

Towards Diverse and Coherent Augmentation for Time-Series Forecasting

Xiyuan Zhang¹, Ranak Roy Chowdhury¹, Jingbo Shang¹,
Rajesh Gupta¹, Dezhi Hong²

¹UC San Diego, ²Amazon



Background and Motivation

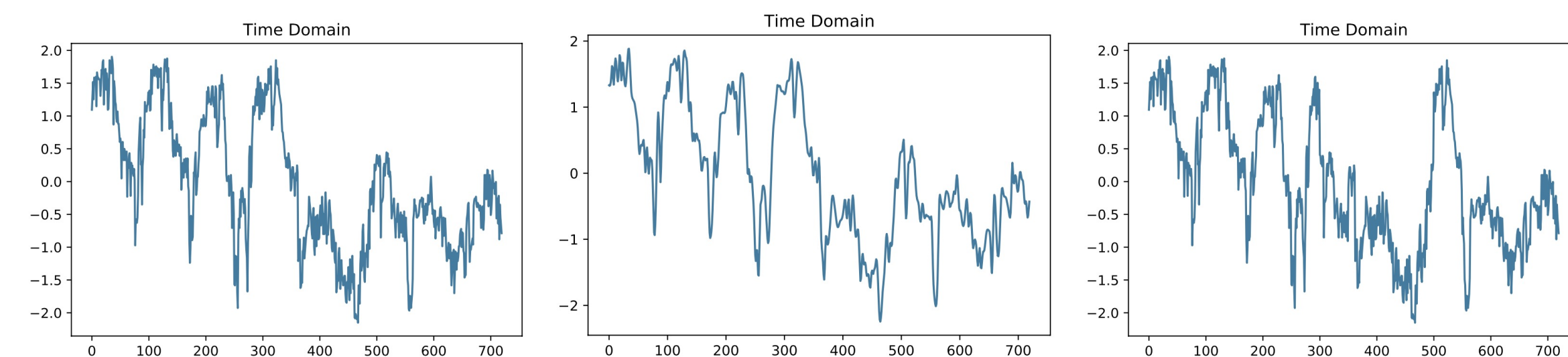
Data scarcity in time series is very common

- Malfunctioning sensors: battery depletion
- Privacy concerns: medical data
- Cold-start prediction: new station, new store



Data augmentation mitigates the data scarcity issue

✗ However, existing augmentation methods are designed for classification, where class labels can be preserved even if augmentation changes the temporal dynamics



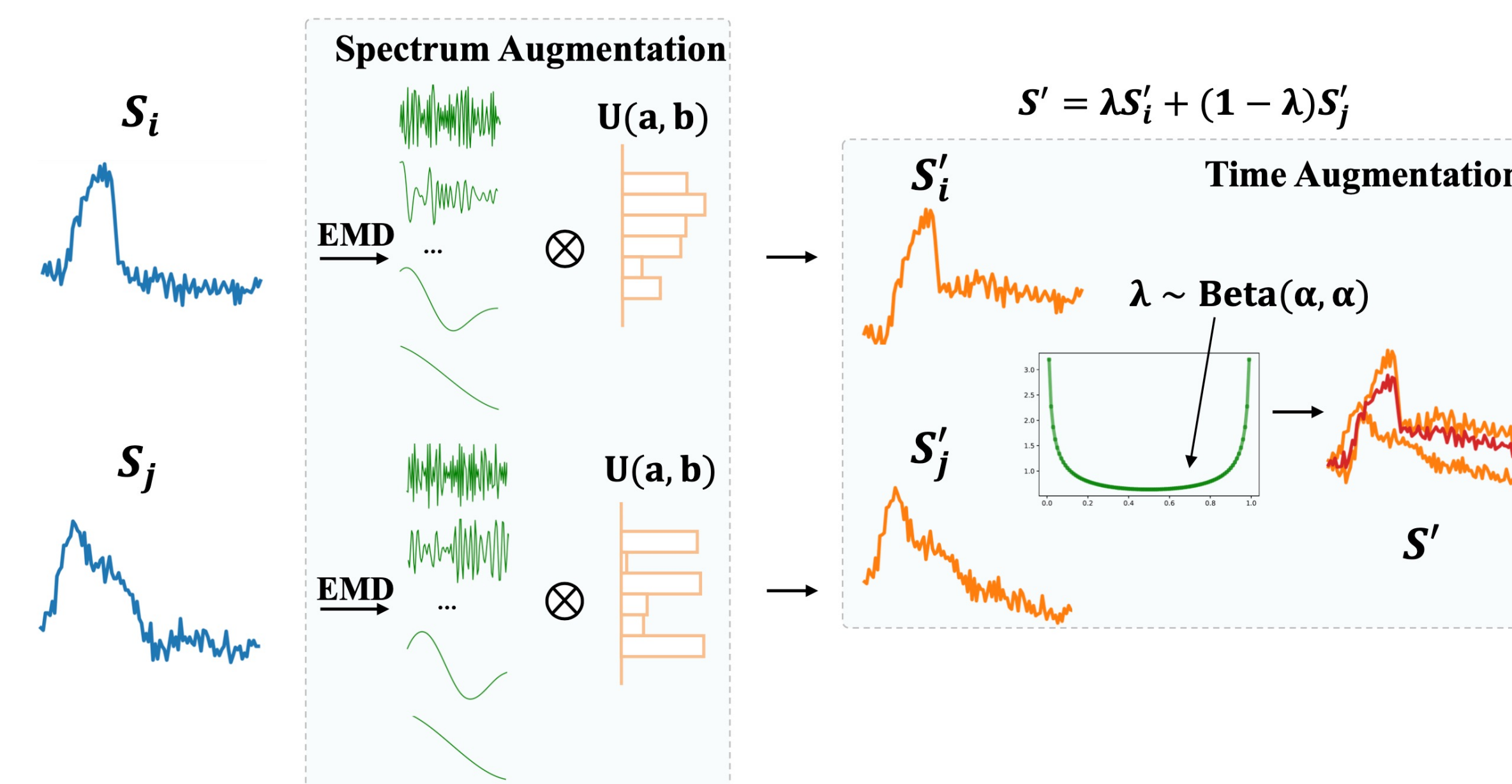
original filtering permutation

✓ Augmentation for forecasting requires **diversity** as well as **coherence** with the temporal dynamics

💡 Time-series data generated by real-life physical processes have characteristics in both **time** and **frequency** domains to capture temporal dynamics

Augmentation Method

Combine frequency-domain augmentation and time-domain augmentation



Frequency-domain augmentation: decompose time series with Empirical Mode Decomposition (EMD), and reassemble subcomponents with random weights sampled from $U(a, b)$

$$S = \sum_{i=1}^n \text{IMF}_i + R \quad S' = \sum_{i=1}^n w_i \cdot \text{IMF}_i$$

Diversity: random weights to combine components

Coherence: augmented samples have the same set of base components

Time-domain augmentation: mix up values at both history and future time steps with $\lambda \sim \text{Beta}(\alpha, \alpha)$

$$H' = \lambda H_i + (1 - \lambda) H_j \quad F' = \lambda F_i + (1 - \lambda) F_j$$

Diversity: random weights to combine series pairs

Coherence: only augment linearly in-between samples by the interpolation nature of “mix up”

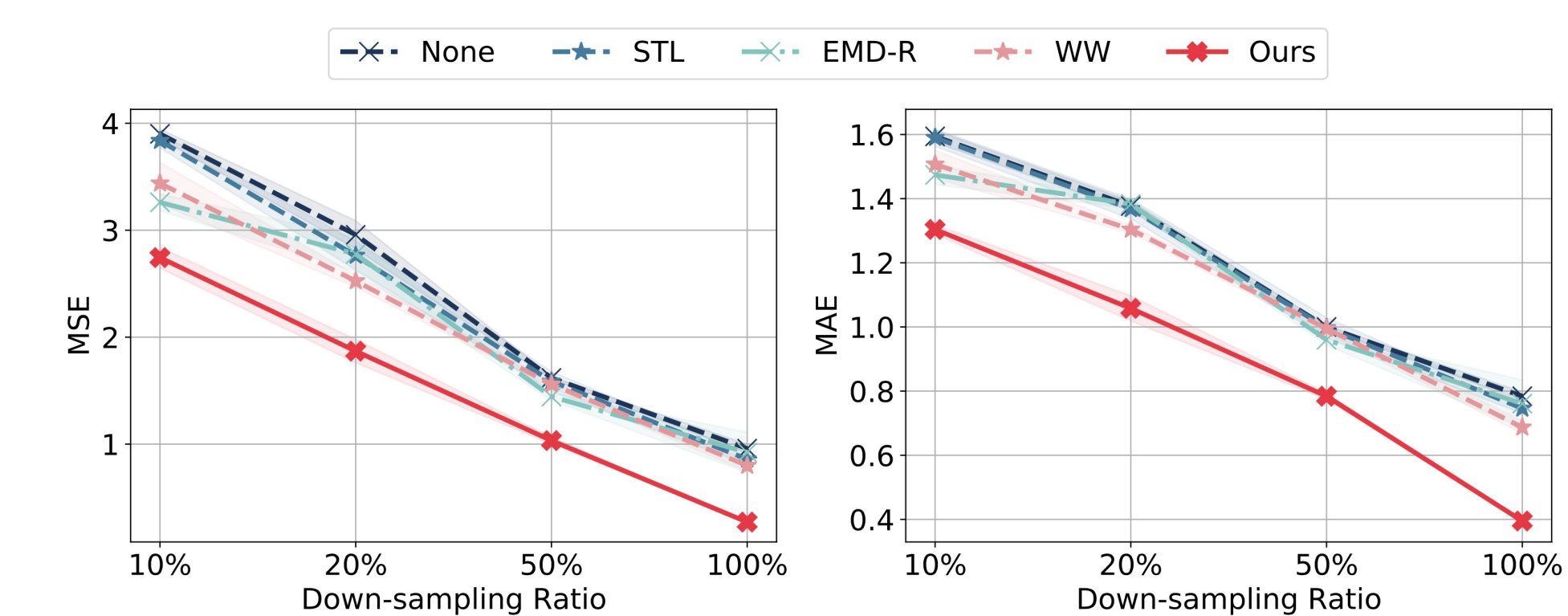
Experiments

Forecasting backbone: Informer (AAAI' 2021 best paper)

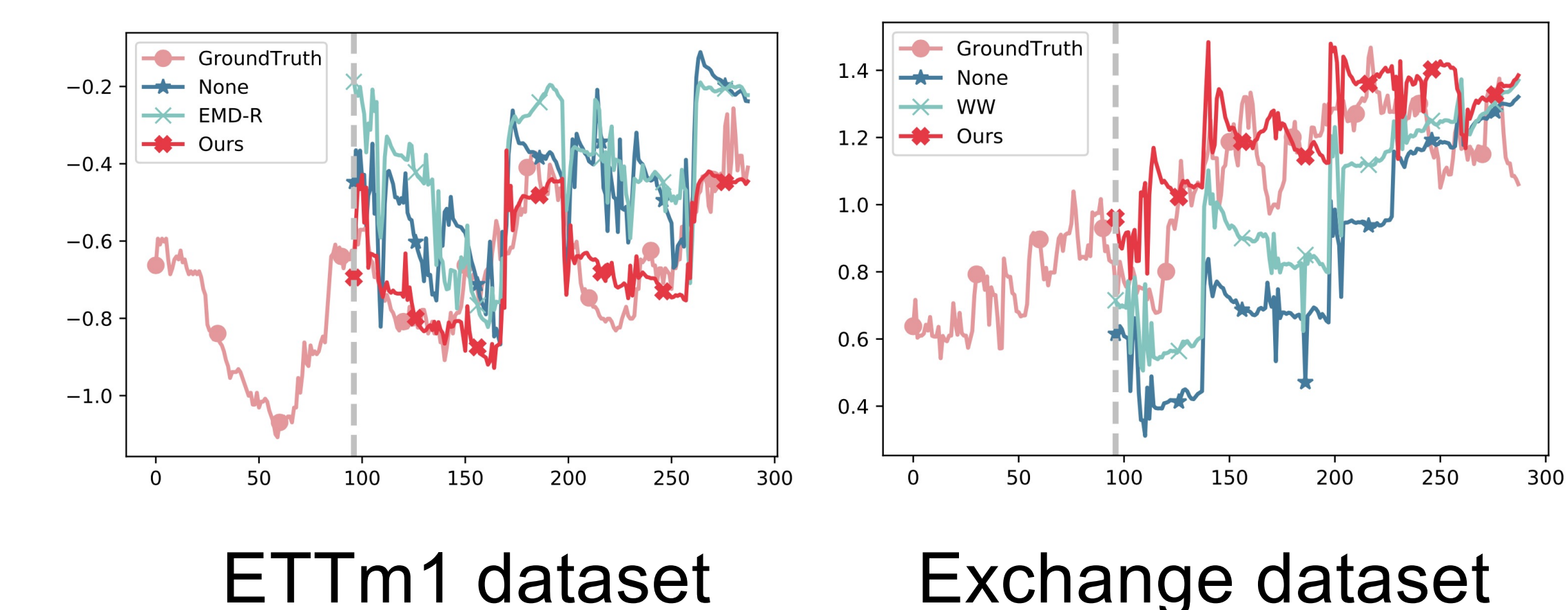
Methods	None		WW		RobustTAD		STL		EMD-R		GAN		DBA		STAug-onTime		STAug-onFreq		STAug		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTm1	96	0.947	0.760	0.894	0.730	1.011	0.789	0.910	0.738	0.838	0.686	0.904	0.734	0.943	0.759	0.655	0.573	0.841	0.693	0.648	0.572
192	0.977	0.767	0.899	0.735	0.946	0.744	0.942	0.740	0.842	0.729	0.985	0.757	0.877	0.760	0.754	0.616	0.862	0.735	0.772	0.626	0.526
336	1.112	0.833	1.058	0.808	1.094	0.816	1.086	0.814	1.056	0.804	1.070	0.816	1.111	0.832	0.860	0.681	1.103	0.829	0.839	0.705	0.589
720	1.187	0.864	1.201	0.804	1.151	0.810	1.085	0.805	1.056	0.802	1.221	0.892	1.104	0.866	0.991	0.723	1.151	0.837	0.772	0.746	0.646
ETTm2	96	1.154	0.819	0.997	0.744	0.843	0.759	0.992	0.733	0.809	0.695	0.862	0.725	1.146	0.836	0.623	0.521	0.900	0.715	0.627	0.559
192	1.128	0.808	1.077	0.786	1.043	0.789	1.034	0.781	1.006	0.785	1.225	0.875	1.113	0.836	0.623	0.617	0.914	0.720	0.744	0.626	0.528
336	1.515	0.946	1.425	0.923	1.390	0.909	1.402	0.908	1.287	0.875	1.453	0.929	1.368	0.962	0.943	0.735	1.287	0.876	0.811	0.692	0.597
720	1.462	1.423	4.370	1.607	4.392	1.654	3.271	1.778	3.247	1.553	3.129	1.355	3.473	1.427	2.646	1.281	3.365	1.509	2.780	1.508	1.036
Exchange	96	0.959	0.784	0.799	0.687	0.962	0.788	0.860	0.746	0.923	0.761	0.942	0.772	0.958	0.784	0.255	0.402	0.809	0.708	0.271	0.305
192	1.113	0.837	1.084	0.787	1.180	0.873	1.143	0.852	1.182	0.853	1.115	0.855	1.109	0.836	0.627	0.615	1.125	0.822	0.606	0.605	0.497
336	1.605	1.009	1.697	0.995	1.562	1.006	1.548	0.987	1.606	1.001	1.465	0.972	1.395	1.007	1.007	0.807	1.438	0.944	0.935	0.779	0.597
720	2.847	1.380	3.198	1.481	2.816	1.380	2.484	1.297	2.914	1.413	2.571	1.306	2.847	1.980	1.489	0.928	2.056	1.153	1.271	1.053	0.628

Average 28.2% MSE reduction and 18.0% MAE reduction compared with the best baseline for each of the five datasets

Our method is the most robust with respect to down-sampling training data



Case study: prediction better aligns with ground truth



ETTm1 dataset

Exchange dataset