DDN: Dual-domain Dynamic Normalization for Non-stationary Time Series Forecasting

NEURAL INFORMATION PROCESSING SYSTEMS

Code Paper

Tao Dai[1,], Beiliang Wu[1,*], Peiyuan Liu[2,†], Naiqi Li[2,†], Yuerong Xue[2], Shu-tao Xia[2], Zexuan Zhu[1]*

1. College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China

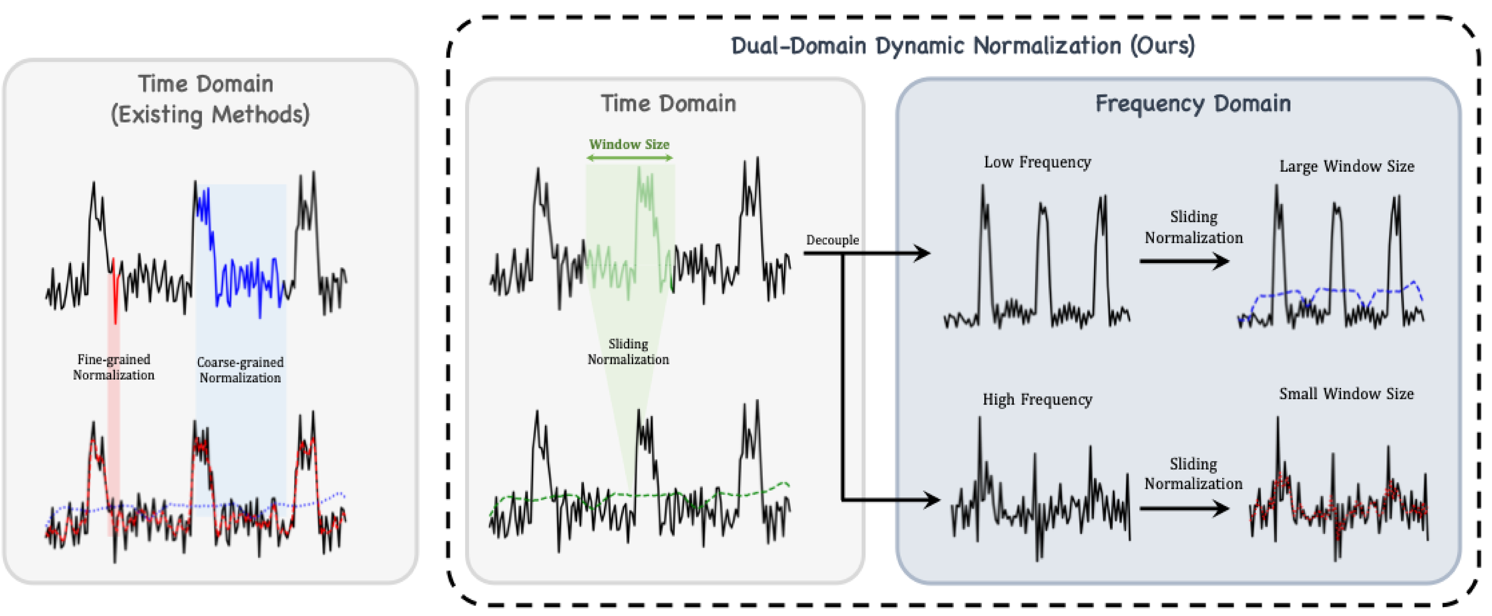
2. Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

* Equal Contributions † Corresponding Author

深圳大学 SHENZHEN UNIVERSITY

清華大學 Tsinghua University

Motivation

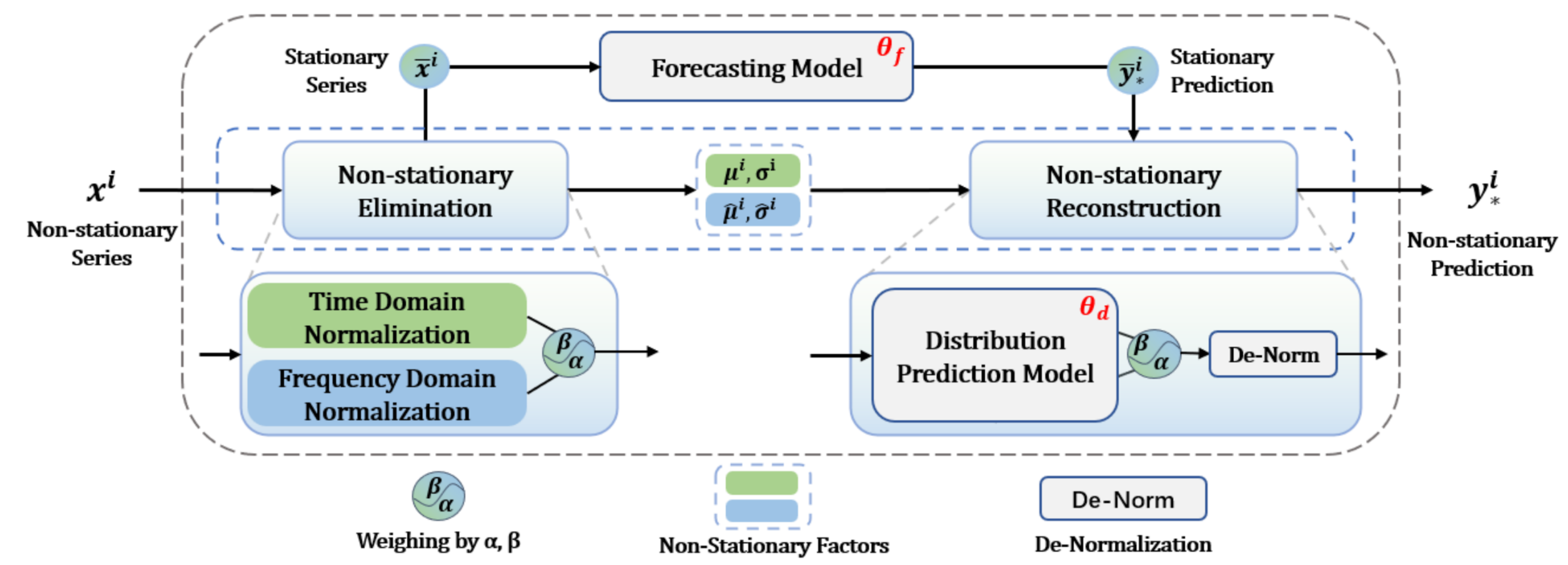


- ➢ Reliable forecasting under rapidly changing and non-stationary data distributions remains a core challenge in time series analysis.
- ➢ Addressing non-stationarity through dynamic normalization in both time and frequency domains enhances prediction accuracy by adapting to time-varying distribution shifts.
- ➢ Frequency domain decomposition captures distribution variations over different periods, while time domain techniques provide fine-grained adjustments for dynamic local patterns.

Contribution

- ➢ We propose Dual-domain Dynamic Normalization (DDN), a novel framework that dynamically normalizes time series in both time and frequency domains, effectively addressing non-stationarity.
- ➢ Our sliding window-based normalization mechanism decomposes time series into fine-grained frequency components and local temporal patterns, enabling precise adaptation to distribution shifts.
- ➢ Extensive experiments across diverse datasets demonstrate DDN's superior accuracy and versatility, significantly improving forecasting performance as a model-agnostic plugin.

Pipeline



- Sliding Normalization

$$\mu_j^i = \frac{1}{2k+1}\sum_{-k}^{k} x_{j+t}^i, \quad \left(\sigma_j^i\right)^2 = \frac{1}{2k+1}\sum_{-k}^{k}\left(x_{j+t}^i - \mu_j^i\right)^2,$$
$$\boldsymbol{\mu}^i = \mathrm{Pad}(\{\mu_{k+1}^i, \cdots, \mu_{L-k}^i\}), \quad \boldsymbol{\sigma}^i = \mathrm{Pad}(\{\sigma_{k+1}^i, \cdots, \sigma_{L-k}^i\}).$$
$$\bar{\boldsymbol{x}}^i = \frac{1}{\boldsymbol{\sigma}^i + \epsilon} \odot \left(\boldsymbol{x}^i - \boldsymbol{\mu}^i\right),$$

- Frequency Domain Normalization

$$\boldsymbol{x}_l^i, \boldsymbol{x}_h^i = \mathrm{DWT}_{\phi_{l,h}}(\boldsymbol{x}^i),$$
$$\bar{\boldsymbol{x}}_l^i, \boldsymbol{\mu}_l^i, \boldsymbol{\sigma}_l^i = \mathrm{SlidingNorm}(\boldsymbol{x}_l^i), \quad \bar{\boldsymbol{x}}_h^i, \boldsymbol{\mu}_h^i, \boldsymbol{\sigma}_h^i = \mathrm{SlidingNorm}(\boldsymbol{x}_h^i),$$
$$\hat{\boldsymbol{x}}^i = \mathrm{IDWT}_{\phi_{l,h}}(\bar{\boldsymbol{x}}_l^i, \bar{\boldsymbol{x}}_h^i), \quad \hat{\boldsymbol{\mu}}^i = \mathrm{IDWT}_{\phi_{l,h}}(\boldsymbol{\mu}_l^i, \boldsymbol{\mu}_h^i), \ \hat{\boldsymbol{\sigma}}^i = \mathrm{IDWT}_{\phi_{l,h}}(\boldsymbol{\sigma}_l^i, \boldsymbol{\sigma}_h^i).$$

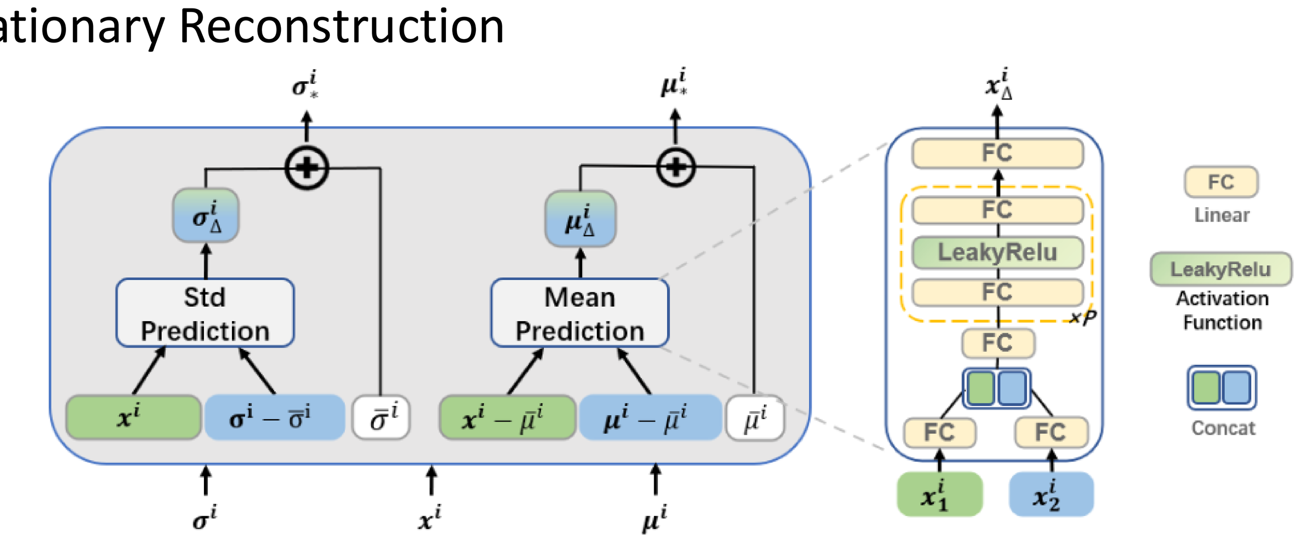
- Time Domain Normalization

$$\bar{\boldsymbol{x}}^i, \boldsymbol{\mu}^i, \boldsymbol{\sigma}^i = \mathrm{SlidingNorm}(\boldsymbol{x}^i),$$

- Stationary Sequences Weighting

$$\bar{\boldsymbol{x}}^i = \bar{\boldsymbol{x}}^i \cdot \beta + \hat{\boldsymbol{x}}^i \cdot \alpha.$$

- Non-stationary Reconstruction



Distribution Prediction Model (DPM) Prediction Branch

- Frequency Domain Prediction

$$\hat{\boldsymbol{\sigma}}_\Delta^i = \mathrm{SP}\left(\hat{\boldsymbol{\sigma}}^i - \sigma_f^i, \boldsymbol{x}^i\right), \quad \hat{\boldsymbol{\sigma}}_*^i = \hat{\boldsymbol{\sigma}}_\Delta^i + \sigma_f^i,$$
$$\hat{\boldsymbol{\mu}}_\Delta^i = \mathrm{MP}\left(\hat{\boldsymbol{\mu}}^i - \mu_f^i, \boldsymbol{x}^i - \mu_f^i\right), \quad \hat{\boldsymbol{\mu}}_*^i = \hat{\boldsymbol{\mu}}_\Delta^i + \mu_f^i.$$

- Time Domain Prediction

$$\boldsymbol{\sigma}_\Delta^i = \mathrm{SP}\left(\boldsymbol{\sigma}^i - \sigma_o^i, \boldsymbol{x}^i\right), \quad \boldsymbol{\sigma}_*^i = \boldsymbol{\sigma}_\Delta^i + \sigma_o^i,$$
$$\boldsymbol{\mu}_\Delta^i = \mathrm{MP}\left(\boldsymbol{\mu}^i - \mu_o^i, \boldsymbol{x}^i - \mu_o^i\right), \quad \boldsymbol{\mu}_*^i = \boldsymbol{\mu}_\Delta^i + \mu_o^i.$$

- De-normalization

$$\boldsymbol{\mu}_*^i = \boldsymbol{\mu}_*^i \cdot \beta + \hat{\boldsymbol{\mu}}_*^i \cdot \alpha, \quad \boldsymbol{\sigma}_*^i = \boldsymbol{\sigma}_*^i \cdot \beta + \hat{\boldsymbol{\sigma}}_*^i \cdot \alpha, \quad \bar{\boldsymbol{x}}^i = \bar{\boldsymbol{x}}^i \cdot \beta + \hat{\boldsymbol{x}}^i \cdot \alpha.$$
$$\boldsymbol{y}_*^i = \bar{\boldsymbol{y}}_*^i \odot \left(\boldsymbol{\sigma}_*^i + \epsilon\right) + \boldsymbol{\mu}_*^i.$$

Experiment

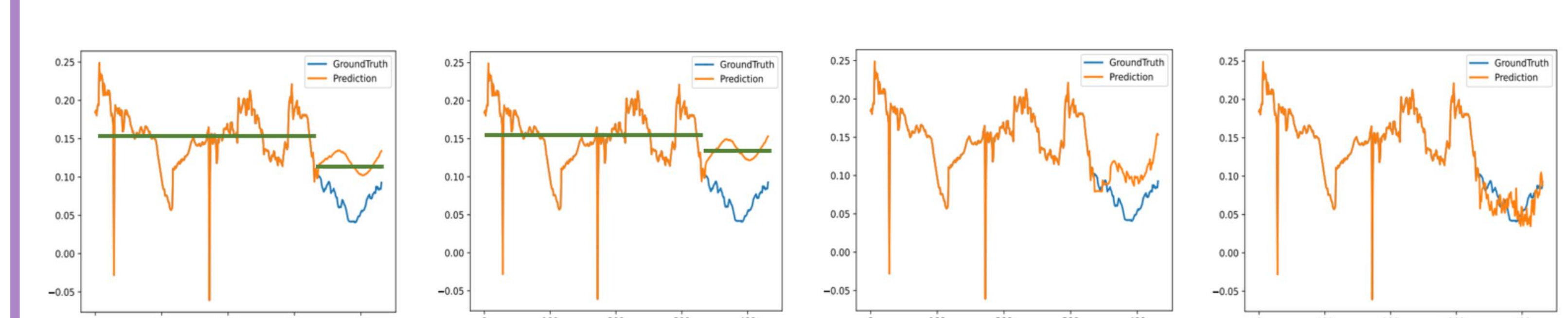
Methods		Autoformer		+DDN		FEDformer		+DDN		DLinear		+DDN		iTransformer		+DDN	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.458	0.448	**0.427**	**0.424**	**0.371**	0.411	0.385	**0.408**	0.377	0.399	**0.372**	**0.396**	0.392	0.422	**0.377**	**0.405**
	192	0.481	0.474	**0.472**	**0.452**	**0.420**	0.443	**0.415**	**0.452**	0.417	0.426	**0.406**	**0.416**	0.428	0.448	**0.414**	**0.430**
	336	0.508	0.485	**0.498**	**0.466**	**0.446**	0.459	0.458	**0.452**	0.464	0.461	**0.432**	**0.434**	0.467	0.475	**0.453**	**0.456**
	720	0.525	0.516	**0.502**	**0.483**	**0.482**	0.495	0.490	**0.479**	0.493	0.505	**0.462**	**0.474**	0.568	0.547	**0.553**	**0.530**
ETTm1	96	0.493	0.470	**0.354**	**0.390**	0.362	0.408	**0.313**	**0.364**	0.301	0.344	**0.288**	**0.342**	0.322	0.371	**0.301**	**0.355**
	192	0.546	0.498	**0.397**	**0.408**	0.395	0.427	**0.361**	**0.396**	0.335	0.366	**0.324**	**0.364**	0.353	0.392	**0.339**	**0.378**
	336	0.658	0.543	**0.429**	**0.433**	0.441	0.454	**0.417**	**0.430**	0.370	0.387	**0.356**	**0.385**	0.385	0.410	**0.370**	**0.396**
	720	0.626	0.532	**0.488**	**0.464**	0.488	0.481	**0.470**	**0.472**	0.425	0.421	**0.415**	**0.419**	0.441	0.443	**0.426**	**0.426**
Weather	96	0.247	0.320	**0.190**	**0.243**	0.246	0.328	**0.174**	**0.237**	0.175	0.237	**0.146**	**0.201**	0.177	0.228	**0.148**	**0.210**
	192	0.302	0.366	**0.231**	**0.282**	0.281	0.341	**0.233**	**0.294**	0.217	0.275	**0.190**	**0.247**	0.223	0.266	**0.191**	**0.252**
	336	0.362	0.394	**0.289**	**0.327**	0.337	0.376	**0.307**	**0.349**	0.263	0.314	**0.239**	**0.288**	0.287	0.310	**0.237**	**0.290**
	720	0.427	0.433	**0.369**	**0.375**	0.414	0.426	**0.399**	**0.405**	0.325	0.366	**0.311**	**0.343**	0.364	0.365	**0.301**	**0.336**
Electricity	96	0.195	0.309	**0.150**	**0.254**	0.185	0.300	**0.146**	**0.251**	0.140	0.237	**0.131**	**0.228**	0.133	0.229	**0.127**	**0.225**
	192	0.215	0.325	**0.173**	**0.275**	0.196	0.310	**0.168**	**0.268**	0.153	0.250	**0.148**	**0.246**	0.154	0.250	**0.146**	**0.246**
	336	0.237	0.344	**0.185**	**0.288**	0.215	0.330	**0.174**	**0.280**	0.168	0.267	**0.164**	**0.264**	0.170	0.266	**0.156**	**0.257**
	720	0.292	0.375	**0.201**	**0.304**	0.244	0.352	**0.216**	**0.312**	0.203	0.301	**0.201**	**0.299**	0.192	0.287	**0.179**	**0.282**
Traffic	96	0.654	0.403	**0.453**	**0.296**	0.579	0.363	**0.442**	**0.288**	0.411	0.283	**0.375**	**0.261**	0.348	0.254	**0.336**	**0.248**
	192	0.654	0.410	**0.462**	**0.304**	0.608	0.376	**0.462**	**0.300**	0.423	0.289	**0.396**	**0.272**	0.364	0.264	**0.347**	**0.254**
	336	0.629	0.391	**0.486**	**0.315**	0.620	0.385	**0.474**	**0.306**	0.437	0.297	**0.411**	**0.279**	0.381	0.272	**0.363**	**0.263**
	720	0.657	0.402	**0.529**	0.344	0.630	0.387	**0.512**	**0.329**	0.467	0.316	**0.448**	**0.298**	0.421	0.290	**0.412**	**0.286**

- ➢ Datasets: ETTh1, ETTm1, Weather, Electricity, Traffic
- ➢ Metric: Mean Square Error (MSE) and Mean Absolute Error (MAE)
- ➢ Integrating DDN into Autoformer, FEDformer, DLinear, and iTransformer achieves MSE reductions of 19.2%, 13.1%, 24.7%, and 22.3%, respectively, demonstrating its effectiveness across diverse forecasting models.

Ablation Studies

Methods	Autoformer +DDN	+RevIN	+NST	+Dish-TS	+SAN	IMP	FEDformer +DDN	+RevIN	+NST	+Dish-TS	+SAN	IMP
ETTh1	**0.475**	0.519	0.521	0.521	0.518	3.7%	**0.437**	0.463	0.456	0.461	0.447	-1.6%
ETTh2	**0.403**	0.489	0.465	1.175	0.411	9.6%	**0.385**	0.465	0.481	1.004	0.404	9.8%
ETTm1	**0.417**	0.562	0.535	0.567	0.406	28.2%	0.390	0.415	0.411	0.422	**0.377**	7.6%
ETTm2	**0.283**	0.325	0.331	0.894	0.311	15.0%	**0.282**	0.310	0.315	0.759	0.287	6.6%
Weather	**0.270**	0.290	0.290	0.433	0.305	19.2%	0.278	0.268	**0.267**	0.398	0.279	13.1%
Electricity	**0.177**	0.219	0.213	0.231	0.204	24.7%	**0.176**	0.200	0.198	0.203	0.191	16.2%
Traffic	**0.483**	0.666	0.664	0.677	0.594	25.6%	**0.473**	0.647	0.649	0.652	0.572	22.3%

- ➢ DDN outperforms other normalization methods: RevIN, NST, Dish-TS, and SAN across all benchmarks, achieving the best results by effectively addressing non-stationarity with finer-grained dynamic normalization.



(a)DLinear (b)+RevIN (c)+SAN (d)+DDN

- ➢ DDN can reconstruct fine-grained variations and rapid local fluctuations, surpassing other reversible normalization methods in precision and adaptability.