Paper ID 101

CMA: A Unified Contextual Meta-Adaptation Methodology for Time-Series Denoising and Prediction

Haiqi Jiang[1], Ying Ding[1], Chenjie Pan[1], Aimin Huang[2], Rui Chen[1], Chenyou Fan[1,*]

[1]South China Normal University, Guangdong, China

[2]Hangzhou Rose Technology Co., Ltd, Zhejiang, China

fanchenyou@scnu.edu.cn

Motivations :

- Current time series forecasting approaches face significant limitations due to **rigid model architectures and static parameters**.
- Standard Auto-Regressive (AR) models and Non-AutoRegressive (Non-AR) models' reliance on static history and future windows creates challenges in capturing real-world complexities such as non-stationary patterns, long-term dependencies, and evolving patterns. Additionally, these methods require **retraining for every prediction horizon**, making them impractical for industrial applications.
- To address these limitations, we introduce three innovative strategies: **in-context learning** for dynamic input adaptation over variable observations, **extend-context learning** to handle variable future horizons and patterns, and **cross-context learning** for effective domain transfer with minimal tuning. Our approaches enhance the adaptability and flexibility of forecasting methodology and seamlessly plug into existing forecasting backbones.

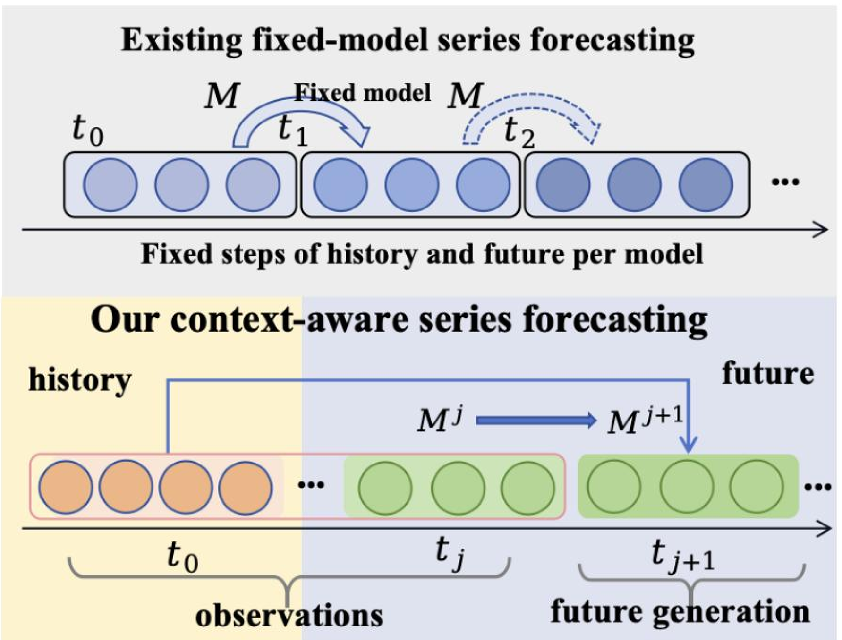

Figure 1: Learning from arbitrary historical steps and extrapolate to varying future horizons with a single unified model.

Our Proposed Method:

- We propose a unified framework to encompass our learning strategies, named Contextual Meta-Adaptation (CMA). CMA seamlessly integrates three elements: a capable generator model which can perform time series denoising, a joint context learning pipeline to adapt model weights to evolving contexts, an efficient model update strategy with meta-learning and test-time adaptation.
- ***Step-1:*** we design a diffusion adapter that embeds the timestep t and feeds it into the QKV projection along with time series patch tokens at each attention layer, allowing alternating prediction and stepwise denoising diffusion.
- ***Step-2:*** we develop a joint context learning pipeline called Con textual Meta-Adaptation (CMA) procedure. As shown in Fig. 1, CMA alternates between two actions to generate a sequence that aligns with historical data while adapting models based on current patterns.
- ***Step-3:*** we develop an efficient model update strategy using gradient-based meta learning for fine-tuning models over evolving patterns during the CMA procedure. This approach enables rapid parameter updates and facilitates testing-time adaptation to better align with test data.

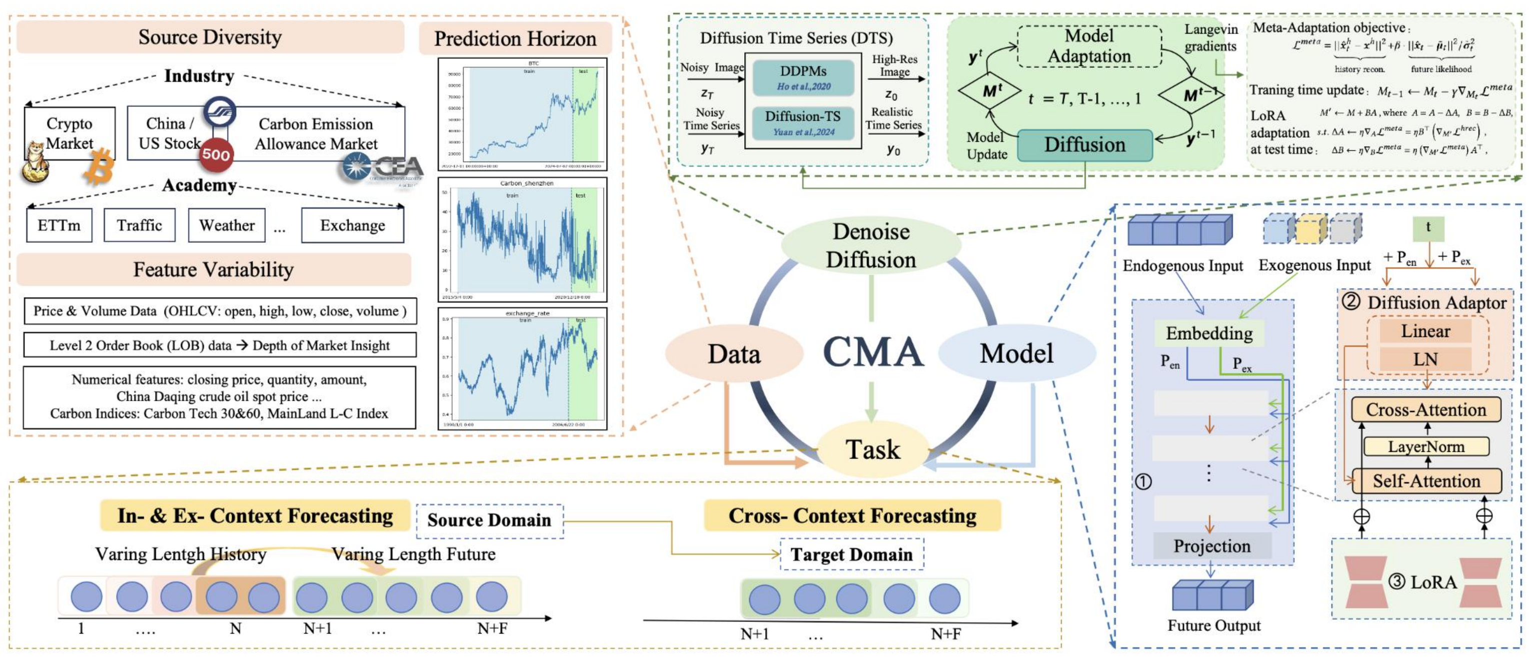

Figure 2: Contextual Meta-Adaptation (CMA) Framework: (1) Unifies continual learning and context test-time adaptation via meta-optimized diffusion processes (Sec. 4); (2) Systematically addresses in-, ex-, and cross-context learning paradigms (Sec. 5); (3) Extensively validated on 6 academic benchmarks and 4 real-world market datasets with diverse features.

Datasets and Baselines:

- **Academic** datasets include (1) ETTm2 Electricity Transformer dataset, (2) Electricity dataset, (3) Exchange dataset, (4) Traffic of California, (5) Weather, and (6) ILI disease data.
- **Industrial** datasets: (7) CEA-Carbon includes Chinese Carbon Emission Allowance daily data from four zones between 2015 and 2022 in China. (8) Crypto datasets. BTC (Bitcoin) and ETH (Ethereum) are widely traded crypto assets. (9) STAR-22 is sourced from China A -share on the Shanghai and Shenzhen Stock Exchanges, covering 4000 stocks from 2022 to 2023. (10) The US S&P500 index includes 5 hourly price-related attributes, with seven timesteps recorded per trading day.
- **Our Methods.** We abbreviate our methods as CMA-{In/Ex/Cr}-{i/T} for three learning tasks In-context (In), Ex-context (Ex) and Cross-context(Cr), with backbone models iTrans (denoted as i) or TimeXer (denoted as T).
- **Baselines.** Autoformer(AF), iTransformer(iTrans) and TimeXer are existing state-of-the-art time series predictors. The Diffusion-TS (DTS) is a recently proposed time-series generation technique. We also compare with a continual online learning framework FSNet and twov test-time adaptation (TTA) methods Tent and CoTTA.

Experimental results:

- **Academic Results**. Table 1 reports the MSE results on academic datasets. CMA-In-T achieves state-of-the-art performance across datasets and horizons. Compared to its backbone TimeXer, it reduces MSE by 7% (0.174 vs. 0.187), 6% (0.211 vs. 0.225), and 10% (0.270 vs. 0.299) on F96/192/336, with larger improvements on extended horizons (e.g., 10% on F336). Similarly, CMA-In-i outperforms iTrans by 4% (0.193 vs. 0.200), 4% (0.238 vs. 0.247), and 9% (0.279 vs. 0.307) across the same horizons.

Table 1: MSEs on academic datasets with varying future horizons. Models with iTrans and TimeXer backbones are highlighted.

	ETTm2			Electricity			Exchange			Traffic			Weather			$\overline{\text{MSE}}$		
Horizon	96	192	336	96	192	336	96	192	336	96	192	336	96	192	336	96	192	336
DTS [39]	1.975	2.075	2.211	0.980	1.077	1.231	0.781	1.247	1.470	2.113	2.450	2.571	1.154	1.544	1.670	1.401	1.679	1.831
AF [5]	0.232	0.295	0.339	0.211	0.227	0.243	0.160	0.496	0.456	0.695	0.768	0.689	0.291	0.320	0.380	0.318	0.422	0.422
FSNet [28]	0.167	0.202	0.373	0.231	0.289	0.222	0.198	0.246	0.337	0.405	0.442	0.588	0.228	0.298	0.318	0.246	0.295	0.368
Tent [34]	0.162	0.218	0.300	0.280	0.271	0.290	0.122	0.241	0.411	0.402	0.422	0.507	0.210	0.259	0.300	0.235	0.282	0.362
CoTTA [35]	0.148	0.214	0.296	0.287	0.300	0.295	0.099	0.202	0.480	0.398	0.415	0.490	0.197	0.255	0.299	0.226	0.277	0.372
iTrans [9]	0.185	0.251	0.314	**0.147**	0.164	0.178	0.099	0.181	0.341	0.393	0.412	0.421	0.176	0.225	0.280	0.200	0.247	0.307
TimeXer [10]	0.114	0.157	**0.185**	0.167	0.177	0.197	0.086	0.177	0.371	0.407	0.410	0.477	0.159	**0.206**	0.263	0.187	**0.225**	0.299
CMA-In-i	0.169	0.243	0.300	0.169	0.172	0.187	0.092	**0.145**	**0.182**	**0.372**	**0.390**	**0.385**	0.165	0.242	0.339	0.193	0.238	0.279
CMA-In-T	**0.104**	**0.150**	0.194	0.164	**0.160**	**0.175**	**0.081**	0.152	0.285	0.377	**0.390**	0.434	**0.145**	**0.204**	**0.262**	**0.174**	**0.211**	**0.270**
CMA-Ex-i	0.169	0.236	0.295	0.169	0.206	0.224	0.092	**0.110**	**0.172**	**0.372**	**0.398**	**0.390**	0.165	0.386	0.527	0.193	0.267	0.322
CMA-Ex-T	**0.104**	**0.155**	0.222	0.164	0.175	0.190	**0.081**	0.239	0.330	0.377	0.407	0.466	**0.145**	0.220	**0.227**	**0.174**	0.239	**0.287**

- **Crypto Market Results**. On Crypto Market. We report the MSE of each horizon in Table 2. Our methods adapt to 6 overlapping historical periods by extending the original history with 30 more preceding steps. Our CMA-In-T reduces MSE by 28% (2.28 vs. 3.17) for BTC compared to TimeXer. CMA-In-i outperforms iTrans with a 25% reduction in MSE. CMA-In-T and CMA-In-i outperform TTA baselines Tent and CoTTA by 23-32%. ETH has a similar trend.

Table 2: Forecasting results for Crypto Market. Models with iTrans and TimeXer backbones are highlighted.

Market	BTC					ETH				
Horizon	30	60	90	$\overline{\text{MSE}}$	$\overline{R^2}$	30	60	90	$\overline{\text{MSE}}$	$\overline{R^2}$
DTS [39]	4.02	3.77	3.78	3.86	0.50	5.37	5.78	5.00	5.38	0.47
AF [5]	3.99	3.58	3.80	3.79	0.55	5.14	5.17	4.91	5.07	0.49
FSNet [28]	2.52	2.58	2.59	2.56	0.54	4.30	4.49	4.50	4.43	0.53
Tent [34]	3.20	3.41	3.46	3.36	0.56	4.38	4.30	4.41	4.36	0.52
CoTTA[35]	3.17	3.17	3.16	3.17	0.56	4.68	4.66	4.67	4.67	0.53
iTrans [9]	3.28	3.19	3.26	3.24	0.60	4.66	4.77	4.68	4.70	0.59
TimeXer[10]	3.18	3.17	3.17	3.17	0.61	4.68	4.66	4.66	4.67	0.64
CMA-In-i	2.39	2.43	2.48	2.43	0.62	4.16	4.28	4.30	4.25	0.64
CMA-In-T	**2.28**	**2.28**	**2.28**	**2.28**	**0.64**	**3.91**	**3.88**	**3.87**	**3.89**	**0.68**
CMA-Ex-i	2.39	**2.28**	**2.38**	**2.35**	**0.62**	4.16	**3.66**	**3.84**	**3.89**	**0.66**
CMA-Ex-T	**2.28**	2.89	2.89	2.69	0.61	**3.91**	4.80	4.86	4.52	0.62

Table 3: China STAR-22 and US S&P500 Stock Market.

	China Stock STAR-22					US Stock S&P500				
Horizon	60	120	240	$\overline{\text{MSE}}$	$\overline{R^2}$	14	28	42	$\overline{\text{MSE}}$	$\overline{R^2}$
DTS [39]	2.84	3.61	2.59	3.01	0.51	7.98	8.11	6.99	7.69	0.67
AF [5]	2.52	3.48	2.56	2.85	0.56	7.60	7.83	6.71	7.38	0.70
FSNet [28]	2.05	2.59	2.20	2.28	0.55	6.11	6.52	6.20	6.28	0.68
Tent [34]	1.99	2.36	2.18	2.18	0.56	6.21	6.49	6.66	6.45	0.69
CoTTA [35]	1.95	2.30	2.04	2.10	0.57	6.20	6.45	6.70	6.45	0.70
iTrans [9]	1.84	2.01	2.15	2.00	0.60	6.26	6.31	6.32	6.30	0.74
TimeXer[10]	**0.88**	**1.47**	1.97	**1.44**	0.64	7.25	7.34	7.12	7.24	0.71
CMA-In-i	1.75	1.89	1.95	1.86	0.63	**5.59**	6.00	6.01	5.87	0.75
CMA-In-T	**0.88**	**1.46**	**1.88**	**1.41**	**0.67**	**5.60**	**5.54**	**5.68**	**5.61**	**0.77**
CMA-Ex-i	1.75	1.98	2.03	1.92	0.63	**5.59**	**5.43**	**5.56**	**5.53**	**0.76**
CMA-Ex-T	**0.88**	1.71	**1.90**	1.53	**0.65**	5.60	7.00	6.97	6.47	0.72

- **Stock Market Results**. Table 3 reports results on China and US Stock Markets. CMA-In-T achieves the lowest average MSE (1.41), 2% lower than TimeXer (1.44) and over 33% better than other baselines. The CMA-In-i comes second with MSE (1.86), falling behind CMA-In-T but reducing iTrans (2.00) by a larger 7%.

Case studies and visualizations:

- We show prediction examples across three contextual learning tasks. Fig. 5(a) illustrates cyclic movement in the Traffic dataset, where CMA-In-T more accurately captures seasonality patterns through in-context learning with extended historical adaptation.

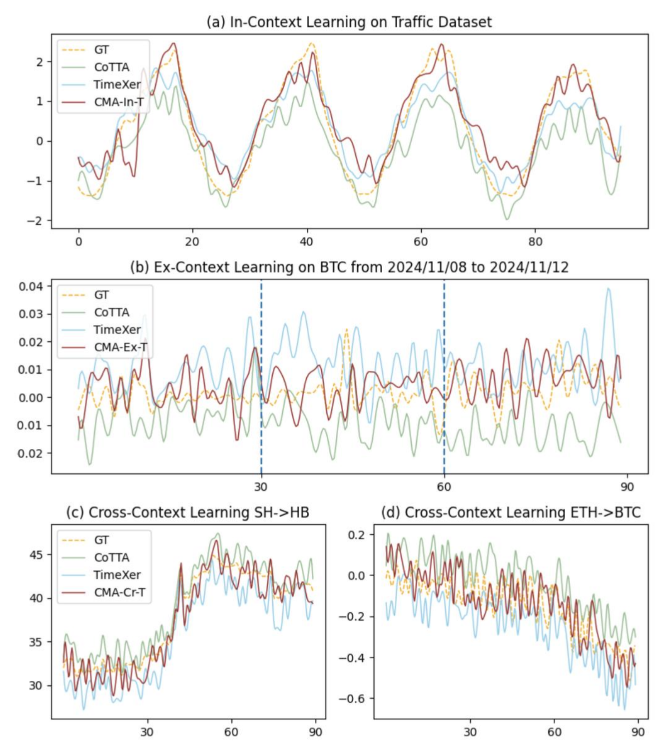

Figure 5: Visualization of various methods.

- Fig. 5(b) shows CMA-Ex-T has stabilized predictions around ground-truth trajectories of BTC in ex-context learning, outperforming CoTTA and TimeXer (specialized for F60/F90 horizons).
- Figs. 5(c) and (d) demonstrate cross-context learning scenarios. In the Carbon market (Shanghai to Hubei, SH->HB), CMA-Cr-T precisely identifies the upward trend aligned with observations.

Comparison with TS foundation model:

- Compared to Time Series foundation models like Moirai (13.8M+) and TimesFM (17M+), our approach simplifies the architecture. Our diffusion adapter introduces only 0.52M additional parameters to the 4.31M TimeXer backbone, with the LoRTA module reducing adaptation overhead to 0.05M (1%) during inference.
- We present the MSE results for Moirai in Table 6, using identical history lengths to our models (96 for ETTm2 and Weather, 36 for BTC). As the table shows, our approach substantially outperforms the zero-shot generalization of these foundation models. Specifically, we achieve a 50–75 % reduction in MSE on ETTm2, a 10–30% reduction on Weather, and a 38–40% reduction on BTC relative to Moirai's zero-shot performance.

Table 6: Comparison with Moirai.

Dataset	ETTm2			Weather			BTC		
Horizon	96	192	336	96	192	336	30	60	90
Moirai [18]	0.400	0.379	0.388	0.207	0.240	0.251	3.78	3.68	3.68
CMA-In-T	**0.104**	**0.150**	**0.194**	**0.145**	**0.204**	0.262	**2.28**	**2.28**	**2.28**
CMA-Ex-T	**0.104**	0.155	0.222	**0.145**	0.220	**0.227**	2.28	2.89	2.89

Ablation studies:

- Ablation of learning with a longer history. We train an even stronger baseline, TimeXer-Long, on the same extended history length as CMA-In-T , which includes 30 more steps than TimeXer. Similarly, we extend iTrans to iTrans-Long and compare with our CMA-In-i .

Table 10: MSE results on various datasets with a history length extended by 30 steps compared to the original.

Datasets	ETTm2			Electricity			Exchange		
Horizon	96	192	336	96	192	336	96	192	336
TimeXer[10]	0.114	0.157	**0.185**	0.167	0.177	0.197	0.086	0.177	0.371
TimeXer-Long	0.121	0.182	0.190	0.188	0.161	0.177	0.112	0.190	0.319
CMA-In-T	**0.104**	**0.150**	0.194	**0.164**	**0.160**	**0.175**	**0.081**	**0.152**	**0.285**
CMA-Ex-T	**0.104**	**0.155**	0.222	0.164	0.175	0.190	**0.081**	0.239	0.330
iTrans[9]	0.185	0.251	0.314	0.147	0.164	**0.178**	0.099	0.181	0.341
iTrans-Long	0.184	0.255	0.310	**0.145**	**0.162**	0.180	0.110	0.180	0.341
CMA-In-i	**0.169**	**0.243**	**0.300**	0.169	0.172	0.187	**0.092**	**0.145**	**0.182**
CMA-Ex-i	0.169	0.236	0.295	0.169	0.206	0.224	0.092	**0.110**	**0.172**

Conclusion

- We demonstrate that our CMA framework effectively captures underlying time series movement patterns, leading to improved predictions.
- Code: https://github.com/FancyAI-SCNU/CMA_KDD_2025