Long-Term Hydrologic Time Series Prediction with LSPM

Sicheng Zhou
PRISMS High School
Princeton, NJ, USA

David C. Anastasiu
Santa Clara University
Santa Clara, CA, USA
danastasiu@scu.edu

ABSTRACT

Predicting multivariate time series has been a topic of interest among researchers for a long time, especially in hydrological prediction. Due to the presence of extreme events, hydrological prediction requires capturing long-range dependencies and modeling rare but significant extreme values. Accurate prediction of these dependencies is often accomplished using complex models, such as stacked RNNs or transformer-based models, which can be computationally expensive and challenging to train. In addition, existing studies have identified a strong correlation between streamflow and rainfall data. However, the use of additional input data in these studies has often been insufficient, resulting in predictions with low accuracy. In this paper, we address these issues and propose LSPM, a **L**ong **S**hort-term **P**olar-Learning time series forecasting **M**odel. LSPM learns polar representations through a feature reuse method called EDDU (Encoder Double-Decoder Unit). EDDU creatively incorporates exogenous input to generate long-term predictions based on these learned representations. To maximize the use of indicator sequences from exogenous data, LSPM enhances short-term predictions by a carefully designed loss function and integrates them into the overall forecast, improving robustness to short-term severe events. Experiments on four real-life hydrologic streamflow datasets demonstrate that LSPM significantly outperforms both state-of-the-art hydrologic time series prediction methods and general methods designed for long-term time series prediction.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; **Supervised learning**; **Sequence prediction**.

KEYWORDS

Long-term, short-term, RNN, time series forecasting, hydrologic prediction, extreme events, streamflow, polar representation.

ACM Reference Format:
Sicheng Zhou and David C. Anastasiu. 2024. Long-Term Hydrologic Time Series Prediction with LSPM. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3627673.3679957

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3679957

1 INTRODUCTION

Time series forecasting is critical in many domains, particularly streamflow prediction [1, 12, 13], where sporadic but major extreme events [4, 22, 28, 29] such as flash floods and droughts provide distinct challenges. Traditional models struggle to account for the huge variation and non-stationarity of water levels, which are influenced by weather, topography, and human activities. Rain data, which is highly volatile and closely associated to extreme events, is a major indicator in these models. However, historical rain data alone is insufficient for predicting long-term streamflows. Many recent studies attempt to use multivariate machine learning models for this purpose [18, 27, 27], however the results have been unsatisfactory.

Traditionally, machine learning and statistics-based models like ARIMA [3] have been foundational for time series prediction, but they struggle with large streamflow variations and are usually limited to short-term forecasting. Neural network architectures have been explored for hydrologic forecasting, yet they mainly focus on short-term horizons. Transformer-based models [30, 31] claim high performance for long-horizon tasks, but imbalanced data or severe events can degrade the performance of long-term predictions. Moreover, many multivariate models align multiple data variables as high-dimensional inputs, limiting the effectiveness of auxiliary variables. The DAN [18] model, designed to utilize rain data, is overly complex and still yields suboptimal results.

In this paper, we propose LSPM, a Long-Range Time Series Forecasting Model enhanced by Short-Term Attention and Polar Learning. LSPM introduces the EDDU (Encoder Double-Decoder Unit), which first uses an encoder to generate high-dimensional features, which are shared and decoded by two decoders, utilizing a multi-loss mechanism to maintain polar features. This succinct mechanism enhances robustness to severe events more effectively. Additionally, LSPM emphasizes the use of auxiliary variables in short-term predictions, creatively integrating them into long-term forecasting.

2 RELATED WORKS

Streamflow forecasting [2, 6, 23] is crucial for enhancing water resource allocation, management, flood warning, and mitigation of flood-related damages. To fully leverage auxiliary variables in prediction, multivariate time series studies have utilized various algorithms, from traditional methods like vector autoregression [24] and multivariate exponential smoothing [9] to advanced deep learning methods. Wang et al. [25] developed a hybrid model that combines Empirical Mode Decomposition (EMD), Ensemble EMD (EEMD), and ARIMA [3] for long-term streamflow forecasting. However, their study did not evaluate the effectiveness of these models on datasets containing extreme values. Recently, transformer-based techniques such as Autoformer [26] and Reformer [14] have been

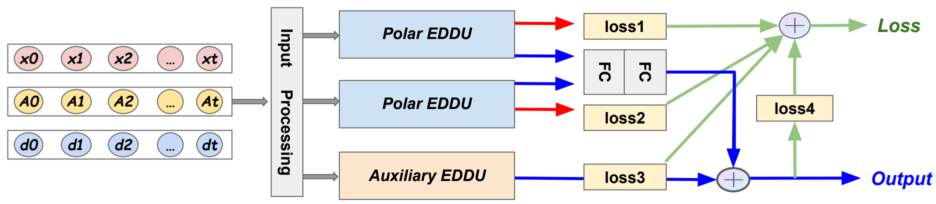

Figure 1: LSPM Architecture and Computing Path.

proposed for long-term forecasting, offering sophisticated dependency discovery and modeling capabilities. Informer [30] introduced a ProbSparse self-attention mechanism and a generative-style decoder to significantly increase the inference speed for long-sequence predictions. FEDFormer [31] improved long-term forecasting by randomly selecting a fixed number of Fourier components to capture the global characteristics of time series. However, recent research suggests that simpler linear models [5, 27] may outperform these complex approaches. Other methods, including representation [18] and hybrid techniques [7, 16, 19], have also been explored for long-term forecasting. NEC+ [17] has shown superior performance in hydrologic time series prediction, especially in the presence of extreme events. Attention-LSTM [15] has been a state-of-the-art model for predicting streamflow using rainfall data.

Few of these prior works have concentrated on addressing both prolonged sequences and used auxiliary variables effectively. To bridge this gap, we propose LSPM to address the challenges of streamflow prediction. Experiments on four real-life streamflow datasets show that LSPM significantly outperforms state-of-the-art methods for hydrologic long-term time series prediction.

3 PROBLEM DEFINITION

3.1 Problem Statement

Assume we have a collection of m ($m > 1$) connected univariate time series, with each row representing a separate time series. We will predict the future h time steps for the first time series (x_1) using past data from various length-t observed series. The problem can be described as:

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,t} \\ x_{2,1} & \cdots & x_{2,t} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,t} \end{bmatrix} \in \mathbb{R}^{m\times t} \rightarrow [x_{1,t+1}, ..., x_{1,t+h}] \in \mathbb{R}^h$$

where $x_{i,j}$ denotes the value of time series i at time j. The matrix on the left and vector on the right are the inputs and output in our method, respectively.

3.2 Dataset Processing

In this project, we used four hydrologic datasets from Santa Clara County, CA–Ross, Saratoga, UpperPen, and SFC–each consisting of a streamflow variable, which we will predict, and a rainfall variable as auxiliary data. Our task was to forecast streamflow during the wet seasons of a hydrologic year, excluding the summer months, specifically from September 2021 to May 2022.

4 METHODS

Figure 1 shows the entire architecture of LSPM. We will explain the details in the following sections.

4.1 Basic EDDU

To maximize feature learning while avoiding bias in application, we propose a novel feature-reuse neural network unit called Encoder Double-Decoder Unit (EDDU), showcased in Figure 2. EDDU first uses an encoder to generate high dimensional hidden features. These features are then decoded by two decoders. The first decoder's output (denoted as the red arrow) can be used to force the encoder to learn specific information in the features by applying a loss penalty. The second decoder reuses the same hidden features generated by the encoder to produce a prediction sequence.

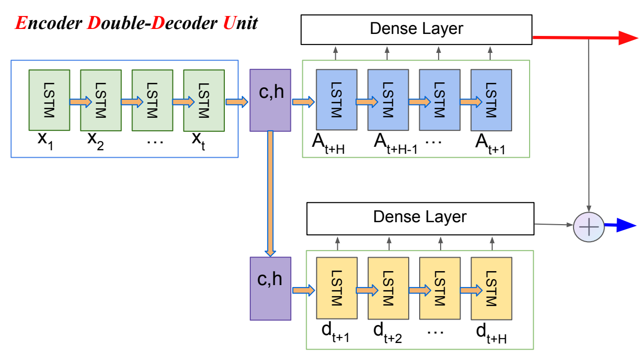

Figure 2: Basic EDDU Dataflow.

For the first decoder, the input is the reversed sequence of auxiliary variables. We believe that the auxiliary variables closer to the start of the prediction sequence are more indicative, so we use the reversed auxiliary variable sequence (denoted as A_{t+H} to A_{t+1}) as a way to learn this rich representation. The second decoder's input is the time features (denoted as d_{t+1} to d_{t+H}) of the prediction interval. These time features are generated using sine and cosine transformations, encoding each month-day date into a feature pair that captures the 365-day periodicity within a range of -1 to 1. This approach strengthens the temporal relationship discovery in the prediction.

The outputs of the first and second decoders are then summed to form the final output of the polar EDDU (denoted as the blue arrow), which is a type of residual design [10]. By combining the outputs in this way, we leverage the strengths of both decoders to improve the overall prediction accuracy.

The basic idea of the EDDU involves using two decoders, effectively creating two hidden functions (neural networks) that make the training process clearer. In a traditional encoder-decoder structure, assuming the encoder is a function h and the decoder is a function f, the expected output $\hat{y}$ can be represented as $\hat{y} = f(h(x, A), d)$. In EDDU, the structure is defined as $\hat{y} = f_3(f_1(h(x), A), f_2(h(x), d))$. By applying loss penalties, we allocate distinct responsibilities to f_1 and f_2. This strategy simplifies the training of complex time series models by making the roles of each component more explicit.

4.2 Polar EDDU

In LSPM, we use two EDDUs to learn polar representations of the prediction variable, referred to as polar EDDUs. One EDDU is designed for extreme feature learning, employing a distance-based weighted loss strategy where weights increase with the distance from the mean. The other EDDU focuses on normal value feature learning. The prediction outputs from the polar EDDUs are merged through two fully connected layers. Unlike the deeper propagation in DAN networks, our feature reuse and double decoder design is a novel approach that achieves higher performance in a simplified manner.

4.3 Auxiliary EDDU

Figure 3 showcases the auxiliary dataflow in the EDDU module. While long-term forecasts depend on trends and distributions in the predicted dataset, short-term fluctuations are more influenced by external variables. Therefore, we designed the auxiliary EDDU based on the basic EDDU, focusing instead on learning both short-term and long-term features rather than polar values. Thus, the first decoder's output does not require polar representation learning and is not reinforced as a separate loss item. Instead, the output is refined through two layers of CNNs, refining local curves and generating a robust prediction output. The second decoder's output serves two purposes: first, it is combined with the outputs of the two polar EDDUs to form the final output; second, its short-term prediction is used as a loss penalty item to further emphasize the importance of auxiliary variables in short-term predictions. In short-term predictions, this means the auxiliary variable EDDU output is more accountable for accuracy.

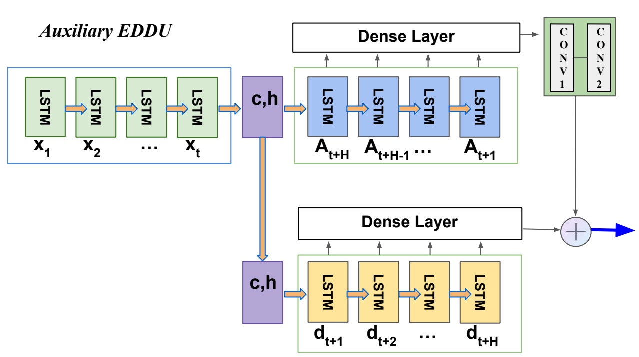

Figure 3: Auxiliary EDDU Dataflow.

4.4 Distance-Based Loss Function

To force the polar EDDUs to learn rich representations, we use multiple loss functions [8, 11, 20] when training the LSPM mode. We define $w_f = (\tanh(y))^2$ to focus on the accuracy of points further away from the mean value of 0 (since the series were standardized) and $w_n = (1 - |\tanh(y)|)^2$ to focus more on the points closer to zero. Based on weights w_f and w_n, we build our polar loss items as follows,

$$\mathcal{L}_1 = RMSE((\hat{y}_f \odot w_f), (y \odot w_f)),$$
$$\mathcal{L}_2 = RMSE((\hat{y}_n \odot w_n), (y \odot w_n)),$$
$$\mathcal{L}_3 = RMSE(\hat{y}_{aux}[:s], y[:s]),$$
$$\mathcal{L}_4 = RMSE(\hat{y}, y),$$

where $\hat{y} = y_p + y_{aux}$ (y_p is the output of polar EDDUs) and $\odot$ is the element-wise multiplication. Similarly, to further emphasize the importance of auxiliary variables in short-term predictions, we use L_3 to regularize the output of auxiliary EDDU $\hat{y}_{aux}$. Here, s represents the length of the short-term interval, which is set to 6 (1.5 hours) in our experiments. Then, the overall loss is composed as,

$$\mathcal{L} = \lambda \times (\mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3) + \mathcal{L}_4,$$

where λ is a multiplier ($\lambda = max(-1 \cdot e^{\frac{epoch}{45}} + 2, 0.2)$ in our experiments) applied on the polar regularization terms, decreasing with each epoch.

4.5 Oversampling

To balance the sparse distribution of extreme events, we use the Kruskal-Wallis test [21], a non-parametric method, to evaluate the normality of training samples and guide our oversampling policy. The test ranks the data, calculates rank sums for each group, and computes the H value to identify significant differences between groups. We include a sample in the training set if $H > \epsilon$ (with ϵ set to 80 in our experiments), or with a probability $p < 1$ (set to 18% in our experiments).

5 EVALUATION

5.1 Baselines

We include seven state-of-the-art models and their experiment results from the latest research DAN [18] for comparison, of which FEDFormer [31], Informer [30], NLinear [27], and DLinear [27] focus on long-term time series forecasting, while NEC+ [17] holds the best performance for hydrologic time series prediction in the presence of extreme events. Attention-LSTM [15] was used as a state-of-the-art hydrologic multivariate model. Finally, DAN [18] learns rich representations along with a representation merging model that makes streamflow predictions in an expandable way.

5.2 Experiment Settings

The training and validation datasets were randomly selected from time series data covering the period from January 1988 to August 2021. Before training, all time series were pre-processed using a log transform ($x_i = \log(1 + x_i) \forall i$) and standardization (subtracting the mean and dividing by the standard deviation). Predictions were post-processed by inverting these transformations. During the inference

Table 1: Multivariate Long-Term (h = 288) Series Forecasting Results on Four Datasets

Methods	Ross		Saratoga		UpperPen		SFC	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
FEDformer	6.01	2.10	6.01	1.55	3.05	1.87	23.54	2.35
Informer	7.84	4.05	5.04	1.43	5.88	4.10	39.89	8.64
Nlinear	6.10	1.99	5.23	0.83	1.57	0.45	18.47	0.92
Dlinear	7.16	3.10	4.33	1.40	3.53	2.35	21.62	2.74
NEC+	9.44	4.80	1.88	0.17	2.22	0.95	17.00	1.07
LSTM-Atten	7.35	3.74	6.49	1.80	6.35	4.76	34.17	9.90
DAN	**4.25**	**0.07**	1.80	0.14	1.10	0.15	15.23	0.26
LSPM	**4.25**	0.10	**1.76**	**0.12**	**1.02**	**0.07**	**14.99**	**0.21**

phase, we predicted streamflow using rolling predictions at intervals of 4 hours. Each prediction, however, inferred 288 data points, i.e., the predicted streamflow over the next 3 days at 15 minute intervals.

We used the Adam optimizer with an initial learning rate of 0.001 that decreased by a factor of 0.9 after each epoch. The training procedure was configured to run for a maximum of 50 epochs, with an early halt triggered if no improvement was seen after four consecutive epochs. All models were trained on a dataset of 20,000 samples, and their performance was validated using a randomly selected set of 120 samples that were not part of the training set. Codes and data for LSPM can be found at https://github.com/davidanastasiu/lspm.

5.3 Model Parameters

For all models, we tested various LSTM layer widths and identified the optimal configurations for each dataset. The best results for Saratoga and Ross were achieved with 2 layer of 256 nodes, for UpperPen and SFC with 2 layers of 400 nodes each. We set the prediction horizon $h = 288$ (3 days) and the input sequence length $t = 1440$, consistent with baseline comparisons. The CNN layers used in these models had kernel sizes of 7 and 3, with padding of 3 and 1, respectively, and a stride of 1 for both layers.

6 RESULTS AND ANALYSIS

Table 1 presents the test root mean squared error (RMSE) and mean absolute percentage error (MAPE) for the models that achieved the best performance on our test dataset. In the forecasting task, our proposed model LSPM outperformed all baselines across the four benchmark datasets. Compared to the second-best results, LSPM achieved equal or better RMSE on all sensors and up to 50% MAPE improvement in the best case.

The visual analysis further supports these findings. In Figure 4, LSPM's predictions closely follow the ground truth, especially in datasets with significant fluctuations. As shown in the figure, LSPM captures short-term fluctuations and extreme values more accurately than DAN, as evidenced by the lower RMSE values.

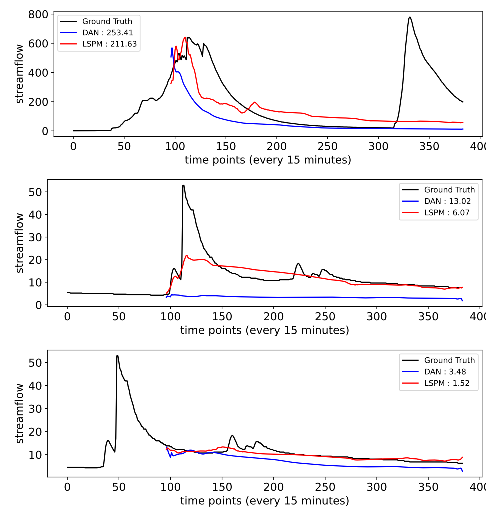

Figure 4: Example comparisons with the second best baselines; the DAN model can capture the general trend of the ground truth, but the LSPM enables it to more accurately predict the streamflow values, both during normal conditions and in the presence of rare events.

7 CONCLUSION

In this paper, we proposed LSPM, a Long Short-term Polar-Learning time series forecasting Model. We first introduced a novel neural network unit called EDDU (Encoder Double-Decoder Unit), which underscores the importance of balancing feature learning and practical application in developing robust and accurate long-term forecasting models. The innovative use of feature reuse and double decoders allows LSPM to achieve superior performance while maintaining a simplified network structure. Additionally, LSPM effectively integrates short-term and long-term feature learning, leveraging external variables to enhance prediction accuracy. Our experimental results across four real-life hydrologic streamflow datasets demonstrate that LSPM significantly outperforms state-of-the-art methods in both RMSE and MAPE metrics.

REFERENCES

- [1] A.A. Masrur Ahmed, Ravinesh C. Deo, Afshin Ghahramani, Qi Feng, Nawin Raj, Zhenliang Yin, and Linshan Yang. 2022. New double decomposition deep learning methods for river water level forecasting. *Science of The Total Environment* 831 (2022), 154722. https://doi.org/10.1016/j.scitotenv.2022.154722
- [2] Imen Aichouri, Azzedine Hani, Nabil Bougherira, Larbi Djabri, Hicham Chaffai, and Sami Lallahem. 2015. River flow model using artificial neural networks. *Energy Procedia* 74 (2015), 1007–1014.
- [3] George.E.P. Box and Gwilym M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. Holden-Day, .
- [4] Z. Chen, H. Yu, Y. Geng, Q. Li, and Y. Zhang. 2020. EvaNet: An Extreme Value Attention Network for Long-Term Air Quality Prediction. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE Computer Society, Los Alamitos, CA, USA, 4545–4552. https://doi.org/10.1109/BigData50022.2020.9378094
- [5] Abhimanyu Das, Weihao Kong, Andrew B. Leach, Shaan Mathur, Rajat Sen, and Rose Yu. 2023. Long-term Forecasting with TiDE: Time-series Dense Encoder. *Transactions on Machine Learning Research* abs/2304.08424 (2023), 21. https://api.semanticscholar.org/CorpusID:258180439
- [6] Lamine Diop, Ansoumana Bodian, Koffi Djaman, Zaher Mundher Yaseen, Ravinesh C Deo, Ahmed El-Shafie, and Larry C Brown. 2018. The influence of climatic inputs on stream-flow pattern forecasting: case study of Upper Senegal River. *Environmental earth sciences* 77 (2018), 1–13.
- [7] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee-Keong Kwoh, Xiaoli li, and Cuntai Guan. 2021. Time-Series Representation Learning via Temporal and Contextual Contrasting, In 30th International Joint Conference on Artificial Intelligence (IJCAI-21). *arXiv preprint arXiv:2106.14112* 1, 2352–2359. https://doi.org/10.24963/ijcai.2021/324
- [8] Sachin Sudhakar Farfade, Mohammad J. Saberian, and Li-Jia Li. 2015. Multi-view Face Detection Using Deep Convolutional Neural Networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval* (Shanghai, China) *(ICMR '15)*. Association for Computing Machinery, New York, NY, USA, 643–650. https://doi.org/10.1145/2671188.2749408
- [9] Everette S Gardner Jr. 1985. Exponential smoothing: The state of the art. *Journal of forecasting* 4, 1 (1985), 1–28.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), Las Vegas, NV, USA, 770–778. https://doi.org/10.1109/CVPR.2016.90
- [11] Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. arXiv:1705.07115 [cs.CV] https://arxiv.org/abs/1705.07115
- [12] Huseyin Cagan Kilinc and Bulent Haznedar. 2022. A hybrid model for streamflow forecasting in the Basin of Euphrates. *Water* 14, 1 (2022), 80.
- [13] Donghyun Kim, Joonseok Lee, Jongsung Kim, Myungjin Lee, Wonjoon Wang, and Hung Soo Kim. 2022. Comparative analysis of long short-term memory and storage function model for flood water level forecasting of Bokha stream in NamHan River, Korea. *Journal of Hydrology* 606 (2022), 127415.
- [14] Nikita Kitaev, Łukasz Kaiser, and Anselm Levskaya. 2020. Reformer: The Efficient Transformer. arXiv:2001.04451 [cs.LG] https://arxiv.org/abs/2001.04451
- [15] Yan Le, Changwei Chen, Ting Hang, and Youchuan Hu. 2021. A stream prediction model based on attention-LSTM. *Earth Science Informatics* 14 (06 2021), 1–11. https://doi.org/10.1007/s12145-021-00571-z
- [16] Hao Li, Jie Shao, Kewen Liao, and Mingjian Tang. 2022. Do Simpler Statistical Methods Perform Better in Multivariate Long Sequence Time-Series Forecasting?. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (Atlanta, GA, USA) *(CIKM '22)*. Association for Computing Machinery, New York, NY, USA, 4168–4172. https://doi.org/10.1145/3511808.3557585
- [17] Yanhong Li, Jack Xu, and David Anastasiu. 2023. An Extreme-Adaptive Time Series Prediction Model Based on Probability-Enhanced LSTM Neural Networks. *Proceedings of the AAAI Conference on Artificial Intelligence* 37 (06 2023), 8684–8691. https://doi.org/10.1609/aaai.v37i7.26045
- [18] Yanhong Li, Jack Xu, and David Anastasiu. 2024. Learning from Polar Representation: An Extreme-Adaptive Model for Long-Term Time Series Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. AAAI Press, Vancouver, British Columbia, Canada, 171–179.
- [19] Yanhong Li, Jack Xu, and David C. Anastasiu. 2023. SEED: An Effective Model for Highly-Skewed Streamflow Time Series Data Forecasting. In *2023 IEEE International Conference on Big Data (Big Data) (IEEE BigData 2023)*. IEEE Computer Society, Los Alamitos, CA, USA.
- [20] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. 2018. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London, United Kingdom) *(KDD '18)*. Association for Computing Machinery, New York, NY, USA, 1930–1939. https://doi.org/10.1145/3219819.3220007
- [21] Patrick E. McKight and Julius Najab. 2010. *Kruskal-Wallis Test*. John Wiley & Sons, Ltd, Hoboken, NJ, USA, 1–1. https://doi.org/10.1002/9780470479216.corpsy0491 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470479216.corpsy0491
- [22] Di Qi and Andrew J Majda. 2020. Using machine learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences* 117, 1 (2020), 52–59.
- [23] Yangho Song, Yoonkyung Park, Jungho Lee, Moojong Park, and Youngseok Song. 2019. Flood forecasting and warning system structures: Procedure and application to a small urban stream in South Korea. *Water* 11, 8 (2019), 1571.
- [24] James H Stock and Mark W Watson. 2001. Vector autoregressions. *Journal of Economic perspectives* 15, 4 (2001), 101–115.
- [25] Zhi-Yu Wang, Jun Qiu, and Fang-Fang Li. 2018. Hybrid Models Combining EMD/EEMD and ARIMA for Long-Term Streamflow Forecasting. *Water* 10, 7 (2018), 14. https://doi.org/10.3390/w10070853
- [26] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems* 34 (2021), 22419–22430.
- [27] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are transformers effective for time series forecasting?. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23)*. AAAI Press, Washington DC, USA, Article 1248, 8 pages. https://doi.org/10.1609/aaai.v37i9.26317
- [28] Mi Zhang, Daizong Ding, Xudong Pan, and Min Yang. 2023. Enhancing Time Series Predictors With Generalized Extreme Value Loss. *IEEE Transactions on Knowledge and Data Engineering* 35, 2 (2023), 1473–1487. https://doi.org/10.1109/TKDE.2021.3108831
- [29] Y. Zhang, J. Li, A. Carlo, A. K. Manda, S. Hamshaw, S. M. Dascalu, F. C. Harris, and R. Wu. 2021. Data Regression Framework for Time Series Data with Extreme Events. In *2021 IEEE International Conference on Big Data (Big Data)*. IEEE Computer Society, Los Alamitos, CA, USA, 5327–5336. https://doi.org/10.1109/BigData52589.2021.9671387
- [30] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 12 (May 2021), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325
- [31] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In *Proceedings of the 39th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 162)*, Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (Eds.). PMLR, Baltimore, MD, USA, 27268–27286. https://proceedings.mlr.press/v162/zhou22g.html