

# An Extreme-Adaptive Time Series Prediction Model Based on Probability-Enhanced LSTM Neural Networks

Yanhong Li, Jack Xu, David C. Anastasiu

## Problem Description

### Problem: Univariate Time Series with Extreme Events Forecasting

$$[X_1, X_2, \dots, X_T] \in R^T \rightarrow [X_{T+1}, \dots, X_{T+H}], \in R^H$$

#### Challenges:

- Most values are normal and contribute significantly to the overall prediction performance
- Few extreme values but they must be precisely forecasted to avoid disastrous events

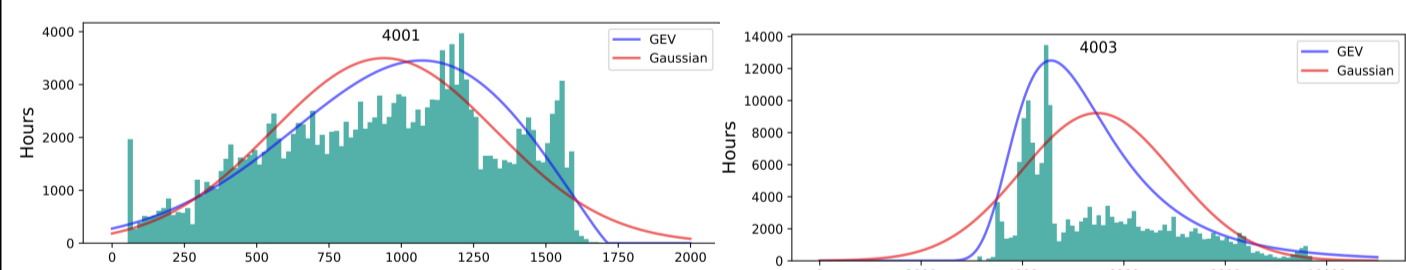
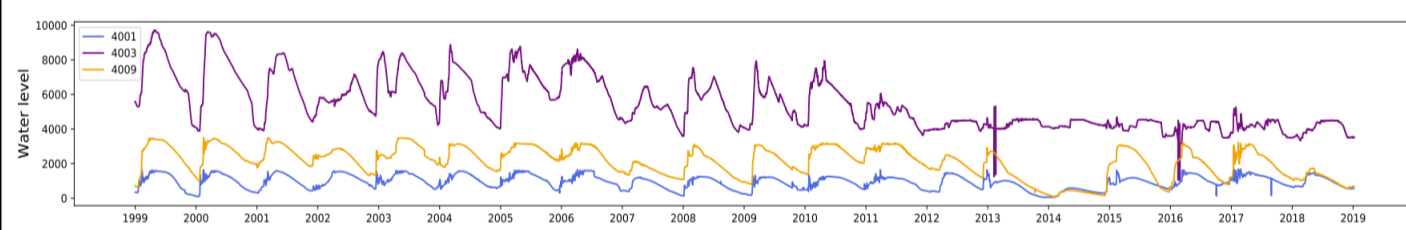
#### Goal:

- A model that concurrently learns extreme and normal prediction functions
- Long sequence forecasting
- Good generalization

#### Dataset:

- We evaluate the proposed model on the difficult 3-day ahead hourly water level prediction task applied to 9 reservoirs in California

## Extreme Events



**GEV** (Generalized Extreme Value) distribution provides a better fit, showing the presence of extreme values in our data,

$$F(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \quad (1)$$

where  $\mu \in R$ ,  $\sigma > 0$ , and  $\xi$  are the location, scale, and shape parameters, respectively, conditioned on  $1 + \xi(x - \mu)/\sigma > 0$ .

## Motivation

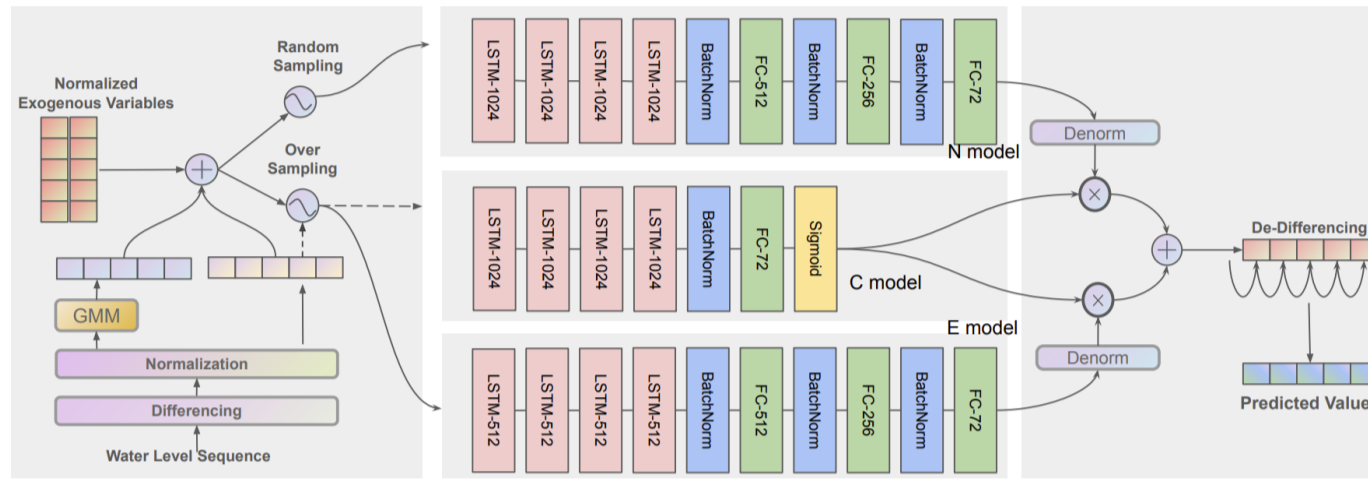
Achieving the best overall prediction performance, without sacrificing either the quality of normal or of extreme predictions.

Root Mean Square Error (*RMSE*)

Mean Absolute Percentage Error (*MAPE*)

## Proposed Framework

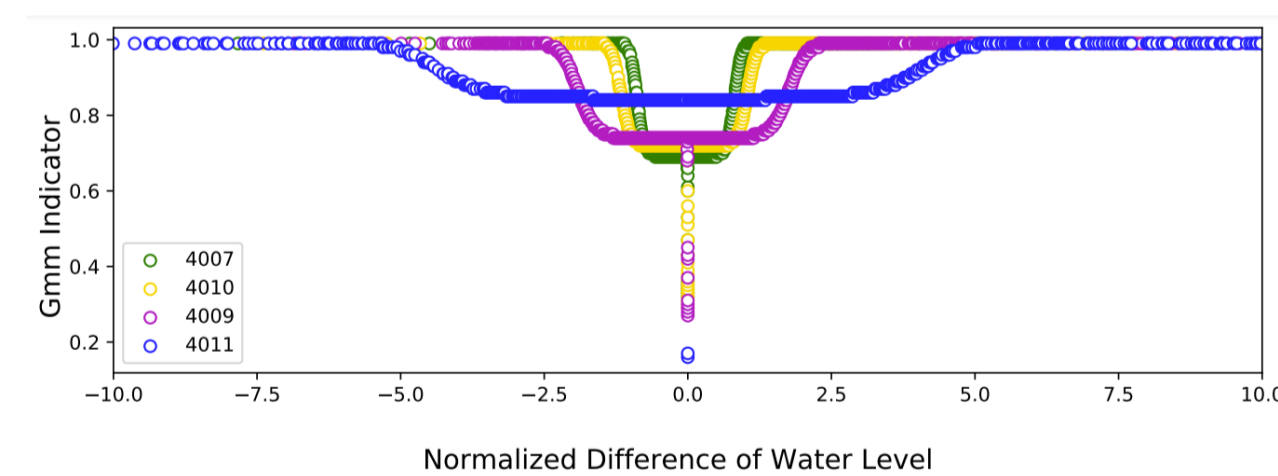
- **NEC framework:** a framework to account for the distribution shift between normal and extreme values. NEC is composed of three separate models, which can be trained in parallel:



- The Normal (N) model is trained to best fit normal values in the time series.
- The Extreme (E) model is trained to best fit extreme time series values.
- The Classifier (C) model is trained to detect when a certain value may be categorized as normal or extreme.
- NEC is a purely single variate model; NEC+ uses optional exogeneous other inputs.

## GMM Indicator Method

- **GMM Indicator:** an unsupervised clustering approach to dynamically produce distribution indicators, which improves the model's robustness to the occurrence of severe events.



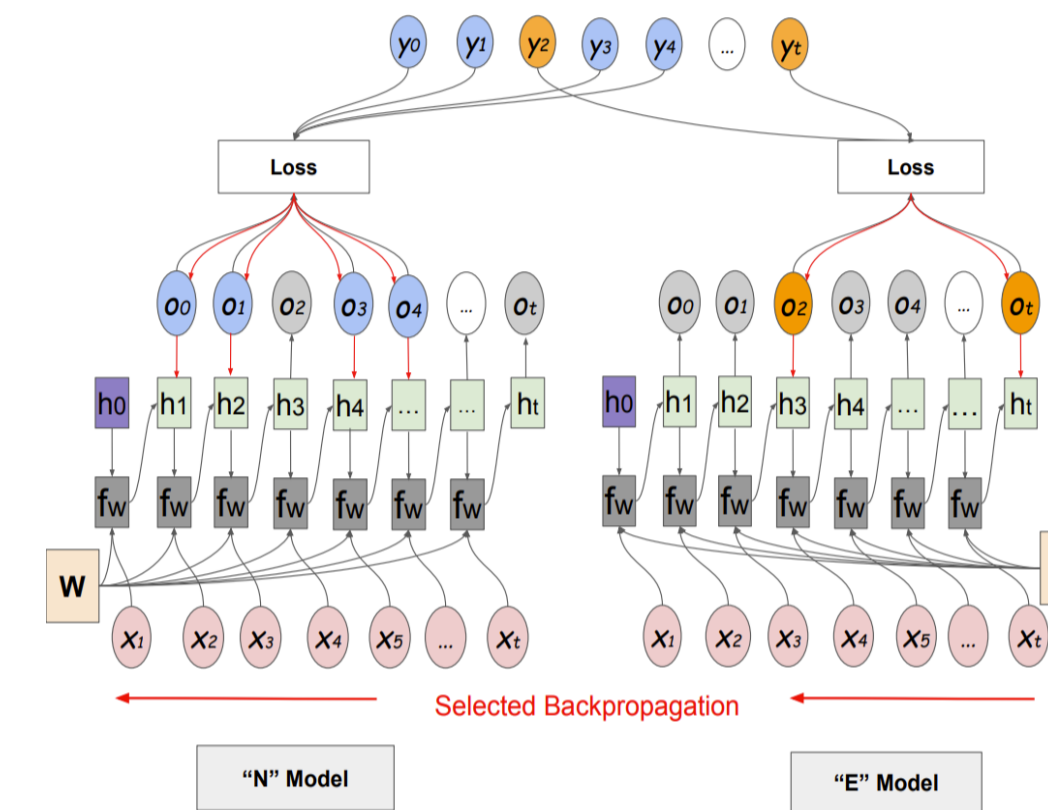
A Gaussian mixture model (GMM) is a weighted sum of M component Gaussian densities,

$$p(x|\lambda) = \sum_{i=1}^M w_i g(x|\mu_i, \Sigma_i),$$

where  $x$  is a D-dimensional continuous-valued vector,  $w_i \forall i = 1, \dots, M$  are the mixture weights, and  $g(x|\mu_i, \Sigma_i)$  are the component Gaussian densities. We compute an indicator feature as the weighted sum of all component probabilities, given the weights learned when fitting the GMM model.

## Selected Backpropagation Method

- Two-stage sampling policy:
  - randomly sample subsections of length  $h + f$  from the series to use in training our models.
  - perform stratified sampling of regions with and without extreme values, allowing the E and C models to oversample up to OS% samples with at least 1 extreme value in the prediction zone.



- Selected Backpropagation:
  - N model: only normal values add to the loss;
  - E model: only extreme values add to the loss.

## Parameterized Loss Function Method

$$BCE(t, p) = -(t \times \log(p) + (1 - t) \times \log(1 - p))$$

$$L = \beta \times BCE(t, p^\alpha) + (1 - \beta) \times RMSE(t, p)$$

where  $\alpha$  and  $\beta$  are parameters that can be tuned. Values  $\alpha > 1$  cause the model to predict  $p$  values that are higher in general in order to minimize the distance between  $t$  and  $p^\alpha$ .

- Problem: for datasets with a high imbalance between the two classes, BCE will favor the prominent class.
- The BCE part: can be thought of as a blunt instrument that grossly exaggerates all miss-classifications in order to more accurately predict the obscure class
- The RMSE part: allows for a more gentle penalty based on the distance between  $t$  and  $p$ .

## Baselines

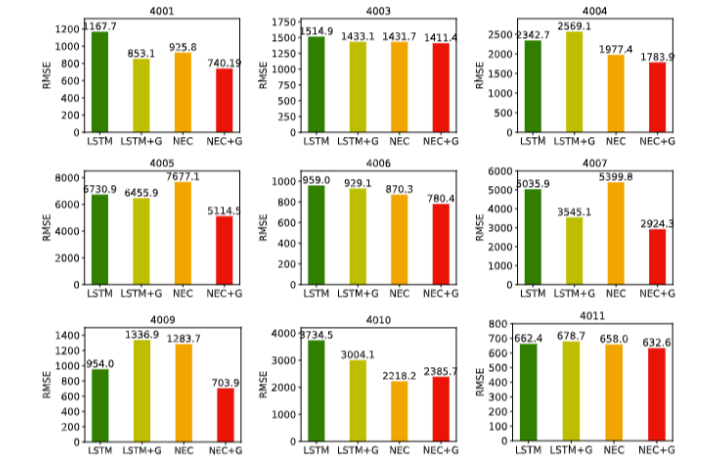
- ARIMA
- Facebook Prophet
- LSTM
- DNN-U: univariate state-of-the-art LSTM-based encoder-decoder hydrologic model.
- Attention-LSTM: a state-of-the-art hydrologic model for stream-flow prediction.
- N-BEATS: a state-of-the-art time series prediction method that outperformed all competitors on the standard M3, M4 and TOURISM datasets.

## Research Questions

- What is the effect of adding the GMM indicator to a model?
- What is the effect of introducing exogenous features?
- How do the loss function parameters affect performance?
- How does NEC+ compare against state-of-the-art baselines?

## Effects of Proposed Methods

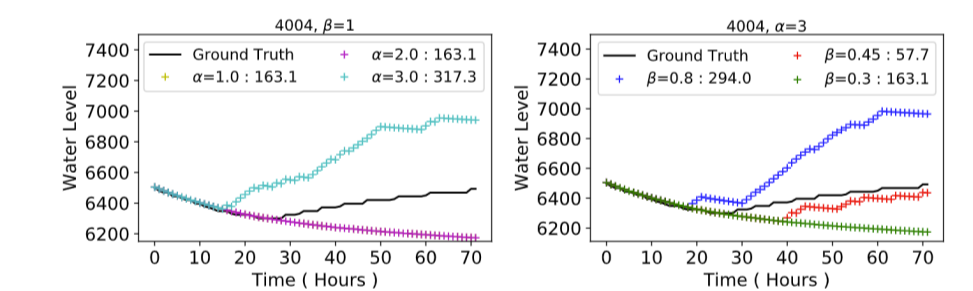
- Effects of GMM indicator



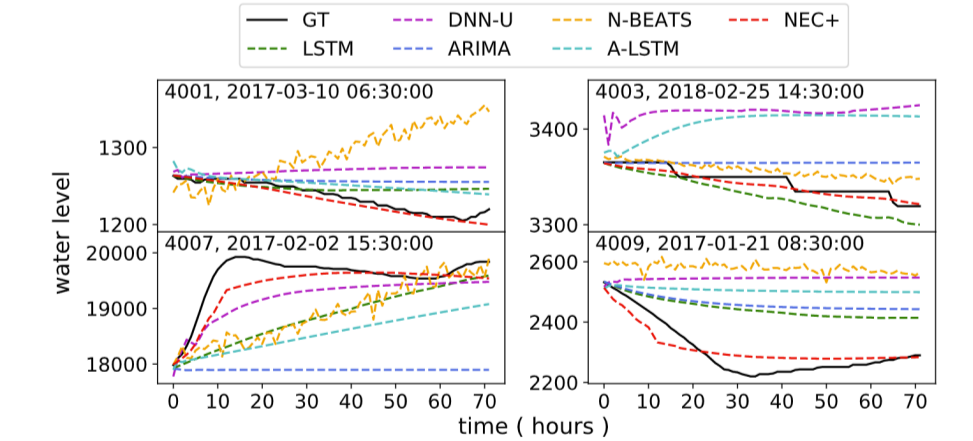
- Effects of introducing exogenous features

Model/Reservoir	4005	4007	4010
LSTM	6730.93	5035.91	3734.53
LSTM+W	7568.68	5728.30	4145.16
LSTM+G	6455.90	3545.19	3004.14
LSTM+G+W	9760.62	4128.37	2602.58
NEC+G	5114.49	2924.30	2385.77
NEC+G+W (NEC+)	4352.74	2092.73	2275.48

- Effects of parameterized loss function



- Effects of NEC+



## Evaluation

Table 1: Effectiveness Comparison (RMSE) of NEC+ Against Baselines for 9 Reservoirs

Model/Reservoir	4001	4003	4004	4005	4006	4007	4009	4010	4011
ARIMA	1016.32	1839.70	2501.97	9692.87	1039.38	5854.48	1060.05	3465.20	690.23
Prophet	8469.74	38827.22	95279.31	181607.50	20904.57	187603.80	28629.44	114115.4	2829.26
LSTM	1167.73	1514.90	2342.71	6730.93	959.05	5035.91	954.04	3734.53	662.48
DNN-U	1162.01	1597.72	3989.20	9878.41	983.27	4320.40	1411.63	4257.58	763.73
A-LSTM	878.71	1536.04	2548.56	8919.33	1638.65	13529.86	1064.15	2914.75	700.50
N-BEATS	937.24	1926.74	2280.83	7153.82	960.42	3153.76	1295.90	3162.17	514.30
NEC+	740.19	1411.44	1783.92	4352.74	780.46	2092.73	703.93	2275.48	632.61

Table 3: MAPE of NEC+ vs. Baselines for 9 Reservoirs

Model/Reservoir	4001	4003	4004	4005	4006	4007	4009	4010	4011
ARIMA	1.3573	0.7626	0.8694	1.2560	1.5401	0.8517	0.9504	1.7871	3.2914
Prophet	16.7877	19.8559	38.9642	35.6662	56.0537	32.9152	31.8069	45.2579	15.3312
LSTM	1.6697	0.6153	0.7450	1.0092	1.3264	0.9253	0.9298	2.5520	3.1282
DNN-U	1.6509	0.6812	1.8738	1.9394	1.4551	0.6509	1.5604	2.1582	3.7131
A-LSTM	1.3533	0.6506	0.8424	1.2060	2.8017	2.1738	0.9705	1.3986	3.4137
N-BEATS	1.3346	0.7972	0.7882	1.1405	2.0061	0.4709	1.4580	1.7146	2.3108
NEC+	1.0319	0.5687	0.6030	0.6350	1.0662	0.3316	0.5992	1.2894	2.9237

