ICML
International Conference
On Machine Learning

SparseTSF

Modeling Long-term Time Series Forecasting with *1k* Parameters

Shengsheng Lin[1], Weiwei Lin[1,2,*], Wentai Wu[3], Haojun Chen[1], Junjie Yang[1]

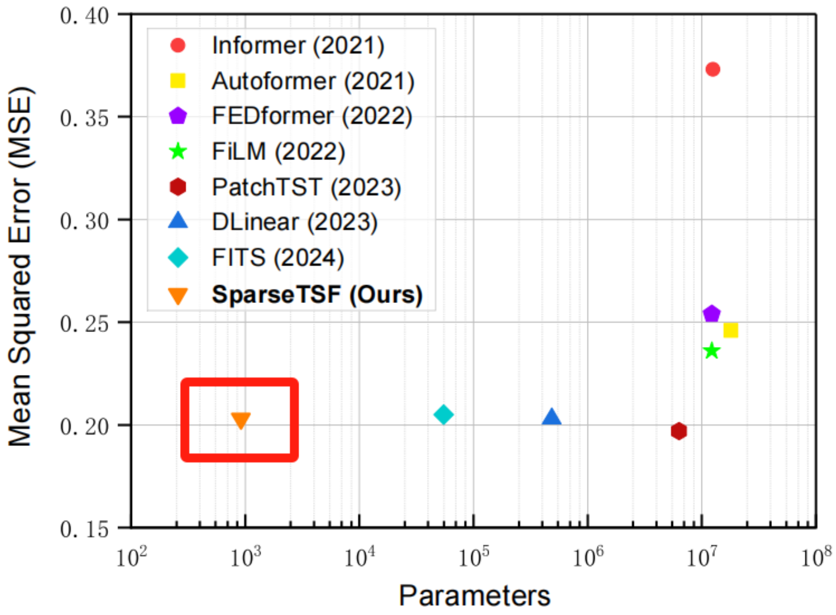
[1]South China University of Technology

[2]Peng Cheng Laboratory

[3]Jinan University

SparseTSF

(*Cross-Period Sparse Forecasting* technique with *Linear* backbone)



- Requires fewer than ***1k* parameters**
- ***1~4 orders of magnitude smaller*** than its counterparts
- **Competitive state-of-the-art** predictive accuracy

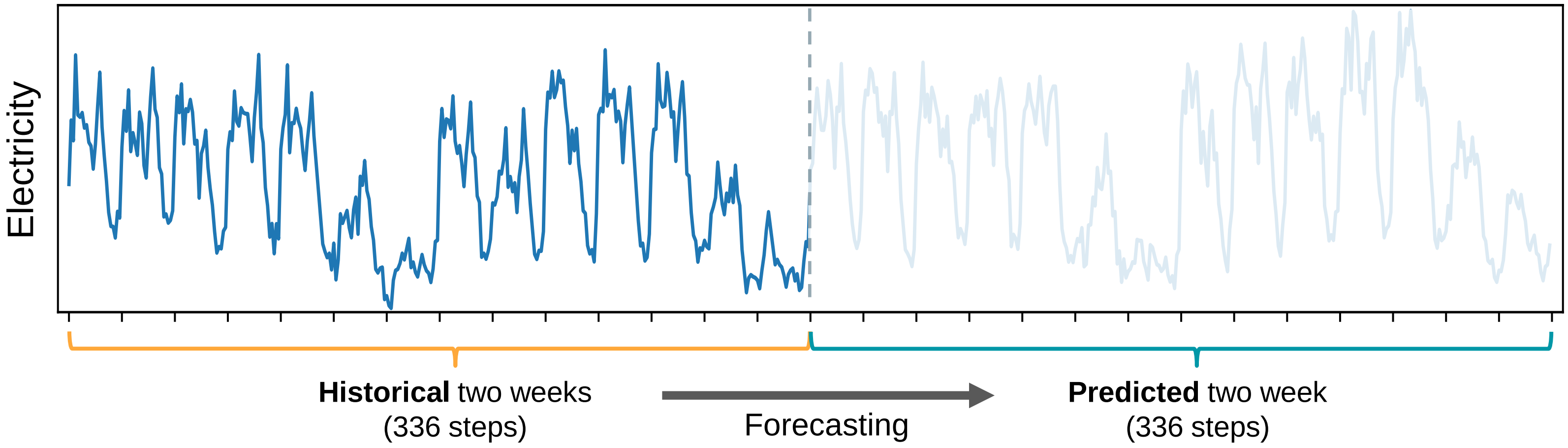
1. Highlights

2. Motivations

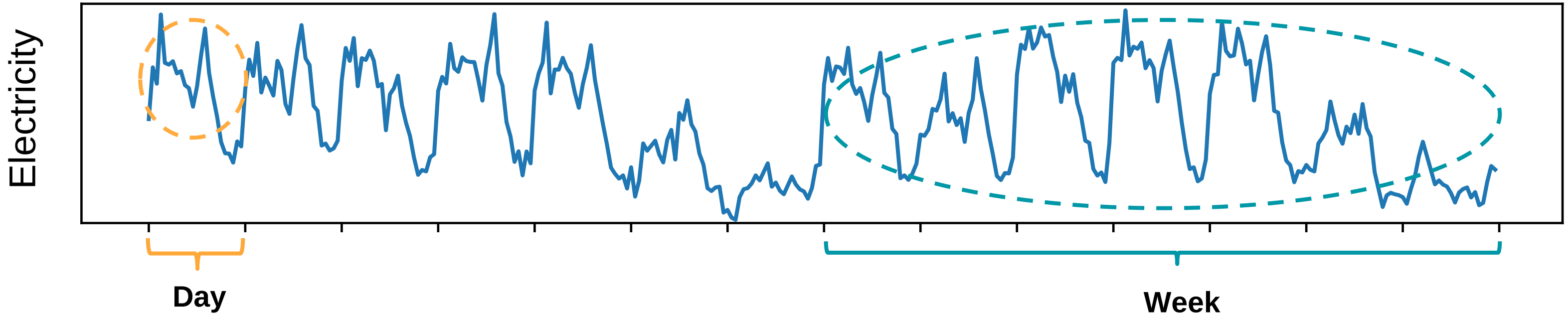
3. Method

4. Results

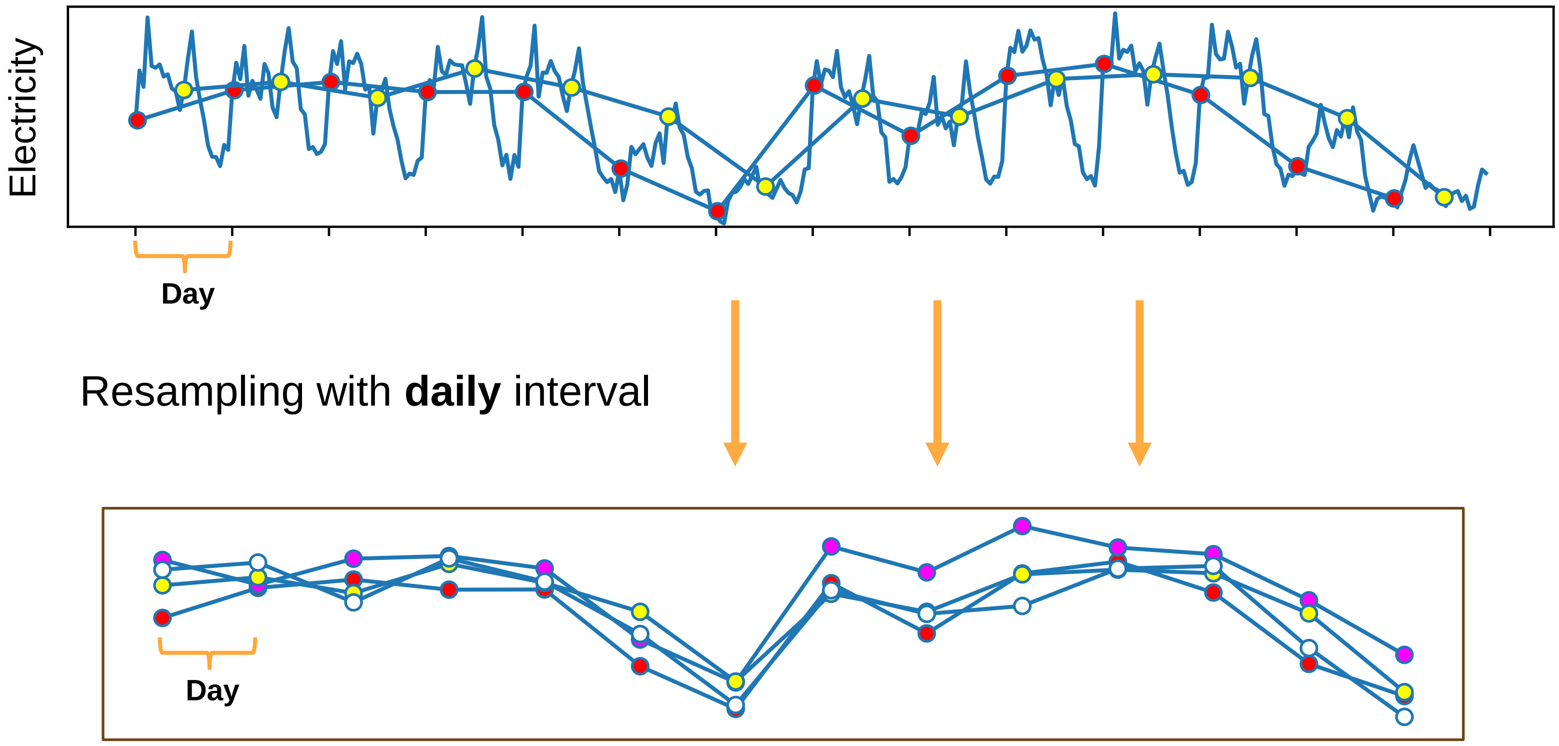
5. Inspirations



- Long-term Time Series Forecasting (LTSF):
  - **Extending forecast horizon** to its *maximum potential* (e.g., up to 720 steps)
  - **Longer lookback windows** are *required for accurate predictions*
  - *Mainstream methods* require ***hundreds of millions of parameters*** to achieve accuracy



- Exhibit significant **daily & weekly** periodicity
- The **realistic basis** for long-term forecasting



Downsampled subsequences

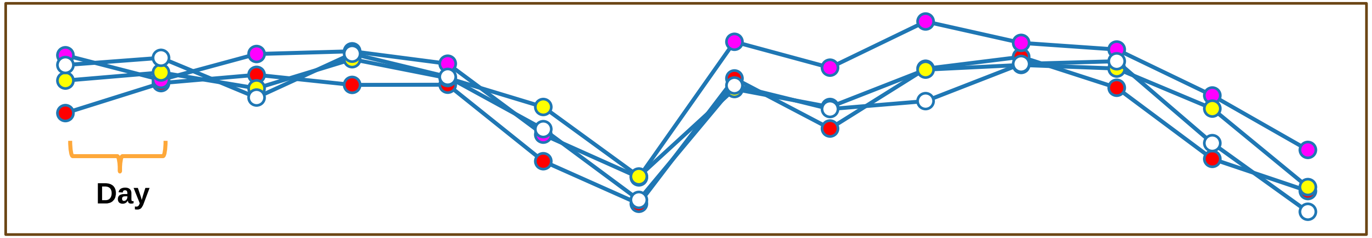
1. Highlights

2. Motivations

3. Method

4. Results

5. Inspirations

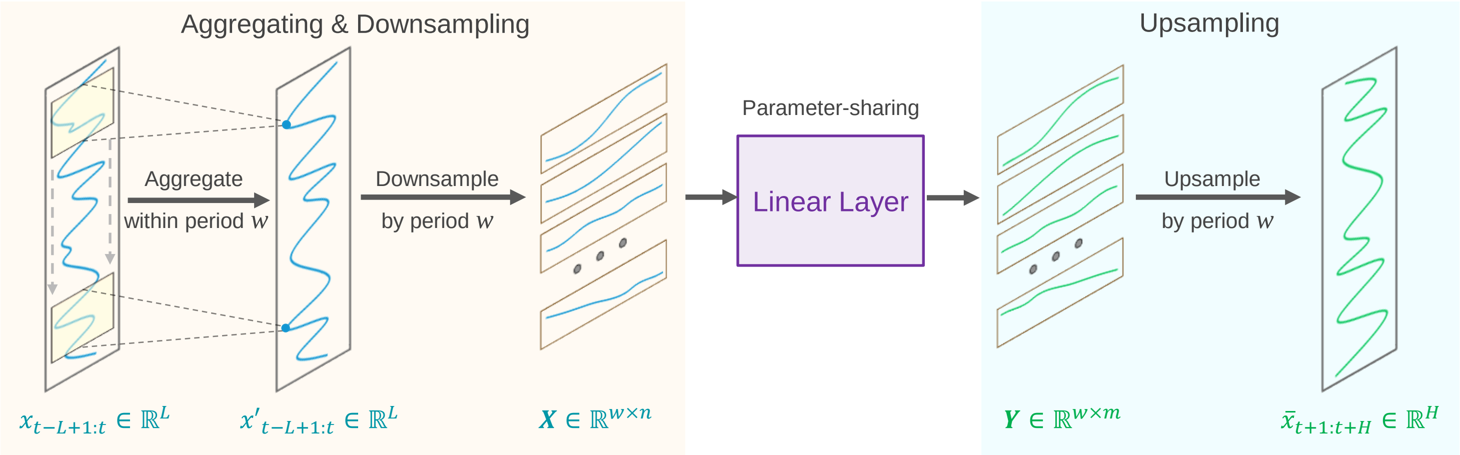


Downsampled subsequences

- Subsequences exhibit **similar** or **consistent** trends
  - Daily periodic patterns → ***Inter***-subsequence patterns
  - Trend patterns → ***Intra***-subsequence patterns
- Subsequence prediction is considerably easier
  - Simplifying into **cross-period trend prediction** task
  - Extremely **compressing** parameter scale

SparseTSF

(*Cross-Period Sparse Forecasting* technique with *Linear* backbone)



- L: Lookback window length
- H: Forecasting horizon
- w: Period length
- $n = \left\lfloor \frac{L}{w} \right\rfloor$: Subsequences lookback
- $m = \left\lfloor \frac{H}{w} \right\rfloor$: Subsequences horizon

1. Highlights

2. Motivations

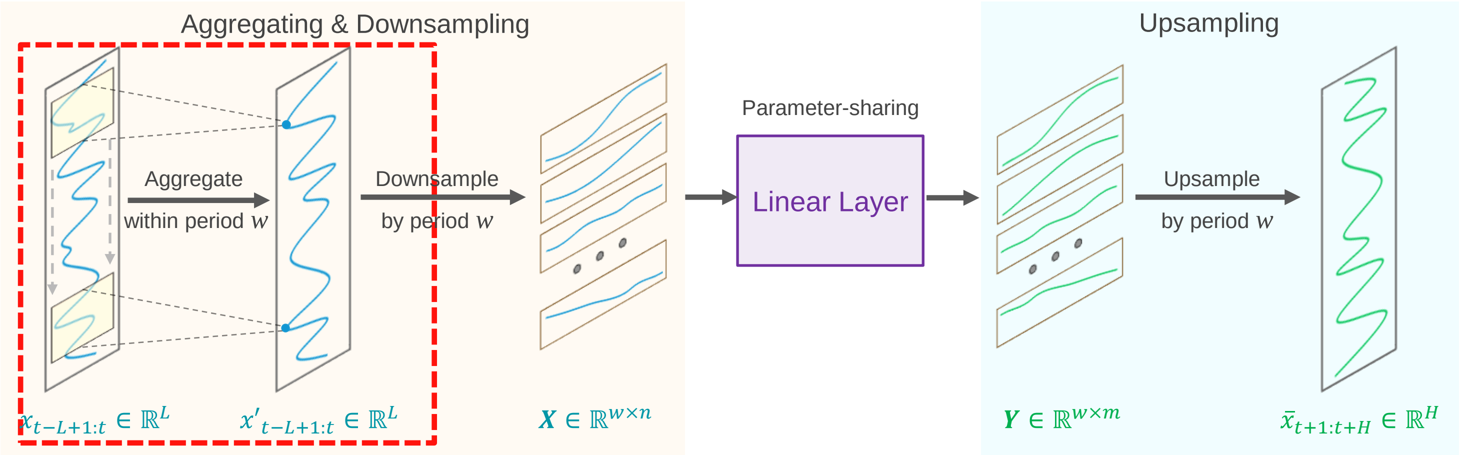
3. Method

4. Results

5. Inspirations

SparseTSF

(*Cross-Period Sparse Forecasting* technique with *Linear* backbone)



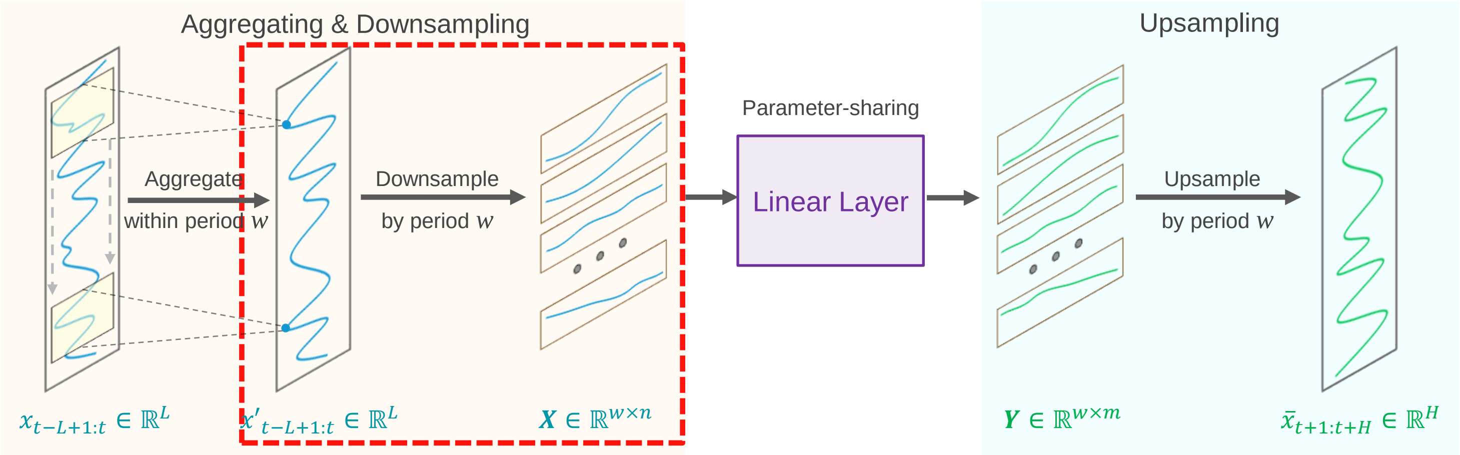
$$x'_{t-L+1:t} = x_{t-L+1:t} + \mathrm{Conv1D}\,(x_{t-L+1:t})$$

Aggregating information within period

Mitigating the impact of outliers

SparseTSF

(*Cross-Period Sparse Forecasting* technique with *Linear* backbone)



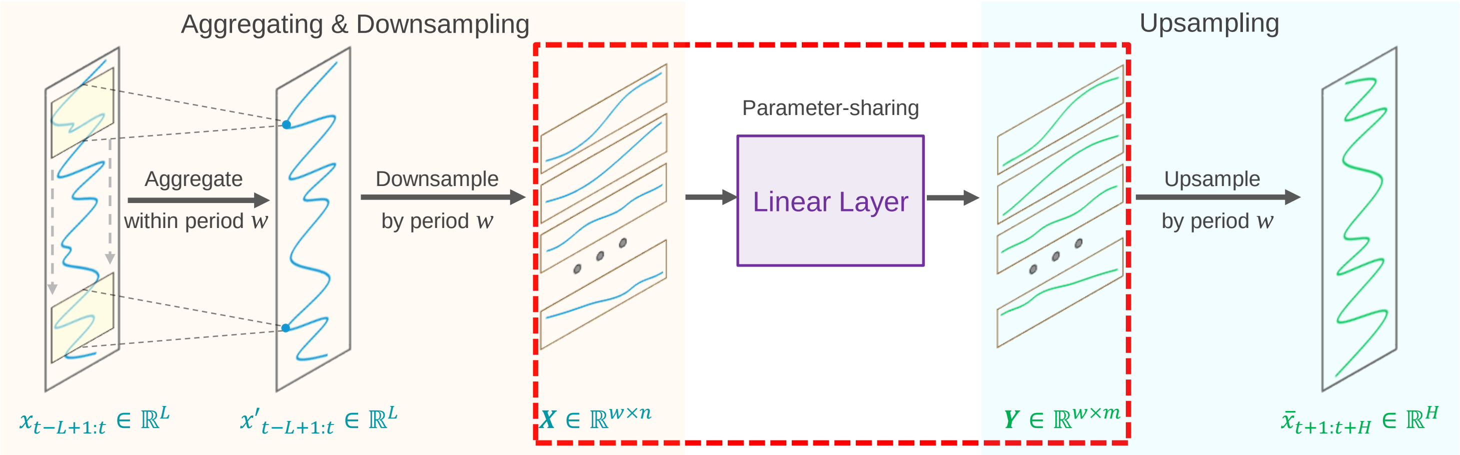
$$\boldsymbol{X} = \text{Downsample}\left(x'_{t-L+1:t}\right) \rightleftarrows \boldsymbol{X} = \text{Reshape}\left(x'_{t-L+1:t}, (n, w)\right)^{\top}$$

* Quickly implementing Downsampling through matrix Reshape and Transpose

1. Highlights
2. Motivations
3. Method
4. Results
5. Inspirations

SparseTSF

(*Cross-Period Sparse Forecasting* technique with *Linear* backbone)



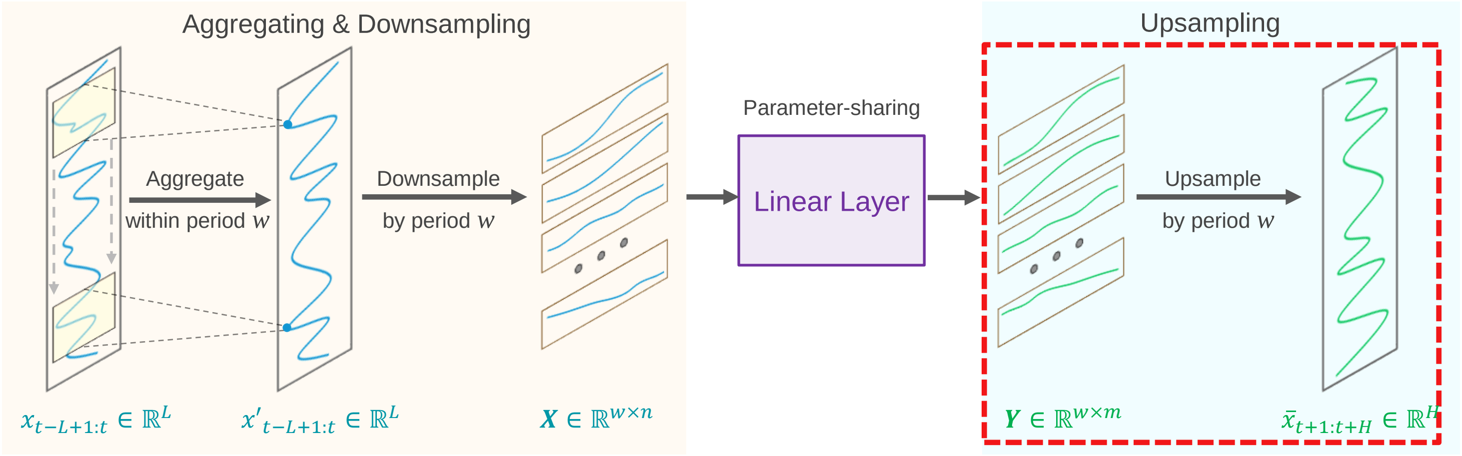
$$\boldsymbol{Y} = \text{Linear}\,(\boldsymbol{X})$$

Parameter-sharing for each subsequence, thus

requiring only $n \times m$ parameters

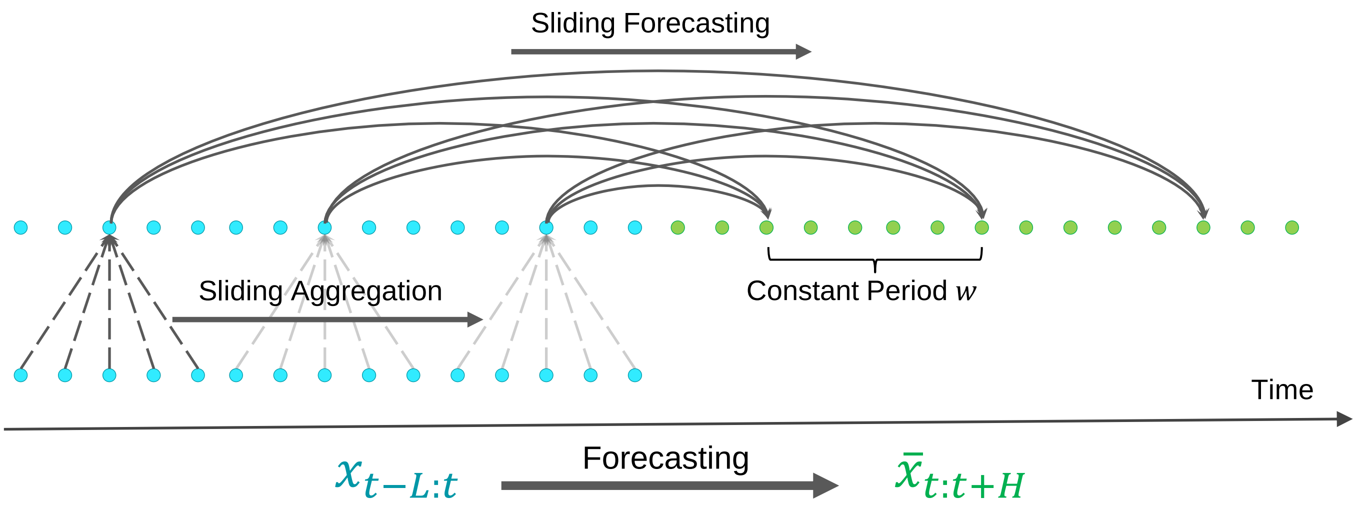
SparseTSF

(*Cross-Period Sparse Forecasting* technique with *Linear* backbone)



$$\bar{x}_{t+1:t+H} = \text{Upsample}\,(\boldsymbol{Y}) \rightleftarrows \bar{x}_{t+1:t+H} = \text{Reshape}\,(\boldsymbol{Y}^\top, (H))$$

* Quickly implementing Upsampling through matrix Reshape and Transpose



Intuitive Workflow: *Cross-Period Sparse Forecasting* on Time Axis

1. Highlights
2. Motivations
3. Method
4. Results
5. Inspirations

Dataset	ETTh1				ETTh2				Electricity				Traffic			
Horizon	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720
Informer (2021)	0.865	1.008	1.107	1.181	3.755	5.602	4.721	3.647	0.274	0.296	0.300	0.373	0.719	0.696	0.777	0.864
Autoformer (2021)	0.449	0.500	0.521	0.514	0.358	0.456	0.482	0.515	0.201	0.222	0.231	0.254	0.613	0.616	0.622	0.660
Pyraformer (2022b)	0.664	0.790	0.891	0.963	0.645	0.788	0.907	0.963	0.386	0.386	0.378	0.376	2.085	0.867	0.869	0.881
FEDformer (2022b)	0.376	0.420	0.459	0.506	0.346	0.429	0.496	0.463	0.193	0.201	0.214	0.246	0.587	0.604	0.621	0.626
FiLM (2022a)	0.371	0.414	0.442	0.465	0.284	0.357	0.377	0.439	0.154	0.164	0.188	0.236	0.416	0.408	0.425	0.520
TimesNet (2023)	0.384	0.436	0.491	0.521	0.340	0.402	0.452	0.462	0.168	0.184	0.198	0.220	0.593	0.617	0.629	0.640
PatchTST (2023)	**0.370**	0.413	**0.422**	0.447	**0.274**	0.341	**0.329**	0.379	**0.129**	**0.147**	**0.163**	**0.197**	**0.360**	**0.379**	**0.392**	**0.432**
DLinear (2023)	0.374	**0.405**	0.429	0.440	0.338	0.381	0.400	0.436	0.140	0.153	0.169	**0.203**	0.410	0.423	0.435	0.464
FITS (2024)	0.375	0.408	0.429	**0.427**	**0.274**	**0.333**	0.340	**0.374**	**0.138**	0.152	0.166	0.205	0.401	0.407	0.420	0.456
SparseTSF (ours)	**0.359** ±0.006	**0.397** ±0.002	**0.404** ±0.001	**0.417** ±0.001	**0.267** ±0.005	**0.314** ±0.003	**0.312** ±0.004	**0.370** ±0.001	**0.138** ±0.001	**0.146** ±0.001	**0.164** ±0.001	**0.203** ±0.001	**0.382** ±0.001	**0.388** ±0.001	**0.402** ±0.001	**0.445** ±0.002
Imp.	+0.011	+0.008	+0.018	+0.010	+0.007	+0.019	+0.017	+0.004	-0.009	+0.001	-0.001	-0.006	-0.022	-0.009	-0.010	-0.013

Comparable to State-of-the-Art with Less Than 1,000 Parameters

Model	Parameters	MACs	Max Mem.(MB)	Epoch Time(s)
Informer (2021)	12.53 M	3.97 G	969.7	70.1
Autoformer (2021)	12.22 M	4.41 G	2631.2	107.7
FEDformer (2022b)	17.98 M	4.41 G	1102.5	238.7
FiLM (2022a)	12.22 M	4.41 G	1773.9	78.3
PatchTST (2023)	6.31 M	11.21 G	10882.3	290.3
DLinear (2023)	485.3 K	156.0 M	123.8	25.4
FITS (2024)	10.5 K	79.9 M	496.7	35.0
SparseTSF (Ours)	**0.92 K**	**12.71 M**	125.2	31.3

- Required parameters:

$$\underbrace{\left\lfloor \frac{L}{w} \right\rfloor \times \left\lfloor \frac{H}{w} \right\rfloor}_{\text{Linear part}} + \underbrace{2 \times \left\lfloor \frac{w}{2} \right\rfloor}_{\text{Conv1D part}} + 1$$

- *1~2 orders of magnitude smaller* than FITS (another lightweight model for LTSF):

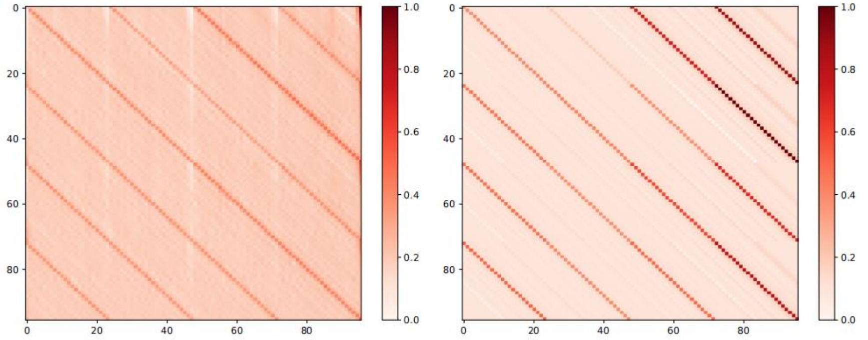
Model	SparseTSF (Ours)				FITS (2024)			
Horizon \ Look-back	96	192	336	720	96	192	336	720
96	41	57	81	145	840	1,218	2,091	5,913
192	57	89	137	265	1,260	1,624	2,542	6,643
336	81	137	221	445	1,890	2,233	3,280	7,665
720	145	265	445	925	3,570	3,857	5,125	10,512

Number of parameters

Equivalent weights of SparseTSF:

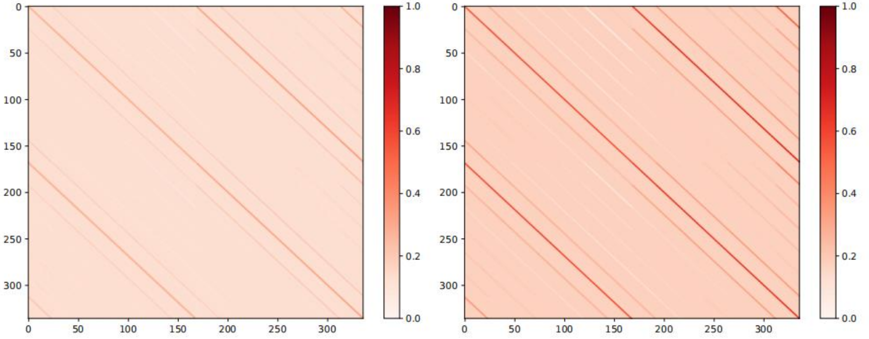
$$weight' = SparseTSF(\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix})^\top$$

- Linear model can *capture enough* **periodic patterns**
- SparseTSF *learns better* (i.e., **more distinct stripes**)
- SparseTSF *pays more attention to* **proper historical items**



(a) Linear (b) SparseTSF

ETTh1



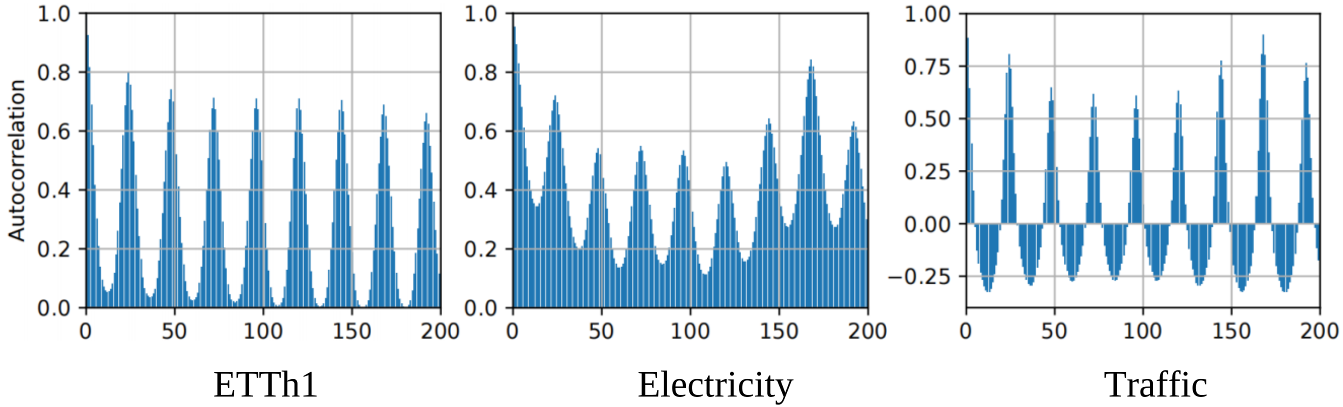
(a) Linear (b) SparseTSF

Traffic

Generalization Ability

Dataset	ETTh2 → ETTh1				Electricity → ETTh1			
Horizon	96	192	336	720	96	192	336	720
Informer (2021)	0.844	0.921	0.898	0.829	\	\	\	\
Autoformer (2021)	0.978	1.058	0.944	0.921	\	\	\	\
FEDformer (2022b)	0.878	0.927	0.939	0.967	\	\	\	\
FiLM (2022a)	0.876	0.904	0.919	0.925	\	\	\	\
PatchTST (2023)	0.449	0.478	0.482	0.476	0.400	0.424	0.475	0.472
DLinear (2023)	0.430	0.478	0.458	0.506	0.397	0.428	0.447	0.470
Fits (2024)	0.419	0.427	0.428	0.445	0.380	0.414	0.440	0.448
SparseTSF (Ours)	**0.370**	**0.401**	**0.412**	**0.419**	**0.373**	**0.409**	**0.433**	**0.439**

- On different datasets with **the *same* length of periodicity**
- SparseTSF demonstrates ***robust* generalization performance**
- Highly **beneficial for scenarios** with *small samples, or low-quality* data



- **Periodicity is fundamental** for *long-term time series forecasting*
- ***Avoid overestimating*** the *complexity of current datasets*
- We *advocate for* **simplifying model design**
- **Future research** should explore techniques to ***better leveraging the periodicity***

Thank You!

Poster location: Hall C 4-9 #309

Contact me: linss2000@foxmail.com

Paper

Code