AdapFAN: A Novel Multichannel Deep Learning Method Driven by Data and Process Representation for Flood Forecasting

Zixuan Wang

Chinese International School, Hong Kong, China

ABSTRACT

Over the last decades, floods have become the most common and deadly natural disaster on the planet. While many countries currently lack effective early warning systems and alerts(India, Malaysia, South Asia etc.). Flood forecasting is a very exploratory project because there is substantial peer-reviewed evidence that flood early warning can prevent between 30% to 50% of both fatalities and economic harms(Perera et al.). Existing challenges in flood forecasting include how to establish relations between more variables, extend the lead-time for actual warning with improved accuracy, and how model-dependent(process-based) and model-independent(data-driven) strategies can be balanced to magnify the long term predictive potential of deep learning models. Given the competitive performance deep learning models have shown in temporal variation modeling, temporal dependency learning, and multivariate representation, this paper aims to explore the potential of deep learning model in forecasting floods and focuses on the adaptation and training of the deep learning network, namely AdapFAN, for flood forecasting. Building on the data-driven basis, AdapFAN incorporates process algorithms(taking multi periodicity, interdimensional, and intradimensional dependencies into account) in separate paths to further its predictive length and reliability. Specifically, AdapFAN adopts TimesNet to transform complex 1D time series into multiple 2D tensors representing interperiod and intraperiod temporal variations, tackling the representation limitation of 1D series. Building on that, AdapFAN adopts channel independence to avoid the loss due to the average of multi periodicity between different dimensions. To take the influence of dimensional dependencies into account, dimensions are aggregated into 3D space and allow conv3d to analyze the hidden inner relations between multivariates. Using AdapFAN, this project achieves reliability in predicting extreme riverine events in ungauged watersheds at up to a 4-day lead time(96 timestamp) that is similar to or better than the reliability of *nowcasts* (0-day lead time) from a current state-of-the-art global modeling system.

Introduction

Flood is a climate-related natural disaster. It is defined as an overflow of water over dry lands, usually followed by fatalities and economic damage. According to the international disaster database, flooding occurs more frequently than all other types of natural hazards across the globe, and accounts for 39% of all disasters arising from natural hazards since 2000, with >94 million people affected worldwide every year(Guha-Sapir et al., 2018). However, substantial peer-reviewed evidence stated that flood early warning can prevent between 30% to 50% of both fatalities and economic harms(Perera et al.). According to WMO, “Recorded economic losses linked to extreme hydro-meteorological events have increased nearly 50 times over the past five decades, but the global loss of life has decreased by a factor of about 10, thus millions of lives are being saved”("Early Warning"). When the rate of runoff-volume exceeds 1000 m^3/s, the likelihood of flood is likely to take place.

HIGH SCHOOL EDITION
Journal of Student Research

Aim of this project is to forecast floods in 4-days lead time(96 timestamp) with improved accuracy measured by mse and mae. However, it's hard to find much reference on streamflow forecasting using hydrological analyzing models(available models in this aspect are limited). Traditional hydrological models including conceptual models, physically based models, and statistical models degrade significantly in performance when calibrated for multiple basins together instead of for a single basin alone. This inability to transfer or extrapolate the hydrologic information from one to another(e.g. From gauged to ungauged watershed) indicates the problem that model-dependent methods with a convoluted process network following pre-assumed patterns can limit a model's scalability.

Recent studies cast doubt on the effectiveness of transformers in addressing forecasting tasks by comparing its performance with linear models or vanilla LSTM. Transformers' problem with low scalability, meaning it can perform well in one scenario but usually perform very poorly or far less competent in another scenarios, are probably due to the same reason – overly process-dependent (being overly process-dependent results in the failure to learn and encode regional differences in catchment characteristics and translate it into appropriately heterogeneous hydrologic behavior uses). Hence, my selected base model in this project, namely Timesnet, is a task general deep learning method not following transformer architecture and is showing strong adaptabilities and competence in all tasks including long term, short term forecasting, anomaly detection, imputation, and classification.

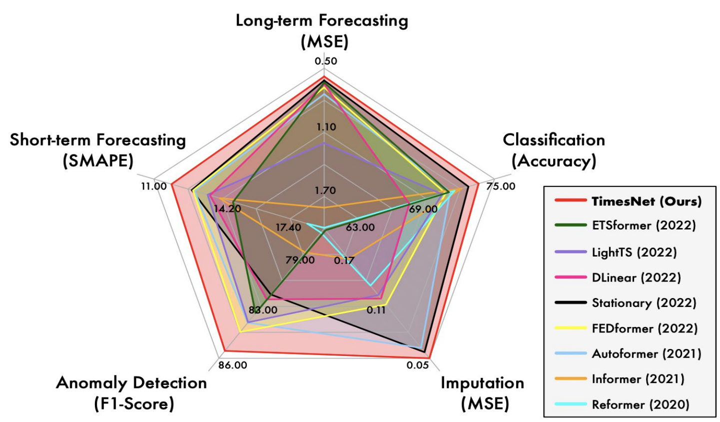

Chart 1. Model performance comparison: As a foundation model, TimesNet achieves consistent state-of-the-art performance on five mainstream analysis tasks compared with other customized models(Wu et al.)

On the other hand, we do need certain process representations when dealing with long term forecasting. Unlike short term forecasting, longer prediction needs stronger structural information that is extendable and is considered inherent in the data feature. In this respect, time series specificities should be taken into account. This includes multivariate and temporal variations.

Multivariate time series refers to data that involves multiple variables (e.g. streamflow rate, precipitation, slope, loam, ET, etc.) and is influenced by the relationship between variables. AdapFAN, by reshaping original 1D time series into 3D space, with length(x) representing the number of timestamps, width(z) representing the number of features in one dimension, and height(y) displaying the number of dimensions enhances the visual representation of multiple dimensions(multivariates, its respective features, and timestamp are holistically rearranged for analyzing). Building on that, by making use of 3D convolutional neural network's competence in addressing high dimensional visual tasks, intradimensional(between multivariate) dependencies can be coherently processed.

For temporal variations, it refers to the common features a time series data possesses. It includes periodicity(this further introduces inter-period, intra-period relationship, and), trend, and autocorrelation. To prevent the setting of too many predefined patterns that may result in scalability reduction of my model, I incorporate only the most long-term-forecasting-related time series specificity, multi periodicity(more implicit relative to trend, more variable, and exist in the long term relative to autocorrelation), as the process representation. I designed AdapFAN which uses TimesNet as backbone and has improved its accuracy by proposing channel independence for distinct variates. In this way, multi periodicity of the original 1D time series is decomposed into its top k prominent period lengths(which can be different for distinct variates). Reshaped and padded based on different period lengths, a set of 2D tensors are produced to show not only temporal variations of different period length, but also the simultaneous representation of interperiod and intraperiod series.

Technically, based on above motivations, my contributions are summaries in three folds:

- Due to the importance of scalability for hydrologic forecasting models, I selected the task general time series method, namely TimesNet, as the backbone of AdapFAN.
- Visual representation of multivariate in higher dimensions: variates, features, and timestamp are holistically processed using 3D convolutional neural networks to achieve holistic interdimensional representation of variants, features, and timestamp.
- Modeling temporal variations in three aspects as process representations: Multi periodicity, interperiod, and intraperiod variations derived from multiperiodicity are decomposed to provide structural reference for long term forecasting.

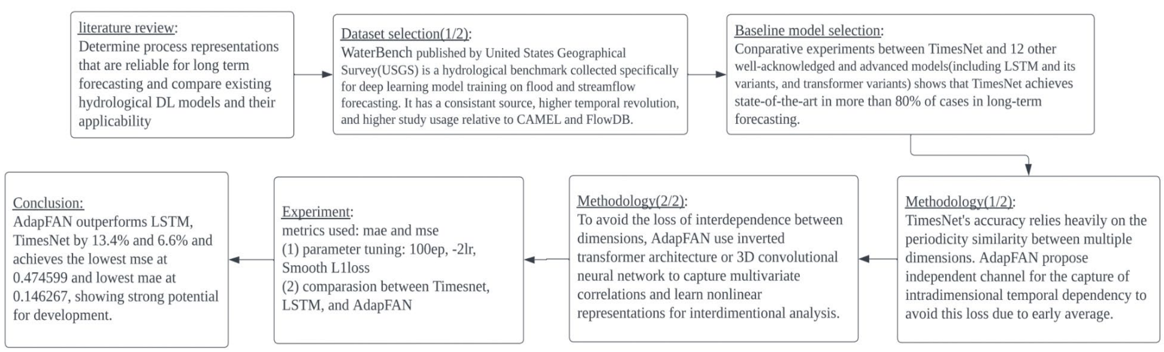

Chart 2. Framework of the project showing all subsections of the procedure

The framework of my overall progression is shown in chart 2. By training AdapFAN on WaterBench data set using meteorological time series data and static catchment attributes, AdapFAN achieves the lowest mse at 0.474599 and lowest mae at 0.146267. This project was able to significantly improve performance compared to a set of several different hydrological benchmark models and can send out signals whenever the rate of run-off volume exceeds 1000 m^3/s.

Related Work

Model-Dependent and Model-Independent Methods for Flood Forecasting

Model-dependent methods use process-based models. These models usually follow predefined patterns and physical assumptions that give specific algorithms to data processing. To give some examples, Samaniego proposed the multiscale parameter regionalization (MPR) method(L et al.), which establishes the model and a regionalization scheme by regressing the parameters of a set of transfer functions that are predefined and map the parameters of hydrological models to ancillary data, such as soil properties. A conceptual model trained on eleven catchments was calibrated by Seibert and regressed on the available features of the catchments(J). When compared to seven other catchments, the regionalization capacity's stated performance varied from a Nash-Sutcliffe Efficiency (NSE) of 1.42 to 1.76 (relatively high, indicating a low reliability)(J). In that sense, most cases using model-dependent methods rely on prior knowledge of the hydrological system and hence can hardly be adapted to the appearance of regional differences. Or to put it in a more common sense way, specificity can be an indication of low elasticity and may lead to rigidity in long term forecasting regardless of variable influencers.

Model-independent methods use data-driven models(this is usually referred to as black box). Most of the published papers that are based on data-driven approaches use vallila long-short-term-memory(LSTM) and its variants. Very few papers explore other RNN-based deep learning models including gated recurrent unit(GRU) and sequence to sequence(S2S) network as flood forecasters, which indicates there has not had much development other than these few vanilla foundations(LSTM and RNN models). In addition, despite its advantage in short term prediction, the existing problem for these process based methods is its low accuracy in long term forecasting. Even *Nowcast*, the current state-of-art model for flood forecasting using LSTM, has 0 lead time(only archives a few hours ahead, not even a day).

Temporal Variation Modeling

As one of the time series specificities, temporal variation modeling has been well explored. Conventional models assume that temporal variance adheres to predetermined patterns. like Holt-Winster, ARIMA, and Prophet. However, the actual usefulness of these classical approaches is limited since the variations of real-world time series are typically too complicated to be covered by these established patterns. However, deep learning methods like RNN-based models, TCN (temporal convolutional network), and MLP (Multilayer Perceptron) are available for temporal modeling. Notably, this work takes into account the temporal 2D-variations based on decomposed periodicity, which are not taken into account by any of the above approaches.

Furthermore, transformers are being passionately explored in this aspect. Primary approach for transformer-based forecasting models to achieve better temporal variation modeling is by enhancing self-attention to a sparse version. (1) locality-sensitive hashing attention (used in reformer); (2) multihead attention; (3) memory-efficient attention; (4) hashing attention; (5) multo-round LSH attention (used in reformer); and (7) prob sparse attention (used in informer). These models still use the point-wise representation aggregation, but performance is much enhanced. Because of the sparse point-wise connections, they will therefore forfeit information utilization in the process of improving efficiency, creating a bottleneck for long-term time series forecasting. My project instead of continuing the use of sparse representation of time series, series level modeling is proposed. AdapFAN is carefully designed regarding the holistic representation of 1D time series reshaped into higher visual dimensions (2D for temporal variation and 3D for multivariate) and processing using effective tools.

Multivariates Modeling

Not much has been done or no significant achievement in this aspect yet due to the inherent complexity of capturing multivariate dependencies. Multivariate modeling has to take extensive assumptions and correlated variables into account, and by introducing one more variable, this may result in double the number of extra related factors, hence its difficulty increases exponentially. However, I would consider a novel method of data representation a beneficial approach to multivariate modeling. Since the convolutional neural network's competence in image processing shows the potential for holistic analysis of data, and if we embed 1D time series

HIGH SCHOOL EDITION
Journal of Student Research

to another spatial arrangement of higher dimensions that can present multivarients separately and simultaneously, better performance may be expected.

In addition, channel design is concerned with the processing of variants. A multivariate time series is a multi-channel signal. Channel-mixing refers to the case where the input token takes the vector of all time series features and projects it to the embedding space to mix information. On the other hand, channel-independence means that each input token only contains information from a single channel. This was proven to work well with CNN (Zheng et al., 2014) and linear models (Zeng et al., 2022).

Methodology

The overall architecture of AdapFAN is shown in figure 3, which is an end-to-end method. AdapFAN contains two modules, one is Timesnet-based method supported with channel independence, fast fourier transform, and inception block determined through comparative experiment. The other one is 3D modeling for multivariate making use of 3D cnn.

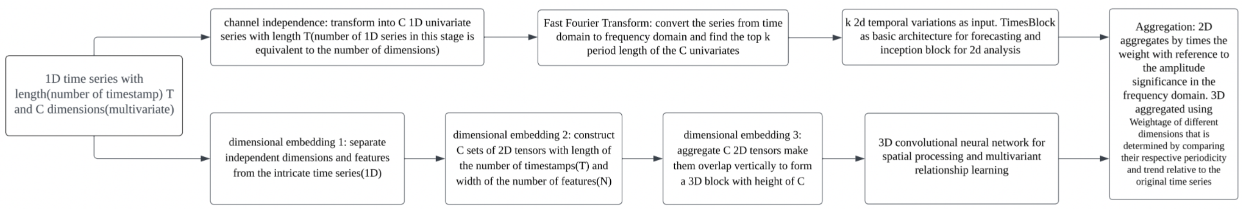

Figure 3. Module overview of AdapFAN(C denotes dimension, T denotes timestamp, P denotes period, and k are parameters set to be 5)

Inter-Dimensional Dependencies Representation Using TimesNet: Transform 1D Series into A Set of 2D Tensors

For a time series, what we see is just a graph of what it looks like when multiple sub-parts, including periods of different lengths, overlap and influence each other in an implicit sense. This results in a time series that is inherently complex and difficult to extract features from. Timestamp possessing the same period length, time points are affected by both intraperiod and interperiod variations. Secondly, for each sub-components in the time series, the time points contained within them are affected not only by the neighboring time points possessing the same period length, but also by the variations of the neighboring periods, i.e. intraperiod and interperiod variations, intraperiod variations can indicate short-term time patterns, and the interperiod variation can reflect the long term aspect. If multiperiodicity is not clearly decomposed, the time series as a whole can be assumed to have infinite period length. In that sense, prediction can only be based on the interaction between each neighboring timepoint, which is short term forecasting. While the long-term reliability would be extremely difficult to find.

Based on the above time series specificities, two features that an effective forecaster have to possess are the ability to decompose entangled multi periodicity into multiple individual period lengths and operate interperiod and intraperiod processing. 2D tensors with different period lengths will display different inter and intra period variations and hence difference hydrological behavior. Therefore, the model needs to further truncate the complex time series according to the top k period length, then interpret the inter and intra period future behavior.

Since it is difficult to capture intra- and inter-period variations simultaneously in one dimension (a long infinitely extended time-series with the time dimension being the x-axis), the time series are truncated into T/P short cuts with single periodicity based on learned C top k period lengths. They are then stitched together chronologically up and down to form a 2D representation (with T/P as the length, and P (all the time points contained in one period) as the width). The final result is k 2d temporal variations generated based on different period lengths. As a result, the ability to represent time series previously limited to 1d time series is tackled, and the learning of interperiod and intraperiod variations can be performed simultaneously.

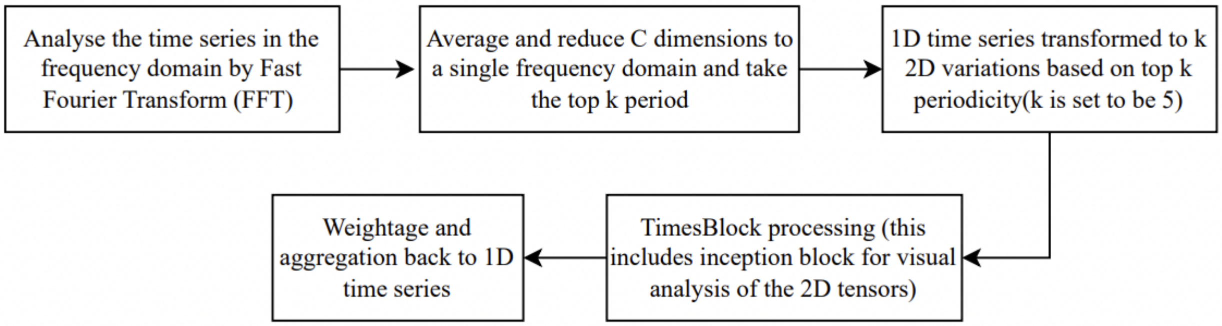

Figure 4. Mechanism of Timesnet architecture

Based on an understanding of the multi-periodic nature of time series and the two forms of intra- and inter-period temporal interaction of a given period length, simultaneous analysis of three forms of temporal variation including point level, series level, and cross series level is achieved by reconstructing the original 1d time series into a set of 2d image according to learned period length, which is a breakthrough in terms of visualization and task universality.

Channel Independence

We consider the following problem: given a collection of multivariate time series samples with lookback window L : (x_1, ..., x_L) where each x_t at time step t is a vector of dimension M, we would like to forecast T future values (x_{L+1}, ..., x_{L+T}). Our model is illustrated in the figure below where the model makes use of the timesnet architecture. For the forward process, we denote a $i\ th$ univariate series of length L starting at time index 1 as $x_{1:L}^{(i)} = (x_1^i, \ldots, x_L^i)$ where i = 1, ..., M. The input $(x_1, \ldots, x_L\ \ )$ is split to M univariate series x (i) ∈ R 1×L, where each of them is fed independently into the TimesNet backbone according to our channel-independence setting. Then the timesnet backbone will provide prediction results xˆ (i) = (ˆx (i) L+1, ..., xˆ (i) L+T) ∈ R 1×T accordingly.

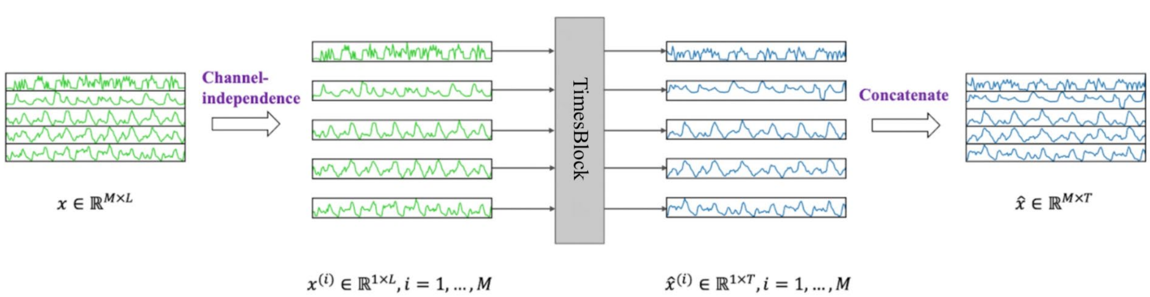

Figure 5. Channel independence between dimensions to avoid the lose or interference of periodicity, Timesblock is the functioning unit of TimesNet

Multivariate time series data is divided into separate channels. They share the same TimesBlock backbone, but the forward processes are independent. For the decomposition of multi periodicity, this project fast fourier transform(FFT). Through which, the time domain of the time series is transformed into the frequency domain. The formula is shown below where X[k] is the expression of a discrete series in the frequency domain, x(t) is the original discrete series in the time domain, k denotes the frequency index, N is the series length.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N}$$

Based on the amplitude values of each frequency component in the frequency domain, the largest k amplitude values are taken. These amplitudes each correspond to one frequency index and are indicated by the magnitude to occupy the k highest significance(prominence) over the other existing periods of the time series. Only k(in my code k is set to be 5) most significant frequencies are taken to avoid noise and overanalyzing (to the extent that some are meaningless) of periodicities beyond the original time series. From the values of these k most significant frequencies, their corresponding period length can be found using the formula below, where p denotes period length, T denotes timestamp, and f denotes frequency.

$$\boldsymbol{p_i} = \frac{\boldsymbol{T}}{\boldsymbol{f_i}}$$

Due to channel independence, predictive reliability no longer depends on the similarity between period length of different dimensions. The C set of 2D tensor is not reshaped based on the averaged top k frequencies from the aggregated time series, instead it is based on separate dimensional levels(keep the correspondence).

Discover Intra-Dimensional Dependency and Association Between Multivariate: Aggregates The C Dimensions in 3D Space

Consider a single timestamp, it is under the influence of inter-period series and intra-period timestamps, but also correlates with other variants/dimensions involved in the scenario. Hence, aiming to capture the dependability between multivariates, this paper proposes to aggregate the C dimensions, with their respective 1 aggregated N features decomposed to form a set of 2D tensors(length is the number of timestamps and width is the number of features in one dimension C x [T, 1] → C x [T, N]), into 3D space and allow conv3d to analyze the hidden inner relation(height is the number of dimensions C x [T, 1, N] → [T, C, N]). The structure is represented in the figure below.

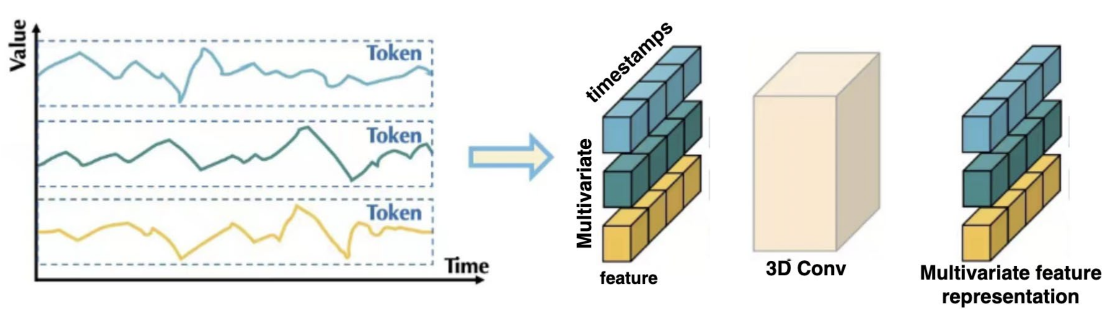

Figure 6. Multivariate modeling and feature representation. Time series decomposed and reshaped into 3 D space. Processing using 3D convolutional neural networks. Length(x-direction), width(z-direction), and height(y-direction) are as indicated

Step one : separate independent dimensions from the intricate time series(1D)
Step two : separate independent features in each dimension
Step three : construct C(the number of independent dimensions) sets of 2D tensors with length of the number of timestamps(T) and width of the number of features(N)
Step four : aggregate C 2D tensors make them overlap vertically to form a 3D block with height of the number of dimensions(C)

Result Aggregation

The learned 3D and 2D representations (variation of time series) will be transformed back into C comparable pairs of 1D representations that are dimension independent for aggregation. For the C times k 2D representations, independent dimension with its k variations, derived from its multi-periodicity, will be aggregated(given periodicity, found by evaluating the significance of its frequency, has its corresponding amplitude. Different variation times the weight/amplitude(A)) to produce one weighted representation of that independent dimension. Repeating aggregation for the other C-1 independent dimensions, we obtain C 1D representations that can be averaged with the outcome produced by the 3D representation.

$$\widehat{X_{3D}^{[T,C,N]}} \text{ and } C \times \widehat{X_{2D}^{[f,T/f]}} \; where \; f \in [1, 2, \ldots, k])$$

$$C \times \widehat{X_{3D\;to\;1D}^{[T,1]}} \text{ and } C \times \widehat{X_{2D\;to\;1D}^{[T,1]}}$$

After each dimension pair is averaged to produce C learned dimensions, we aggregate these dimensions to obtain the learned time series. For the last step : Weightage of different dimensions can be determined by comparing their respective periodicity and trend relative to the original time series (high similarity α high correlation α weights more).

Smooth L1 Loss function

$$l_n = \begin{cases} 0.5(x_n - y_n)^2/beta, & \text{if } |x_n - y_n| < beta \\ |x_n - y_n| - 0.5 * beta, & \text{otherwise} \end{cases}$$

The loss function used in AdapFAN is smooth L1. The formula is shown above where $l_n$denotes loss, $X_n$denotes the predicted value, $Y_n$denotes the actual value, beta is a hyper-parameter(in my code the default value of beta is 1). Smooth L1 is determined in experiment. It can prevent gradient explosion. In addition, Smooth L1 Loss combines the advantage of L2 Loss of fitting faster and having a derivative at point 0 for easy convergence. It also combines the advantage of L1 Loss in the boundary region, making the network more robust against outliers and able to pull back when the offset is large.

Experiment

Datasets Descriptions

This paper adapted the real-world scenario that is relevant to streamflow rate forecasting. We conduct experiments on the benchmark published by the United States Geographical Survey(USGS) to evaluate the performance of the proposed AdapFAN. The State of Iowa, the study area, is situated in the country's Midwest. With 71,655 miles of rivers and streams extending from border to border, it boasts a wealth and diversity of water resources (Iowa DNR, 2004). In 2008, catastrophic flooding in Eastern Iowa resulted in property losses exceeding $6 billion. Therefore, streamflow forecasting and monitoring are essential for Iowa's improved water resource management and disaster recovery. Furthermore, Iowa's agricultural sector has a low paving rate and little human involvement, which makes it a good place to conduct rainfall-runoff research. Below table is a description of different categories the dataset is concerned with. Dimension denotes the variate number of each dataset. In the dimensions row, (v) denotes variable, while (c) denotes constant. Frequency refers to the sampling interval of time point, that is the time interval in which the data is recorded and updated(15-60 minutes is a fairly small interval).

Table 1. Detailed datasets descriptions. Dataset Size denotes the total number of time points in (Train, Validation, Test) split respectively. Prediction Length denotes the future time points to be predicted and three prediction settings are included in each dataset

	1	2	3	4	5	6	7	8	9	10
Dimensions	(v) Streamflow rate	(v) Precipitation	(v) Evapotranspiration	(c) Travel time	(c) Area	(c) Slope	(c) loam	(c) Slit	(c) Silty-clay loam	(c) Sandy-clay loam
Unit	ft^3/s	mm/hour	mm/month	/	/	/	/	/	/	/
Frequency	15-60 minutes	Hourly	Monthly	constant	constant	constant	constant	constant	constant	constant
Dataset size [train, valid, test] : [25982, 3644, 7383]										
Dataset id (ifc_ID) : 519										

Another point to note, WaterBench, CAMEL, and DeepDP are the three major hydrological dataset available for flood research. However, the WaterBench dataset is chosen among CAMEL and DeepDP Because the time series in the CAMEL dataset are gathered from various sources, there might be a wide range of temporal variations, which makes predicting considerably more difficult. While deepDP is relatively new compared to WaterBench, it has less historical data and focuses more on flash floods instead of riverine floods, so WaterBench is considered the most suitable one and it, in addition, explicitly stated in its dataset overview that WaterBench is published specifically for deep learning model training and hydrological forecasting tasks.

HIGH SCHOOL EDITION
Journal of Student Research Volume 13 Issue 4 (2024)

Deep Learning Model Selection

This project extensively compare the well-acknowledged and advanced models in all five tasks, including the RNN-based models: LSTM (1997), and LSSL (2022); CNN-based Model: TCN (2019); MLP-based models: LightTS (2022) and DLinear (2023); Transformer-based models: Reformer (2020), Informer (2021), Pyraformer (2021a), Autoformer (2021), FEDformer (2022), Non-stationary Transformer (2022a) and ETSformer (2022).

Table 2. Full results for the long-term forecasting task. This project compares extensive competitive models under different prediction lengths. The input sequence length is set to 96 for the weather dataset. Avg is averaged from all four prediction lengths.

Models		**TimesNet (Ours)**		ETSformer (2022)		LightTS* (2022)		DLinear* (2023)		FEDformer (2022)		Stationary (2022a)		Autoformer (2021)		Pyraformer (2021a)		Informer (2021)		LogTrans (2019)		Reformer (2020)		LSSL (2022)		LSTM (1997)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	**0.172**	**0.220**	0.197	0.281	0.182	0.242	0.196	0.255	0.217	0.296	0.173	0.223	0.266	0.336	0.622	0.556	0.300	0.384	0.458	0.490	0.689	0.596	0.174	0.252	0.369	0.406
	192	**0.219**	**0.261**	0.237	0.312	0.227	0.287	0.237	0.296	0.276	0.336	0.245	0.285	0.307	0.367	0.739	0.624	0.598	0.544	0.658	0.589	0.752	0.638	0.238	0.313	0.416	0.435
	336	**0.280**	**0.306**	0.298	0.353	0.282	0.334	0.283	0.335	0.339	0.380	0.321	0.338	0.359	0.395	1.004	0.753	0.578	0.523	0.797	0.652	0.639	0.596	0.287	0.355	0.455	0.454
	720	0.365	**0.359**	0.352	0.288	0.352	0.386	**0.345**	0.381	0.403	0.428	0.414	0.410	0.419	0.428	1.420	0.934	1.059	0.741	0.869	0.675	1.130	0.792	0.384	0.415	0.535	0.520
	Avg	**0.259**	**0.287**	0.271	0.334	0.261	0.312	0.265	0.317	0.309	0.360	0.288	0.314	0.338	0.382	0.946	0.717	0.634	0.548	0.696	0.602	0.803	0.656	0.271	0.334	0.444	0.454

Mse or mae in red indicates it is the lowest value among all others over the same row. While values in blue are the second best results. Whether using mse or mae as the metric, it is clearly shown that Timesnet on average achieves the lowest loss value for all four prediction lengths.

Model Training

Table 3. Parameter tuning using dataset 519 (hyper-parameters : input 96, predict 96, encoder layer - 2, decoder layer - 1, Period - 2, Top k - 5, early stop is employed to prevent overfitting)

Trial	Methodology	ep	lr	Loss function	Epoch loss (test | train)	mse	mae
1	Timesnet	10	-5	Smooth L1 loss	0.09083 | 0.08715	0.5601865649223328	0.17039169371128082
2	Timesnet	50	-5	Smooth L1 loss	0.09135 | 0.08717	0.56008937201923	0.1703198302938217
3	Timesnet	100	-5	Smooth L1 loss	0.09245 | 0.08707	0.519901237487793	0.1702250838279724
9	Timesnet	200	-5	Smooth L1 loss	0.09452 | 0.9006	0.523419381294792	0.17269392311573029
3	Timesnet	100	-4	Smooth L1 loss	0.08109 | 0.07577	0.5293264389038086	0.15653079748153687
4	Timesnet	100	-2	Smooth L1 loss	0.07744 | 0.06121	0.48386871814727783	0.14621685445308685
6	Timesnet	100	-1	Smooth L1 loss	0.085 | 0.07006	0.5227121114730835	0.14834637939929962

5	Timesnet	100	-2	L1 loss	0.5912 | 0.21938	0.6329827391048472	0.178294738297421
5	Timesnet	100	-2	Mse	0.5505 | 0.27625	0.5062459707260132	0.1692022979259491
7	Timesnet and 3D	100	-2	Smooth L1	0.08356 | 0.05541	0.4745990037918091	0.1420362399101261
8	Timesnet(channel independence) and 3D	100	-2	Smooth L1	0.08653 | 0.05601	0.482191801071167	0.1403442712545395

The performance determined using mse and mae values(the lower the mse or mae, meaning lower loss, denotes a higher predicting accuracy) the epoch is determined to be 100 the optimum, learning rate to the power of -5 the optimum, smooth L1 loss the optimum. Text in red corresponds to the best result among the entire column.

Comparative Experiments

Table 4. Comparison between convolutional neural networks in TimesNet. Input length is 96, prediction length is 96(equivalent to 4 days lead time)

	Inception Block	Vgg	ResNet	ResNeXt	ConvNext	SwinBlock
MSE	0.169263934	0.17619237	0.18259283	0.19189992	0.19983721	0.18102931
MAE	0.219832912	0.22913782	0.22991807	0.24010027	0.23819213	0.22130982

By comparing different cnn for the analysis of 2D tensors in TimesNet, I went over six cnn options and determined to use inception block as it achieves the lowest mse and mae relative to the other five.

After building the model, I firstly compared AdapFAN to the two foundation models: S2S network and vanilla LSTM. the blue line is LSTM, the S2S is in red. Below these two lines, AdapFAN has been run for three times with small changes in its parameters, but the average value is clearly shown. AdapFAN has much lower value in mae or mse relative to LSTM and S2S as indicated in the y-axis.

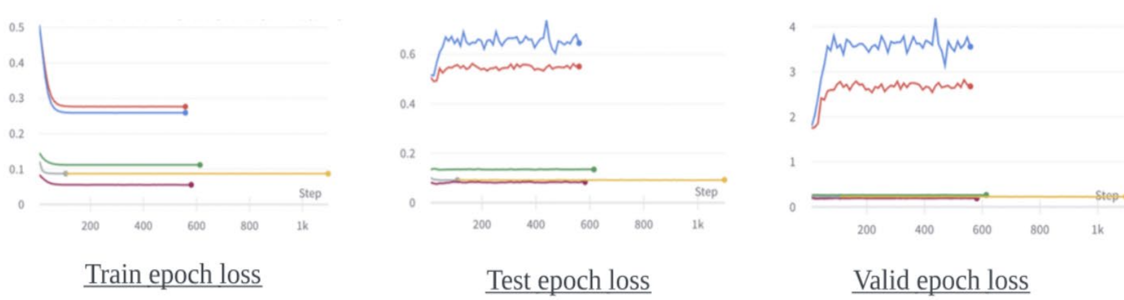

Figure 7. Training data downloaded from Wandb for comparative purpose

Specifically, AdapFAN outperforms LSTM, S2S by 13.4% and 26.6% and achieves the lowest mse at 0.474599 and lowest mae at 0.1403, showing strong potential for operational usage. More experiments have been conducted on AdapFAN, Dlinear, Autoformer, and FEDformer for comparative references(shown in figure 7).

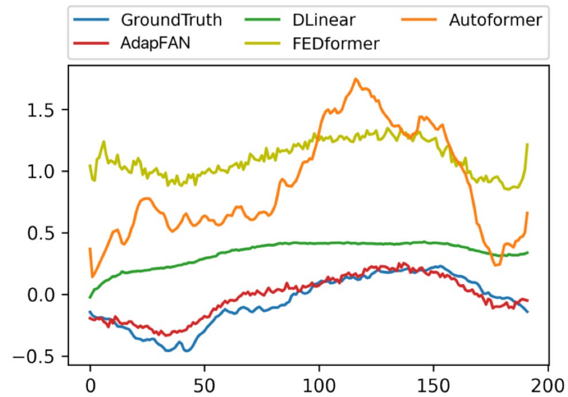

Figure 8. Visual comparison between Dlinear, Autoformer, FEDformer and AdapFAN (input-96, predict-96)

AdapFAN achieves the closest simulation relative to the ground truth curve in the experiments when compared to the recently published and well-acknowledged transformer-based or linear models. The prediction length is 96 which is equivalent to 4 days lead time. Three more visual comparisons between AdapFAN, linear, and vanilla transformer are shown in figure 9 and 10.

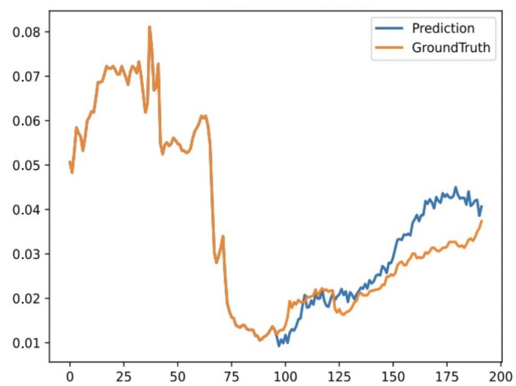

Figure 9. Visualization of prediction made by AdapFAN on WaterBench dataset (input-96, predict-96)

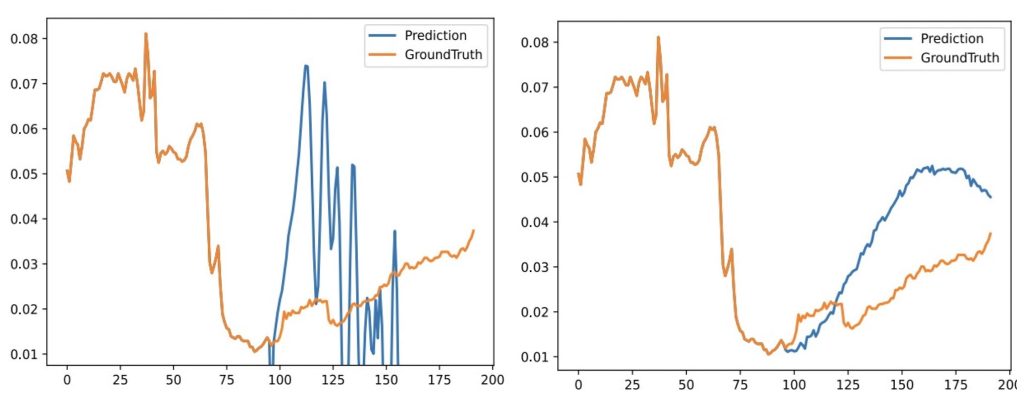

Figure 10. Visualization of predictions made by vanilla transformer and Dlinear model on the same dataset (input-96, predict-96)

Results

Compared to both data-driven methods using linear, MLP methods, and LSTM models, and process-based methods using vanilla transformer and its recent variants designed specifically for time series forecasting, AdapFAN performed admirably (Figure7-10). Mse and mae are comparatively lower on average. In terms of long-term forecasting, AdapFAN achieves four days lead time while with improved accuracy compared to aforementioned models (Table 8-10).

Conclusion and Evaluation

As shown in the experiment section, AdapFAN outperforms LSTM, S2S by 13.4% and 26.6% and achieves the lowest mse at 0.474599 and lowest mae at 0.1403, showing strong potential for operational usage. In addition, AdapFAN achieves the closest simulation relative to the ground truth curve in the experiments when compared to the recently published and well-acknowledged transformer-based or linear models with the prediction length of 96 which is equivalent to 4 days lead time. This lead time is 4 days longer than the lead time achieved in nowcast from a current state-of-the-art global modeling system.

For future direction of improvement, I would consider, if possible, putting the proposed AdapFAN into practice in South China or India. The model can send out warning signals whenever the predicted value of rate of run-off volume exceeds 1000 m^3/s. The ultimate value of the real-world-scenario-based model is determined by its practical use. I believe AI technology, in addition to technical evolution, should be utilized in practice for global good.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

1. Demir, Ibrahim, and Zhongrun Xiang. "WaterBench-Iowa: A Large-scale Benchmark Dataset for Data-driven Streamflow Forecasting." *Earth System Science Data*, 2022, essd.copernicus.org/articles/14/5605/2022/.
2. "Early Warning Systems Saves Millions of Lives." *World Meteorological Organization*, library.wmo.int/viewer/53697/download?file=MHEWS.pdf&type=pdf&navigator=1.
3. Ellery, Marni. "'Nowcasting' Model Enables Forecasting of Extreme Precipitation." *Barkeley Engineering*, 12 July 2023, engineering.berkeley.edu/news/2023/07/nowcasting-model-enables-forecasting-of-extreme-precipitation/#:~:text=With%20climate%20change%20driving%20more,of%20flash%20floods%20and%20landslides.
4. J, Seibert. "Regionalisation of Parameters for a Conceptual Rainfall-runoff Mode." *Agricultural and Forest Meteorology*, 1999.
5. Kitaev, Nikita, and Anselm Levskaya. "Reformer: The Efficient Transformer." *Arxiv*, 18 Feb. 2020, arxiv.org/pdf/2001.04451.pdf.

6. Kose, Neslihan, and Ahmet Gunduz. "Resource Efficient 3D Convolutional Neural Network." *Arxiv*, 18 Oct. 2021, arxiv.org/pdf/1904.02422.pdf.
7. Kratzert, Frederik, and Daniel Klotz. "Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine-Learning Applied to Large-Sample Datasets." *Arxiv*, 11 Nov. 2019, arxiv.org/pdf/1907.08456.pdf.
8. Kratzert, Frederik, et al. "Towards Improved Predictions in Ungauged Basins: 2 Exploiting the Power of Machine Learning." *Arxiv*, file:///Users/vanessaw/Downloads/wrr.pdf.
9. L, Samaniego, et al. "Multiscale Parameter Regionalization of a Grid-based Hydrologic Model at the Mesoscale." *Water Resources Research*, 2010.
10. "A Large-scale Benchmark Dataset for Data-Driven Streamflow Forecasting." *A Large-scale Benchmark Dataset for Data-Driven Streamflow Forecasting*, 2018, github.com/uihilab/WaterBench. Accessed 10 Nov. 2023.
11. Liu, Yong, et al. "ITransformer." *Arxiv*, 10 Oct. 2023, arxiv.org/pdf/2310.06625.pdf.
12. Nie, Yuqi, and Nam H. Nguyen. "A Time Series Is worth 64 Words." *Arxiv*, 5 Mar. 2023, arxiv.org/pdf/2211.14730.pdf.
13. Perera, Duminda, et al. "Flood Early Warning Systems: A Review of Benefits, Challenges and Prospects." *UNU-INWEH*, 2019, inweh.unu.edu/wp-content/uploads/2019/08/Flood-Early-Warning-Systems-A-Review-Of-Benefits-Challenges-And-Prospects.pdf.
14. Ravuri, Suman, et al. "Skillful Precipitation Nowcasting Using Deep Generative Models of Radar." *Arxiv*, Apr. 2021, Skillful Precipitation Nowcasting using Deep Generative Models of Radar.
15. Sorkhabi, Omid Memarian, et al. "Deep Learning of Sea-level Variability and Flood for Coastal City Resilience." *Science Direct*, Jan. 2023, www.sciencedirect.com/science/article/pii/S2590252022000204.
16. Soydaner, Derya. "Attention Is All You Need." *Arxiv*, 27 Apr. 2022, arxiv.org/pdf/2204.13154.pdf.
17. "TimesNet Code Source." *Science in Context*, Apr. 2023, github.com/thuml/TimesNet.
18. Wu, Haixu, et al. "TimesNet." *Arxiv*, 12 Apr. 2023, arxiv.org/pdf/2210.02186.pdf.
19. Yamashita, Rikiya. "Convolutional Neural Networks." *SpringerOpen*, 22 June 2018, insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9#:~:text=The%20CNN%20architecture%20includes%20several,or%20more%20fully%20connected%20layers.
20. Zeng, Ailing, et al. "Are Transformers Effective for Time Series Forecasting?" *Arxiv*, 2022.
21. Zhao, Shan, et al. "Exploring Geometric Deep Learning For Precipitation Nowcasting." *Arxiv*, 2022, arxiv.org/pdf/2309.05828.pdf.
22. Zheng, Yi, et al. *In International Conference on web-age information management*. 2014.