Two-Stage Framework for Seasonal Time Series Forecasting

Qingyang Xu[1,2], Qingsong Wen[3], Liang Sun[3]

[1]MIT Operations Research Center

[2]MIT Laboratory for Financial Engineering

[3]Machine Intelligence Technology, DAMO Academy, Alibaba Group

Overview

- **Background: seasonal time series forecast and self-supervised learning**
- Two-stage framework for seasonal time series forecasting
- Prediction performance on M4 Hourly dataset and ablation study
- Conclusion and future work

Seasonal Time Series Forecasting

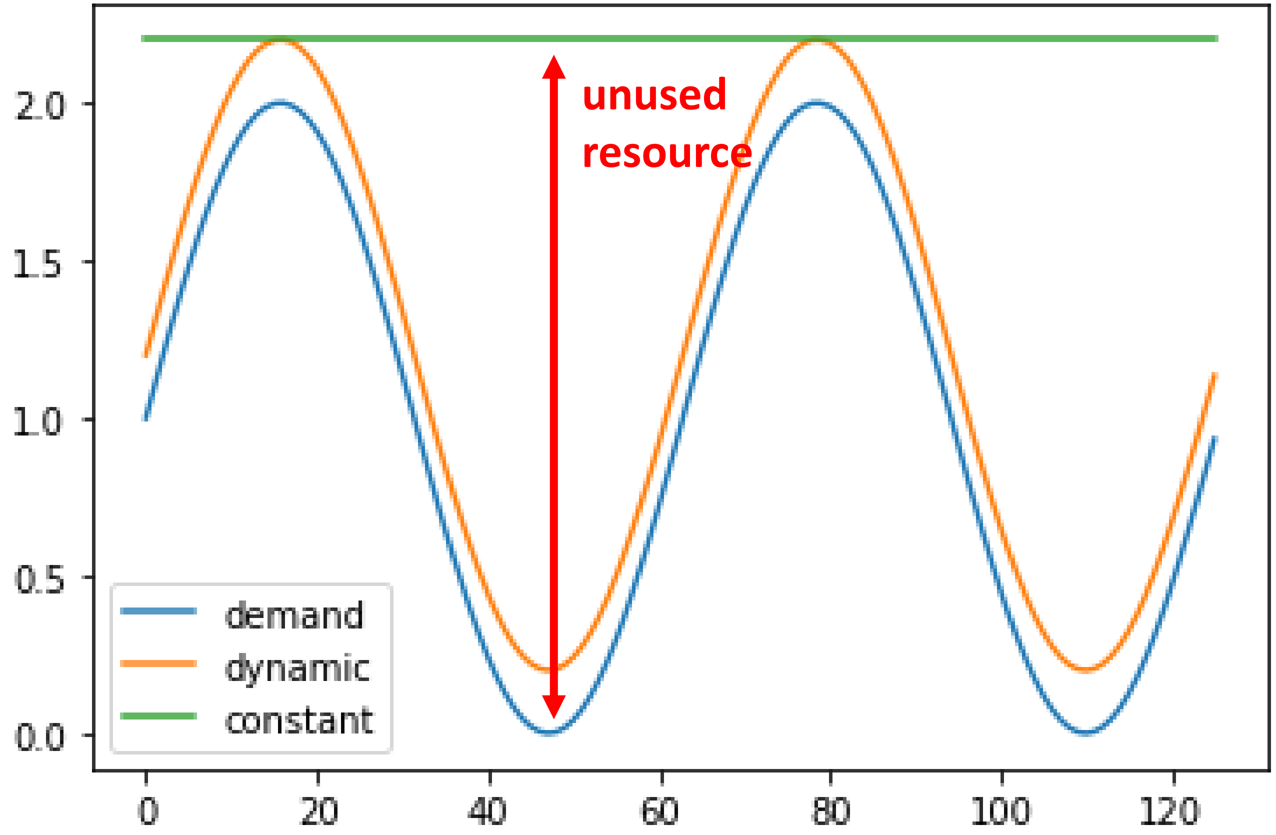
- **Auto-scaling** of cloud computing resources to minimize unused resource

- Goal: predict next horizon

  $x_f = (x_{t+1}, \dots, x_{t+h})$

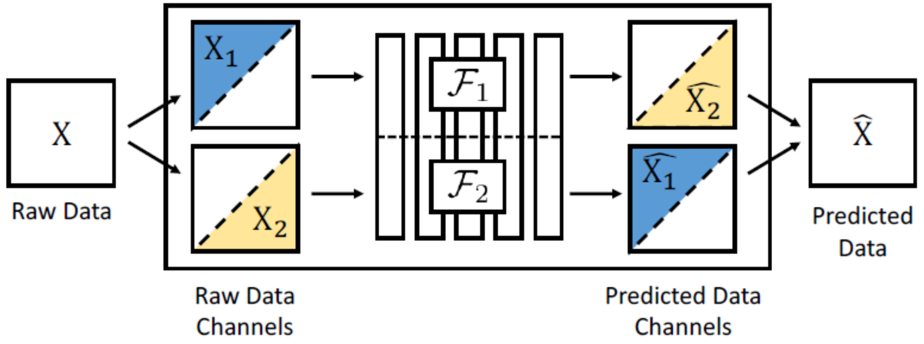
  from historical TS values

$x_{his} = (x_{t-L+1}, \dots, x_t)$



Self-supervised Learning (SSL)

- LeCun (ICML 2019): model learns to predict one part of its input data from other parts of its input
  - Split-brain autoencoder (Zhang et al. 2017)



- Two-stage procedure of general SSL
  - **Pre-training** – learn latent representation of data $z = \text{enc}(x)$
  - **Fine-tuning** – use latent representation for downstream tasks $y = f(z)$
- Advantages
  - Pre-train on **large unlabeled** data and fine-tune to tasks with **small labeled** data
  - Learn latent representation of data rather than memorizing outcomes

Zhang, R., Isola, P., & Efros, A. A. (2017). Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In Proceedings of the IEEE CVPR pp. 1058-1067, 2017.

Challenges of SSL on univariate seasonal TS

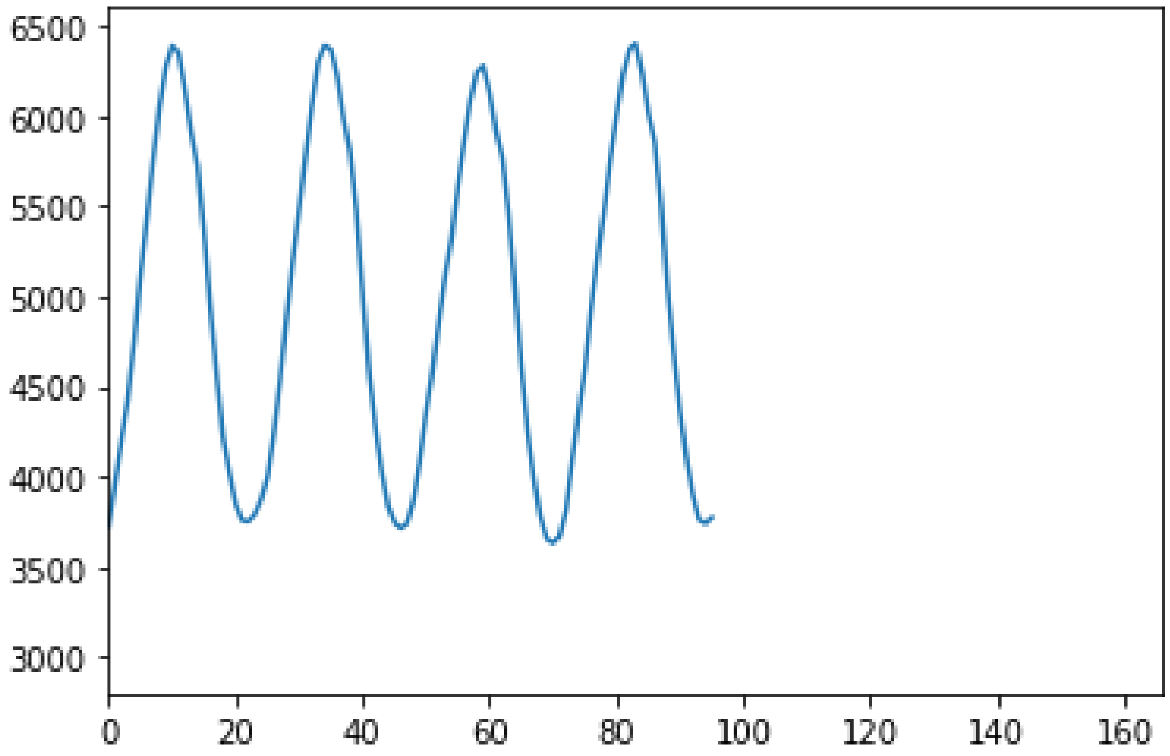
- Simple models (e.g., MLP, auto-regression) yield **strong baselines**
- Univariate time series $\rightarrow O(T)$ **with seasonality** $\rightarrow$ small data!
- Random masking in pre-training does not **utilize periodic structure**
- Black-box deep learning models **overfit** and are difficult to **interpret**

Overview

- Background: seasonal time series forecast and self-supervised learning
- **Two-stage framework for seasonal time series forecasting**
- Prediction performance on M4 Hourly dataset and ablation study
- Conclusion and future work

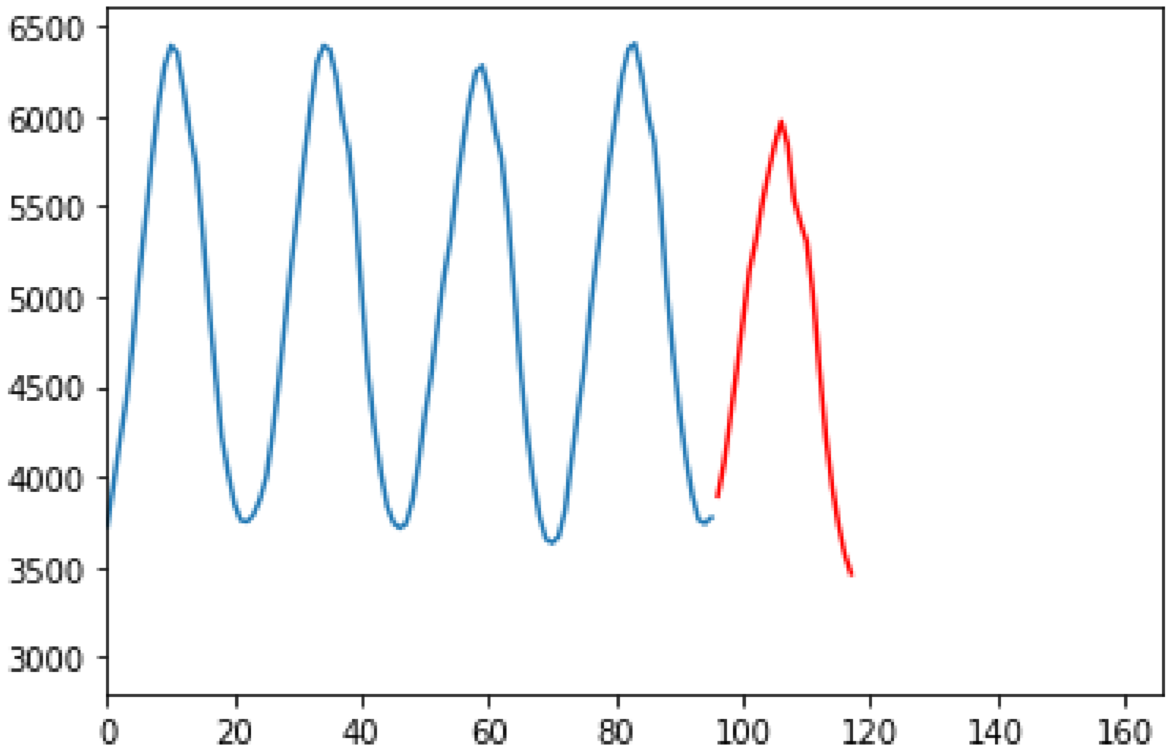
Motivation

- Traditionally: use **yesterday** to predict **today**



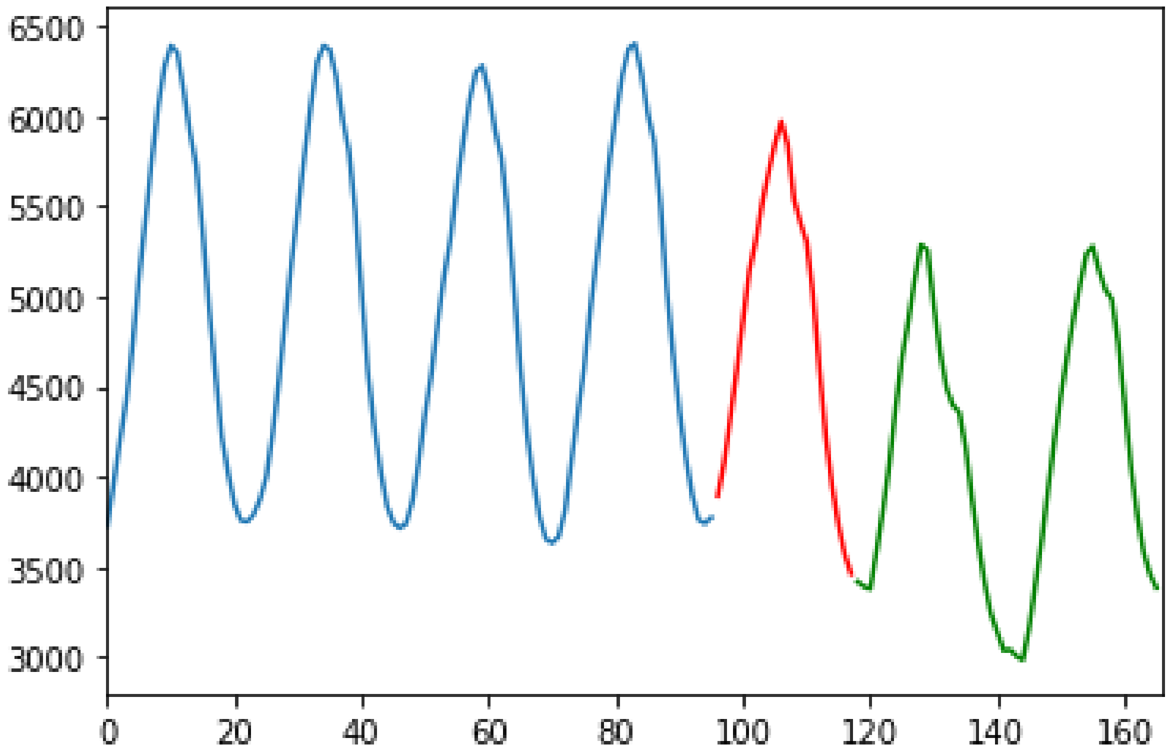
Motivation

- Traditionally: use **yesterday** to predict **today (surprise!)**



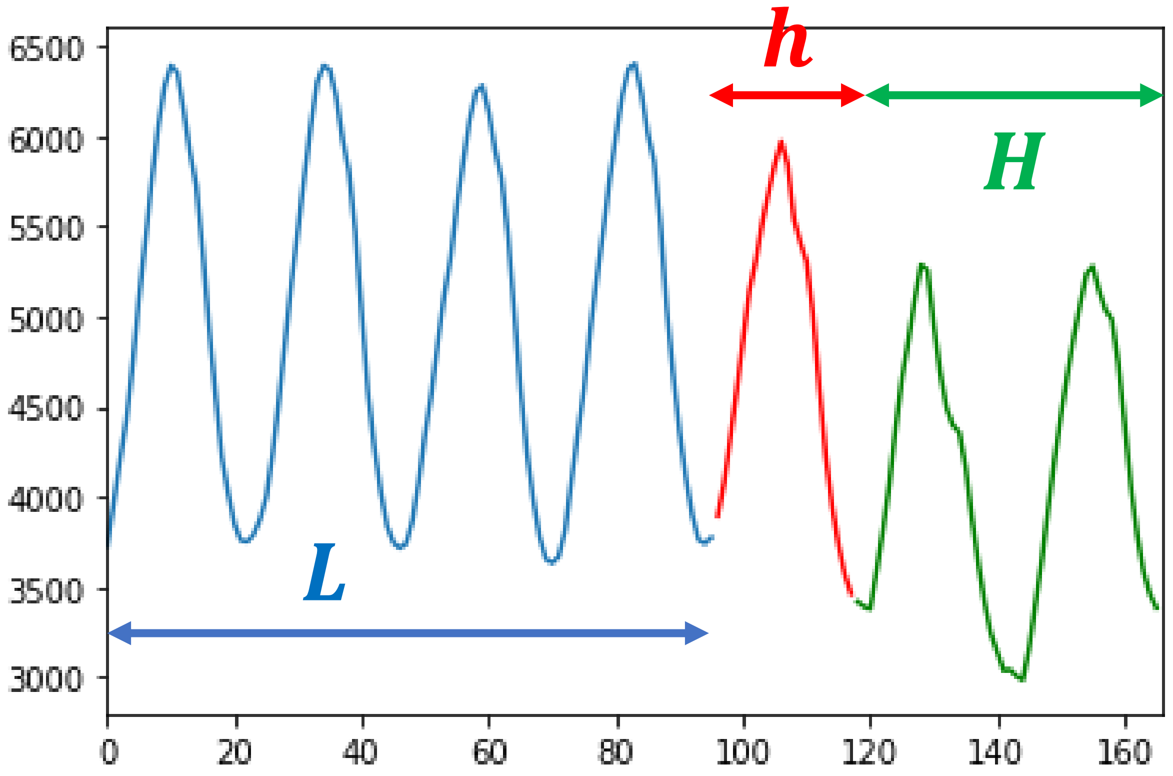
Motivation

- Question: what if we also know **tomorrow**?



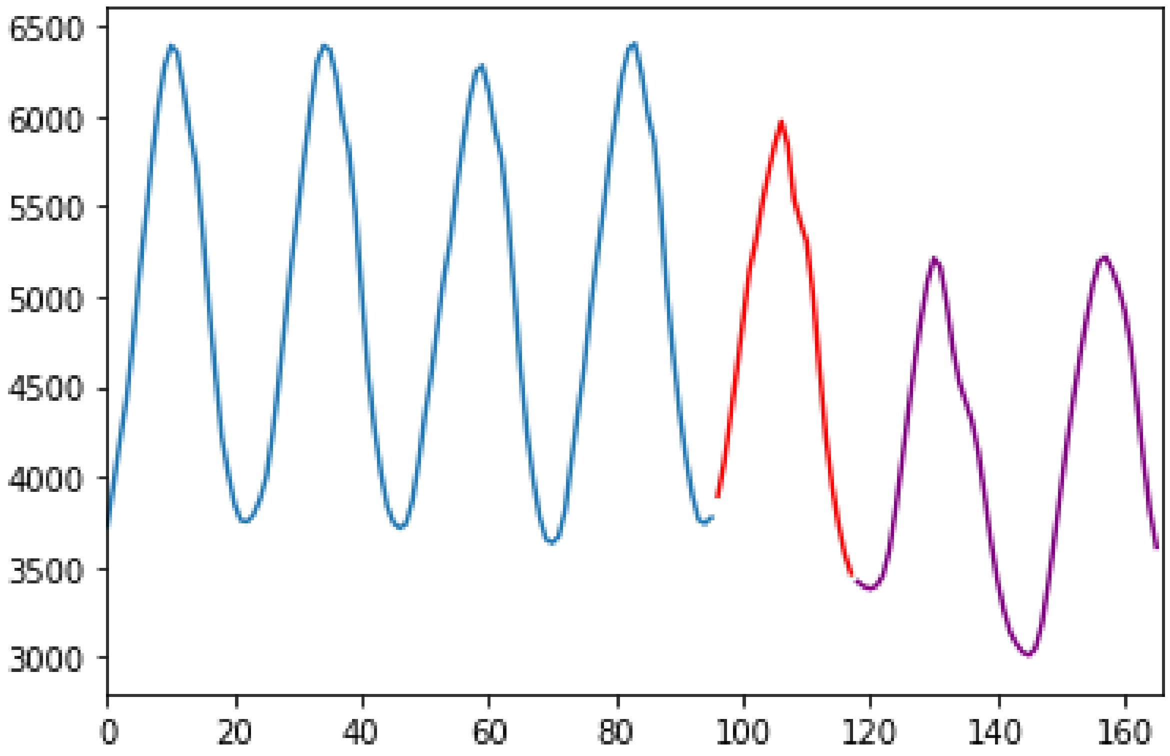
Motivation

- Question: what if we also know **tomorrow**?



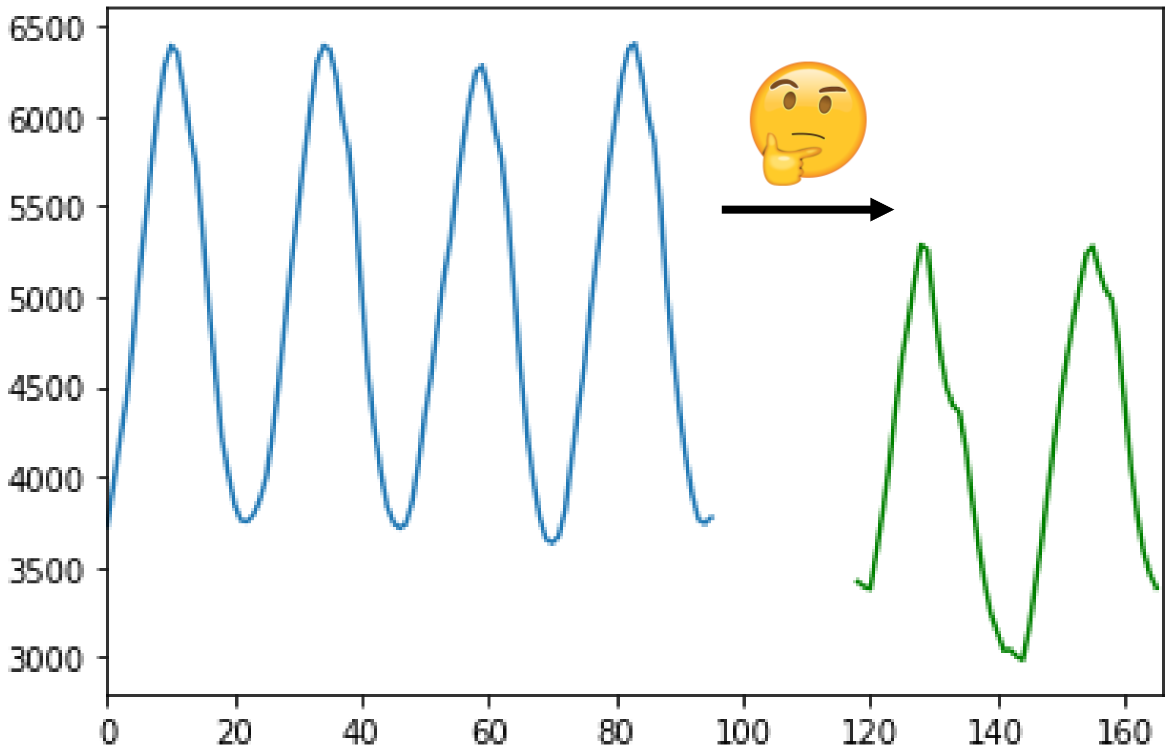
Motivation

- Idea: what if we first *predict* **tomorrow**?



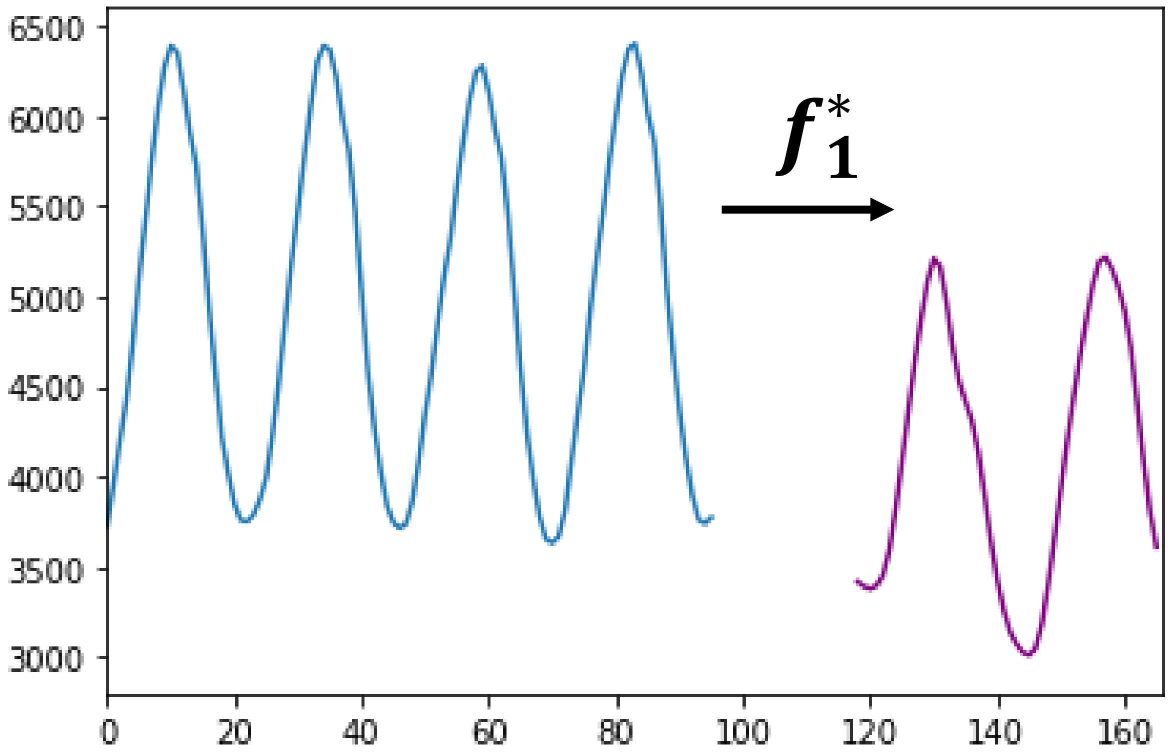
Stage 1

- Predict **tomorrow** with **yesterday**



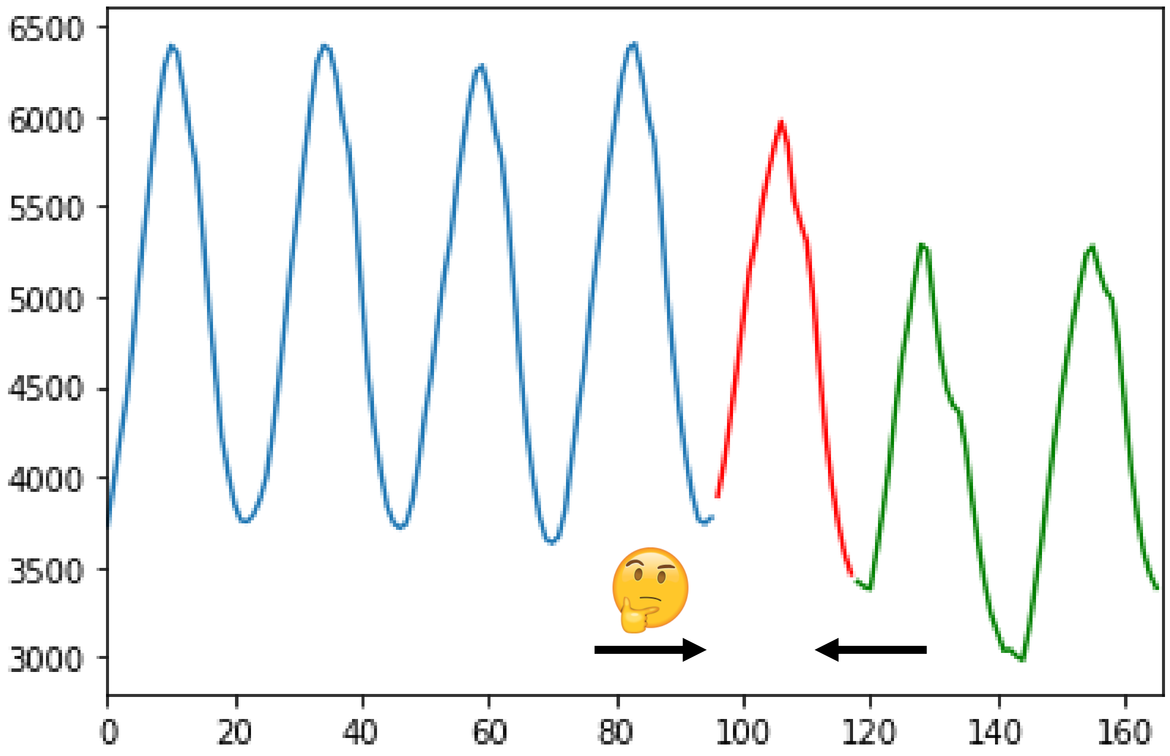
Stage 1

- Predict **tomorrow** with **yesterday**



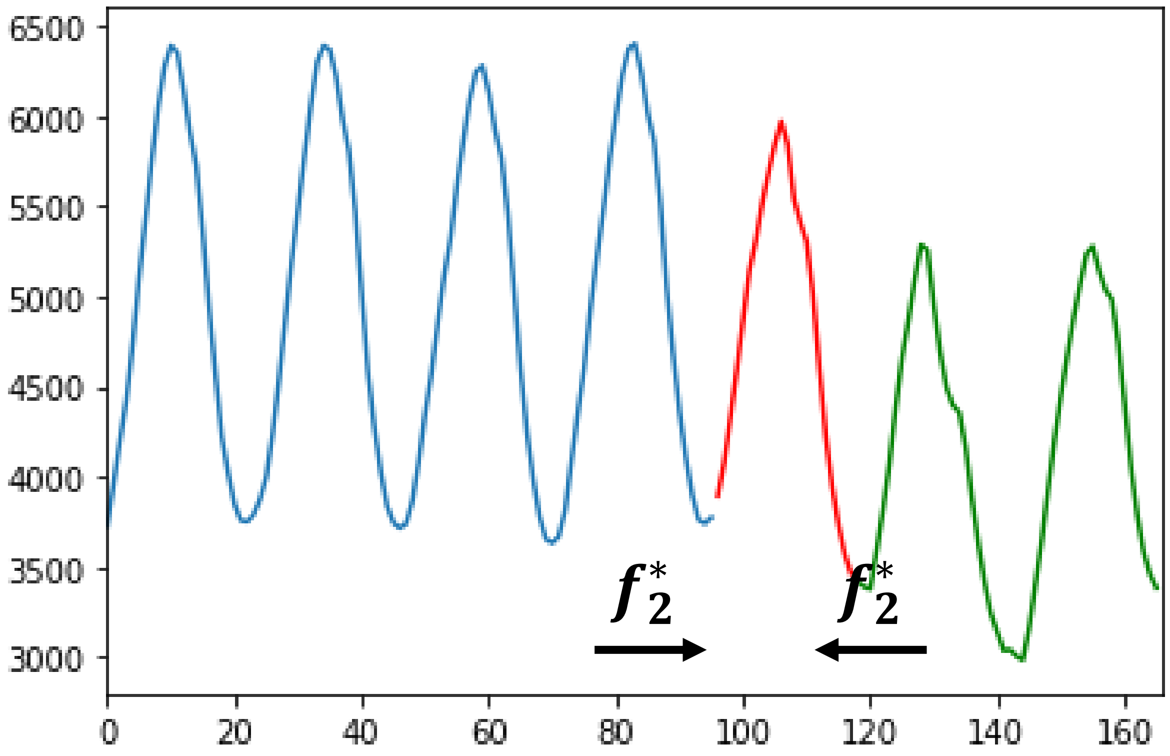
Stage 2 - Training

- Predict **today** with **yesterday** and (true) **tomorrow**



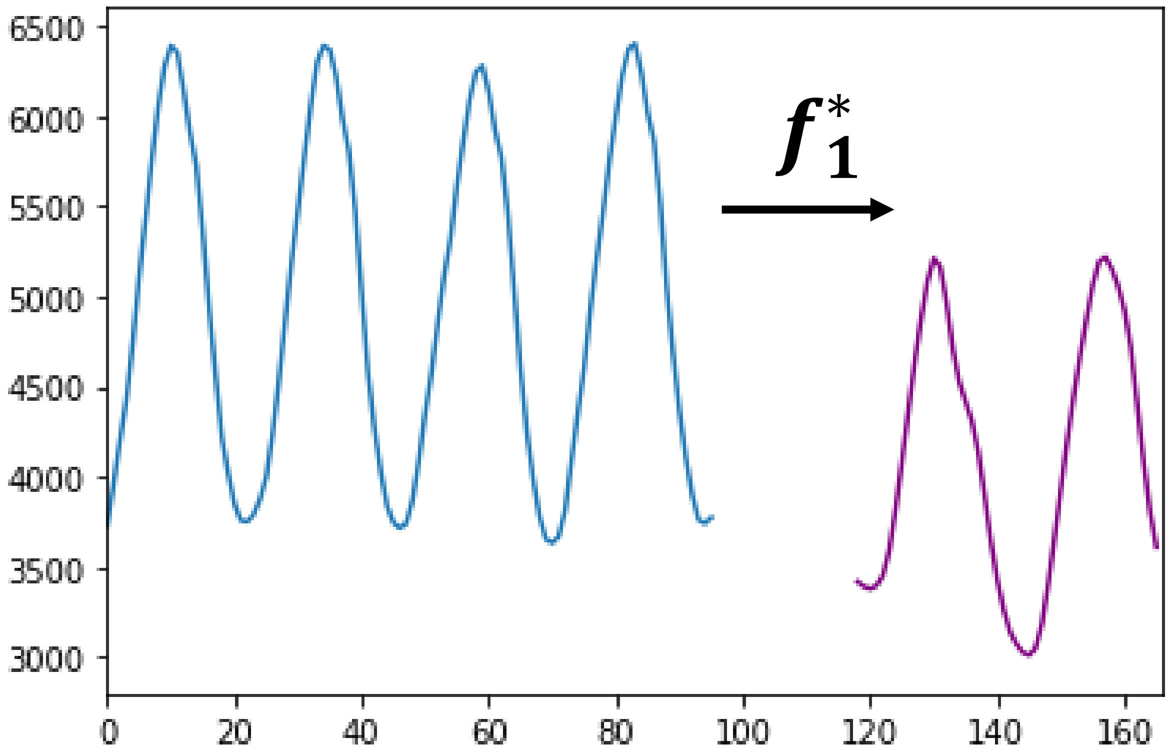
Stage 2 - Training

- Predict **today** with **yesterday** and (true) **tomorrow**



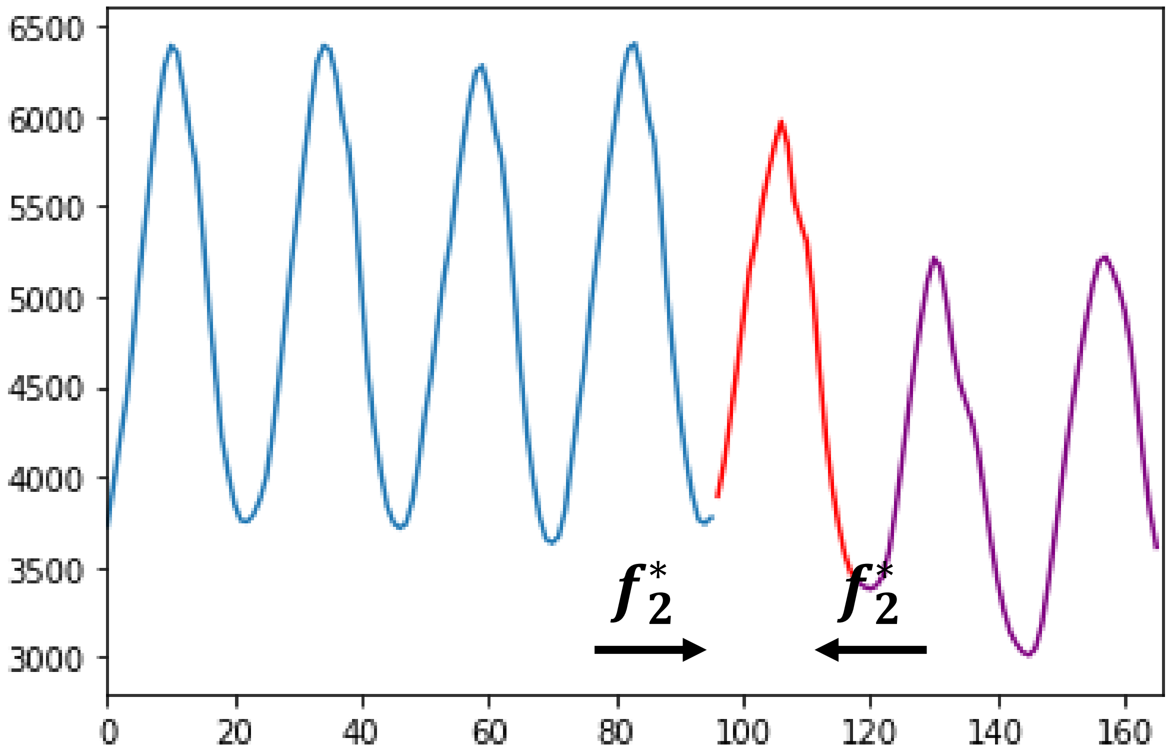
Stage 2 - Prediction

- Predict **tomorrow** with **yesterday**

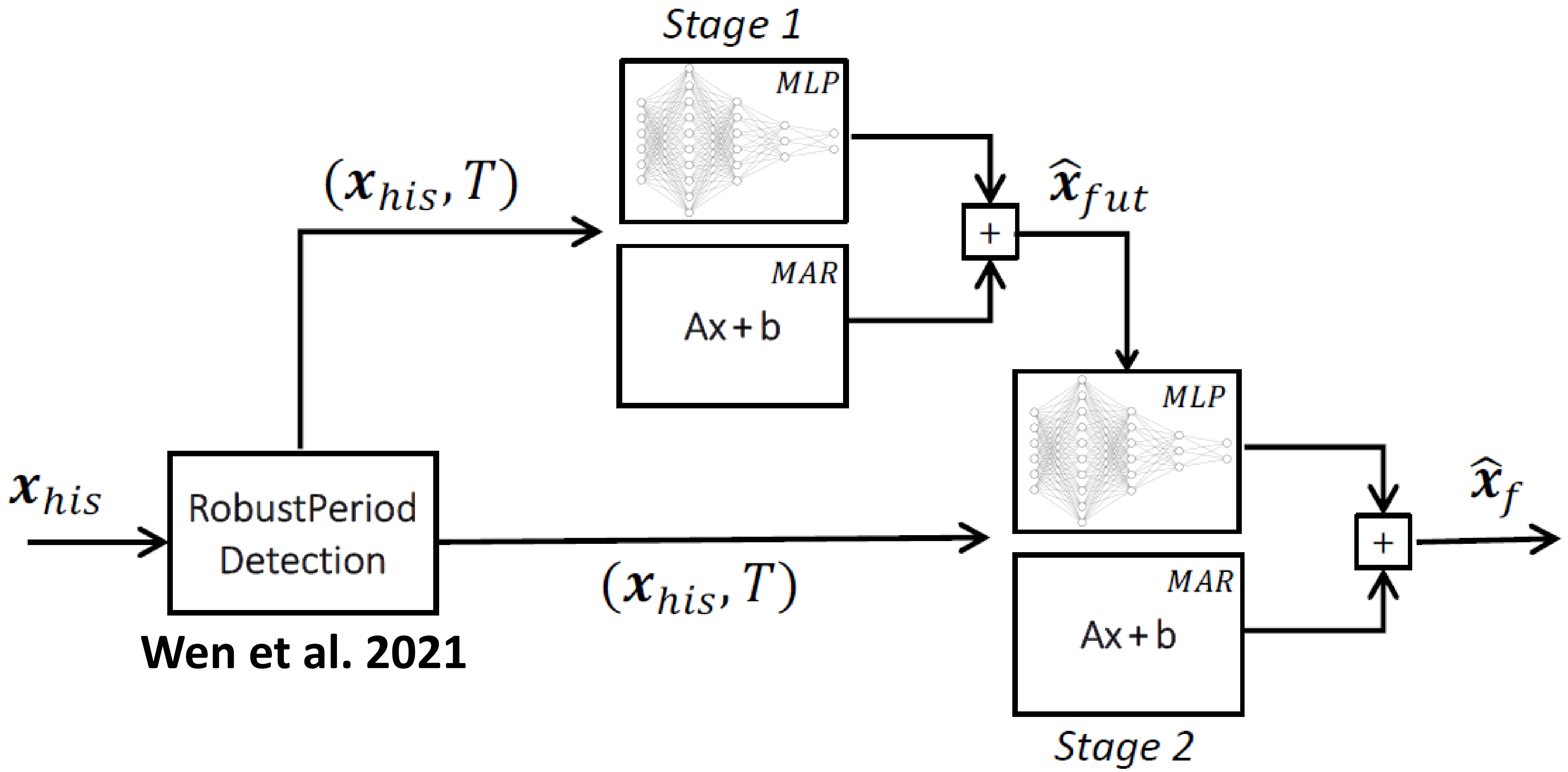


Stage 2 - Prediction $\hat{\mathbf{x}}_f = f_2^*(\mathbf{x}_{his}, \hat{\mathbf{x}}_{fut}) = f_2^*(\mathbf{x}_{his}, f_1^*(\mathbf{x}_{his}))$

- Predict **today** with **yesterday** and (predicted) **tomorrow**



Schematic of 2S Framework

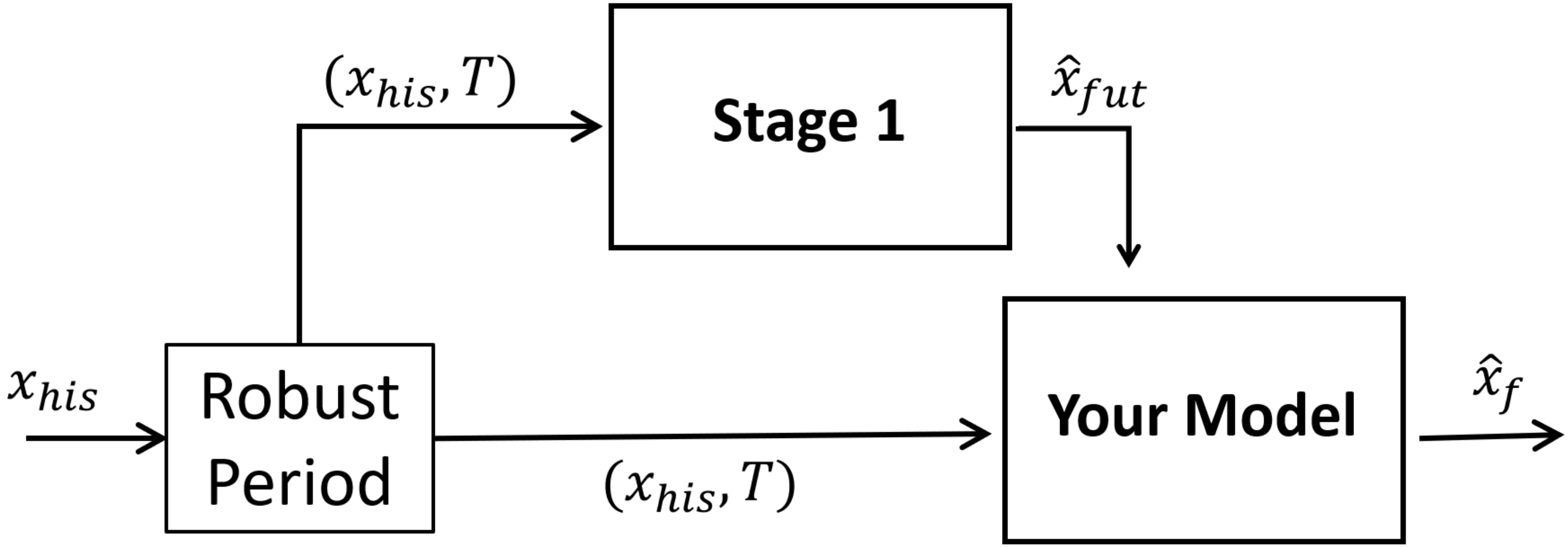


Wen, Q., He, K., Sun, L., Zhang, Y., Ke, M., & Xu, H. (2021). RobustPeriod: Time-frequency mining for robust multiple periodicity detection. *in Proc. SIGMOD, 2021.*

Advantages of 2S on univariate seasonal TS

- Simple models (e.g., MLP, auto-regression) yield **strong baselines**
  - Reliable predictions for Stage 1 (yesterday → tomorrow) models
- Univariate time series → $O(T)$ **with seasonality** → small data!
  - Use **both past and future** to predict the present
- Black-box deep learning models **overfit** and are difficult to **interpret**
  - Stage 1 (yesterday → tomorrow) outcomes are transparent and easy to evaluate
- Random masking in pre-training does not **utilize periodic structure**
  - Use seasonality to choose the future horizon length H

2S as Model Augmentation Technique



Wen et al. 2021

Wen, Q., He, K., Sun, L., Zhang, Y., Ke, M., & Xu, H. (2021). RobustPeriod: Time-frequency mining for robust multiple periodicity detection. *in Proc. SIGMOD, 2021.*

Overview

- Background: seasonal time series forecast and self-supervised learning
- Two-stage framework for seasonal time series forecasting
- **Prediction performance on M4 Hourly dataset and ablation study**
- Conclusion and future work

Performance Metrics

- x_t true time series values
- $\hat{x}_t$ predicted values
- Also compute each metrics after removing the **worst 5%** predictions (e.g., MAPE-95) to reduce the effect of outliers

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|x_t - \hat{x}_t|}{|x_t|}$$

$$\text{RMSPE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{|x_t - \hat{x}_t|^2}{|x_t|^2}}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}|x_t - \hat{x}_t|^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|x_t - \hat{x}_t|$$

*https://mofc.unic.ac.cy/m4/

Performance on M4 Competition* Hourly data

- 414 seasonal TS with 700 or 960 time points. $h = 12, H = 12, L = 72$

Model	MAPE	MAPE-95	RMSPE	RMSPE-95	RMSE	RMSE-95	MAE	MAE-95
Two-Stage	1.399	0.346	11.629	0.562	0.305	0.232	0.214	0.179
MLP+MAR	1.417	0.379	11.058	0.610	0.330	0.255	0.235	0.199
MLP	1.423	0.410	11.197	0.605	0.385	0.305	0.281	0.242
Deep-LSTM	1.539	0.459	11.857	0.652	0.422	0.341	0.320	0.278
MAR	1.551	0.416	12.275	0.672	0.349	0.275	0.253	0.216
RESTFul	1.642	0.451	11.808	0.721	0.375	0.301	0.276	0.238
PrevPeriod	1.776	0.435	14.365	0.733	0.391	0.292	0.263	0.217
SPAR-h12	2.077	0.570	15.825	0.869	0.447	0.364	0.340	0.297

Enhance existing model performance

- Augment Stage 1 outputs $f_1^*(x_{his})$ as inputs to baseline (BL) model

h	Model	MAPE	MAPE-95	RMSPE	RMSPE-95	RMSE	RMSE-95	MAE	MAE-95
6	BL	1.265	0.332	10.008	0.542	0.285	0.219	0.201	0.169
	2S	1.214	0.305	10.102	0.496	0.266	0.201	0.185	0.155
12	BL	1.454	0.384	11.423	0.617	0.331	0.258	0.237	0.201
	2S	1.399	0.346	11.629	0.562	0.305	0.232	0.214	0.179
24	BL	1.511	0.405	11.833	0.651	0.349	0.273	0.251	0.214
	2S	1.489	0.374	11.900	0.614	0.319	0.247	0.226	0.191

Optimizing future horizon length $\boldsymbol{H}^*$

- Larger H $\rightarrow$ more structure, harder to learn. **S1 MSE** is good diagnostic.

H	MAPE	MAPE-95	RMSPE	RMSPE-95	RMSE	RMSE-95	MAE	MAE-95	S1 MSE
0	1.454	0.384	11.423	0.617	0.331	0.258	0.237	0.201	N/A
6	1.417	0.351	11.715	0.570	0.308	0.235	0.216	0.182	0.151
12	1.399	0.346	11.629	0.562	0.305	0.232	0.214	0.179	0.147
18	1.462	0.356	12.052	0.582	0.309	0.237	0.218	0.183	0.158
24	1.470	0.356	11.985	0.585	0.306	0.234	0.215	0.180	0.166

Overview

- Background: seasonal time series forecast and self-supervised learning
- Two-stage framework for seasonal time series forecasting
- Prediction performance on M4 Hourly dataset and ablation study
- **Conclusion and future work**

Conclusion & Future Work

- Two-stage framework (motivated by SSL) achieves state-of-the-art performance of horizon forecast on M4 Hourly data
  - Address the challenges of applying SSL to univariate seasonal TS forecasting
  - Enhance prediction accuracy of baseline models
  - Important to optimize the future horizon length $\boldsymbol{H}^*$
- Future work
  - Joint training for Stage 1 and Stage 2 parameters
  - Explore combinations of Stage 1 and Stage 2 models

Thanks!
Q&A