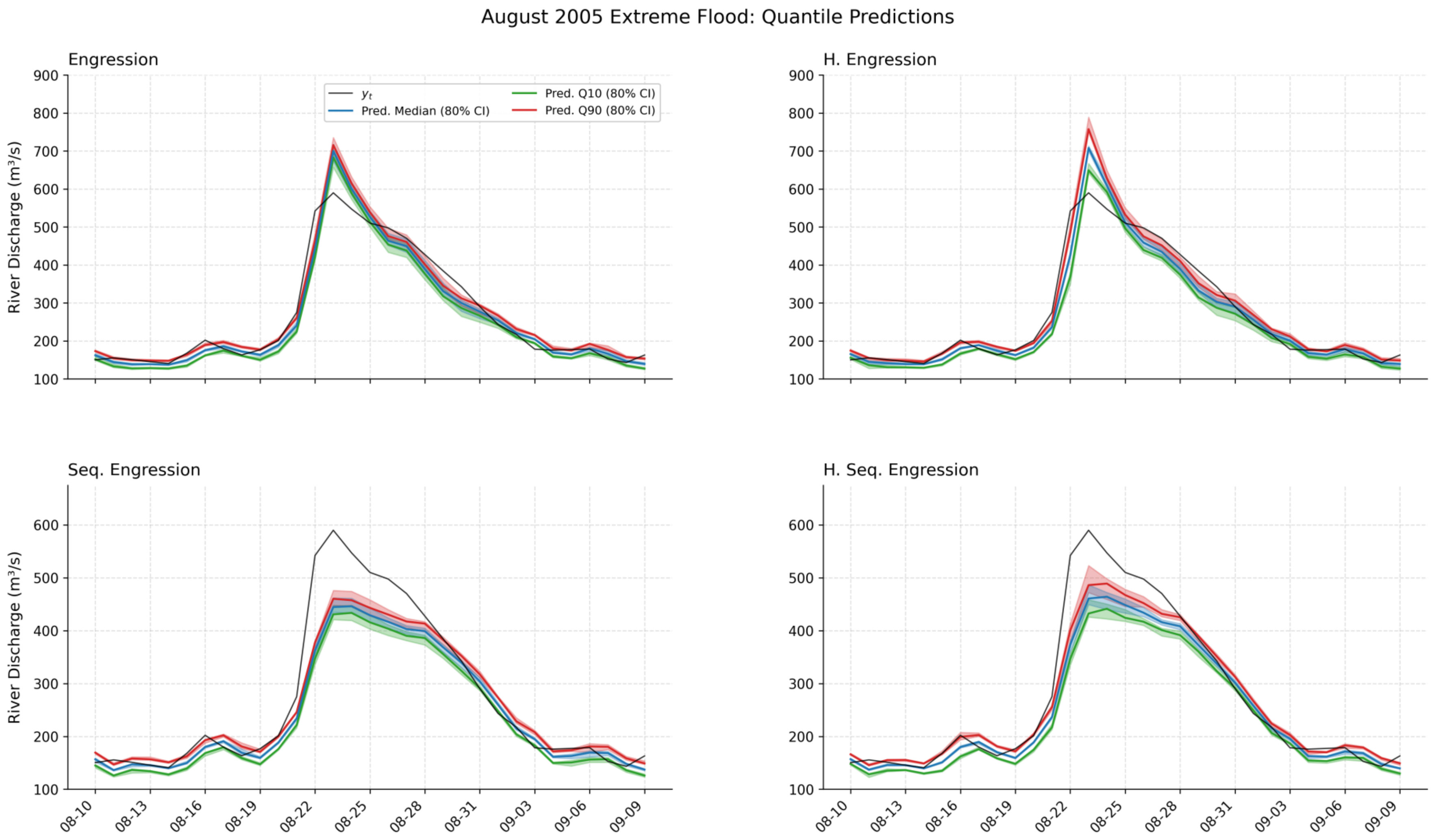
Heteroskedastic Extensions Improve Distributional Quality:

- Best when combined with temporal encoder (Heteroskedastic Sequential Engression), competitive standalone (Heteroskedastic Engression).
- Critical for capturing time-varying volatility and adaptive uncertainty quantification.

Experiments: River Discharge Application



Experiments: River Discharge Application



Conclusion

Main Findings:

- Successfully extended Engression to temporal forecasting via temporal encoder and heteroskedastic noise mechanisms.
- Engression-based models achieve comparable mean prediction with superior extreme event forecasting compared to the respective deterministic baselines.
- Extensions show context-dependent benefits by improving asymmetric tail forecasting and adaptive uncertainty quantification.

Conclusion

Key Limitations:

- Limited empirical validation with two case studies. Broader temporal domains required for generalizability.
- Complex simulation design complicated isolation of individual architectural contributions.
- Computational resource limitations constrained experimental design and hyperparameter exploration.