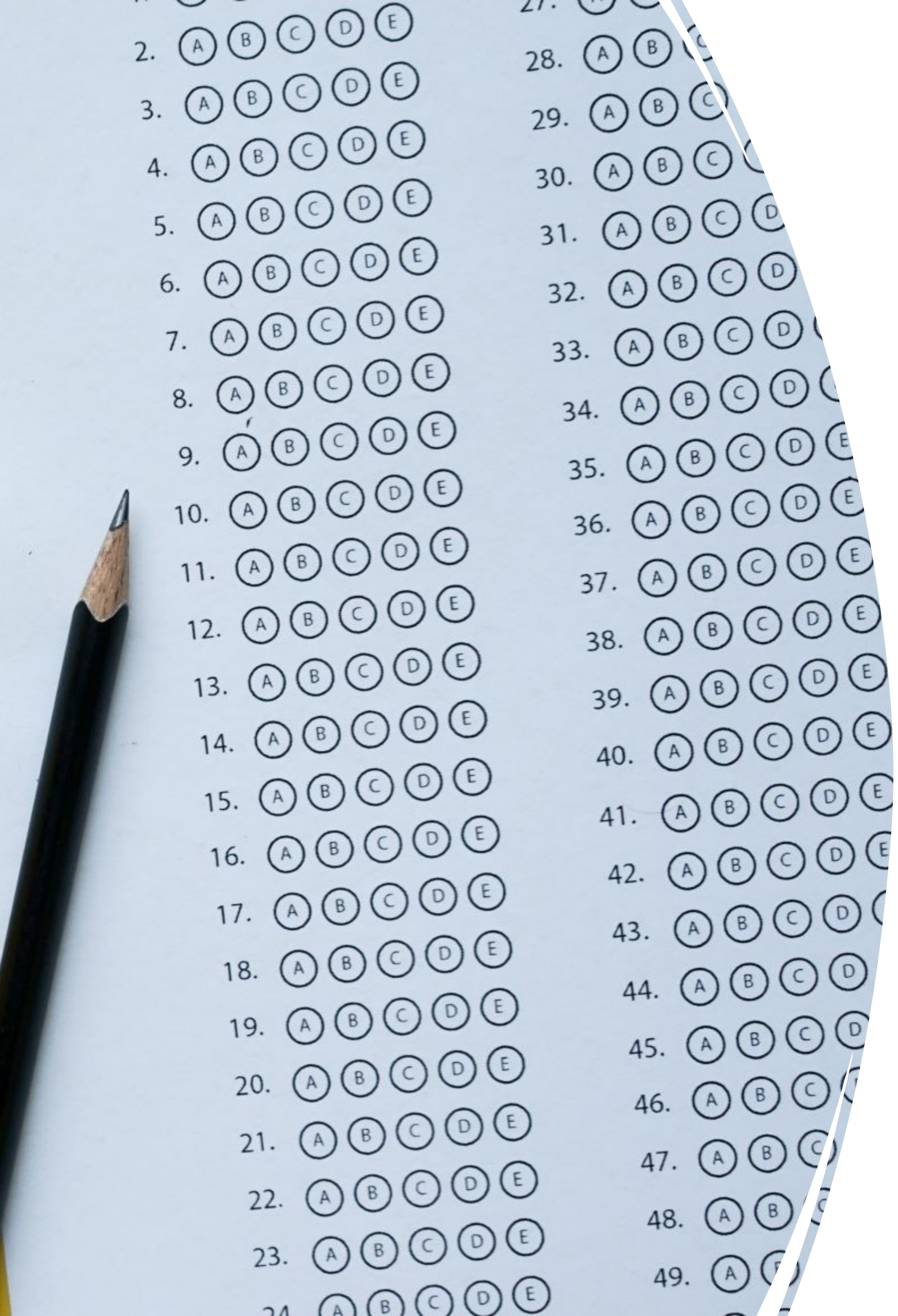
Week 4

Wentao Gao



Content

- ***Paper***
- ***Method***
- ***Data***
- ***Test result***

Paper reading – Drought Prediction

Drought Prediction

Evaluation of Time Series Models in Simulating Different Monthly Scales of Drought Index for Improving Their Forecast Accuracy

Estimation of SPEI Meteorological Drought using Machine Learning Algorithms

Prediction of Impending Drought Scenarios Based on Surface and Subsurface Parameters in a Selected Region of Tropical Queensland, Australia

Monthly drought prediction based on ensemble models

A novel intelligent deep learning predictive model for meteorological drought forecasting

Paper reading – Methods in Drought prediction

Drought modeling – A review **

A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions *

Explainable AI in drought forecasting ***

Big Earth data: disruptive changes in Earth observation data management and analysis?

Drought prediction using a wavelet based approach to model the temporal consequences of different types of droughts Author links open overlay panel **

A Contemporary Review on Drought Modeling Using Machine Learning Approaches (researchgate.net)

A Contemporary Review on Deep Learning Models for Drought Prediction *

Drought forecasting through statistical models using standardised precipitation index: a systematic review and meta-regression analysis *

Characterising the seasonal nature of meteorological drought onset and termination across Australia *

Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting *

Improved Transformer Model for Enhanced Monthly Streamflow Predictions of the Yangtze River *

I believe that some advanced time series forecasting method will be a very helpful for my research.

Prediction in Time series using Transformer

Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecastin

Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting

PYRAFORMER: LOW-COMPLEXITY PYRAMIDAL ATTENTION FOR LONG-RANGE TIME SERIES MODELING AND FORECASTING

FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting (mlr.press)

Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting

LightTS: Less Is More: Fast Multivariate Time Series Forecasting with Light Sampling-oriented MLP Structures (arxiv.org)

ETSformer: Exponential Smoothing Transformers for Time-series Forecasting

DLinear: Are Transformers Effective for Time Series Forecasting?

Long-term Forecasting with TiDE: Time-series Dense Encoder

TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis

Fourier-Mixed Window Attention: Accelerating Informer for Long Sequence Time-Series Forecasting

Drought

- Drought may be classified as follows:
  - i) meteorological drought: usually illustrated by precipitation anomaly;
  - ii) agricultural drought: soil moisture depletion resulting from insufficient available water; or
  - iii) hydrological drought: surface and subsurface water shortage or deviation from normal conditions over a long period of time (Dalezios, and Spyropoulosand, 2017).
- We gonna focus on the **meteorological drought prediction.**

Standardized Precipitation Evapotranspiration Index (SPEI)

- The **calculation method** of SPEI is as follows:

1. First, the evaporative precipitation difference (i.e., precipitation minus evaporation) is calculated for each time period (e.g., monthly or yearly).
2. The evapotranspiration-precipitation difference is then converted to a probability distribution, usually using the Gamma distribution.
3. Finally, the probability distribution is converted to a standard normal distribution to obtain the SPEI value.

Long-term Forecasting:

- The aim of long-term forecasting is to predict the **values of a time series over a longer future period**. For this type of problem, we usually use a segment of **historical data as input**, and the output is the predicted time series values for a longer future period (e.g., the next year).

Problem Defining

The time series forecasting problem can be formulated as:

Predicting future M steps(day or month) precipitation $Z_{t+1:t+M}$. Given the previous L steps of observations $Z_{t-L:t}$ and the covariates(air temperature …) $X_{t-L+1:t}$.

Input: $Z_{t-L:t}, X_{t-L+1:t}$

Output: $Z_{t+1:t+M}$

Methods:

- ***Statistical method: ARIMA***
- ***RNN baded: RNN, LSTM***
- ***Transformer based:***
  - Transformer
  - Transformer-LSTM
  - Sparse Trans: LogTrans, Reformer, Informer
  - Decomposition Trans: Autoformer
- ***Linear_based: DLinear***
- ***<u>TimesNet</u>***

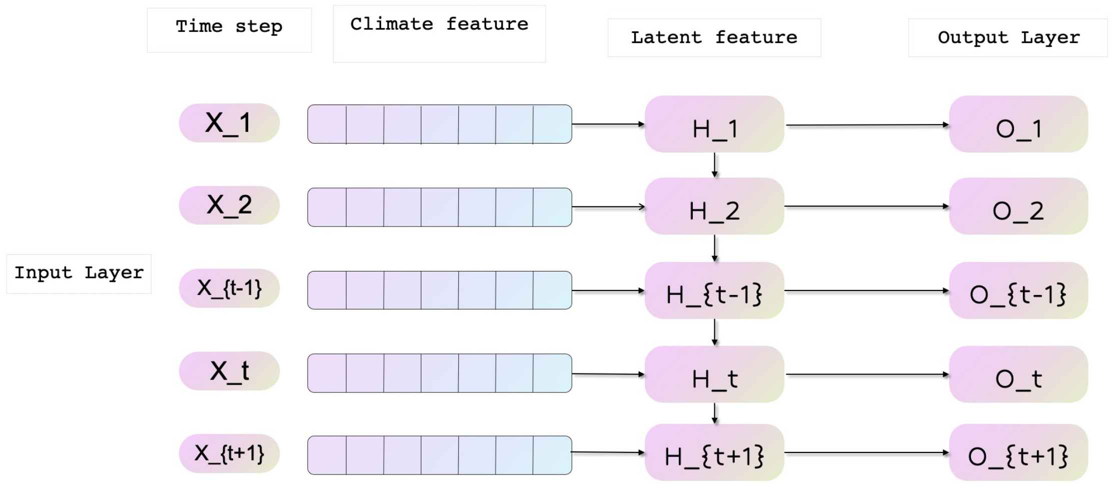
ARIMA

The core idea of the ARIMA model is **to transform the time series data into a stationary series** and then **capture the correlation and trend** in the series through a combination of autoregressive (AR) and moving average (MA) components.

$Y'(t) = c + \varphi_1 Y'(t-1) + \varphi_2 Y'(t-2) + ... + \varphi_p Y'(t-p) + \theta_1 \varepsilon(t-1) + \theta_2 \varepsilon(t-2) + ... + \theta\_q \varepsilon(t-q) + \varepsilon(t)$

- Autoregressive (AR)： $Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \varepsilon_t$
- Moving Average (MA)： $Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_\varphi \varepsilon_{t-\varphi}$
- Differencing (Integrated)：differencing the time series d times to make it stationary. Y'(t) is the time series after performing dth order differencing and ε(t) is the white noise error term

RNN Architecture



$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$$

$$y_t = W_{hy}h_t + b_y$$

LSTM Architecture

Forget gate: $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$

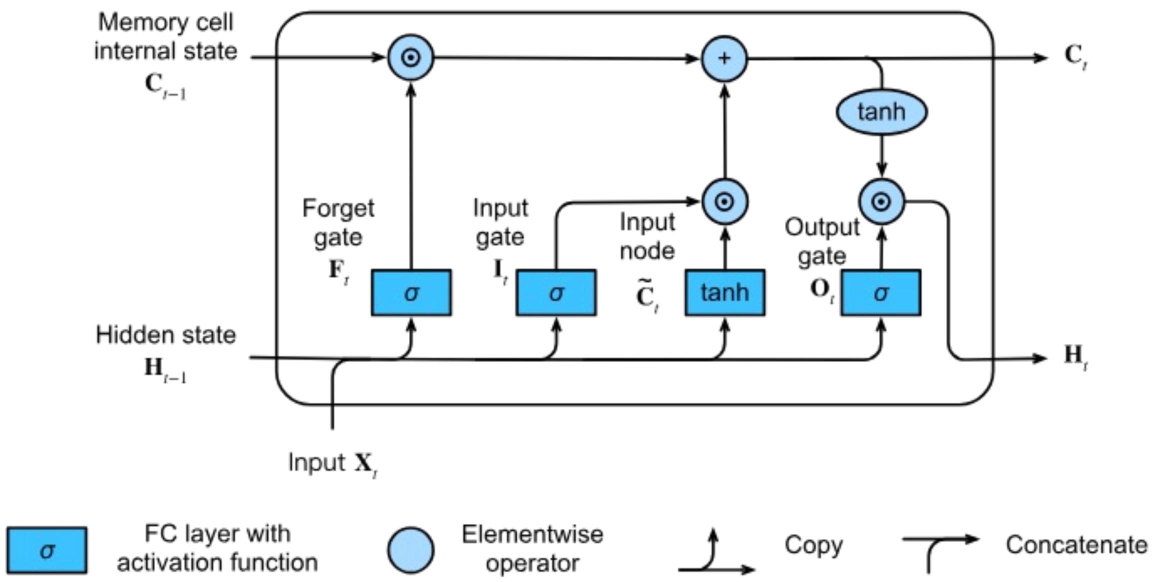
Input gate: $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$

Candidate cell state: $\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$

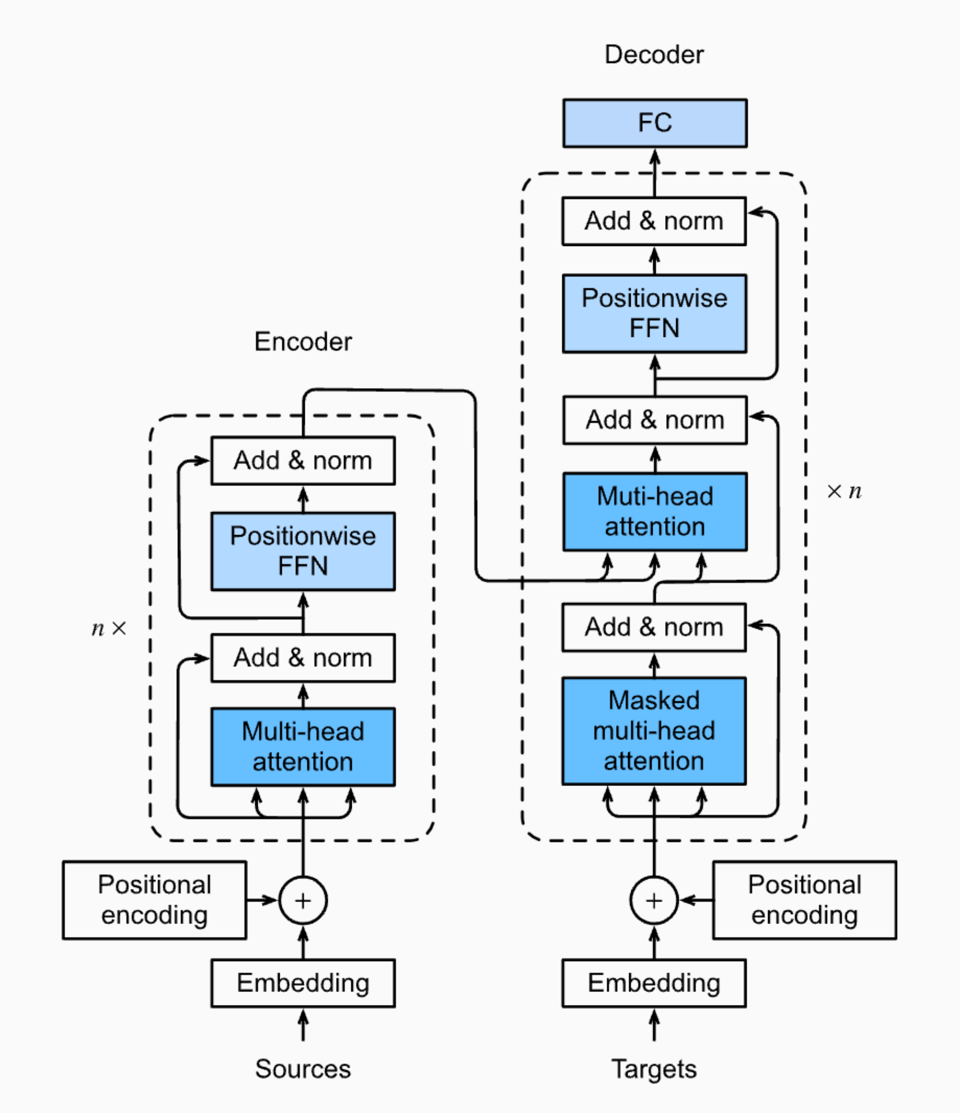
Update cell state: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$

Output gate: $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$

Hidden state: $h_t = o_t \odot \tanh(C_t)$



Transformer Architecture



- Scaled Dot-Product Attention:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

- Multi-Head Attention:

$$MultiHead(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O$$

- Position-wise Feed-Forward Networks:

$$\text{head}_i = \text{Attention}(QW_{Qi}, KW_{Ki}, VW_{Vi})$$

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

- Positional Encoding:

$$PE_{(\text{pos},2i)} = \sin\left(\frac{\text{pos}}{10000^{2i/d_{\text{model}}}}\right)$$

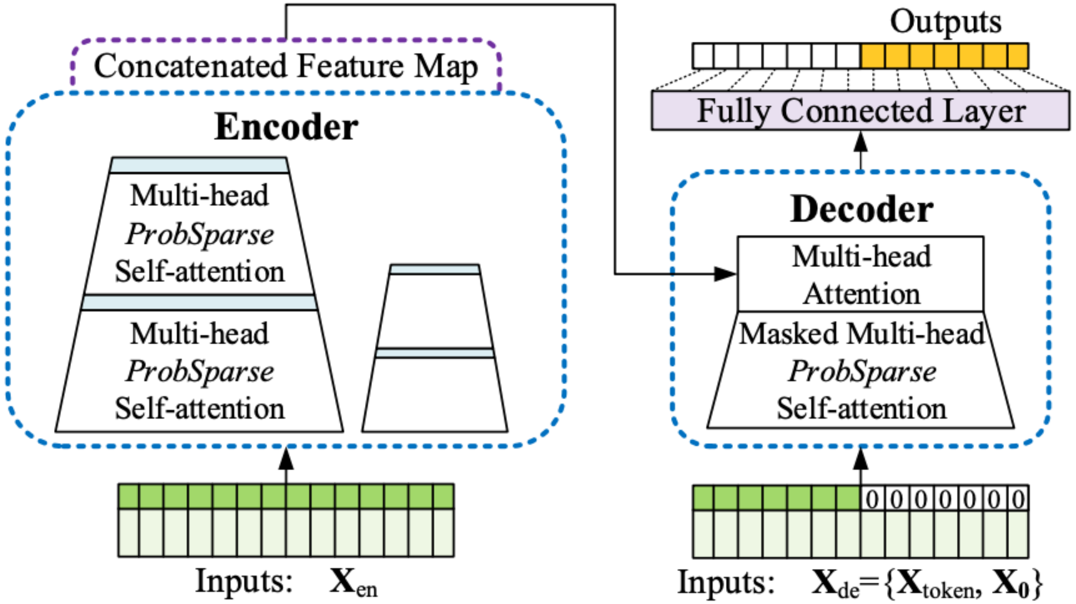
$$PE_{(\text{pos},2i+1)} = \cos\left(\frac{\text{pos}}{10000^{2i/d_{\text{model}}}}\right)$$

- Layer Normalization:

$$y = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

Informer

- ***ProbSparse self-attention,***
- ***Self-attention Distilling （KL divergence）***
- ***Generative-discriminative training***



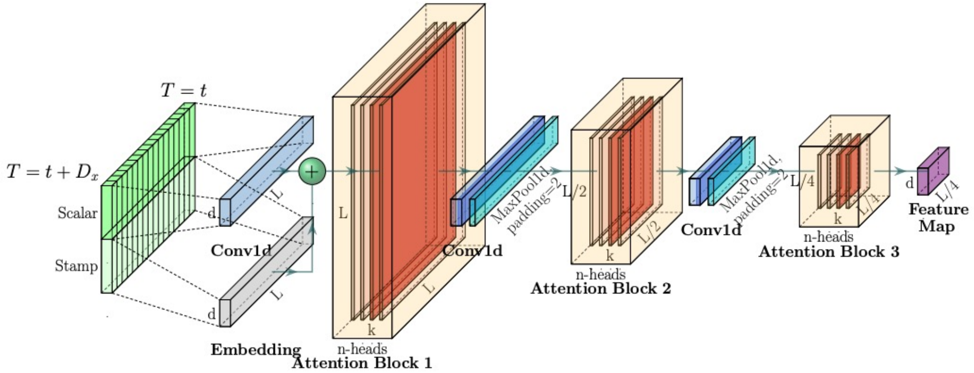
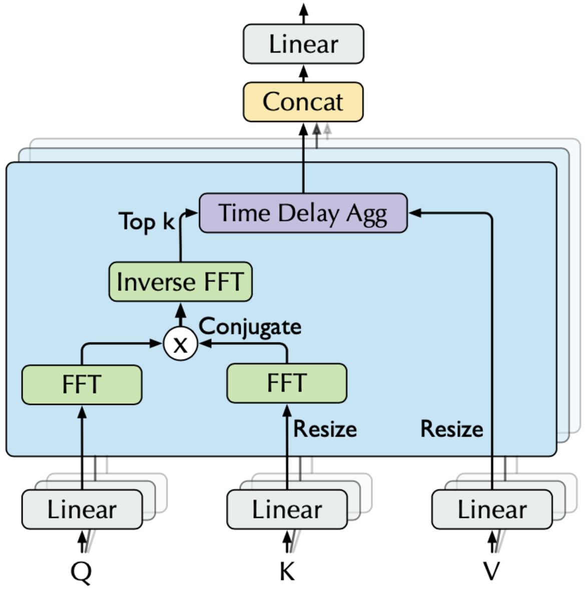
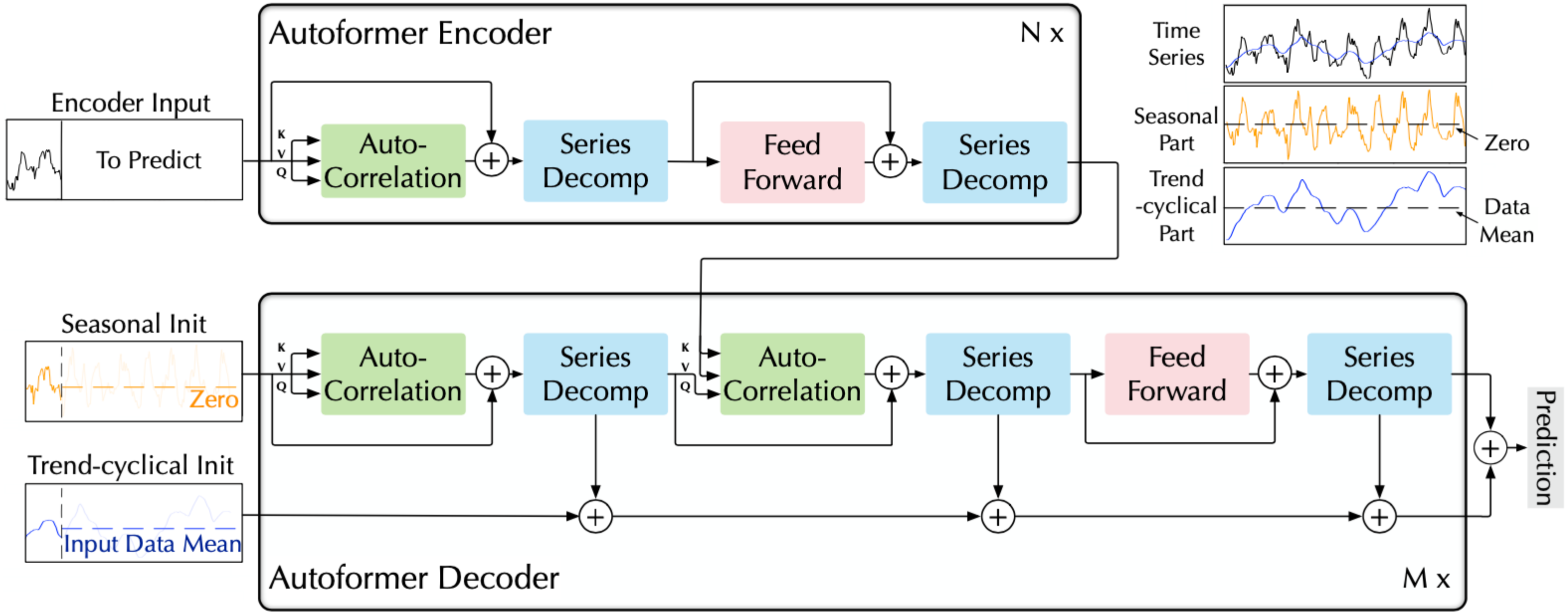


Figure 3: The single stack in Informer's encoder. (1) The horizontal stack stands for an individual one of the encoder replicas in Fig.(2). (2) The presented one is the main stack receiving the whole input sequence. Then the second stack takes half slices of the input, and the subsequent stacks repeat. (3) The red layers are dot-product matrixes, and they get cascade decrease by applying self-attention distilling on each layer. (4) Concatenate all stacks' feature maps as the encoder's output.

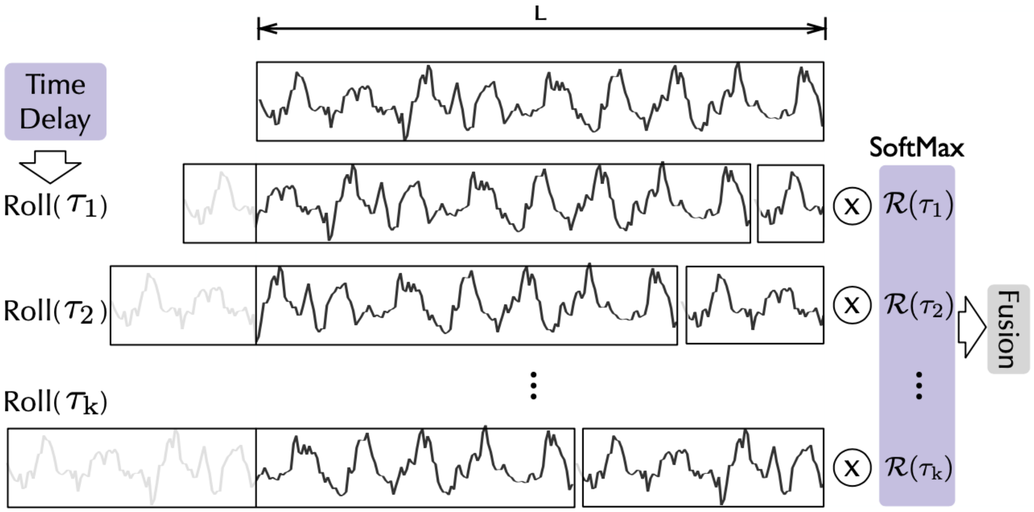




AutoFormer

Decomposition architecture

Auto-Correlation mechanism



TimesNet

- ***Transforms one-dimensional time series into two-dimensional space for analysis***
- ***Task-general temporal basis model - TimesNet***

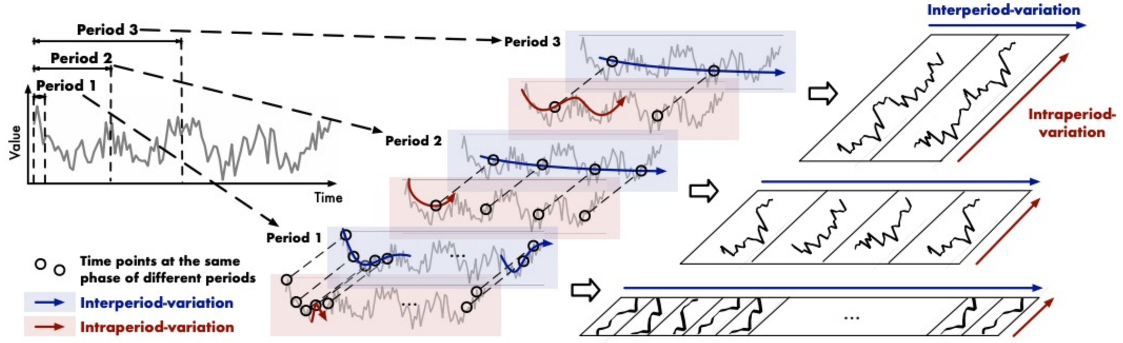
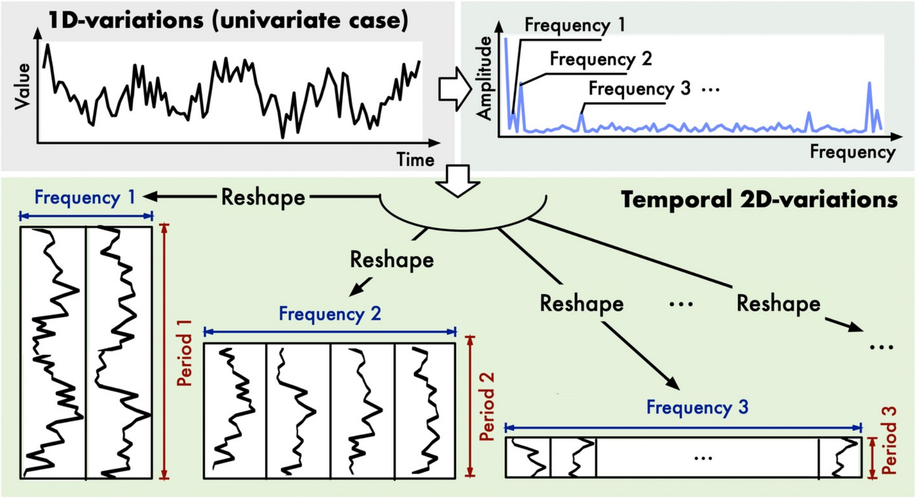
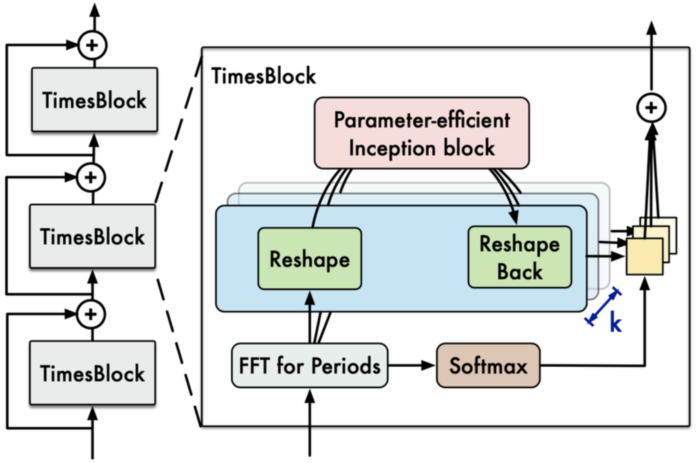
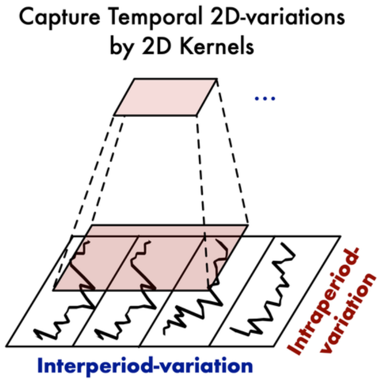
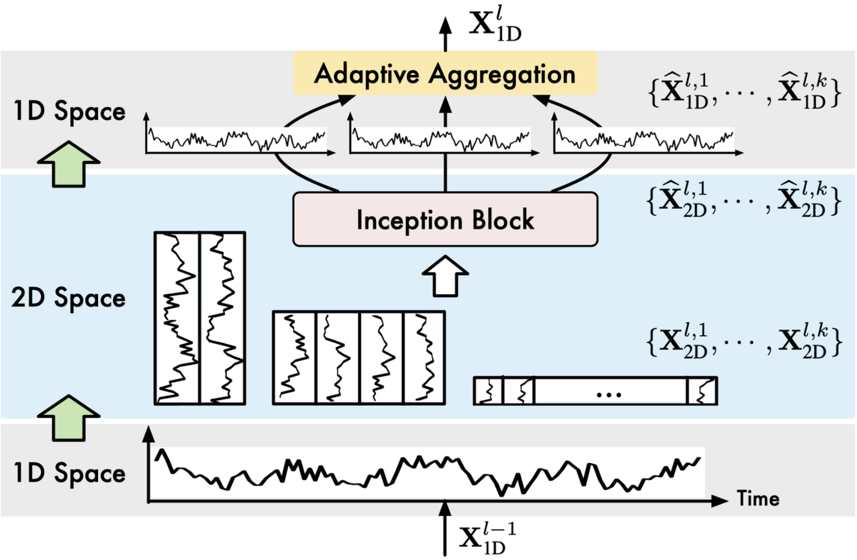


Figure 1: Multi-periodicity and temporal 2D-variation of time series. Each period involves the intraperiod-variation and interperiod-variation. We transform the original 1D time series into a set of 2D tensors based on multiple periods, which can unify the intraperiod- and interperiod-variations.

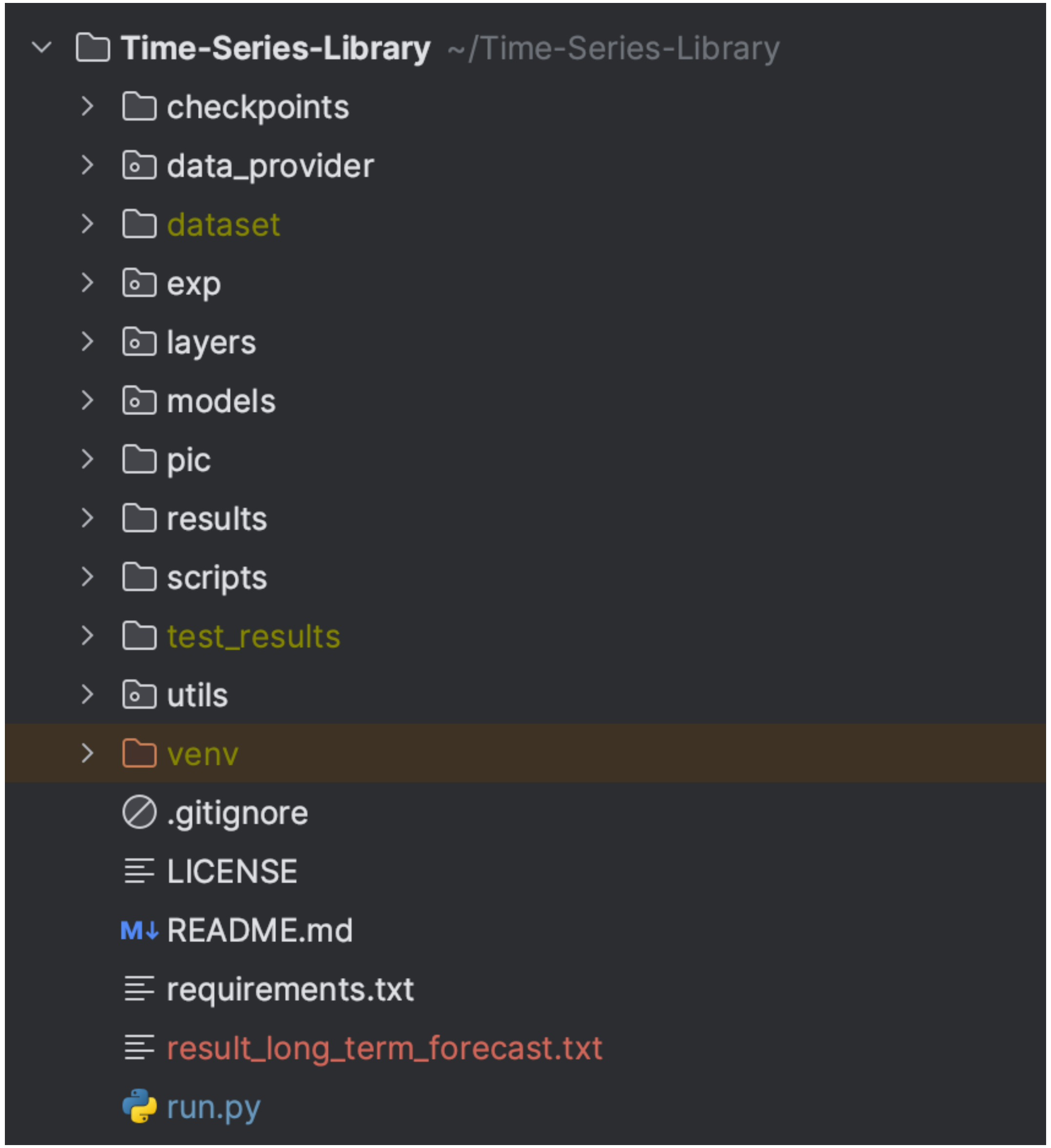








Time series library



Code Package

```
- TSGao
    - data
    - TSdeepl
        - RNN
        - LSTM
        - Transformer_full
        - Transformer_simple
        - Transformer_encoder
        - Transformer_lstm
    - TStrad
        - AR
        - MA
        - ARMA
        - ARIMA
        - SARIMA
    - testdata # Time series analysis including stationary
```

DATA

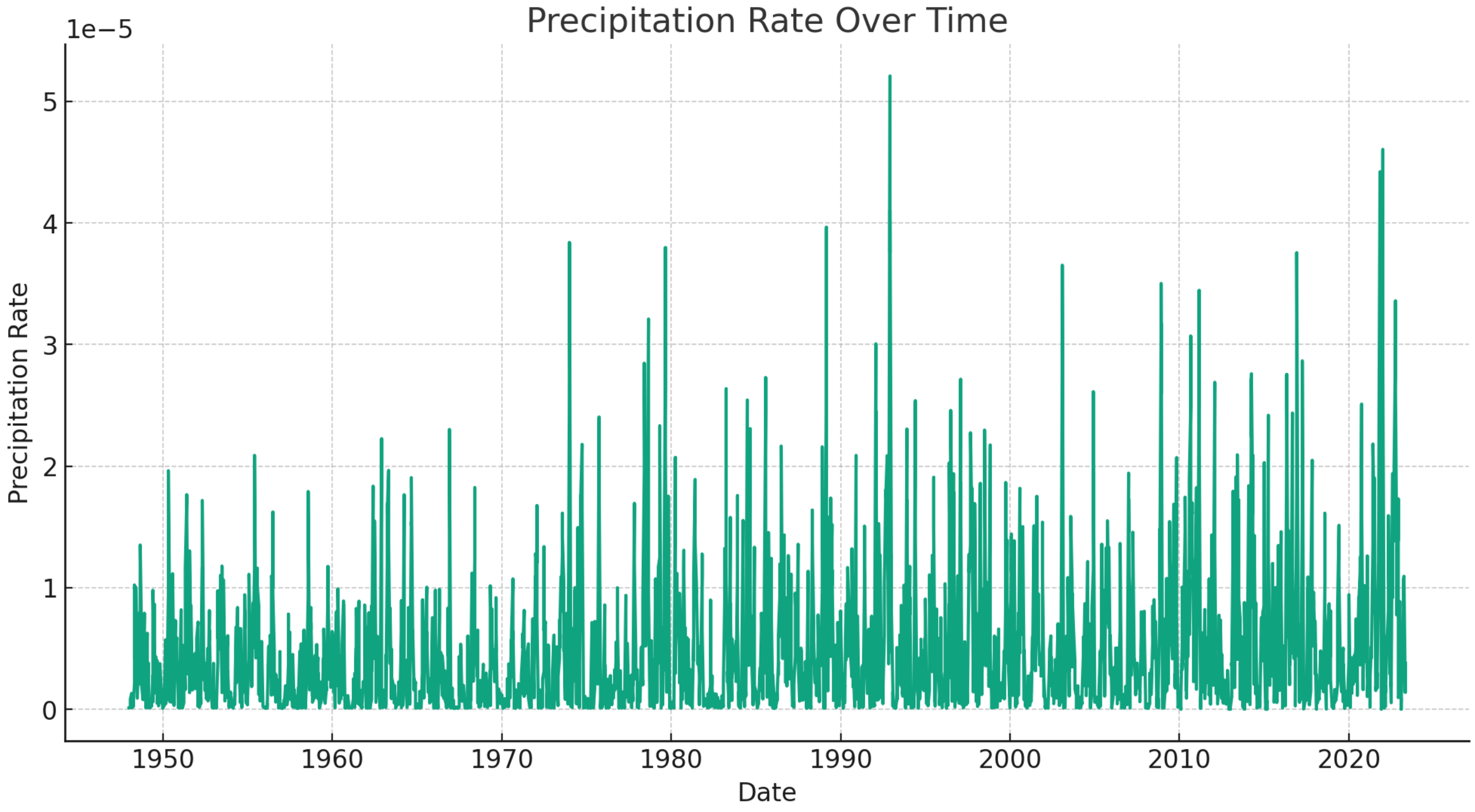
- NCEP-NCAR Reanalysis 1([NCEP-NCAR Reanalysis 1: NOAA Physical Sciences Laboratory NCEP-NCAR Reanalysis 1]())
- Reform the NetCDF4 data into CSV format.
- Choose South Australia, Monthly data as our start up.

date	lat	lon	prate	pres	air	skt	dswrf	csusf	csdlf	cfnlf	wspd	tmin	season
1948-01-01	-33.332801818847700	135.0	1.1278767E-07	100720.0	295.78036	23.49548	366.30005	92.0032	322.62268	13.258009	4.1900024	288.16	1
1948-01-01	-33.332801818847700	136.875	9.9884346E-08	99760.0	296.8249	24.216444	369.09985	99.44191	322.01938	12.274146	3.3000183	287.24	1
1948-01-01	-33.332801818847700	138.75	9.9884346E-08	99170.0	297.76096	24.798058	373.59985	110.18707	323.3807	11.8515625	2.3900146	287.19	1
1948-02-01	-33.332801818847700	138.75	1.4816058E-07	98870.0	296.86453	23.872753	327.0	99.3586	328.62762	14.586173	2.25	287.03998	1
1948-02-01	-33.332801818847700	136.875	1.6195378E-07	99480.0	295.3697	22.73862	337.0	89.73791	323.1035	11.810294	3.0899963	286.33002	1
1948-02-01	-33.332801818847700	135.0	8.240278E-07	100470.0	293.9055	21.373447	345.3999	83.06895	319.55524	9.73445	3.8399963	286.76	1
1948-03-01	-33.332801818847700	135.0	1.3127893E-06	101040.0	291.67615	19.018711	275.69995	70.974144	310.41617	12.56769	3.4300232	285.09	2
1948-03-01	-33.332801818847700	136.875	1.1923933E-07	100070.0	291.60712	18.794512	276.3999	77.2161	307.69684	13.119298	2.9200134	283.66998	2
1948-03-01	-33.332801818847700	138.75	1.1278767E-07	99500.0	291.9874	18.94871	280.69995	85.8161	307.4517	12.470908	2.1900024	283.53998	2
1948-04-01	-33.332801818847700	138.75	1.8655167E-07	99710.0	288.54367	15.411002	211.80005	69.95993	291.5634	14.8266115	2.100006	281.43	2
1948-04-01	-33.332801818847700	136.875	8.365527E-07	100280.0	288.94	15.970336	206.80005	63.189926	293.9667	15.999943	2.880005	282.65	2
1948-04-01	-33.332801818847700	135.0	7.9322297E-07	101220.0	289.70132	16.738667	201.59985	58.279922	298.29672	17.009953	3.3600159	284.31	2
1948-05-01	-33.332801818847700	138.75	2.109566E-06	99730.0	283.47775	10.249034	150.30005	56.754765	276.27744	17.91285	1.8200073	276.95	2
1948-05-01	-33.332801818847700	135.0	1.0206349E-05	101300.0	285.7474	12.424517	142.3999	48.34187	286.92905	21.832197	2.5	280.53998	2
1948-05-01	-33.332801818847700	136.875	5.064412E-06	100320.0	284.30002	10.974517	146.0	51.625732	280.37747	19.745102	2.3200073	278.13	2
1948-06-01	-33.332801818847700	138.75	3.1732257E-06	99920.0	281.35565	8.068333	116.09985	50.50993	265.48	29.839941	1.5400085	275.46	3
1948-06-01	-33.332801818847700	136.875	5.5165574E-06	100450.0	282.68835	9.265003	111.69995	46.37327	270.46335	31.94327	2.350006	277.8	3
1948-06-01	-33.332801818847700	135.0	1.0029894E-05	101370.0	284.64667	11.277002	106.69995	43.75659	277.78668	34.35661	3.0299988	281.18	3
1948-07-01	-33.332801818847700	136.875	1.6321503E-06	100230.0	283.5755	10.286451	125.8999	48.32251	271.56134	25.7322	3.3700256	278.72998	3

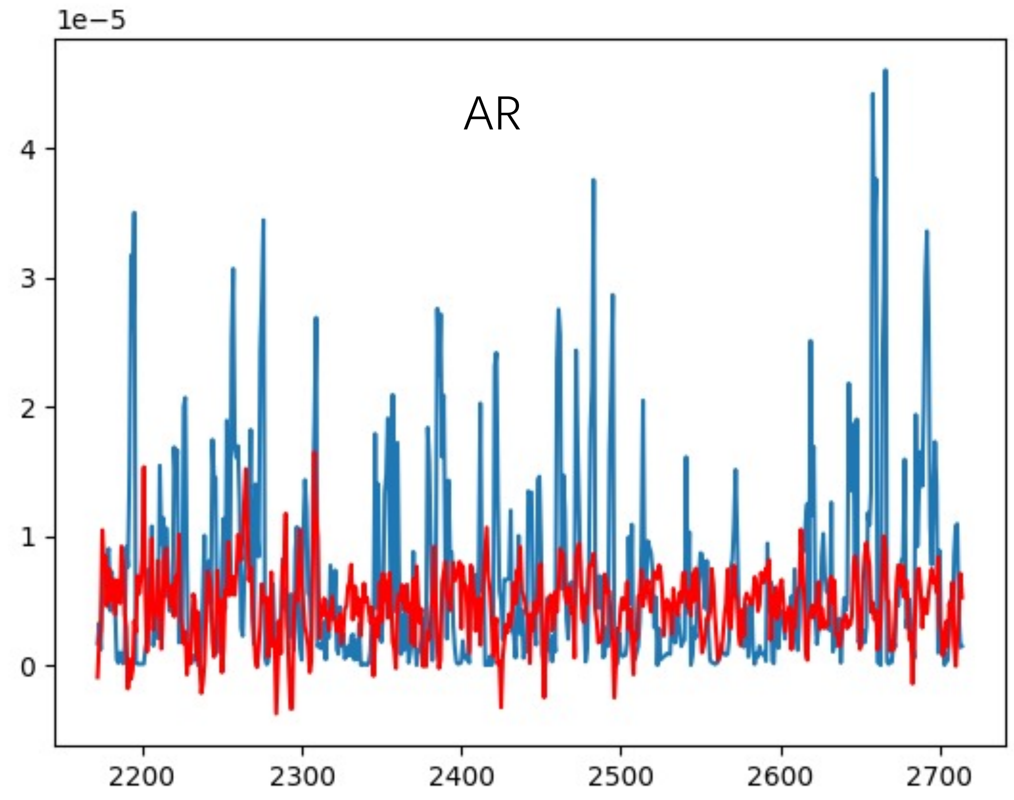
Description of variable

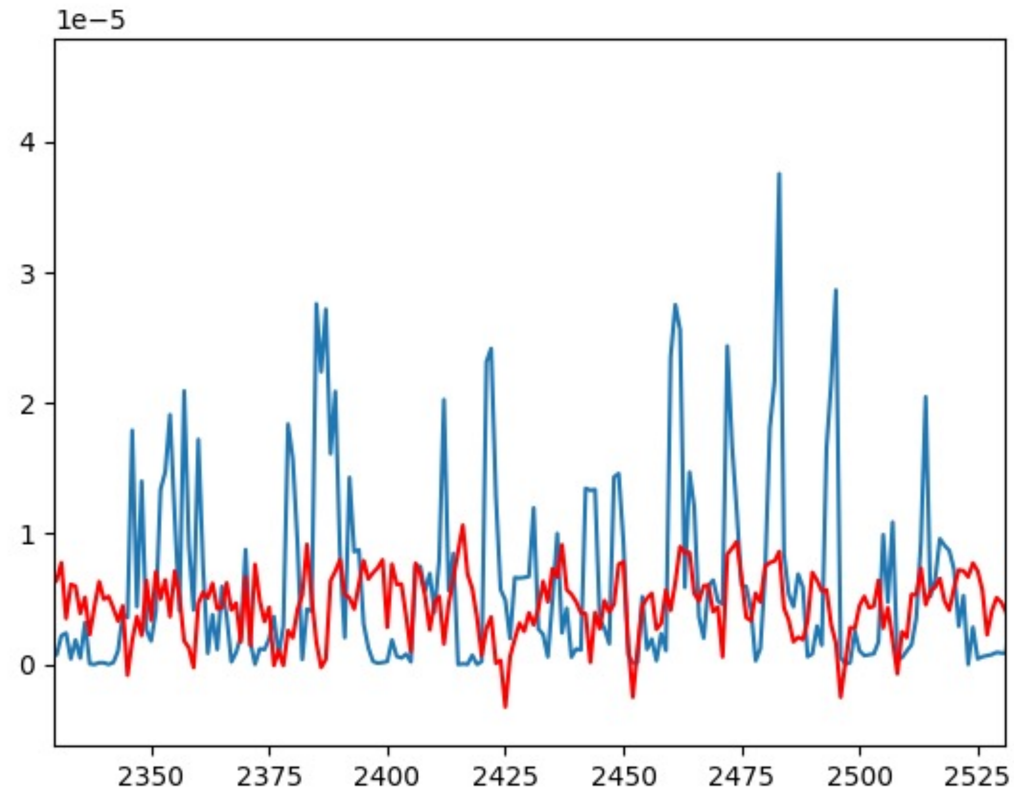
- prate: Precipitation rate (mm/s). This is the amount of rainfall expressed in millimeters per second.
- pres: Pressure level (Pa). This is atmospheric pressure, usually given in Pascals.
- air: Air temperature (K). This is the temperature of the air, given in Kelvin.
- skt: Skin temperature (K). This is the temperature at the surface of the Earth, given in Kelvin.
- dswrf: Downward short-wave radiation flux (W/m^2). This is the amount of solar energy that is directly reaching the Earth's surface, expressed in Watts per square meter.
- csusf: Clear-sky upward solar flux (W/m^2). This is the amount of solar radiation energy that is reflected back into the atmosphere under clear-sky conditions, expressed in Watts per square meter.
- csdlf: Clear-sky downward longwave flux (W/m^2). This is the amount of energy that radiates down to the ground from the atmosphere under clear-sky conditions, expressed in Watts per square meter.
- cfnlf: Cloud forcing net longwave flux (W/m^2). This represents the effect of clouds on the radiation balance, expressed in Watts per square meter.
- wspd: Wind speed (m/s). This is the speed of the wind, given in meters per second.
- tmin: Minimum temperature (K). This is the minimum temperature, given in Kelvin.
- season: Season. This variable represents the four seasons (spring, summer, fall, winter) and is likely a categorical variable rather than a numerical one.

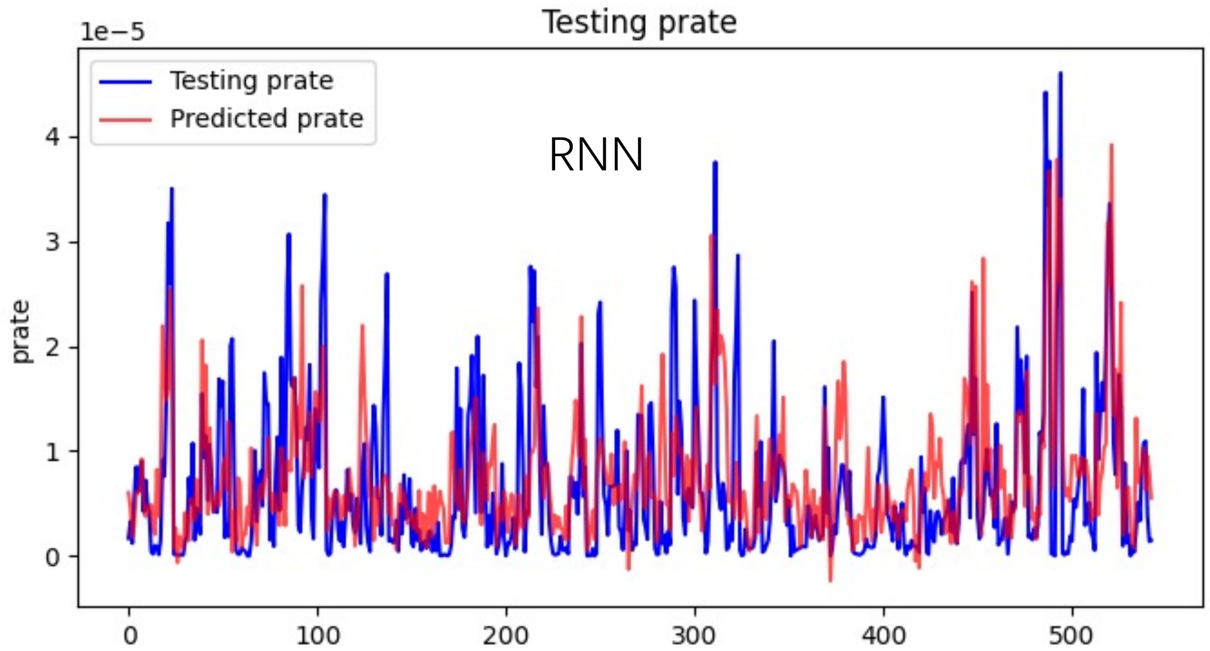
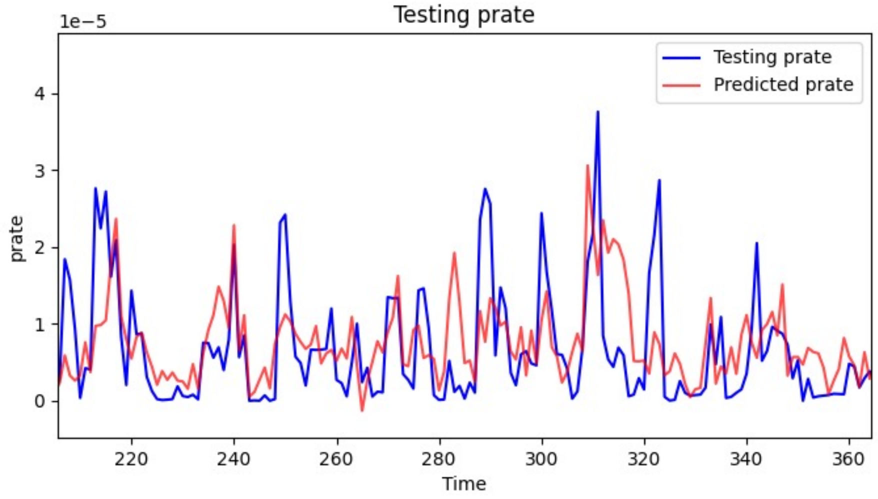
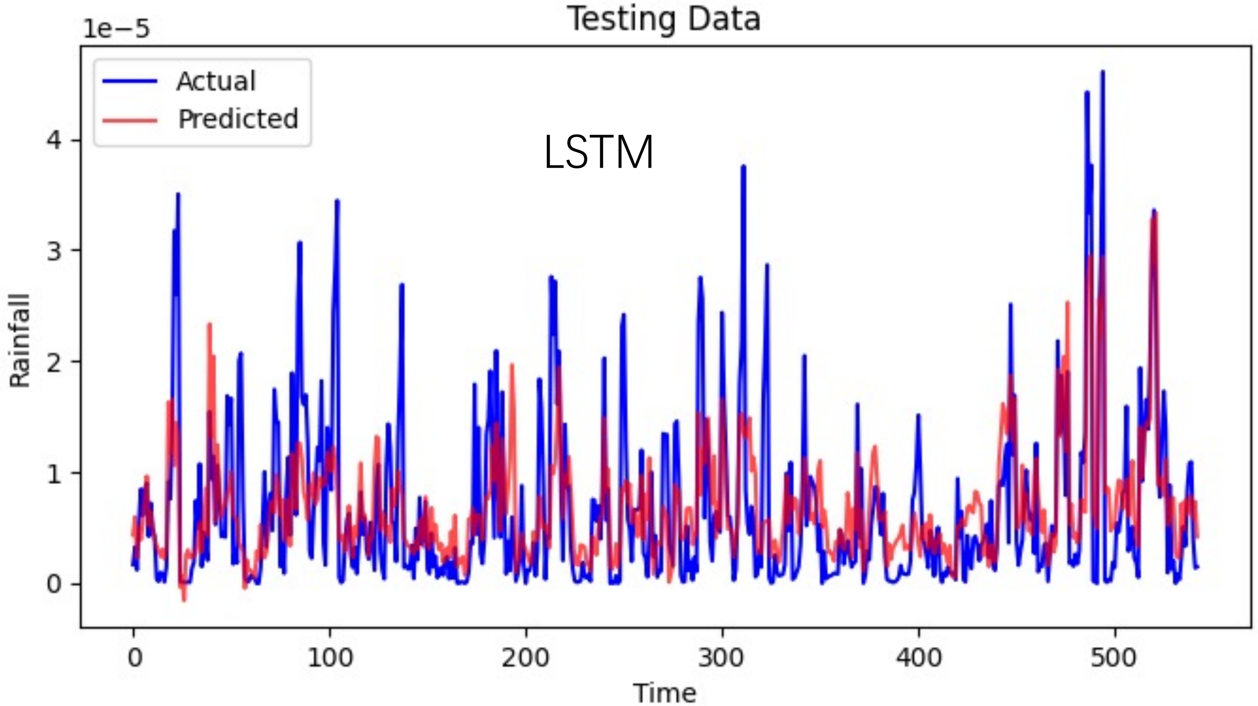
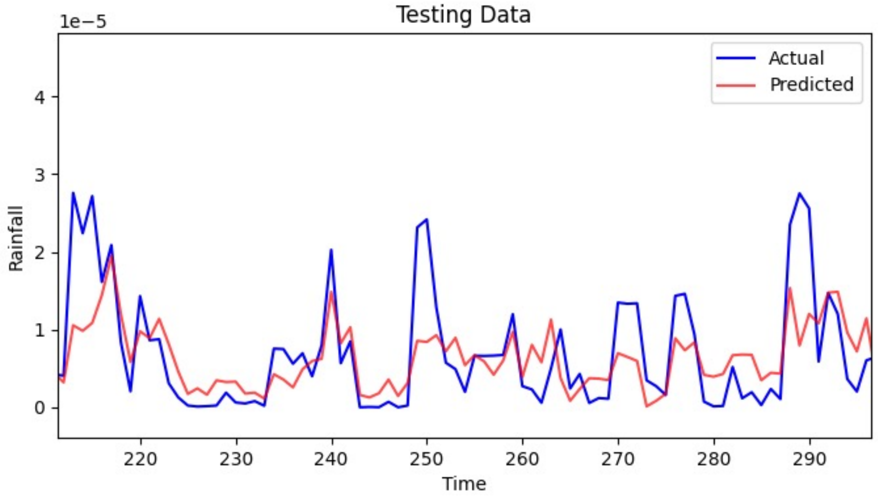
Precipitation rate over time

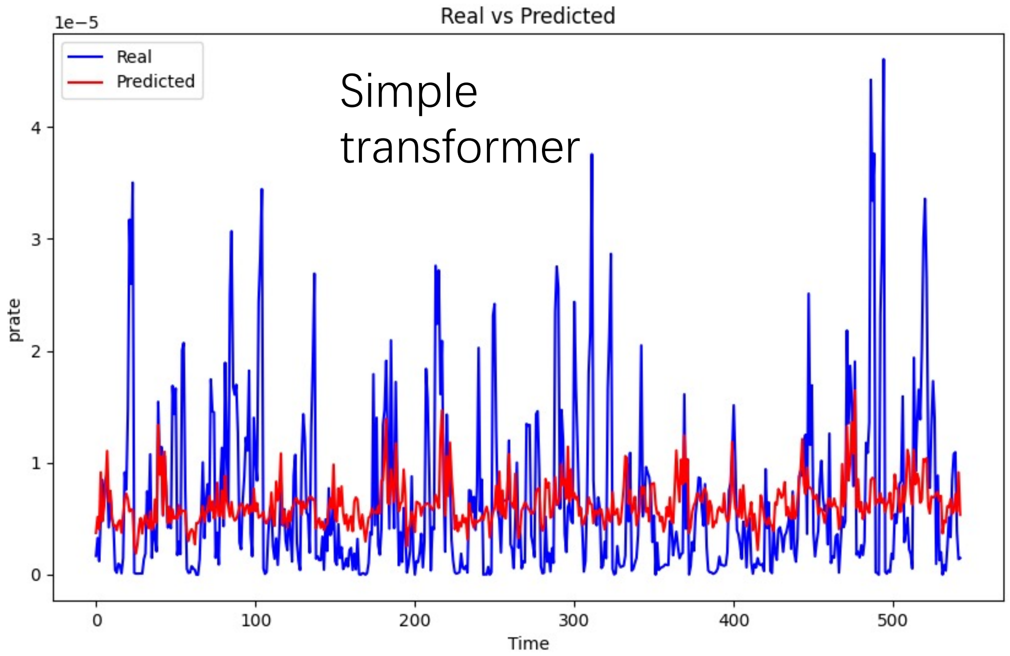


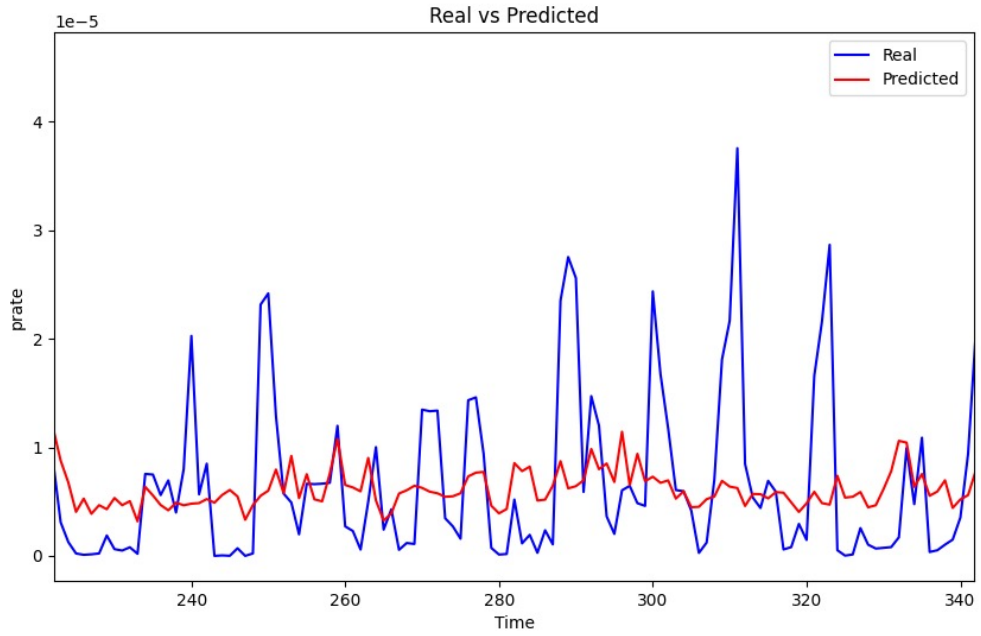
Test Results

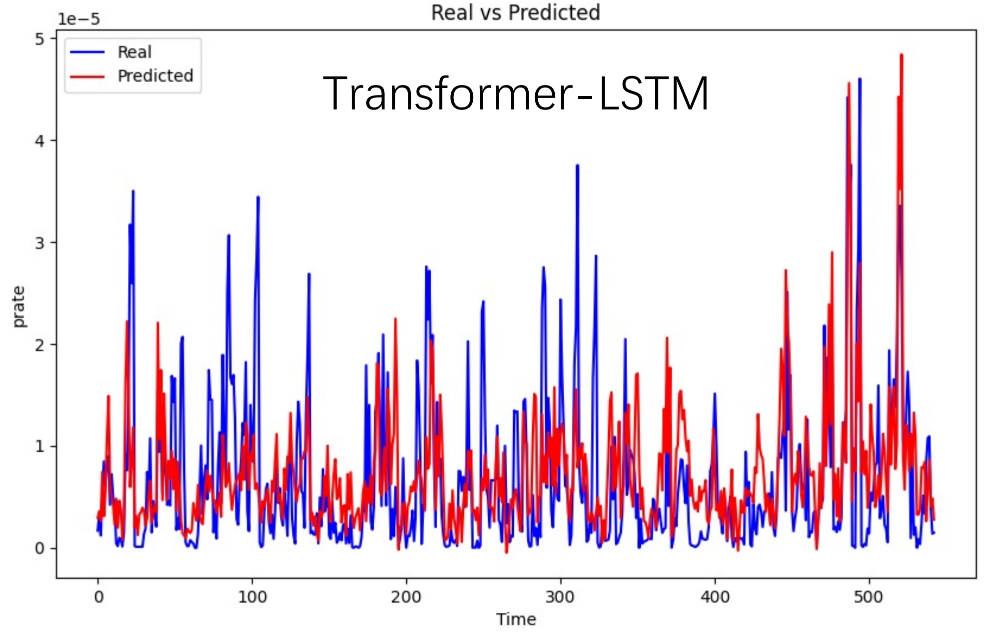


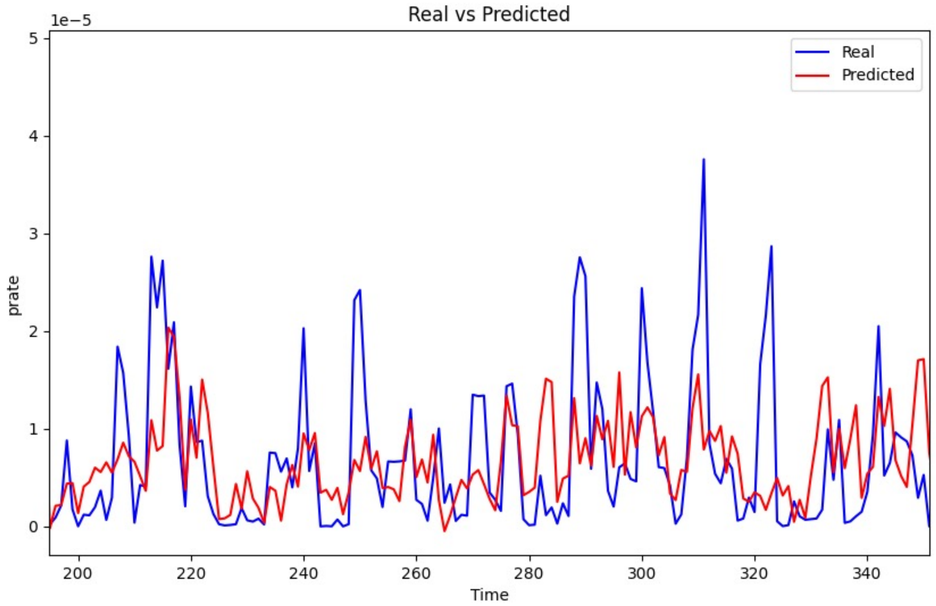



Testing prate
1e−5
Testing prate
Predicted prate
RNN
prate

Testing prate
1e−5
Testing prate
Predicted prate
prate
Time

Testing Data
1e−5
Actual
Predicted
LSTM
Rainfall
Time

Testing Data
1e−5
Actual
Predicted
Rainfall
Time


Real vs Predicted
1e−5
Real
Predicted
Simple transformer
prate
Time


Real vs Predicted
1e−5
Real
Predicted
prate
Time


Real vs Predicted
1e−5
Real
Predicted
Transformer-LSTM
prate
Time


Real vs Predicted
1e−5
Real
Predicted
prate
Time

Model	**MSE**	**MAE**
Pyraformer	0.527003	0.528641
Crossformer	1.278597	0.919760
Transformer	0.467064	0.470900
TimesNet	0.461482	0.452023
LightTS	0.864506	0.634235
PatchTST	0.468976	0.454848
Nonstationary_Transformer	0.457582	0.436677

Next Month

Main

- Start Causality (Reading books, related papers)
- Keep trying to understand Drought prediction deeper(Code)
- Start build a code package for causality in drought prediction.
- Trying to summarize Math in some most common improvements in transformer and other architecture. Like module maths which could be used easily in our own module.

by the way,

Trying to build a code library structure using pytorch.