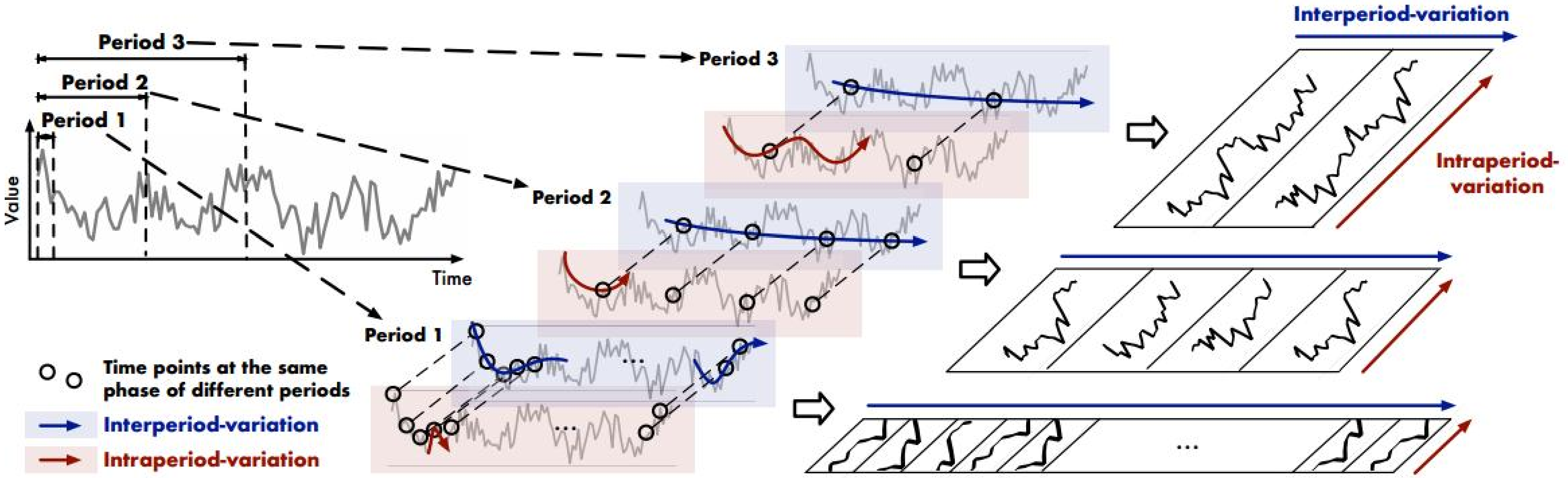
《MSGNet: Learning Multi-Scale Inter-Series Correlations for Multivariate Time Series Forecasting》

四川大学， AAAI2024

报告时间　　2024.6.14

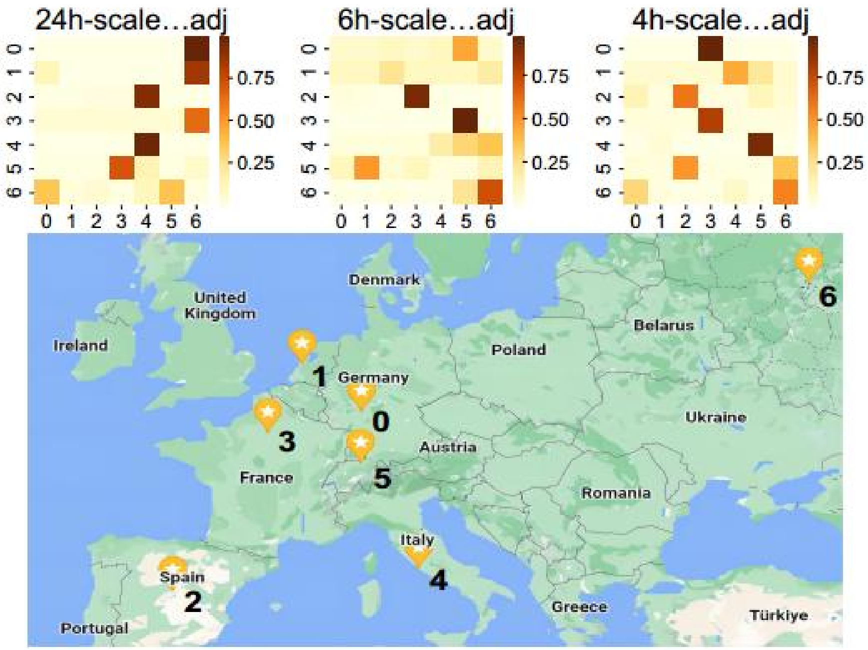
研究方向　　时间序列

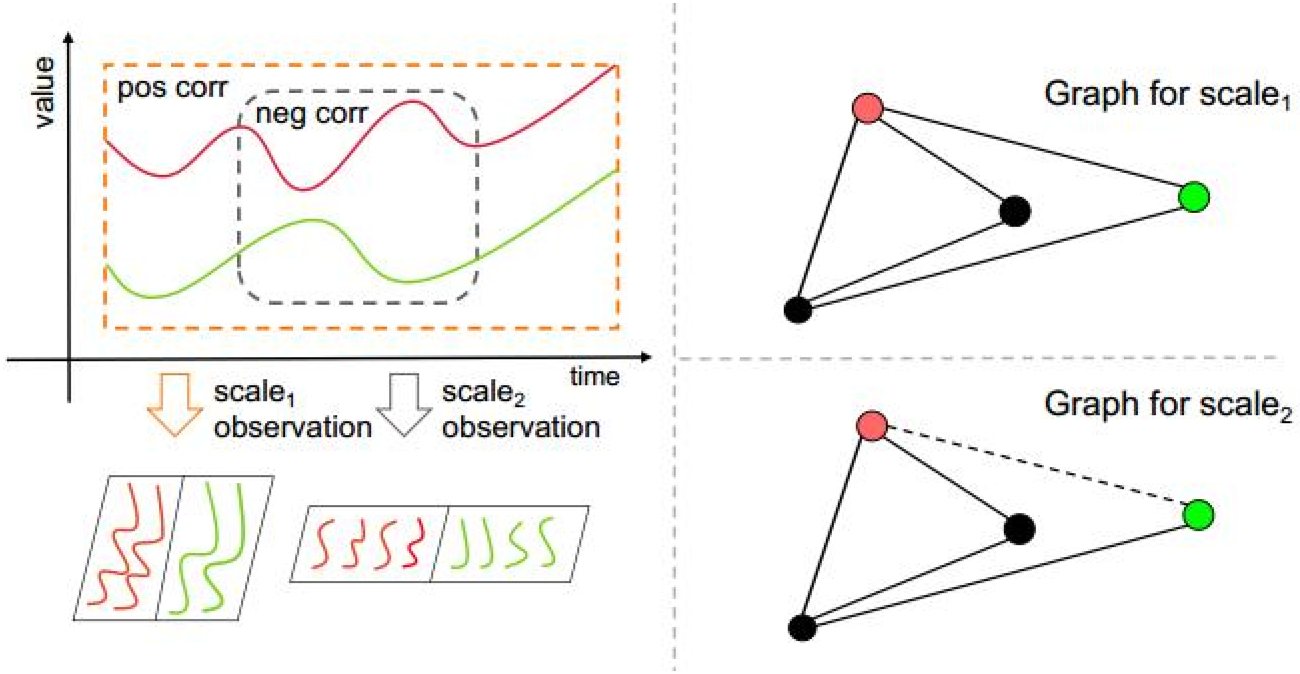


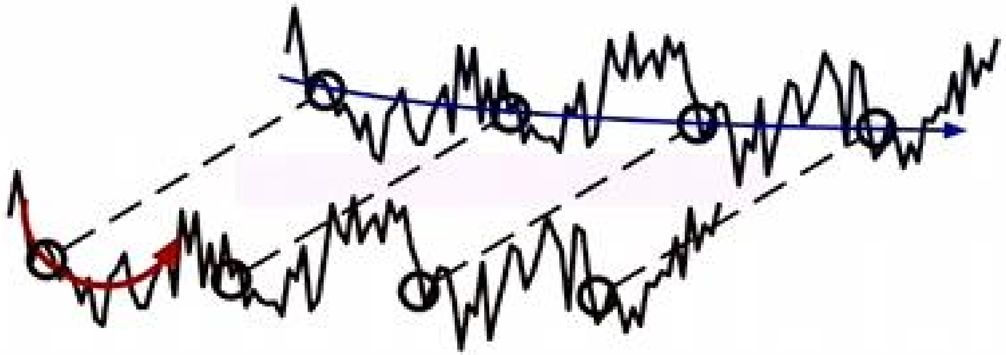
TimesNet，2023ICLR

Shandong Jianzhu University

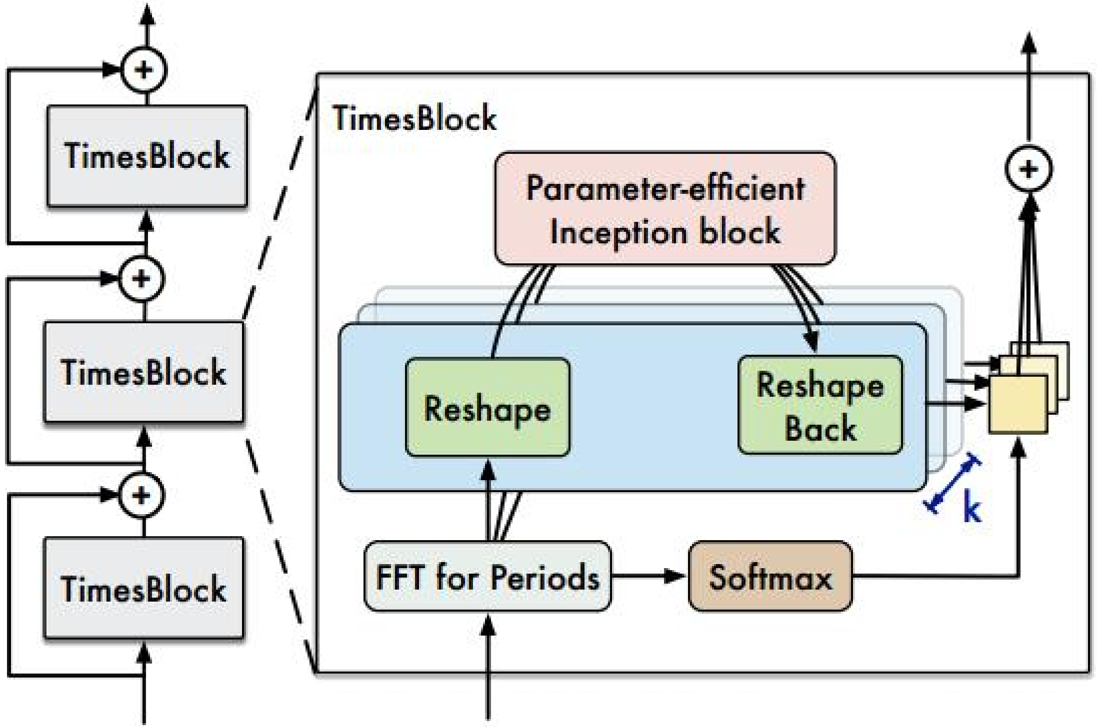
序列内、序列间相关性





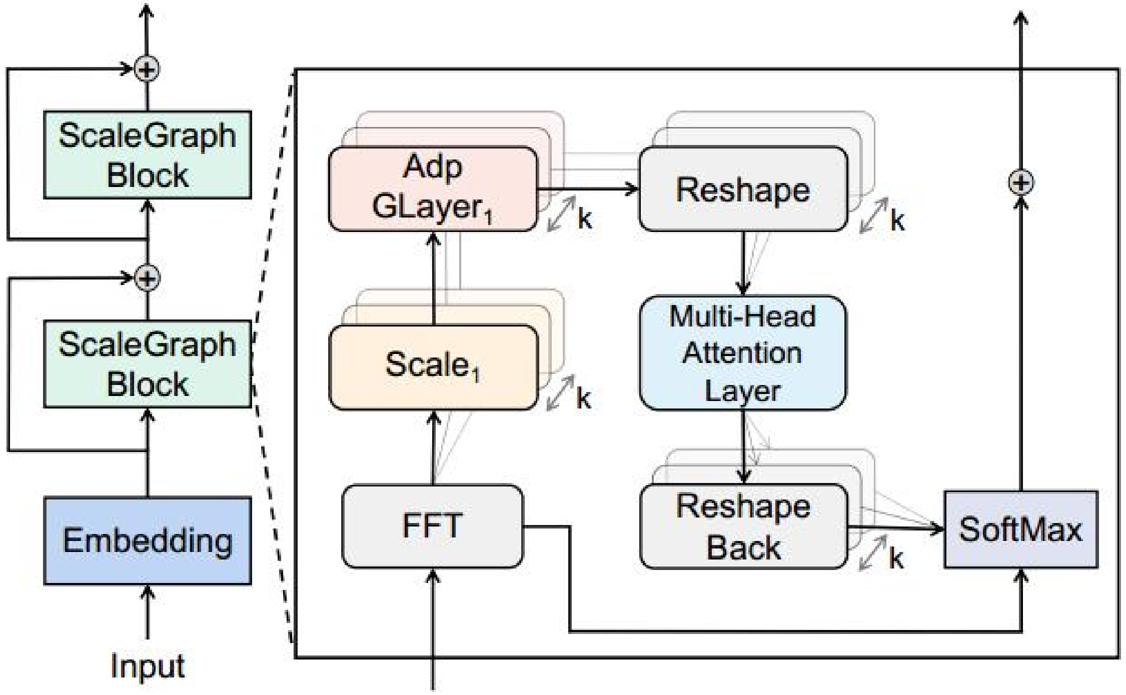


TimesNet

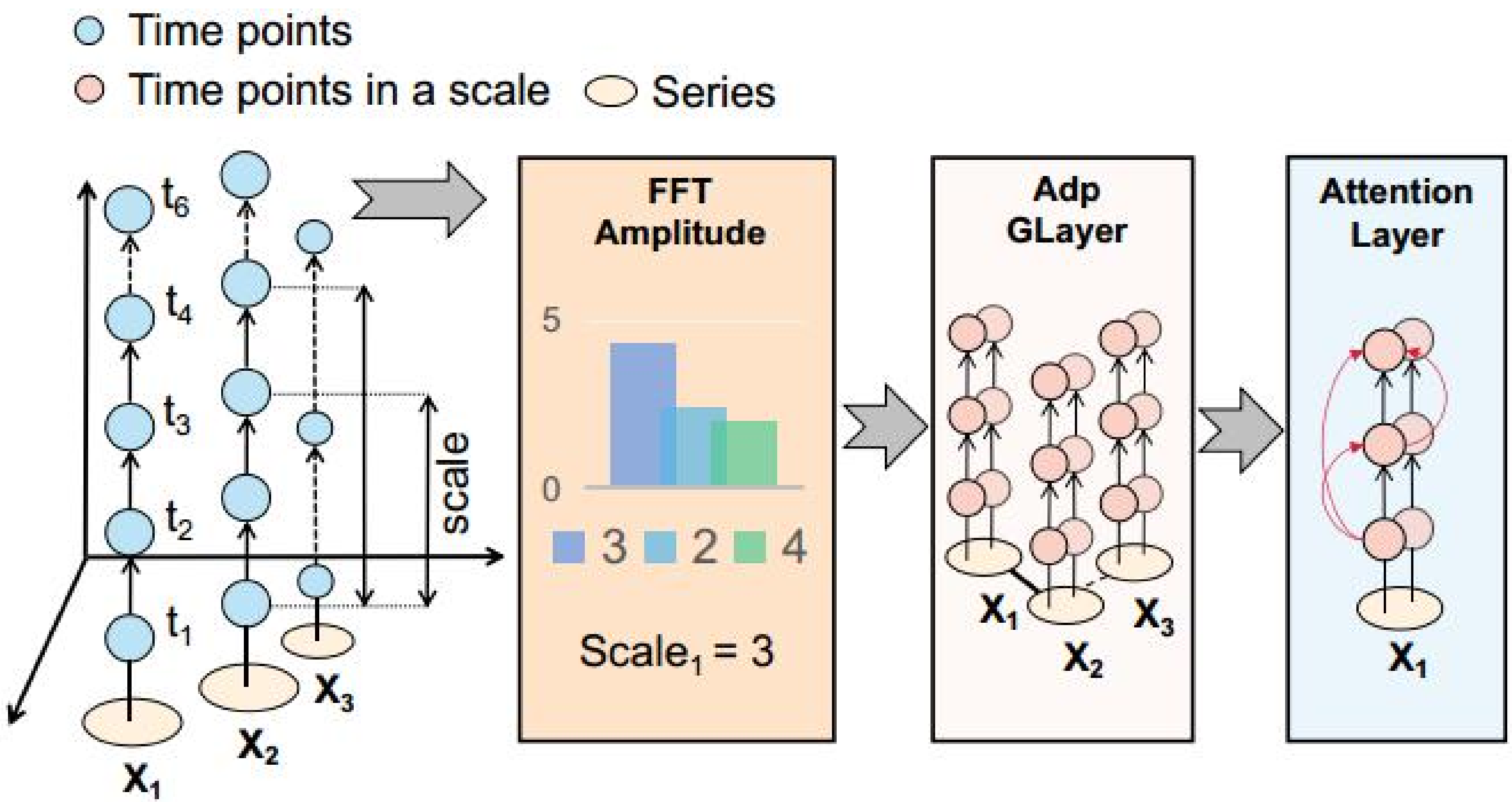


ICLR-2023，清华THUML

MSGNet



四川大学， AAAI2024


Time points
Time points in a scale
Series
t6
t4
t3
t2
t1
scale
X1
X2
X3
FFT
Amplitude
5
0
3
2
4
Scale1 = 3
Adp
GLayer
X1
X2
X3
Attention
Layer
X1

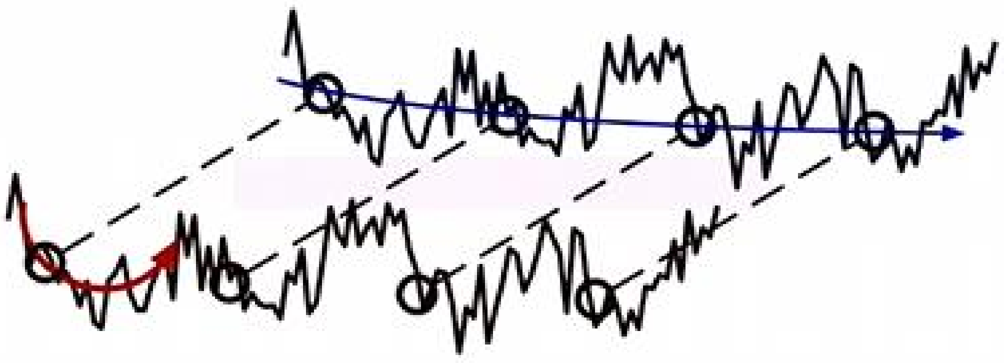
《Water-wave Information Transmission and Recurrent Acceleration Network for Long-range Time Series Forecasting》

北京交通大学，NeurIPS 2023

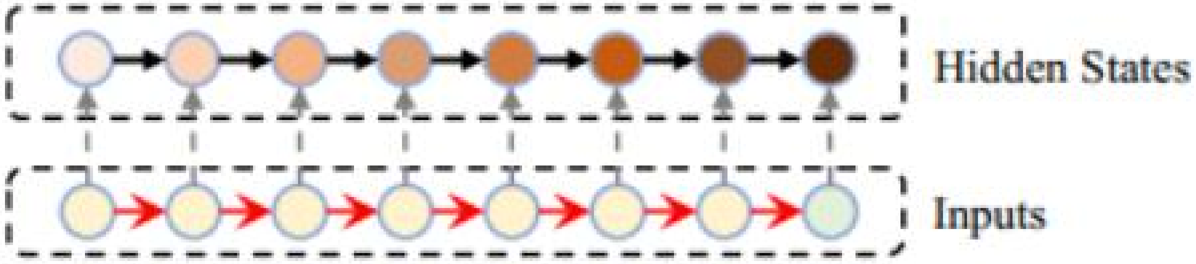
报告时间 2024.6.14

研究方向 时间序列

任务：使用更长的历史序列作为输入



全局和局部相关性：短期变化和长期趋势
不同尺度上的周期性语义信息



(a) RNN

Table 1: Advantages of **WITRAN** compared to other methods.

Advantages	(a) RNN	(b) CNN	(c) Full Attention	(d) LogTrans	(e) Pyraformer	(f) MICN	(g) PatchTST	(h) TimesNet	**(i) WITRAN (ours)**
Non point-wise semantic information capture	✓	✓	✗	✓	✓	✓	✓	✓	✓
Special design to capture long-term repetitive patterns	✗	✗	✗	✗	✓	✗	✗	✓	✓
Efficiently (1 or 2 layers) model global correlations	✓(1)	✗	✓(1)	✗	✗	✓(2)	✓(2)	✗	✓(1)
Well solve the gradient vanishing/exploding problem of RNN	✗	–	–	–	–	–	–	–	✓

1）Transformer，informer，不依赖逐点信息，考虑周期性信息（Autoformer ,Pyraformer）

Methods	RNN	CNN	Transformer	LogTrans	Informer	Autoformer	Pyraformer	FEDformer	FiLM	PatchTST	MICN	**WITRAN (ours)**
Time	$\mathcal{O}(L)$	$\mathcal{O}(L)$	$\mathcal{O}(L^2)$	$\mathcal{O}(L\log L)$	$\mathcal{O}(L\log L)$	$\mathcal{O}(L\log L)$	$\mathcal{O}(L)$	$\mathcal{O}(L)$	$\mathcal{O}(L)$	$\mathcal{O}((L/S)^2)$	$\mathcal{O}(L)$	$\mathcal{O}(\sqrt{L})$
Memory	$\mathcal{O}(L)$	$\mathcal{O}(L)$	$\mathcal{O}(L^2)$	$\mathcal{O}(L^2)$	$\mathcal{O}(L\log L)$	$\mathcal{O}(L\log L)$	$\mathcal{O}(L)$	$\mathcal{O}(L)$	$\mathcal{O}(L)$	$\mathcal{O}((L/S)^2)$	$\mathcal{O}(L)$	$\mathcal{O}(L)$

水波纹信息传输WIT

1）输入序列根据自然周期进行排列
2）在水平和垂直两个方向上分别结合门控选择单

TimesNet用的傅里叶提取周期，

WITran用的自然周期

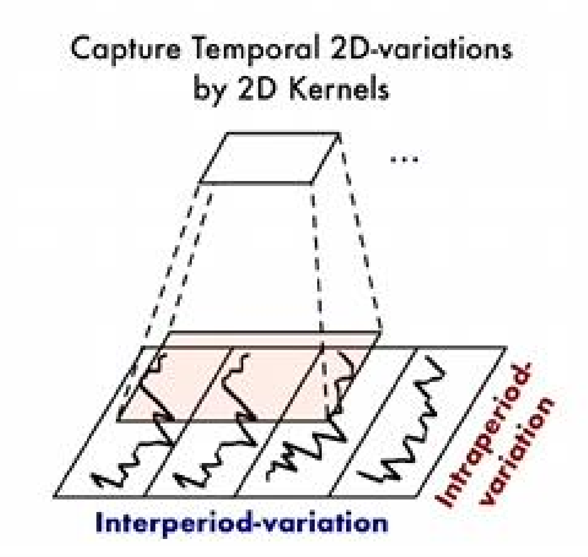
$$\mathbf{X}_{1\mathrm{D}} \in \mathbb{R}^{T \times C}.$$

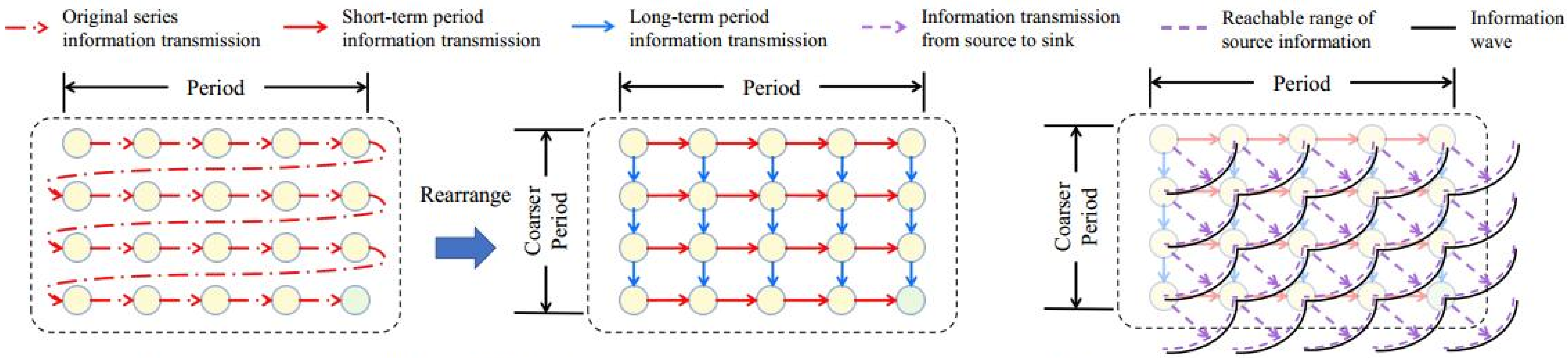
$$\mathbf{A} = \mathrm{Avg}\left(\mathrm{Amp}\left(\mathrm{FFT}(\mathbf{X}_{1\mathrm{D}})\right)\right),\ \{f_1, \cdots, f_k\} = \underset{f_* \in \{1, \cdots, [\frac{T}{2}]\}}{\arg\mathrm{Topk}} (\mathbf{A}),\ p_i = \left\lceil \frac{T}{f_i} \right\rceil, i \in \{1, \cdots, k\}.$$

$$\{p_1, \cdots, p_k\}.\cdot_k\},$$

循环加速网络（RAN）
3）并行处理两个方向的信息传输。O（L）内存复杂度，O（根号L）的时间复杂度

TimesNet转到2维度之后用卷积来做，
WITRAN用的RNN来做





(a) Information transmission before and after the input sequence rearrangement (b) The water-wave information transmission process

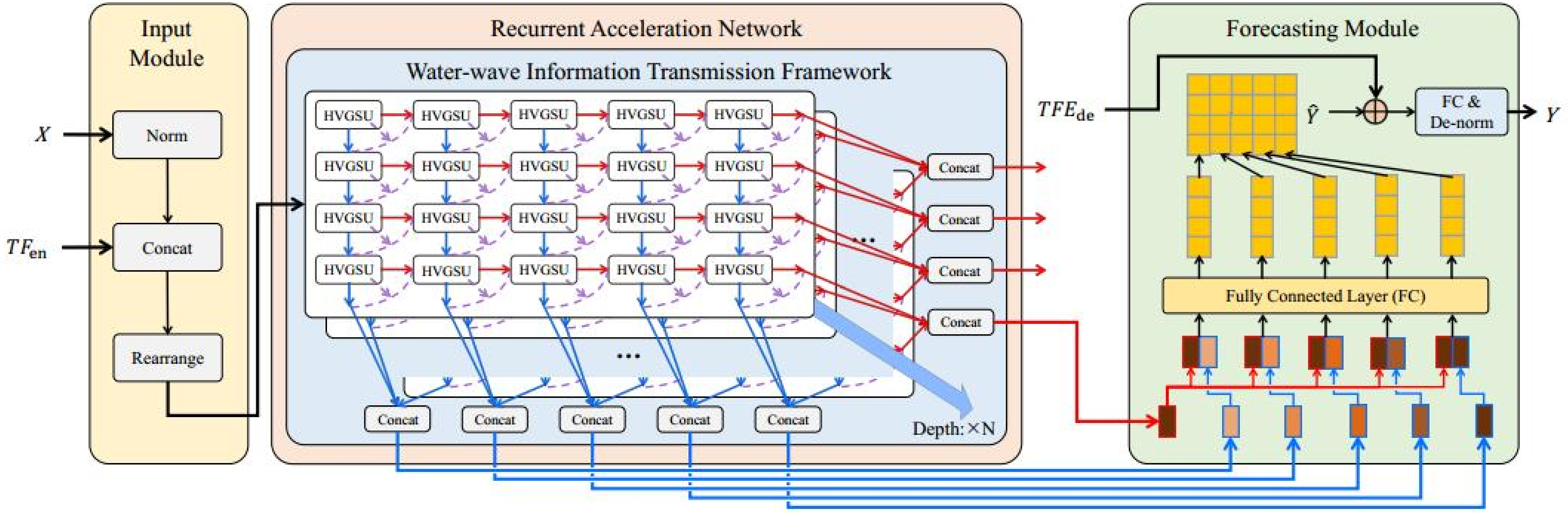
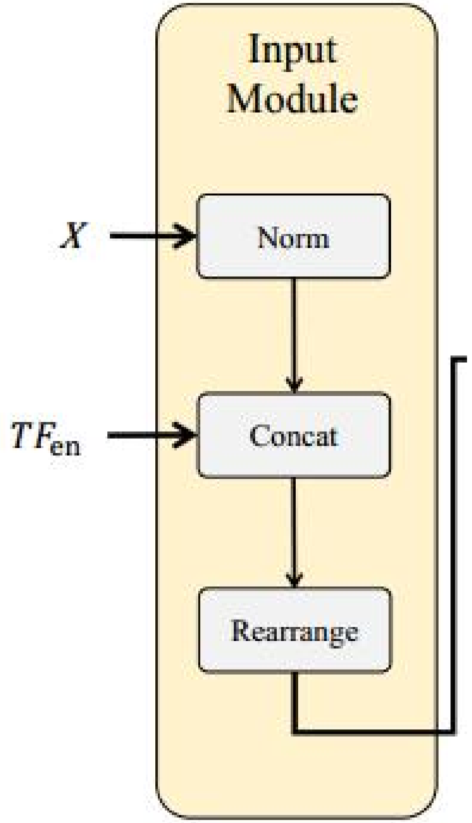


Figure 3: Overall structure of WITRAN.

序列重排：一维到二维

$X_{1D} \in \mathbb{R}^{H \times c_{in}}$ $TF_{en} \in \mathbb{R}^{H \times c_{time}}$

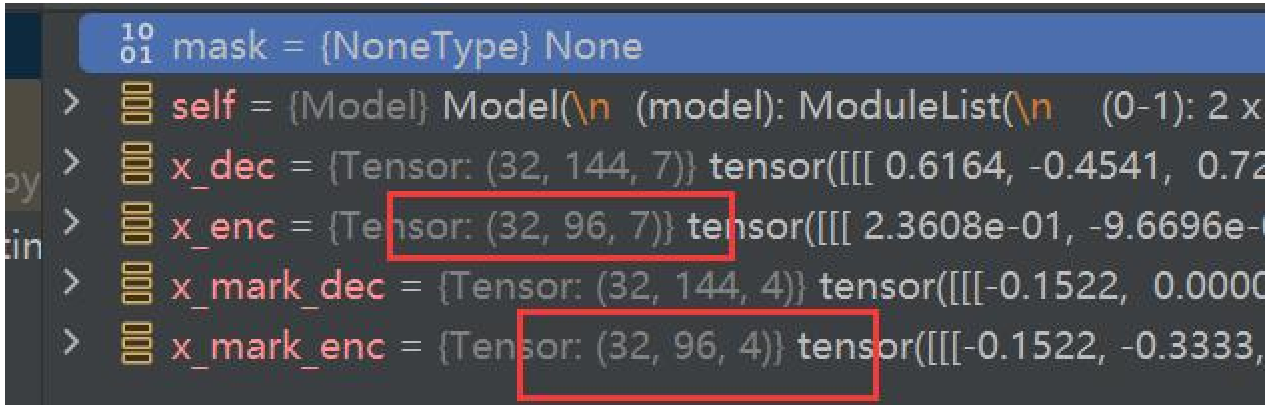
HourOfDay, DayOfWeek, DayOfMonth and DayOfYear)



$$X_{1D} = \begin{cases} X & , norm = 0 \\ X - x_H & , norm = 1 \end{cases}$$

$$X_{2D} = \mathrm{Rearrange}([X_{1D}, TF_{en}]),$$

$$X_{2D} \in \mathbb{R}^{R \times C \times (c_{in} + c_{time})}$$



$$X_{1\mathrm{D}} = \begin{cases} X & , norm = 0 \\ X - x_H & , norm = 1 \end{cases}$$

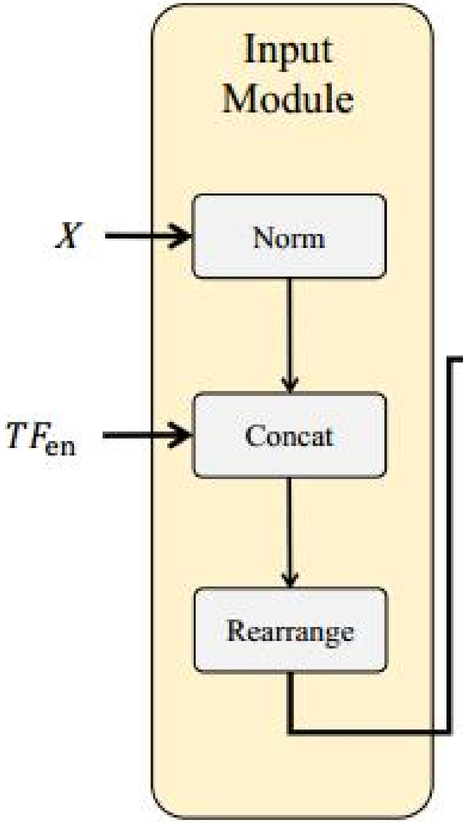
$$X_{2\mathrm{D}} = \mathrm{Rearrange}([X_{1\mathrm{D}}, TF_{\mathrm{en}}]),$$

减去最后一个时间点的值

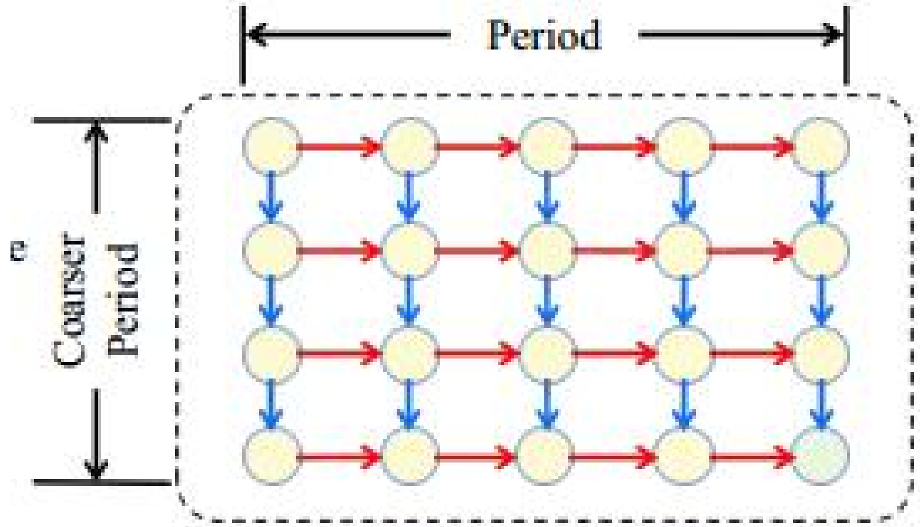
训练集和预测集分布差异不明显时，说明两组数据的波动是相似的norm是0

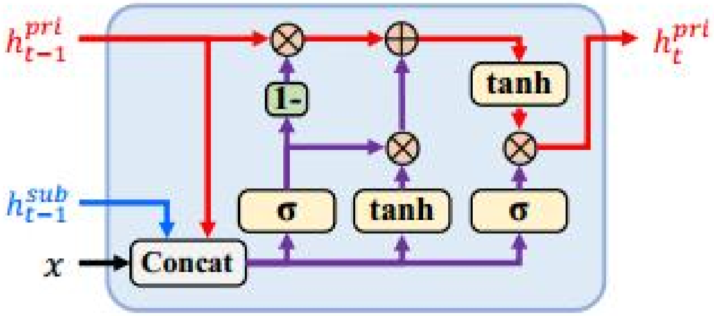
Table 10: The distribution of data in the training and validation sets (Mean and STD) and the value of $norm$.

Datasets	ECL			Traffic			ETTh1			ETTh2			Weather		
Tasks	training set	validation set	$norm$	training set	validation set	$norm$	training set	validation set	$norm$	training set	validation set	$norm$	training set	validation set	$norm$
168-168	3425.733±564.8776	3036.397±388.2128		0.029±0.0170	0.034±0.0201		16.880±8.2921	6.667±4.1794		28.959±12.0653	18.680±9.0427		0.500±6.6321	1.143±7.7659	
168-336	3427.480±566.9556	3036.291±388.0110		0.029±0.0170	0.035±0.0202		16.606±8.0735	6.258±3.8462		28.767±12.0604	17.922±8.5800		0.536±6.6444	1.017±7.8162	
336-336	3428.455±569.1108	3036.291±388.0110	0	0.029±0.0170	0.035±0.0202	1	16.207±7.5364	6.258±3.8462	1	28.434±11.8740	17.922±8.5800	0	0.585±6.6409	1.017±7.8162	0
336-720	3434.150±573.2660	3037.919±387.2758		0.029±0.0170	0.035±0.0203		15.446±6.6217	5.583±3.4658		27.774±11.6120	16.360±7.7594		0.700±6.6410	0.721±7.9234	
720-720	3437.773±578.4705	3037.919±387.2758		0.029±0.0170	0.035±0.0203		14.832±5.9927	5.583±3.4658		27.111±11.3299	16.360±7.7594		0.825±6.6270	0.721±7.9234	
720-1440	3439.817±586.5029	3046.877±397.7761		0.029±0.0170	0.035±0.0204		14.044±5.5077	4.273±2.7600		26.403±11.4257	13.646±6.5725		0.978±6.6791	-0.547±7.5174	
1440-1440	3452.135±594.6857	3046.877±397.7761	0	0.029±0.0169	0.035±0.0204	1	13.722±5.5456	4.273±2.7600	1	26.355±11.8918	13.646±6.5725	1	1.029±6.7679	-0.547±7.5174	0
1440-2880	3458.328±610.2118	3093.128±446.4128		0.029±0.0171	0.036±0.0208		14.195±5.5780	2.623±2.5005		28.303±12.1275	9.130±5.5363		0.725±6.8812	-3.859±5.5129	



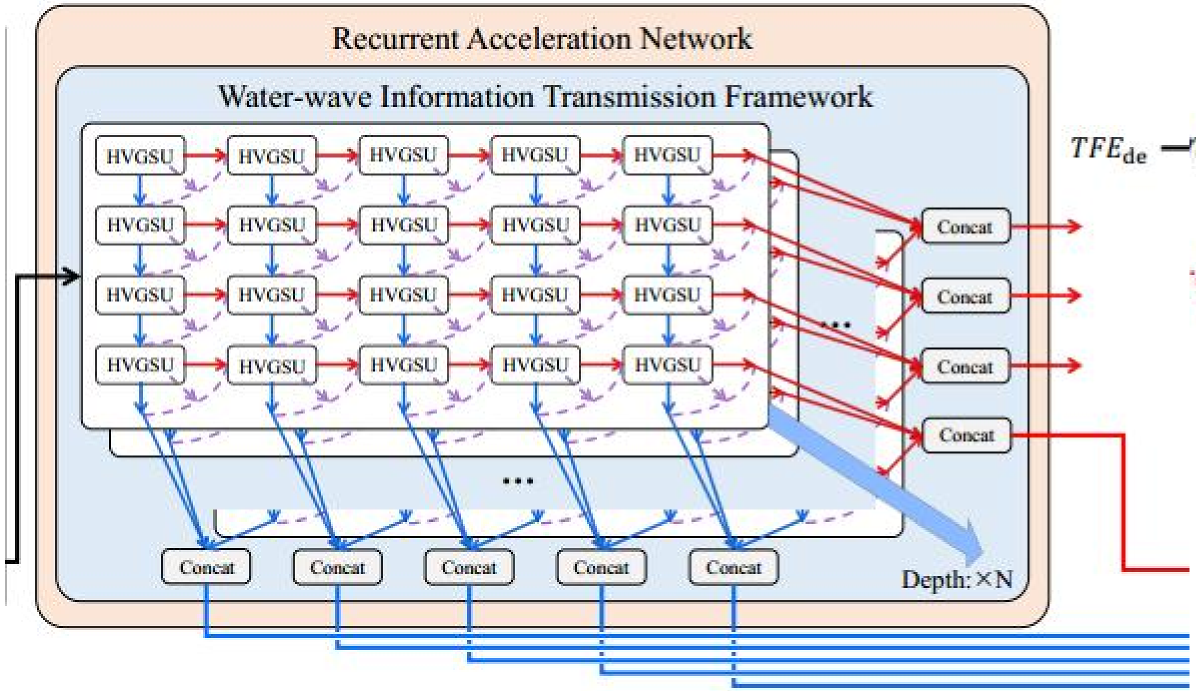
分析时序的自然周期C，做序列重排
L=R行*C列
L能被C整除避免了padding的出现
C-12,24,48......

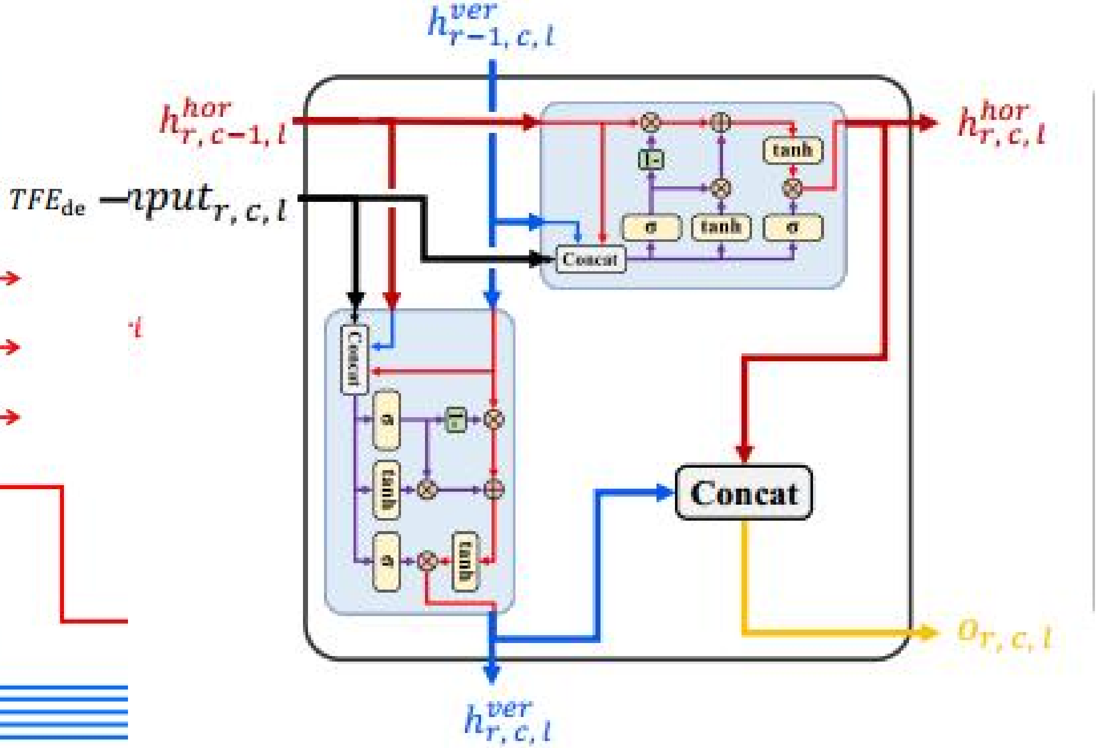




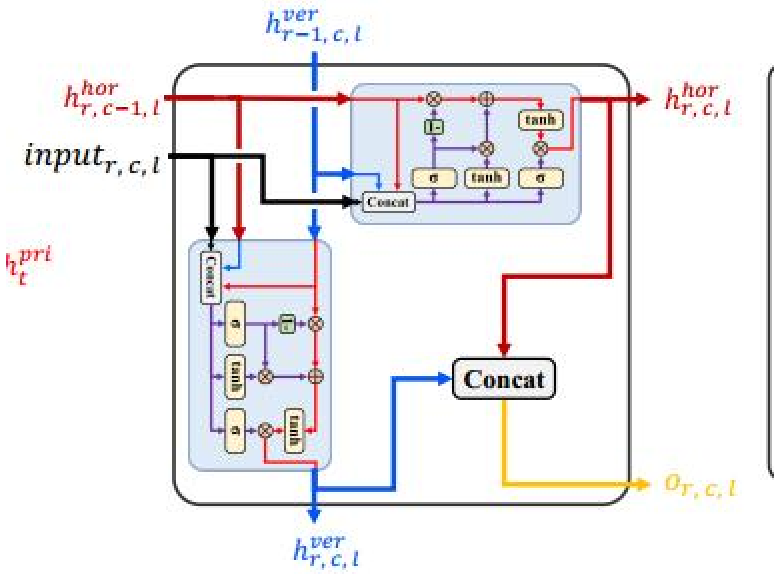
$$h_{r,\,c,\,l}^{\mathrm{hor}} = \mathrm{GSC}_{\mathrm{hor}}(input_{r,\,c,\,l},\ \boxed{h_{r,\,c-1,\,l}^{\mathrm{hor}}},\ h_{r-1,\,c,\,l}^{\mathrm{ver}})$$
$$h_{r,\,c,\,l}^{\mathrm{ver}} = \mathrm{GSC}_{\mathrm{ver}}(input_{r,\,c,\,l},\ \boxed{h_{r-1,\,c,\,l}^{\mathrm{ver}}},\ h_{r,\,c-1,\,l}^{\mathrm{hor}})$$
$$o_{r,\,c,\,l} = [h_{r,\,c,\,l}^{\mathrm{hor}},\ h_{r,\,c,\,l}^{\mathrm{ver}}],$$

主隐藏态，副隐藏态

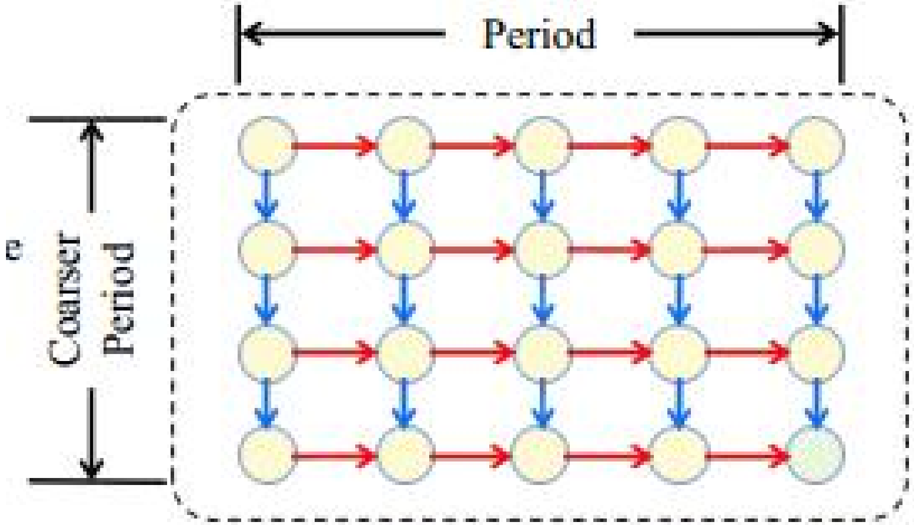




循环加速网络RAN

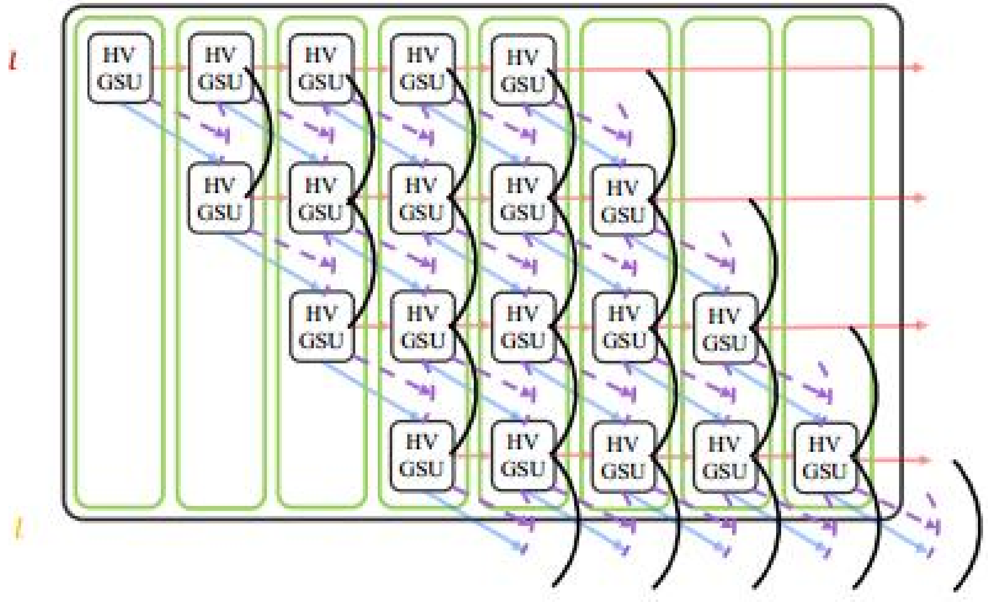


(b) HVGSU structure



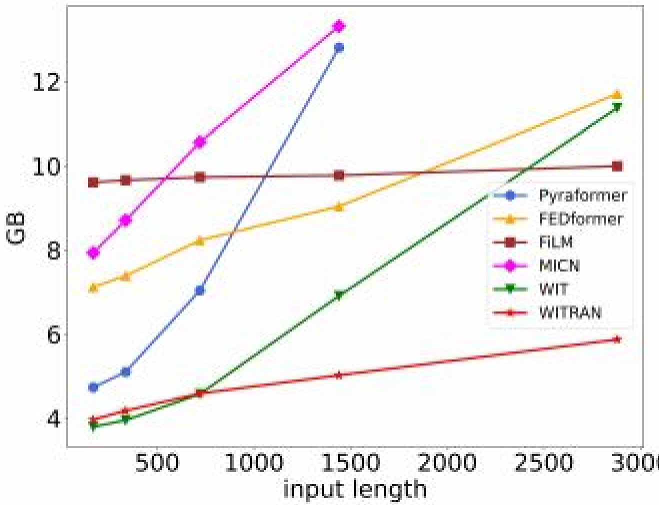
数据点并行运算

切片长度=R+C-1

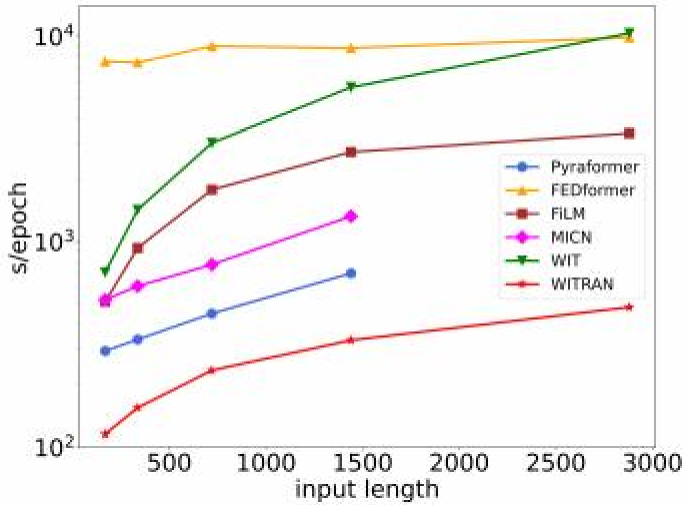


(c) RAN structure

固定预测长度
变化输入长度

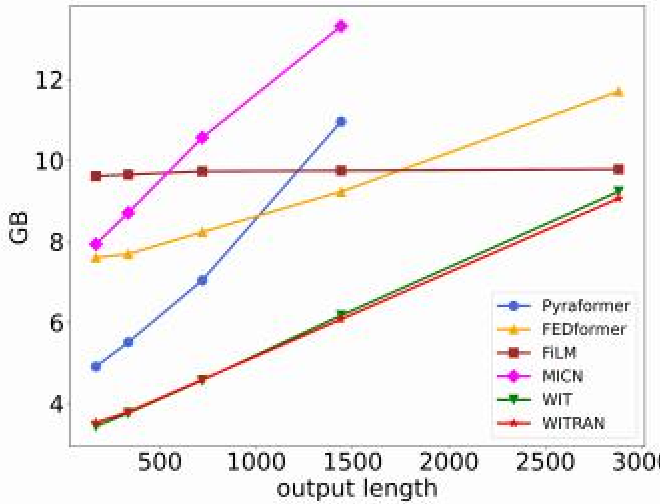


(c) Memory consumption (output length is fixed as 720)

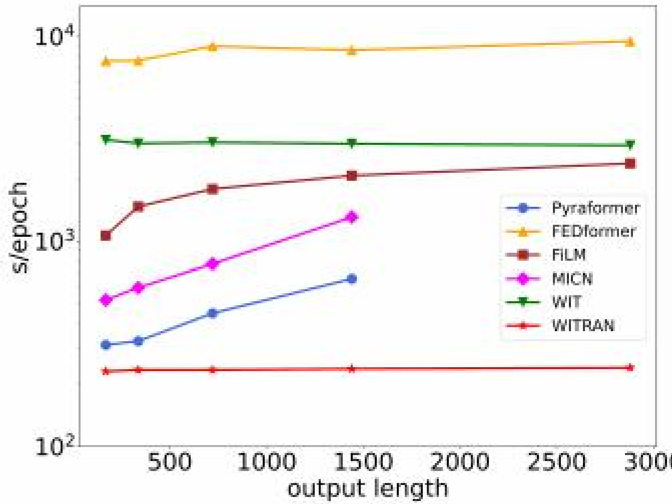


(d) Time consumption (output length is fixed as 720)

固定输入长度
变化预测长度



(a) Memory consumption (input length is fixed as 720)



(b) Time consumption (input length is fixed as 720)

山东建筑大学
Shandong Jianzhu University

《First De-Trend then Attend: Rethinking Attention for Time-Series Forecasting》
NeurIPS 2022， 亚马逊AI实验室

报告时间　　2024.6.14

研究方向　　时间序列

基于attention和Transformer的时间序列预测模型

1）时域（informer）

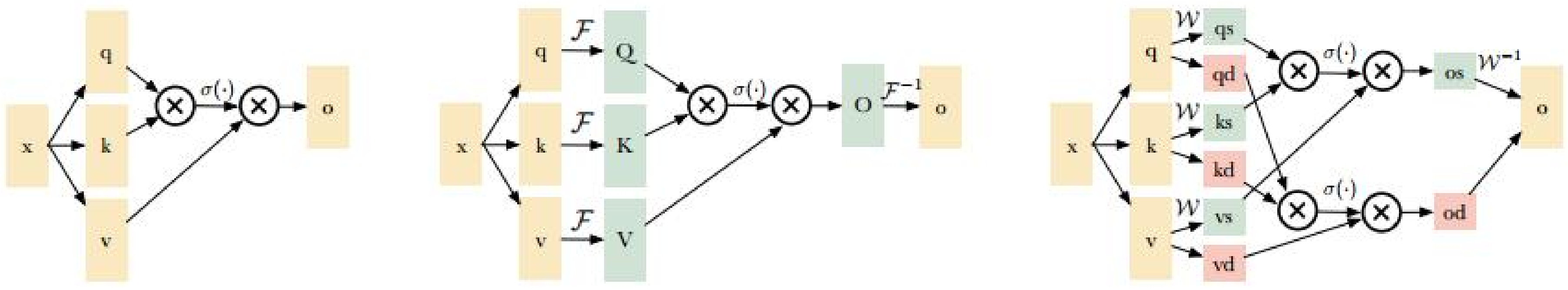
2）傅里叶频域（FEDformer）

3）小波变换频域

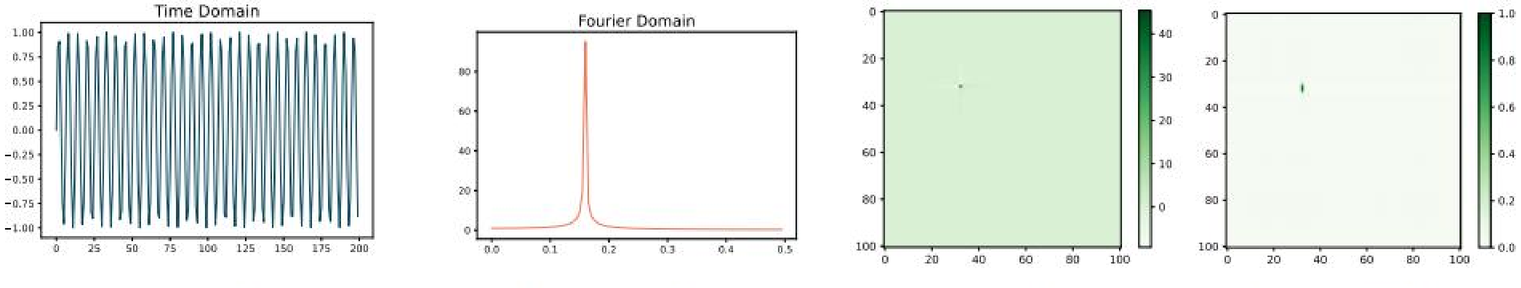
$$Attention(Q,K,V) = Softmax(\frac{Q \cdot K^T}{\sqrt{d_k}}) \cdot V$$

$$\mathbf{o}(\mathbf{q},\mathbf{k},\mathbf{v}) = \mathcal{F}^{-1}\Big(\sigma\big(\mathcal{F}(\mathbf{q})\overline{\mathcal{F}(\mathbf{k})}^T/\sqrt{d_q}\big)\mathcal{F}(\mathbf{v})\Big).$$

$$\mathbf{o}(\mathbf{q},\mathbf{k},\mathbf{v}) = \mathcal{W}^{-1}\Big(\sigma\left(\mathcal{W}(\mathbf{q})\mathcal{W}(\mathbf{k}^T)/\sqrt{d_q}\right)\mathcal{W}(\mathbf{v})\Big).$$



(a) (b) (c)

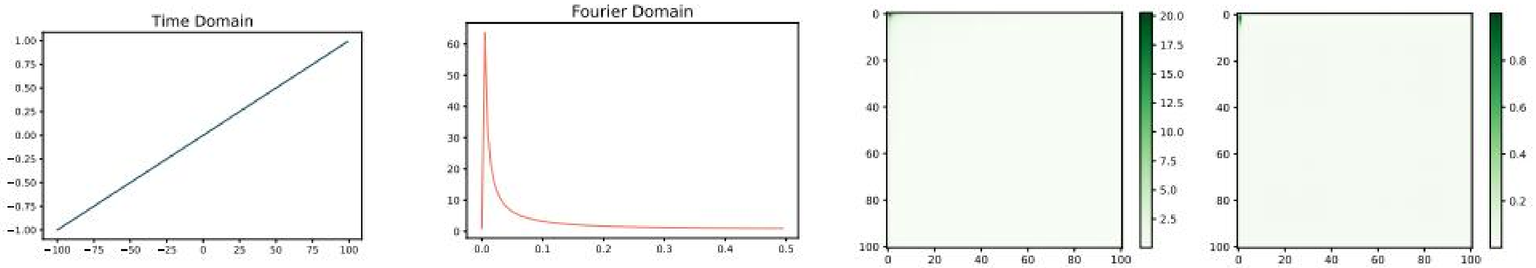


(a) sin(x) Time (b) sin(x) Freq (c) sin(x) Linear (d) sin(x) Softmax

softmax的极化作用能够更好利用季节性数据

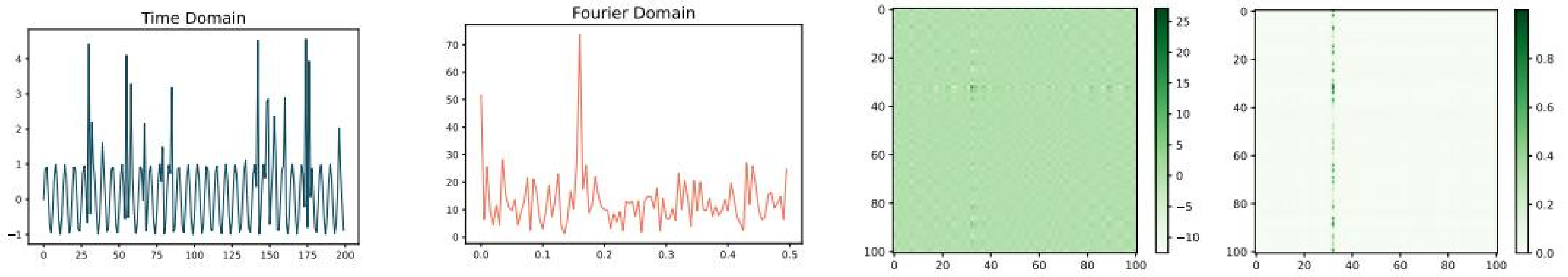
Table 1: MSE and MAE of attention models and MLP with linear-trend data.

Metric	Time	Fourier	Wavelet	MLP
MSE	3.157 ± 0.435	8.567 ± 0.487	2.327 ± 0.689	$\mathbf{0 \pm 0}$
MAE	1.741 ± 0.121	2.880 ± 0.073	1.477 ± 0.239	$\mathbf{0.006 \pm 0.003}$



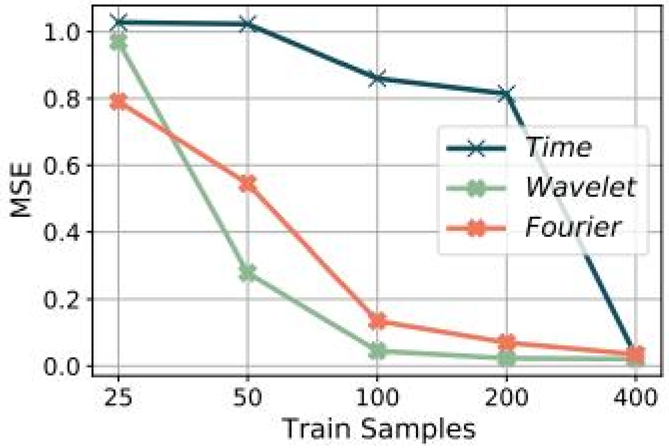
(i) Trend Time (j) Trend Freq (k) Trend Linear (l) Trend Softmax

MLP能够更好利用趋势数据



(m) Spike Time (n) Spike Freq (o) Spike Linear (p) Spike Softmax

傅里叶变化对尖峰数据的鲁棒性强



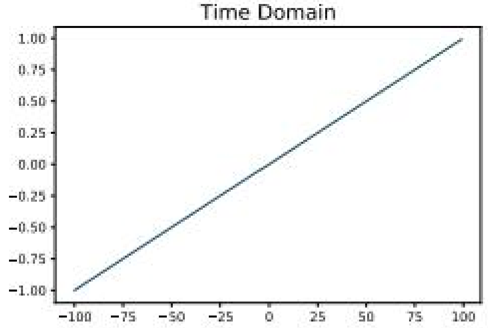
(b) Attention Models

对于周期性强的数据，
傅里叶频域变换效果更好

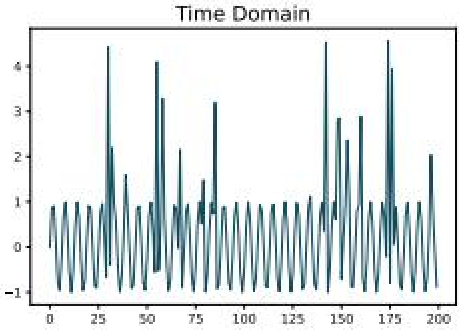
主要贡献

1）证明在线性条件下，时域、傅里叶、小波变换的注意模型有相同的表征能力

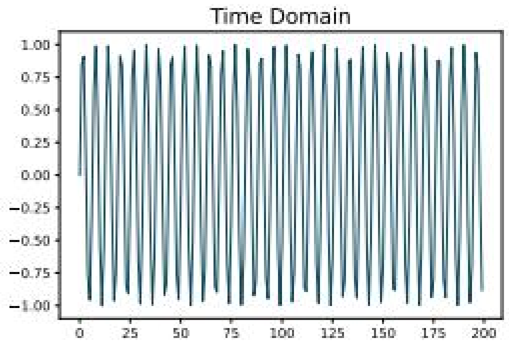
2）考虑softmax的非线性



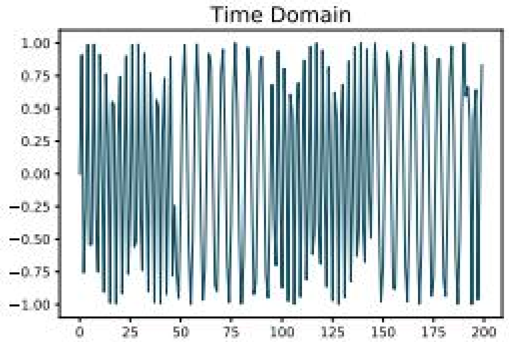
(i) Trend Time



(m) Spike Time

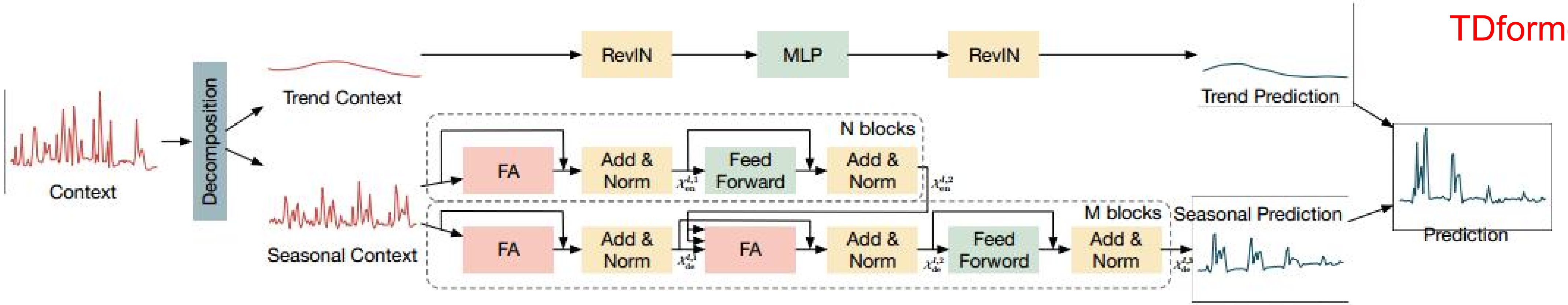


(a) sin(x) Time

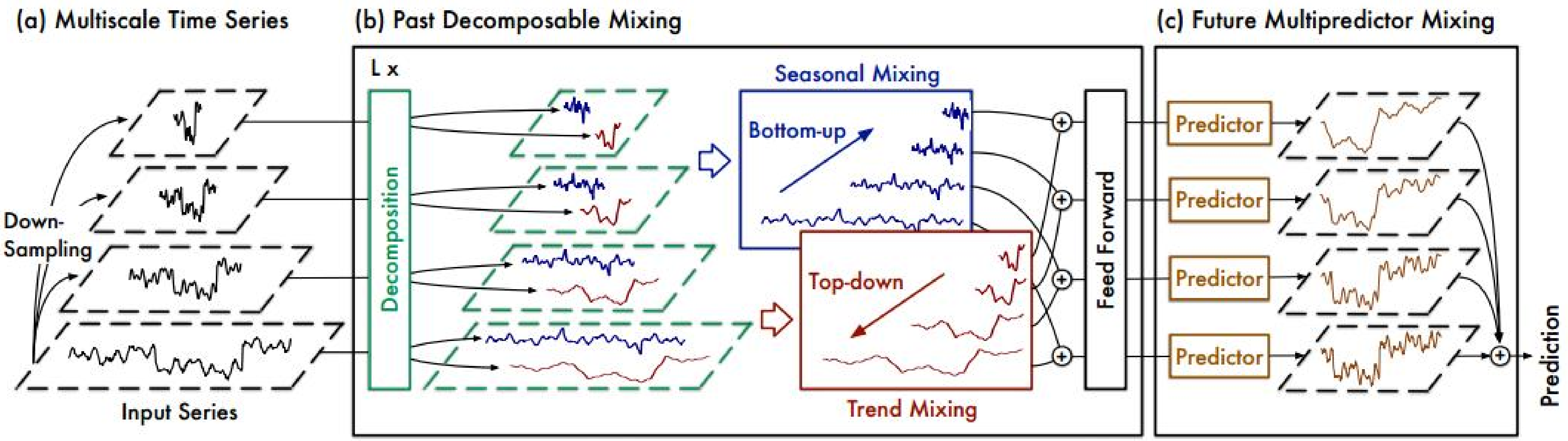


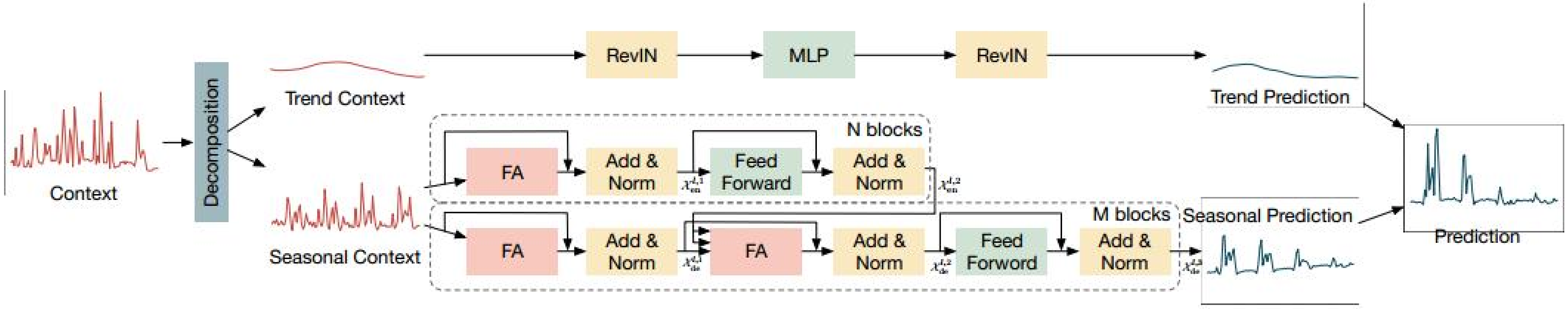
(e) Vary Time

TDformer



TimeMixer





- 趋势项用MLP，Revin（去除序列中的非平稳项）
- 季节项用傅里叶频域注意力