Supplementary Materials: *SimMTM: A Simple Pre-Training Framework for Masked Time-Series Modeling*

Anonymous Author(s)
Affiliation
Address
email

A Implementation Details

All the experiments are repeated five times, implemented in PyTorch [12] and conducted on NVIDIA A100 SXM4 40GB GPU. We implement the baselines based on their official implementation and follow the configuration from their original papers. For the metrics, we adopt the mean square error (MSE) and mean absolute error (MAE) for the time series forecasting. As for the classification, accuracy, precision, recall, F1 score, and their average value are recorded.

A.1 Dataset Description

We conduct experiments to evaluate the effect of our method under in-domain and cross-domain settings on twelve real-world datasets for two typical time series analysis tasks: forecasting and classification, covering diverse application scenarios (electricity system, neurological healthcare, human activity recognition, mechanical fault detection, and physical status monitoring), different types of signals (ECG, EMG, acceleration, vibration, power load, weather, and transoirtation), multivariate channel dimensions (from 1 to 862), varying times series lengths (from 96 to 5120) and large span sampling ratio (from 100 Hz to 4000 Hz). The detailed descriptions of these datasets are summarized in Table 1.

(1) **ETT (4 subsets)** [26] contains the time series of oil temperature and power load collected by electricity transformers from July 2016 to July 2018. ETT is a group of four subsets with different recorded frequencies: ETTh1/ETTh2 are recorded every hour, and ETTm1/ETTm2 are recorded every 15 minutes.

(2) **WEATHER** [18] includes meteorological time series with 21 weather indicators collected every 10 minutes from the Weather Station of the Max Planck Biogeochemistry Institute in 2020.

(3) **ELECTRICITY** [15] records the hourly electricity consumption of 321 clients from 2012 to 2014. Values are in kW of each 15 min. All time labels report to Portuguese hour. However, all days present 96 measures (24×4). For every year in March, time change day (which has only 23 hours), values between 1:00 am and 2:00 am are zero for all points. For every year in October, time change day (which has 25 hours), the values between 1:00 am and 2:00 am aggregated consumption of two hours.

(4) **TRAFFIC** [13] encompasses the hourly measures of road occupancy rates obtained from 862 sensors situated in the San Francisco Bay area freeways. These measurements were carried out between January 2015 and December 2016.

(5) **SLEEPEEG** [6] contains 153 whole-night sleeping electroencephalography (EEG) recordings from 82 healthy subjects. We follow the same data preprocessing approach as [25] to segment the EEG signals without overlapping and get 371,055 univariate brainwaves. Each brainwave is sampled

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

Table 1: Dataset descriptions. *Samples* are organized in (Train/Validation/Test).

Tasks	Datasets	Channels	Length	Samples	Classes	Information	Frequency
Forecasting	ETTh1,ETTh2	7	{96,192,336,720}	8545/2881/2881	-	Electricity	1 Hour
	ETTm1,ETTm2	7	{96,192,336,720}	34465/11521/11521	-	Electricity	15 Mins
	Weather	21	{96,192,336,720}	36792/5271/10540	-	Weather	10 Mins
	Electricity	321	{96,192,336,720}	18317/2633/5261	-	Electricity	1 Hour
	Traffic	862	{96,192,336,720}	12185/1757/3509	-	Transportation	1 Hour
Classification	SleepEEG	1	200	371005/-/-	5	EEG	100 Hz
	Epilepsy	1	178	60/20/11420	2	EEG	174 Hz
	FD-B	1	5120	60/21/135599	3	Faulty Detection	64K Hz
	Gesture	3	315	320/120/120	8	Hand Movement	100 Hz
	EMG	1	1500	122/41/41	3	Muscle responses	4K Hz

at a frequency of 100 Hz and associated with one of five sleeping stages: Wake, Non-rapid eye movement (3 sub-states), and Rapid Eye Movement.

(6) **EPILEPSY** [1] monitors the brain activities of 500 subjects with a single-channel EEG sensor. Every subject is recorded for 23.6 seconds of brain activities. The dataset is sampled at 178 Hz and contains 11,500 samples in total. We follow the procedure described by [25]. The first four classes (eyes open, eyes closed, EEG measured in the healthy brain region, and EEG measured in the tumor region) of the original five categories of each sample are classified as positive, and the remaining classes (whether the subject has a seizure episode) are used as negative.

(7) **FD-B** [7] is generated by electromechanical drive systems. It monitors the condition of rolling bearings and detects their failures based on the monitoring conditions, which include speed, load torque, and radial force. Concretely, FD-B has 13,640 samples in total. Each recording is sampled at 64k Hz with 3-class labels: undamaged, inner damaged, and outer damaged.

(8) **GESTURE** [9] are collected from 8 hand gestures based on the paths of hand movement recorded by an accelerometer. The eight gestures are hand swiping left, right, up, and down, hand waving in a counterclockwise or clockwise circle, hand waving in a square, and waving a right arrow, respectively. This dataset contains 440 examples of balanced classification labels that can be used, and each sample includes eight different kinds of gesture categories.

(9) **EMG** [14] is sampled with 4K Hz and consists of 163 single-channel EMG recordings from the anterior tibialis muscle of three healthy volunteers suffering from neuropathy and myopathy. Each patient is a classification category, so each sample is associated with one of three classes.

A.2 Baselines Implementation

We compare our proposed SimMTM against six state-of-the-art baselines. To make a fair and comprehensive comparison, we tried two baseline implementation approaches for forecasting and classification tasks: the unified encoder and reproduced with their official implementation encoder. Notably, the designs of LaST [17] and TF-C [25] are closely related to model structures. We directly report results from their papers or reproduce codes with official implementation.

(1) **Unified encoder**. We attempt to unify the encoder for these pre-training methods. Specifically, we adopt the vanilla Transformer [16] with channel independent [11] for forecasting to accomplish cross-domain transfer between datasets with different variate numbers. As for the classification, we use 1D-ResNet [5] as the encoder following [25]. Besides, we do a comprehensive hyperparameter

Table 2: Baselines implementation details.

Baselines	Task	Encoder	Performance Comparison	Report
Ti-MAE [8]	Forecasting	Channel-independent Transformer	better	Main text
		Official implementation		Section E
	Classification	1D-ResNet	better	Main text
		Official implementation		Section E
TST [24]	Forecasting	Channel-independent Transformer	better	Main text
		Official implementation		Section E
	Classification	1D-ResNet	better	Main text
		Official implementation		Section E
LaST [17]	Forecasting	Official implementation	/	Main text
	Classification	Official implementation	/	Main text
TF-C [25]	Forecasting	Official implementation	/	Main text
	Classification	Official implementation	/	Main text
CoST [20]	Forecasting	Channel-independent Transformer	better	Main text
		Official implementation		Section E
	Classification	1D-ResNet	better	Main text
		Official implementation		Section E
TS2Vec [23]	Forecasting	Channel-independent Transformer	better	Main text
		Official implementation		Section E
	Classification	1D-ResNet	better	Main text
		Official implementation		Section E

search for all baselines. For the Transformer encoder, we vary the number of Transformer layers in $\{1, 2, 3, 4\}$, select the model dimension from $\{16, 32, 64, 128, 256\}$, and the attention head from $\{4, 8, 16, 32\}$. For the 1D-ResNet, we search the number of 1D-ResNet layers from $\{1, 2, 3, 4\}$, the kernel size from $\{3, 5, 8\}$ respectively. Additionally, for the masked modeling methods TST [24], Ti-MAE [8], we also searched the masked ratio $r = \{0.125, 0.25, 0.5, 0.75\}$ for better performance.

(2) **Official implementation**. We also implement the baselines following the corresponding official codes, including encoder, hyperparameters, etc. The comparisons are included in Section E of this supplementary material. We directly report the results from their original papers for the same set. For mismatched settings, the results are from our implementation.

Finally, for baselines Ti-MAE [8], TST [24], CoST [20], and TS2Vec [23], we report the results based on the unified encoder in the main text. But for baselines LaST [17] and TF-C [25], we report the results of the official code implementation or their original paper, which are limited by their model structures. As a result, the performances of all baselines with unified encoder (that we reported in the main text) generally surpass their official implementation and results reported in their own paper. Table 2 shows more details. Full experimental results are in Section E.

A.3 Pre-training and Fine-tuning Configuration

We built two types of pre-training and fine-tuning scenarios, in-domain and cross-domain, based on the benchmarks of forecasting and classification tasks to compare the effectiveness of our method and other time series pre-training methods.

We pre-train a model on one subset for forecasting tasks and fine-tune it to the same dataset to build seven in-domain transfer evaluation scenarios. In cross-domain evaluation, we pre-train a model on one specific dataset and then use the other datasets for fine-tuning. Based on the above settings, we constructed fifteen in-domain and cross-domain pre-training and fine-tuning experiments, covering the same dataset with the same sampled frequency, different datasets with the same sampled frequency, and different datasets with different sampled frequencies.

Table 3: Pre-training and fine-tuning scenarios for time series forecasting (Fore.) and classification (Class.) tasks, including the same and different datasets and in- and cross-domain settings.

Tasks	Evaluation	Scenarios	Characteristic
Fore.	In-domain	ETTh1 → ETTh1	The same dataset with the same frequency
		ETTh2 → ETTh2	
		ETTm1 → ETTm1	
		ETTm2 → ETTm2	
		Weather → Weather	
		Electricity → Electricity	
		Traffic → Traffic	
	Cross-domain	ETTh2 → ETTh1	Different datasets with the same frequency.
		ETTm2 → ETTm1	
		{ETTm1, ETTm2, Weather} → ETTh1	Different datasets with different frequencies.
		{ETTh1, ETTh2, Weather} → ETTm1	
Class.	In-domain	Epilepsy → Epilepsy	The same dataset with the same frequency.
	Cross-domain	SleepEEG → {Epilepsy, FD-B, Gesture, EMG}	Different datasets with different frequencies.

We use the same dataset, Epilepsy, to construct the in-domain setting for classification tasks. For the cross-domain setting, we pre-train a model for classification tasks on a univariate time series dataset SleepEEG with the most complex temporal dynamics and the most samples. And then fine-tune the model separately on Epilepsy, FD-B, Gesture, and EMG. Furthermore, we constructed four cross-domain evaluation scenarios by pre-training from SleepEEG and fine-tuning to Epilepsy, FD-B, Gesture, and EMG because of fewer commonalities and the enormous gap among these datasets. Table 3 shows detailed pre-training and fine-tuning settings.

A.4 Model and Training Configuration

Following the previous convention, we choose the encoder part of Transformer [16] with channel independent as the feature extractor for forecasting tasks. For the classification tasks, we adopt 1D-ResNet [5] as the encoder following [25]. In the pre-training stages, we pre-train the model with different learning rates and batch sizes according to the pre-train datasets. Then we fine-tune it to downstream forecasting and classification tasks supervised by L2 and Cross-Entropy losses, respectively. The configuration details are in Table 4.

Table 4: Model and training configuration in Forecasting (*Fore.*) and Classification (*Class.*) tasks.

Tasks	Encoder		Pre-training			Fine-tuning			
	e_{layers}	d_{model}	learning rate	batch size	epochs	learning rate	loss function	batch size	epochs
Fore.	2	16	1e-3	32	50	1e-4	L2	{16,32}	10
Class.	3	128	1e-4	128	10	1e-4	Cross-Entropy	32	300

B Hyperparameter Sensitivity

We verify the hyperparameter sensitivity of the proposed time series pre-training method SimMTM on ETTh1 in Table 5, including masked ratio (r), the number of masked series (M), temperature (τ), masked function (Mask), encoder depth (e_{layers}), and the hidden dimension (d_{model}). Lower MSE and MAE represent better performance.

As shown in Table 5 (a) and 5 (b), we can observe the effect of the method is closely related to the trade-off of the masked ratio and the number of masked series. Hence, a reasonable balance between the two kinds of parameters is critical. For the temperature hyperparameter of softmax normalization (τ), we use an appropriately small τ that leads to higher differences and diversity of masked sequences. For the masked methods, we chose two masked methods for verification: masking following random distribution and masking following geometric distribution [24]. The results show that the method based on geometric masking is better than random masking modeling. Besides, we can find that 2 encoder layers are enough for reconstruction tasks. Note our method SimMTM consistently performs better than training from scratch under various hyperparameter changes.

Table 5: Hyperparameter sensitivity experiments on ETTh1 for the in-domain setting. The entries marked in bold are the same which specify the default settings. This table format follows [4].

(a) Masked ratio

Ratio	MSE	MAE
12.5%	0.429	0.440
25%	0.427	0.434
50%	**0.409**	**0.428**
75%	0.422	0.434

(b) Masked numbers

Numbers	MSE	MAE
1	0.429	0.437
2	0.416	0.429
3	**0.409**	**0.428**
4	0.419	0.431

(c) Temperature

Value	MSE	MAE
0.02	**0.409**	**0.428**
0.2	**0.409**	0.429
2	0.416	0.428

(d) Masked function

Type	MSE	MAE
Random	0.409	0.431
Geometric	**0.409**	**0.428**

(e) Encoder depth

Layers	MSE	MAE
1	0.420	0.426
2	**0.409**	**0.428**
3	0.421	0.430
4	0.426	0.436

(f) Hidden layer dimension

Dim	MSE	MAE
16	**0.409**	**0.428**
32	0.420	0.432
64	0.422	0.434
128	0.428	0.444

C Ablations on Aggregation Setting

SimMTM proposes to recover masked time points by the weighted aggregation of multiple neighbors outside the manifold. We explored two types of aggregation settings.

(1) **Positive Samples Aggregation (PSA)**: only aggregate multiple positive neighbors (the masked series of the same sample) to reconstruct masked time points.

(2) **Positive and Negative Samples Aggregation (PNSA)**: aggregate both positive and negative neighbors (the masked series of all samples) to reconstruct masked time points.

As shown in Table 6, although PSA made good progress compared to training from scratch (Random Init.), PNSA is consistently better than SimMTM PSA in all ablation settings. In masked time-series modeling, masking can be viewed as adding noise to the original data, and masked modeling is to project masked data from the neighborhood back to the original manifold. We use positive and negative masked time series as the reconstruction candidates to drive the model to select the positive samples adaptively, which can make the model learn the structure of the manifold better. Therefore, as stated in the Method Section of the main text, we choose positive and negative sample aggregation (PNSA) as the standard aggregation setting of SimMTM.

Table 6: Ablations on aggregation setting in forecasting (*MSE*) and classification (*Acc*) tasks for in- and cross-domain settings. A smaller MSE or a higher Accuracy indicates a better result (↑).

Tasks	Evaluation	Scenarios	Aggregation	Metric
Forecasting	In-domain	ETTh1 → ETTh1	Random init.	0.431
			SimMTM (**PSA**)	0.420 ↑
			SimMTM (**PNSA**)	**0.409** ↑
	Cross-domain	ETTh2 → ETTh1	Random init.	0.431
			SimMTM (**PSA**)	0.426 ↑
			SimMTM (**PNSA**)	**0.415** ↑
Classification	In-domain	Epilepsy → Epilepsy	Random init.	89.83
			SimMTM (**PSA**)	92.56 ↑
			SimMTM (**PNSA**)	**94.75** ↑
	Cross-domain	SleepEEG → EMG	Random init.	77.80
			SimMTM (**PSA**)	87.80 ↑
			SimMTM (**PNSA**)	**97.56** ↑

D Comparison of Masked Modeling

To investigate the reconstruction process of different masked modeling methods, we plot both original and reconstructed time series from TST and SimMTM in Figure 1, where TST [24] follows the canonical masked modeling paradigm and learns to predict removed time points based on the remaining time points. In Figure 1, we can find that direct reconstruction is too difficult in time series, even for the 12.5% masking ratio. As for the 75% masking ratio, TST degenerates more seriously. Because of this poor reconstruction effect, direct reconstruction is difficult to provide reliable guidance to model pre-training. In contrast, our proposed SimMTM can precisely reconstruct the original time series, benefiting the representation learning. These results also support our design in neighborhood reconstruction.

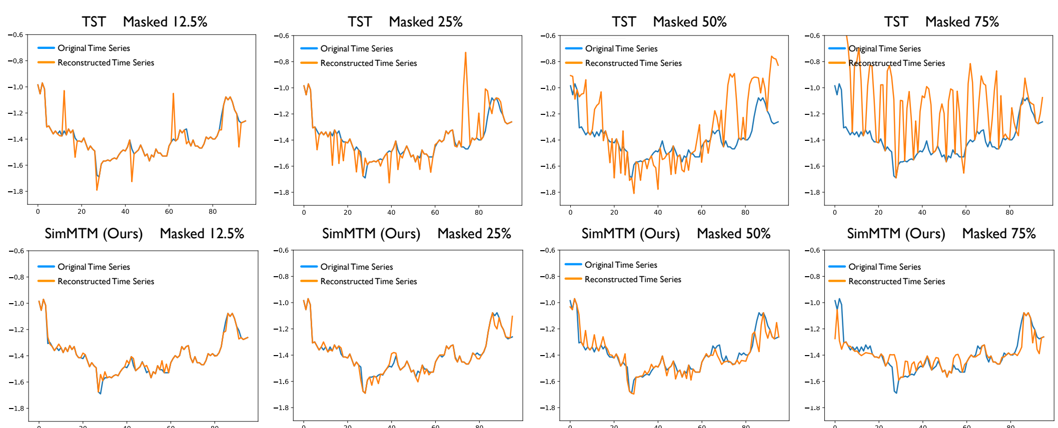

Figure 1: Comparison of the canonical masked modeling paradigm TST and neighborhood aggregation masked modeling SimMTM in reconstructing time series. All the cases are shown from ETTh1.

E Full Results

Due to the limited length of the text, we summarize all the experiments in the main text into two parts: the main experiment and the analytical experiment. We categorize and index them in Tabel 7, 8.

Table 7: The main results of pre-training and fine-tuning scenarios for time series forecasting and classification tasks, including the same and different encoder for in- and cross-domain settings.

Tasks	Evaluation	Encoder	Tabels Name
Forecasting	In-domain	The model utilized in the original papers	Table 9
		Transformer with channel independent	Table 10
	Cross-domain	The model utilized in the original papers	Table 11
		Transformer with channel independent	Table 12
Classification	In-domain	The model utilized in the original papers	Table 17
		1D-ResNet	Table 18
	Cross-domain	The model utilized in the original papers	Table 17
		1D-ResNet	Table 18

Table 8: The model analysis results of pre-training and fine-tuning scenarios for time series forecasting and classification tasks with the unified encoder for in- and cross-domain settings.

Tasks	Evaluation	Analysis	Tabels Name
Forecasting	In-domain	Ablation study	Table 13
		Model generality	Table 15
	Cross-domain	Ablation study	Table 14
		Limited data	Table 16
Classification	In-domain	Ablation study	Table 19
	Cross-domain	Ablation study	Table 19

F Limitations

SimMTM is inspired by the manifold perspective of masked modeling. Although we have provided relatively comprehensive results to verify the model's effectiveness, the model performance still needs theoretical guarantees. In fact, the most high-impact works in the self-supervised pre-training community are without theoretical analysis, such as BERT [2], GPT-3 [3], MAE [4] and SimMIM [22]. Thus, we would like to leave this problem as a future work.

The masking ratio of masked modeling methods is an essential hyper-parameter. Although we have provided a chosen principle to masking ratio r and the number of masked time series M as $M \propto r$ in the main text, we still need to tune these two hyperparameters for different datasets to achieve the best performance. Notably, previous methods also chose the masking ratio solely based on the empirical results [2, 4]. Thus, despite there exist limitations of SimMTM in choosing hyperparameters, the principle of $M \propto r$ can somewhat ease this problem. And the chosen strategy of the masking ratio can also be a potential topic in masked modeling [19].

G Social Impacts

This paper presents SimMTM as a new masked modeling method for time series. SimMTM achieves state-of-the-art in two mainstream time series analysis tasks, which can be a good supplement for the self-supervised pre-training community. We will also publish the codebase of time-series pre-training to facilitate future research.

This paper only focuses on the algorithm design. Using all the codes and datasets strictly follows the corresponding licenses (Appendix A.1). There is no potential ethical risk or negative social impact.

Table 9: Complete results of long-term forecasting tasks for the in-domain setting of forecasting the future $O \in \{96, 192, 336, 720\}$ time points based on the past 336 time points. **All the results of baselines are based on the encoder utilized in their original papers.** The standard deviations of SimMTM are within 0.005 for MSE and within 0.004 for MAE.

	Models	**SimMTM**		Random init.		Ti-MAE [8]		TST [24]		LaST [17]		TF-C [25]		CoST [20]		TS2Vec [23]	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.379**	**0.407**	0.380	0.412	0.708	0.570	0.503	0.527	0.399	0.412	0.463	0.406	0.514	0.512	0.709	0.650
	192	**0.412**	**0.424**	0.416	0.434	0.725	0.587	0.601	0.552	0.484	0.468	0.531	0.540	0.655	0.590	0.927	0.757
	336	**0.421**	**0.431**	0.448	0.458	0.713	0.589	0.625	0.541	0.580	0.533	0.535	0.545	0.790	0.666	0.986	0.811
	720	**0.424**	0.449	0.481	0.487	0.736	0.618	0.768	0.628	0.432	**0.432**	0.577	0.562	0.880	0.739	0.967	0.790
	Avg	**0.409**	**0.428**	0.431	0.448	0.721	0.591	0.624	0.562	0.474	0.461	0.527	0.513	0.710	0.627	0.897	0.752
ETTh2	96	**0.293**	**0.347**	0.325	0.374	0.443	0.465	0.335	0.392	0.331	0.390	0.463	0.521	0.465	0.482	0.506	0.477
	192	**0.355**	**0.386**	0.400	0.424	0.533	0.516	0.444	0.441	0.451	0.452	0.525	0.561	0.671	0.599	0.567	0.547
	336	**0.370**	**0.401**	0.405	0.433	0.445	0.472	0.455	0.494	0.460	0.478	0.850	0.883	0.848	0.776	0.694	0.628
	720	**0.395**	**0.427**	0.451	0.475	0.507	0.498	0.481	0.504	0.552	0.509	0.930	0.932	0.871	0.811	0.728	0.838
	Avg	**0.353**	**0.390**	0.395	0.427	0.482	0.488	0.429	0.458	0.449	0.457	0.692	0.724	0.714	0.667	0.624	0.623
ETTm1	96	**0.288**	0.348	0.295	**0.346**	0.647	0.497	0.454	0.456	0.316	0.355	0.419	0.401	0.376	0.420	0.563	0.551
	192	**0.327**	0.373	0.333	0.374	0.597	0.508	0.471	0.490	0.349	**0.366**	0.471	0.438	0.420	0.451	0.599	0.558
	336	**0.363**	**0.395**	0.370	0.398	0.699	0.525	0.457	0.451	0.429	0.407	0.540	0.509	0.482	0.494	0.685	0.594
	720	**0.412**	**0.424**	0.427	0.431	0.786	0.596	0.594	0.488	0.496	0.464	0.552	0.548	0.628	0.578	0.831	0.698
	Avg	**0.348**	**0.385**	0.356	0.387	0.682	0.532	0.494	0.471	0.398	0.398	0.496	0.474	0.477	0.486	0.670	0.600
ETTm2	96	0.172	0.261	0.175	0.268	0.304	0.357	0.363	0.301	**0.163**	**0.255**	0.401	0.477	0.276	0.384	0.448	0.482
	192	**0.223**	**0.300**	0.240	0.312	0.334	0.387	0.342	0.364	0.239	0.303	0.422	0.490	0.500	0.532	0.545	0.536
	336	0.282	**0.331**	0.298	0.351	0.420	0.441	0.414	0.361	**0.259**	0.366	0.513	0.508	0.680	0.695	0.681	0.744
	720	**0.374**	0.388	0.403	0.413	0.508	0.481	0.580	0.456	0.397	**0.382**	0.523	0.772	0.925	0.914	0.691	0.837
	Avg	**0.263**	**0.320**	0.279	0.336	0.392	0.417	0.425	0.371	0.265	0.327	0.465	0.562	0.595	0.631	0.591	0.650
Weather	96	0.158	**0.211**	0.166	0.216	0.216	0.280	0.292	0.370	**0.153**	**0.211**	0.215	0.296	0.327	0.359	0.433	0.462
	192	**0.199**	**0.249**	0.208	0.254	0.303	0.335	0.410	0.473	0.207	0.250	0.267	0.345	0.390	0.422	0.508	0.518
	336	**0.246**	0.286	0.257	0.290	0.351	0.358	0.434	0.427	0.249	**0.264**	0.299	0.360	0.477	0.446	0.545	0.549
	720	**0.317**	0.337	0.326	0.338	0.425	0.399	0.539	0.523	0.319	**0.320**	0.361	0.395	0.551	0.586	0.576	0.572
	Avg	**0.230**	0.271	0.239	0.275	0.324	0.343	0.419	0.448	0.232	**0.261**	0.286	0.349	0.436	0.453	0.516	0.525
Electricity	96	**0.133**	**0.223**	0.190	0.279	0.399	0.412	0.292	0.370	0.166	0.254	0.366	0.436	0.230	0.353	0.322	0.401
	192	**0.147**	**0.237**	0.195	0.285	0.400	0.460	0.270	0.373	0.178	0.278	0.366	0.433	0.253	0.371	0.343	0.416
	336	**0.166**	**0.265**	0.211	0.301	0.564	0.573	0.334	0.323	0.186	0.275	0.358	0.428	0.197	0.287	0.362	0.435
	720	**0.203**	0.297	0.253	0.333	0.880	0.770	0.344	0.346	0.213	**0.288**	0.363	0.431	0.230	0.328	0.388	0.456
	Avg	**0.162**	**0.256**	0.212	0.300	0.561	0.554	0.310	0.353	0.186	0.274	0.363	0.432	0.228	0.335	0.354	0.427
Traffic	96	**0.368**	**0.262**	0.471	0.309	0.431	0.482	0.559	0.454	0.706	0.385	0.613	0.340	0.751	0.431	0.321	0.367
	192	**0.373**	**0.251**	0.475	0.308	0.491	0.346	0.583	0.493	0.709	0.388	0.619	0.516	0.751	0.424	0.476	0.367
	336	**0.395**	**0.254**	0.490	0.315	0.502	0.384	0.637	0.469	0.714	0.394	0.785	0.497	0.761	0.425	0.499	0.376
	720	**0.432**	**0.290**	0.524	0.332	0.533	0.543	0.663	0.594	0.723	0.421	0.850	0.472	0.780	0.433	0.563	0.390
	Avg	**0.392**	**0.264**	0.490	0.316	0.489	0.399	0.611	0.503	0.713	0.397	0.717	0.456	0.761	0.428	0.501	0.375

Table 10: Complete results of long-term forecasting tasks for the in-domain setting of forecasting the future $O \in \{96, 192, 336, 720\}$ time points based on the past 336 time points. **All the results of baseline are based on the unified channel-independent Transformer encoder.** The standard deviations of SimMTM are within 0.005 for MSE and within 0.004 for MAE.

Models		**SimMTM**		Random init.		Ti-MAE [8]		TST [24]		LaST [17]		TF-C [25]		CoST [20]		TS2Vec [23]	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.379	**0.407**	0.380	0.412	**0.356**	0.420	0.401	0.425	-	-	-	-	0.422	0.436	0.392	0.420
	192	**0.412**	**0.424**	0.416	0.434	0.421	0.434	0.427	0.432	-	-	-	-	0.520	0.487	0.445	0.452
	336	**0.421**	**0.431**	0.448	0.458	0.447	0.446	0.519	0.487	-	-	-	-	0.472	0.462	0.453	0.455
	720	**0.424**	**0.449**	0.481	0.487	0.469	0.482	0.515	0.504	-	-	-	-	0.525	0.501	0.495	0.496
	Avg	**0.409**	**0.428**	0.431	0.448	0.423	0.446	0.466	0.462	-	-	-	-	0.485	0.472	0.446	0.456
ETTh2	96	**0.293**	**0.347**	0.325	0.374	0.339	0.378	0.322	0.358	-	-	-	-	0.321	0.374	0.365	0.509
	192	**0.355**	**0.386**	0.400	0.424	0.380	0.402	0.448	0.435	-	-	-	-	0.380	0.403	0.396	0.422
	336	**0.370**	**0.401**	0.405	0.433	0.388	0.323	0.420	0.440	-	-	-	-	0.430	0.451	0.399	0.436
	720	**0.395**	**0.427**	0.451	0.475	0.414	0.442	0.424	0.452	-	-	-	-	0.466	0.480	0.508	0.503
	Avg	**0.353**	**0.390**	0.395	0.427	0.380	0.386	0.404	0.421	-	-	-	-	0.399	0.427	0.417	0.468
ETTm1	96	**0.288**	0.348	0.295	0.346	0.305	0.351	0.310	0.348	-	-	-	-	0.291	**0.343**	0.681	0.689
	192	**0.327**	0.373	0.333	0.374	0.343	0.374	0.362	0.380	-	-	-	-	0.330	**0.370**	0.689	0.551
	336	**0.363**	**0.395**	0.370	0.398	0.387	0.407	0.389	0.402	-	-	-	-	0.382	0.401	0.704	0.559
	720	**0.412**	**0.424**	0.427	0.431	0.428	0.432	0.433	0.427	-	-	-	-	0.422	0.425	0.721	0.571
	Avg	**0.348**	**0.385**	0.356	0.387	0.366	0.391	0.373	0.389	-	-	-	-	0.356	0.385	0.699	0.557
ETTm2	96	**0.172**	0.261	0.175	0.268	0.174	**0.258**	0.215	0.296	-	-	-	-	0.242	0.333	0.224	0.303
	192	**0.223**	**0.300**	0.240	0.312	0.257	0.303	0.259	0.323	-	-	-	-	0.283	0.345	0.273	0.331
	336	0.282	**0.331**	0.298	0.351	**0.277**	0.333	0.319	0.364	-	-	-	-	0.303	0.349	0.399	0.402
	720	0.374	**0.388**	0.403	0.413	**0.360**	0.404	0.395	0.405	-	-	-	-	0.431	0.431	0.406	0.408
	Avg	**0.263**	**0.320**	0.279	0.336	0.267	0.325	0.297	0.347	-	-	-	-	0.314	0.365	0.326	0.361
Weather	96	0.158	0.211	0.166	0.216	**0.153**	**0.196**	0.162	0.214	-	-	-	-	0.216	0.280	0.154	0.205
	192	**0.199**	0.249	0.208	0.254	0.214	0.253	0.203	0.252	-	-	-	-	0.303	0.335	0.200	**0.243**
	336	0.246	0.286	0.257	0.290	**0.243**	**0.272**	0.260	0.297	-	-	-	-	0.351	0.358	0.252	0.286
	720	**0.317**	0.337	0.326	0.338	0.324	0.349	0.330	0.342	-	-	-	-	0.425	0.343	0.324	**0.335**
	Avg	**0.230**	0.271	0.239	0.275	0.234	**0.265**	0.239	0.276	-	-	-	-	0.324	0.329	0.233	0.267
Electricity	96	**0.133**	**0.223**	0.190	0.279	0.163	0.255	0.186	0.268	-	-	-	-	0.197	0.277	0.195	0.275
	192	**0.147**	**0.237**	0.195	0.285	0.194	0.288	0.193	0.276	-	-	-	-	0.197	0.279	0.195	0.277
	336	**0.166**	**0.265**	0.211	0.301	0.201	0.298	0.206	0.289	-	-	-	-	0.211	0.295	0.210	0.294
	720	**0.203**	**0.297**	0.253	0.333	0.263	0.343	0.250	0.324	-	-	-	-	0.255	0.330	0.252	0.327
	Avg	**0.162**	**0.256**	0.212	0.300	0.205	0.296	0.209	0.289	-	-	-	-	0.215	0.295	0.213	0.293
Traffic	96	**0.368**	**0.262**	0.471	0.309	0.448	0.298	0.595	0.360	-	-	-	-	0.378	0.365	0.480	0.357
	192	0.373	**0.251**	0.475	0.308	0.445	0.301	0.576	0.353	-	-	-	-	**0.371**	0.352	0.439	0.336
	336	**0.395**	**0.254**	0.490	0.315	0.492	0.320	0.569	0.362	-	-	-	-	0.467	0.354	0.460	0.344
	720	**0.432**	**0.290**	0.524	0.332	0.514	0.321	0.603	0.372	-	-	-	-	0.525	0.378	0.499	0.364
	Avg	**0.392**	**0.264**	0.490	0.316	0.475	0.310	0.586	0.362	-	-	-	-	0.435	0.362	0.470	0.350

Table 11: Complete results of long-term forecasting tasks for the cross-domain setting of forecasting the future $O \in \{96, 192, 336, 720\}$ time points based on the past 336 time points. **All the results of baselines are based on the encoder utilized in their original papers.** The standard deviations of SimMTM are within 0.005 for MSE and within 0.004 for MAE.

Models		**SimMTM**		Random init.		Ti-MAE [8]		TST [24]		LaST [17]		TF-C [25]		CoST [20]		TS2Vec [23]	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2 → ETTh1	96	0.372	**0.401**	0.380	0.412	0.703	0.562	0.653	0.468	**0.362**	0.420	0.596	0.569	0.378	0.421	0.849	0.694
	192	**0.414**	**0.425**	0.416	0.434	0.715	0.567	0.658	0.502	0.426	0.478	0.614	0.621	0.424	0.451	0.909	0.738
	336	**0.429**	**0.436**	0.448	0.458	0.733	0.579	0.631	0.561	0.522	0.509	0.694	0.664	0.651	0.582	1.082	0.775
	720	**0.446**	**0.458**	0.481	0.487	0.762	0.622	0.638	0.608	0.460	0.478	0.635	0.683	0.883	0.701	0.934	0.769
	Avg	**0.415**	**0.430**	0.431	0.448	0.728	0.583	0.645	0.535	0.443	0.471	0.635	0.634	0.584	0.539	0.944	0.744
ETTm1 → ETTh1	96	0.367	0.398	0.380	0.412	0.715	0.581	0.627	0.477	**0.360**	**0.374**	0.666	0.647	0.423	0.450	0.991	0.765
	192	0.396	0.421	0.416	0.434	0.729	0.587	0.628	0.500	**0.381**	**0.371**	0.672	0.653	0.641	0.578	0.829	0.699
	336	0.471	**0.437**	**0.448**	0.458	0.712	0.583	0.683	0.554	0.472	0.531	0.626	0.711	0.863	0.694	0.971	0.787
	720	**0.454**	**0.463**	0.481	0.487	0.747	0.627	0.642	0.600	0.490	0.488	0.835	0.797	1.071	0.805	1.037	0.820
	Avg	**0.422**	**0.430**	0.431	0.448	0.726	0.595	0.645	0.533	0.426	0.441	0.700	0.702	0.750	0.632	0.957	0.768
ETTm2 → ETTh1	96	0.388	0.421	0.380	**0.412**	0.699	0.566	0.559	0.489	0.428	0.454	0.968	0.738	**0.377**	0.419	0.783	0.669
	192	0.419	**0.423**	**0.416**	0.434	0.722	0.573	0.600	0.579	0.427	0.497	1.080	0.801	0.422	0.450	0.828	0.691
	336	**0.435**	**0.444**	0.448	0.458	0.714	0.569	0.677	0.572	0.528	0.540	1.091	0.824	0.648	0.580	0.990	0.762
	720	**0.468**	**0.474**	0.481	0.487	0.760	0.611	0.694	0.664	0.527	0.537	1.226	0.893	0.880	0.699	0.985	0.783
	Avg	**0.428**	**0.441**	0.431	0.448	0.724	0.580	0.632	0.576	0.503	0.507	1.091	0.814	0.582	0.537	0.896	0.726
Weather → ETTh1	96	0.477	0.444	**0.380**	**0.412**	-	-	-	-	-	-	-	-	-	-	-	-
	192	0.454	0.522	**0.416**	**0.434**	-	-	-	-	-	-	-	-	-	-	-	-
	336	**0.424**	**0.434**	0.448	0.458	-	-	-	-	-	-	-	-	-	-	-	-
	720	**0.468**	**0.469**	0.481	0.487	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	0.456	0.467	**0.431**	**0.448**	-	-	-	-	-	-	-	-	-	-	-	-
ETTh1 → ETTm1	96	0.290	0.348	0.295	0.346	0.667	0.521	0.425	0.381	0.295	0.387	0.672	0.600	**0.248**	**0.332**	0.605	0.561
	192	**0.327**	**0.372**	0.333	0.374	0.561	0.479	0.495	0.478	0.335	0.379	0.721	0.639	0.336	0.391	0.615	0.561
	336	**0.357**	0.392	0.370	0.398	0.690	0.533	0.456	0.441	0.379	**0.363**	0.755	0.664	0.381	0.421	0.763	0.677
	720	0.409	**0.423**	0.427	0.431	0.744	0.583	0.554	0.477	**0.403**	0.431	0.837	0.705	0.469	0.482	0.805	0.664
	Avg	**0.346**	**0.384**	0.356	0.387	0.666	0.529	0.482	0.444	0.353	0.390	0.746	0.652	0.359	0.407	0.697	0.616
ETTh2 → ETTm1	96	0.322	0.347	0.295	0.346	0.658	0.505	0.449	0.343	0.314	0.396	0.677	0.603	**0.253**	**0.342**	0.466	0.480
	192	**0.332**	**0.372**	0.333	0.374	0.594	0.511	0.477	0.407	0.587	0.545	0.718	0.638	0.367	0.392	0.557	0.532
	336	0.394	0.391	**0.370**	**0.398**	0.732	0.532	0.407	0.519	0.631	0.584	0.755	0.663	0.388	0.431	0.646	0.576
	720	**0.411**	**0.424**	0.427	0.431	0.768	0.592	0.557	0.523	0.368	0.429	0.848	0.712	0.498	0.488	0.752	0.638
	Avg	0.365	0.384	**0.356**	**0.387**	0.688	0.535	0.472	0.448	0.475	0.489	0.750	0.654	0.377	0.413	0.606	0.556
ETTm2 → ETTm1	96	0.297	0.348	0.295	0.346	0.647	0.497	0.471	0.422	0.304	0.388	0.610	0.577	**0.239**	**0.331**	0.586	0.515
	192	**0.332**	**0.370**	0.333	0.374	0.597	0.508	0.495	0.442	0.429	0.494	0.725	0.657	0.339	0.371	0.624	0.562
	336	**0.364**	**0.393**	0.370	0.398	0.700	0.525	0.455	0.424	0.499	0.523	0.768	0.684	0.371	0.421	1.035	0.806
	720	**0.410**	**0.421**	0.427	0.431	0.786	0.596	0.498	0.532	0.422	0.450	0.927	0.759	0.467	0.481	0.780	0.669
	Avg	**0.351**	**0.383**	0.356	0.387	0.682	0.531	0.480	0.455	0.414	0.464	0.758	0.669	0.354	0.401	0.756	0.638
Weather → ETTm1	96	0.304	0.354	**0.295**	**0.346**	-	-	-	-	-	-	-	-	-	-	-	-
	192	0.338	0.375	**0.333**	**0.374**	-	-	-	-	-	-	-	-	-	-	-	-
	336	0.371	**0.397**	**0.370**	0.398	-	-	-	-	-	-	-	-	-	-	-	-
	720	**0.417**	**0.426**	0.427	0.431	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	0.358	0.388	**0.356**	**0.387**	-	-	-	-	-	-	-	-	-	-	-	-

Table 12: Complete results of long-term forecasting tasks for the in-domain setting. **All the results of baseline are based on the unified channel-independent Transformer encoder.** The past sequence length is set as 336. The unified channel-independent transformer model can perform the transfer experiment between datasets with different variables. The standard deviations of SimMTM are within 0.005 for MSE and within 0.004 for MAE.

Models		**SimMTM**		Random init.		Ti-MAE [8]		TST [24]		LaST [17]		TF-C [25]		CoST [20]		TS2Vec [23]	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2 ↓ ETTh1	96	**0.372**	0.401	0.380	0.412	0.399	0.424	0.401	0.425	-	-	-	-	0.376	**0.362**	0.436	0.430
	192	0.414	0.425	0.416	0.434	0.454	0.440	0.531	0.484	-	-	-	-	**0.376**	**0.362**	0.455	0.440
	336	**0.429**	**0.436**	0.448	0.458	0.497	0.469	0.474	0.459	-	-	-	-	0.444	0.444	0.689	0.584
	720	**0.446**	0.458	0.481	0.487	0.515	0.492	0.471	0.469	-	-	-	-	0.517	0.510	0.489	0.490
	Avg	**0.415**	**0.430**	0.431	0.448	0.466	0.456	0.469	0.459	-	-	-	-	0.428	0.433	0.517	0.486
ETTm1 ↓ ETTh1	96	**0.367**	**0.398**	0.380	0.412	0.400	0.418	0.443	0.440	-	-	-	-	0.465	0.456	0.413	0.443
	192	**0.396**	**0.421**	0.416	0.434	0.434	0.445	0.471	0.455	-	-	-	-	0.722	0.588	0.459	0.465
	336	0.471	**0.437**	**0.448**	0.458	0.510	0.467	0.462	0.455	-	-	-	-	0.712	0.586	0.614	0.554
	720	0.454	**0.463**	0.481	0.487	0.636	0.544	0.525	0.503	-	-	-	-	0.581	0.533	**0.450**	0.464
	Avg	**0.422**	**0.430**	0.431	0.448	0.495	0.469	0.475	0.463	-	-	-	-	0.620	0.541	0.484	0.482
ETTm2 ↓ ETTh1	96	0.388	0.421	**0.380**	**0.412**	0.433	0.431	0.389	0.413	-	-	-	-	0.403	0.426	0.483	0.480
	192	0.419	**0.423**	**0.416**	0.434	0.474	0.458	0.463	0.452	-	-	-	-	0.457	0.468	0.579	0.537
	336	**0.435**	0.444	0.448	0.458	0.515	0.448	0.492	0.465	-	-	-	-	0.794	0.682	0.673	0.563
	720	**0.468**	**0.474**	0.481	0.487	0.496	0.488	0.468	0.468	-	-	-	-	0.739	0.617	0.729	0.620
	Avg	**0.428**	**0.441**	0.431	0.448	0.464	0.456	0.453	0.450	-	-	-	-	0.598	0.548	0.616	0.550
Weather ↓ ETTh1	96	0.477	0.444	**0.380**	0.412	0.397	0.440	0.428	0.429	-	-	-	-	0.421	0.410	0.393	**0.410**
	192	0.454	0.522	**0.416**	**0.434**	0.458	0.466	0.461	0.451	-	-	-	-	0.539	0.503	0.440	0.437
	336	**0.424**	**0.434**	0.448	0.458	0.479	0.458	0.463	0.456	-	-	-	-	0.568	0.514	0.450	0.451
	720	**0.468**	**0.469**	0.481	0.487	0.515	0.492	0.507	0.489	-	-	-	-	0.544	0.522	0.567	0.541
	Avg	0.456	0.467	**0.431**	**0.448**	0.462	0.464	0.465	0.456	-	-	-	-	0.518	0.487	0.463	0.460
ETTh1 ↓ ETTm1	96	**0.290**	0.348	0.295	**0.346**	0.311	0.355	0.315	0.354	-	-	-	-	0.308	0.355	0.681	0.545
	192	0.327	**0.372**	0.333	0.374	0.337	**0.372**	0.365	0.391	-	-	-	-	0.357	0.390	0.689	0.551
	336	**0.357**	**0.392**	0.370	0.398	0.372	0.398	0.384	0.400	-	-	-	-	0.396	0.402	0.705	0.560
	720	**0.409**	**0.423**	0.427	0.431	0.422	0.433	0.428	0.426	-	-	-	-	0.419	0.423	0.722	0.571
	Avg	**0.346**	**0.384**	0.356	0.387	0.360	0.390	0.373	0.393	-	-	-	-	0.370	0.393	0.699	0.557
ETTh2 ↓ ETTm1	96	0.322	0.347	**0.295**	**0.346**	0.323	0.362	0.338	0.383	-	-	-	-	0.322	0.351	0.679	0.546
	192	0.332	**0.372**	0.333	0.374	0.370	0.395	0.394	0.408	-	-	-	-	**0.331**	0.373	0.673	0.551
	336	0.394	**0.391**	**0.370**	0.398	0.397	0.413	0.401	0.412	-	-	-	-	0.382	0.397	0.703	0.557
	720	**0.411**	**0.424**	0.427	0.431	0.442	0.439	0.434	0.432	-	-	-	-	0.417	0.428	0.722	0.573
	Avg	0.365	**0.384**	**0.356**	0.387	0.383	0.402	0.391	0.409	-	-	-	-	0.363	0.387	0.694	0.557
ETTm2 ↓ ETTm1	96	0.297	0.348	**0.295**	**0.346**	0.333	0.378	0.327	0.364	-	-	-	-	0.320	0.364	0.422	0.434
	192	**0.332**	**0.370**	0.333	0.374	0.381	0.398	0.362	0.389	-	-	-	-	0.367	0.386	0.387	0.371
	336	**0.364**	**0.393**	0.370	0.398	0.394	0.413	0.401	0.418	-	-	-	-	0.374	0.394	0.402	0.444
	720	**0.410**	**0.421**	0.427	0.431	0.455	0.453	0.437	0.437	-	-	-	-	0.479	0.503	0.481	0.432
	Avg	**0.351**	**0.383**	0.356	0.387	0.390	0.410	0.382	0.402	-	-	-	-	0.385	0.412	0.423	0.420
Weather ↓ ETTm1	96	**0.294**	0.354	0.295	**0.346**	0.338	0.380	0.324	0.366	-	-	-	-	0.324	0.360	0.329	0.359
	192	**0.318**	**0.355**	0.333	0.374	0.473	0.457	0.349	0.377	-	-	-	-	0.359	0.387	0.392	0.392
	336	**0.361**	**0.397**	0.370	0.398	0.402	0.415	0.378	0.398	-	-	-	-	0.395	0.399	0.372	0.400
	720	0.427	**0.426**	0.427	0.431	0.432	0.438	**0.422**	0.427	-	-	-	-	0.450	0.467	0.434	0.429
	Avg	**0.350**	**0.383**	0.356	0.387	0.411	0.423	0.368	0.392	-	-	-	-	0.382	0.403	0.382	0.395

Table 13: Full ablation studies for the in-domain setting of forecasting. The standard deviations of SimMTM are within 0.005 for MSE and within 0.004 for MAE.

Input-336		Supervised		W/o $\mathcal{L}_{\text{reconstruction}}$		W/o $\mathcal{L}_{\text{constraint}}$		**SimMTM**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.380	0.412	**0.377**	0.408	0.381	0.409	0.379	**0.407**
	192	0.416	0.434	0.419	0.443	**0.409**	0.443	0.412	**0.424**
	336	0.448	0.458	0.423	0.434	0.432	0.444	**0.421**	**0.431**
	720	0.481	0.487	0.437	0.454	0.447	0.454	**0.424**	**0.449**
	Avg	0.431	0.448	0.414	0.435	0.417	0.438	**0.409**	**0.428**
ETTh2	96	0.325	0.374	**0.288**	**0.344**	0.312	0.365	0.293	0.347
	192	0.400	0.424	0.356	0.391	0.389	0.418	**0.355**	**0.386**
	336	0.405	0.433	**0.368**	0.406	0.396	0.432	0.370	**0.401**
	720	0.451	0.475	0.409	0.432	0.448	0.479	**0.395**	**0.427**
	Avg	0.395	0.427	0.355	0.393	0.386	0.424	**0.353**	**0.390**
ETTm1	96	0.295	0.346	0.291	0.343	**0.282**	**0.337**	0.288	0.348
	192	0.333	0.374	0.330	0.390	**0.324**	0.388	0.327	**0.373**
	336	0.370	0.398	0.369	0.399	0.366	0.397	**0.363**	**0.395**
	720	0.427	0.431	0.417	0.429	0.424	0.435	**0.412**	**0.424**
	Avg	0.356	0.387	0.352	0.390	0.349	0.389	**0.348**	**0.385**
ETTm2	96	0.175	0.268	0.174	0.265	**0.170**	**0.261**	0.172	**0.261**
	192	0.240	0.312	0.232	0.303	0.244	0.320	**0.223**	**0.300**
	336	0.298	0.351	0.313	0.365	**0.279**	0.334	0.282	**0.331**
	720	0.403	0.413	0.376	0.451	0.376	**0.378**	**0.374**	0.388
	Avg	0.279	0.336	0.274	0.346	0.267	0.323	**0.263**	**0.320**
Weather	96	0.166	0.216	0.164	**0.209**	0.160	0.212	**0.158**	0.211
	192	0.208	0.254	0.203	0.258	0.203	0.251	**0.199**	**0.249**
	336	0.257	0.290	0.244	0.289	0.253	0.290	**0.246**	**0.286**
	720	0.326	0.338	0.322	0.343	0.325	0.340	**0.317**	**0.337**
	Avg	0.239	0.275	0.233	0.275	0.235	0.273	**0.230**	**0.271**
Electricity	96	0.190	0.279	0.177	0.270	0.134	**0.220**	**0.133**	0.223
	192	0.195	0.285	0.184	0.279	0.163	0.274	**0.147**	**0.237**
	336	0.211	0.301	0.202	0.300	0.223	0.311	**0.166**	**0.265**
	720	0.253	0.333	0.250	0.337	0.241	0.321	**0.203**	**0.297**
	Avg	0.212	0.300	0.203	0.397	0.190	0.282	**0.162**	**0.256**
Traffic	96	0.471	0.309	**0.366**	**0.257**	0.457	0.301	0.368	0.262
	192	0.475	0.308	**0.373**	0.266	0.468	0.325	**0.373**	**0.251**
	336	0.490	0.315	0.401	0.249	0.487	0.302	**0.395**	**0.254**
	720	0.524	0.332	0.472	0.312	0.485	0.315	**0.432**	**0.290**
	Avg	0.490	0.316	0.403	0.271	0.474	0.311	**0.392**	**0.264**

Table 14: Full ablation studies on transfer to ETTh1 and ETTm1 for the cross-domain setting of forecasting. The standard deviations of SimMTM are within 0.005 for MSE and within 0.004 for MAE.

Input-336		Supervised		W/o $\mathcal{L}_{\text{reconstruction}}$		W/o $\mathcal{L}_{\text{constraint}}$		**SimMTM**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2 ↓ ETTh1	96	0.380	0.412	0.377	**0.400**	0.402	0.411	**0.372**	0.401
	192	0.416	0.434	0.417	0.424	0.417	**0.420**	**0.414**	0.425
	336	0.448	0.458	0.437	0.439	0.437	**0.435**	**0.429**	0.436
	720	0.481	0.487	0.448	0.463	0.456	0.467	**0.446**	**0.458**
	Avg	0.431	0.448	0.420	0.432	0.423	**0.430**	**0.415**	**0.430**
ETTm1 ↓ ETTh1	96	0.380	0.412	0.382	**0.397**	0.375	0.399	**0.367**	0.398
	192	0.416	0.434	0.418	**0.418**	0.413	0.422	**0.396**	0.421
	336	0.448	0.458	0.437	**0.434**	**0.434**	0.438	0.471	0.437
	720	0.481	0.487	0.459	0.469	0.467	0.475	**0.454**	**0.463**
	Avg	0.431	0.448	0.424	0.430	0.422	0.434	**0.422**	**0.430**
ETTm2 ↓ ETTh1	96	**0.380**	**0.412**	0.388	0.418	0.384	0.415	0.388	0.421
	192	**0.416**	0.434	0.429	0.444	0.423	0.439	0.419	**0.423**
	336	0.448	0.458	0.467	0.472	0.458	0.465	**0.435**	**0.444**
	720	0.481	0.487	0.521	0.507	0.501	0.497	**0.468**	**0.474**
	Avg	0.431	0.448	0.451	0.460	0.441	0.454	**0.428**	**0.441**
Weather ↓ ETTh1	96	**0.380**	0.412	0.385	**0.400**	0.394	0.406	0.477	0.444
	192	**0.416**	0.434	0.417	0.429	0.425	**0.424**	0.454	0.522
	336	0.448	0.458	0.434	**0.434**	0.441	0.439	**0.424**	**0.434**
	720	0.481	0.487	**0.444**	**0.464**	0.446	0.468	0.468	0.469
	Avg	0.431	0.448	**0.420**	**0.432**	0.427	0.434	0.456	0.467
ETTh1 ↓ ETTm1	96	0.295	0.346	**0.286**	**0.341**	0.290	0.346	0.290	0.348
	192	0.333	0.374	**0.322**	**0.362**	0.353	0.388	0.327	0.372
	336	0.370	0.398	0.362	0.418	0.362	0.412	**0.357**	**0.392**
	720	0.427	0.431	0.417	0.431	0.422	0.432	**0.409**	**0.423**
	Avg	0.356	0.387	0.347	0.388	0.357	0.395	**0.346**	**0.384**
ETTh2 ↓ ETTm1	96	**0.295**	**0.346**	0.299	0.348	0.301	0.352	0.322	0.347
	192	0.333	0.374	**0.324**	0.366	0.332	**0.359**	0.332	0.372
	336	**0.370**	0.398	0.374	0.401	0.389	**0.382**	0.394	0.391
	720	0.427	0.431	0.415	**0.419**	0.421	0.442	**0.411**	0.424
	Avg	0.356	0.387	**0.353**	0.386	0.361	**0.384**	0.365	**0.384**
ETTm2 ↓ ETTm1	96	0.295	0.346	0.299	0.351	**0.285**	**0.336**	0.297	0.348
	192	0.333	0.374	0.334	0.372	0.343	**0.366**	**0.332**	0.370
	336	0.370	0.398	0.362	**0.388**	**0.360**	0.399	0.364	0.393
	720	0.427	0.431	0.417	0.431	0.422	0.432	**0.410**	**0.421**
	Avg	0.356	0.387	0.353	0.386	0.353	**0.383**	**0.351**	**0.383**
Weather ↓ ETTm1	96	0.295	**0.346**	0.322	0.361	0.309	0.354	**0.294**	0.354
	192	0.333	0.374	0.344	0.378	0.343	0.365	**0.318**	**0.355**
	336	0.370	0.398	0.371	0.399	0.401	0.411	**0.361**	**0.397**
	720	0.427	0.431	0.426	**0.422**	**0.425**	0.427	0.427	0.426
	Avg	0.356	0.387	0.366	0.390	0.370	0.389	**0.350**	**0.383**

Table 15: Full results for applying SimMTM to four advanced time series forecasting models under the in-domain setting. The gray mark represents negative transfer (↓).

Models		Transformer [16]				Autoformer [21]				Ns Transformer [10]				PatchTST [11]					
		Random init.		**+SimMTM**		Random init.		**+SimMTM**		Random init.		**+SimMTM**		Random init.		+Sub-serie Masking		**+SimMTM**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.847	0.731	0.775	0.691	0.536	0.548	0.526	0.536	0.513	0.491	0.490	0.489	0.375	0.399	0.366	0.398	0.373	0.399
	192	1.084	0.841	0.918	0.763	0.543	0.551	0.523	0.548	0.534	0.504	0.517	0.499	0.414	0.421	0.431	0.443	0.406	0.428
	336	1.350	0.956	1.079	0.845	0.615	0.592	0.595	0.591	0.588	0.535	0.552	0.520	0.431	0.436	0.450↓	0.456↓	0.422	0.431
	720	1.069	0.817	0.935	0.761	0.599	0.600	0.600	0.597	0.643	0.616	0.614	0.598	0.449	0.466	0.472↓	0.484↓	0.436	0.452
	Avg	1.088	0.836	**0.927**	**0.761**	0.573	0.573	**0.561**	**0.568**	0.570	0.537	**0.543**	**0.527**	0.417	0.431	0.430↓	0.445↓	**0.409**	**0.428**
ETTh2	96	2.029	1.150	1.879	1.104	0.492	0.517	0.488	0.514	0.476	0.458	0.445	0.448	0.274	0.336	0.284↓	0.343↓	0.274	0.337
	192	6.785	2.099	5.054	1.771	0.556	0.551	0.547	0.549	0.512	0.493	0.482	0.502	0.339	0.379	0.355↓	0.387↓	0.339	0.377
	336	4.568	1.711	4.242	1.658	0.572	0.578	0.563	0.570	0.552	0.551	0.512	0.537	0.331	0.380	0.379↓	0.411↓	0.327	0.381
	720	3.030	1.486	2.815	1.413	0.580	0.588	0.575	0.588	0.562	0.560	0.531	0.568	0.379	0.422	0.400↓	0.435↓	0.375	0.423
	Avg	4.103	1.612	**3.498**	**1.487**	0.550	0.559	**0.543**	**0.555**	0.526	0.516	**0.493**	**0.514**	0.331	0.379	0.355↓	0.394↓	**0.329**	0.379
ETTm1	96	0.562	0.520	0.513	0.497	0.523	0.488	0.482	0.465	0.386	0.398	0.340	0.376	0.290	0.342	0.289↓	0.344↓	0.288	0.343
	192	0.810	0.668	0.686	0.606	0.543	0.498	0.499	0.476	0.459	0.444	0.423	0.445	0.332	0.369	0.323	0.368	0.329	0.367
	336	1.096	0.814	1.003	0.760	0.675	0.551	0.601	0.524	0.495	0.464	0.423	0.459	0.366	0.392	0.353	0.387	0.361	0.387
	720	1.136	0.813	1.032	0.790	0.720	0.528	0.629	0.555	0.585	0.516	0.539	0.499	0.420	0.424	0.398	0.416	0.413	0.417
	Avg	0.901	0.704	**0.809**	**0.663**	0.615	0.528	**0.553**	**0.505**	0.481	0.456	**0.431**	**0.445**	0.352	0.382	**0.341**	0.379	0.348	**0.378**
ETTm2	96	0.508	0.539	0.336	0.425	0.255	0.339	0.255	0.340	0.192	0.274	0.188	0.277	0.165	0.255	0.166↓	0.256↓	0.163	0.253
	192	0.972	0.721	0.713	0.610	0.281	0.340	0.276	0.332	0.280	0.339	0.277	0.336	0.220	0.292	0.221↓	0.295↓	0.219	0.292
	336	1.419	0.897	1.517	0.942	0.339	0.372	0.309	0.359	0.334	0.361	0.325	0.355	0.278	0.329	0.278	0.333↓	0.275	0.328
	720	3.598	1.445	2.720	1.254	0.422	0.419	0.420	0.410	0.417	0.413	0.414	0.412	0.367	0.385	0.365	0.388↓	0.359	0.381
	Avg	1.624	0.901	**1.322**	**0.808**	0.324	0.368	**0.315**	**0.360**	0.306	0.347	**0.301**	**0.345**	0.258	0.317	0.258	0.318↓	**0.254**	**0.313**

Table 16: Full results for fine-tuning to limited data scenarios. We fine-tune the model pre-trained from ETTh2 to ETTh1 with different data proportions $\{10\%, 25\%, 50\%, 75\%, 100\%\}$.

Models		**SimMTM**		Random init.		Ti-MAE[8]		TST[24]		LaST[17]		TF-C[25]		CoST[20]		TS2Vec[23]	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MAE	MSE
ETTh2 ↓ ETTh1	10%	**0.591**	**0.523**	0.653	0.558	0.660	0.517	0.783	0.588	0.645	0.507	0.799	0.783	0.784	0.604	0.655	0.550
	25%	**0.535**	**0.490**	0.632	0.502	0.594	0.518	0.641	0.578	0.610	0.611	0.736	0.725	0.624	0.539	0.632	0.543
	50%	**0.491**	**0.473**	0.512	0.479	0.550	0.504	0.525	0.509	0.540	0.513	0.731	0.704	0.540	0.499	0.599	0.526
	75%	**0.466**	**0.458**	0.499	0.488	0.475	0.465	0.516	0.488	0.479	0.470	0.697	0.689	0.494	0.475	0.577	0.534
	100%	**0.415**	**0.430**	0.431	0.448	0.466	0.456	0.469	0.459	0.443	0.471	0.635	0.634	0.428	0.433	0.517	0.486

Table 17: In- and cross-domain settings of classification, where **all the baselines are based on the encoder utilized in their original papers**. For in-domain setting, we pre-train and fine-tune on the same dataset: Epilepsy. For cross-domain setting, we pre-train the model on SleepEEG and then fine-tune it on different datasets: Epilepsy, FD-B, Gesture, and EMG.

	Scenarios	Models	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Avg (%)
In-Domain	Epilepsy ↓ Epilepsy	Random init.	89.83	92.13	74.47	79.59	84.00
		TS2vec [23]	92.17	93.84	81.19	85.71	88.23
		CoST[20]	88.07	91.58	66.05	69.11	78.70
		LaST [17]	92.11	93.12	81.47	85.74	88.11
		TST [24]	80.21	40.11	50.00	44.51	53.71
		Ti-MAE [8]	90.09	93.90	77.24	78.21	84.86
		TF-C [25]	93.96	94.87	85.82	89.46	91.03
		SimMTM	**94.75**	**95.60**	**89.93**	**91.41**	**92.92**
Cross-Domain	SleepEEG ↓ Epilepsy	Random init.	89.83	92.13	74.47	79.59	84.00
		TS2vec [23]	93.95	90.59	90.39	90.45	91.35
		CoST[20]	88.40	88.20	72.34	76.88	81.45
		LaST [17]	86.46	90.77	66.35	70.67	78.56
		TST [24]	80.21	40.11	50.00	44.51	53.71
		Ti-MAE [8]	89.71	72.36	67.47	68.55	74.52
		TF-C [25]	94.95	**94.56**	89.08	91.49	92.52
		SimMTM	**95.49**	93.36	**92.28**	**92.81**	**93.49**
	SleepEEG ↓ FD-B	Random init.	47.36	48.29	52.35	49.11	49.28
		TS2vec [23]	47.90	43.39	48.42	43.89	45.90
		CoST[20]	47.06	38.79	38.42	34.79	39.76
		LaST [17]	46.67	43.90	47.71	45.17	45.86
		TST [24]	46.40	41.58	45.50	41.34	43.71
		Ti-MAE [8]	60.88	66.98	68.94	66.56	65.84
		TF-C [25]	69.38	**75.59**	72.02	74.87	72.97
		SimMTM	**69.40**	74.18	**76.41**	**75.11**	**73.78**
	SleepEEG ↓ Gesture	Random init.	42.19	47.51	49.63	48.86	47.05
		TS2vec [23]	69.17	65.45	68.54	65.70	67.22
		CoST[20]	68.33	65.30	68.33	66.42	67.09
		LaST [17]	64.17	70.36	64.17	58.76	64.37
		TST [24]	69.17	66.60	69.17	66.01	67.74
		Ti-MAE [8]	71.88	70.35	76.75	68.37	71.84
		TF-C [25]	76.42	77.31	74.29	75.72	75.94
		SimMTM	**80.00**	**79.03**	**80.00**	**78.67**	**79.43**
	SleepEEG ↓ EMG	Random init.	77.80	59.09	66.67	62.38	66.49
		TS2vec [23]	78.54	80.40	67.85	67.66	73.61
		CoST[20]	53.65	49.07	42.10	35.27	45.02
		LaST [17]	66.34	79.34	63.33	72.55	70.39
		TST [24]	78.34	77.11	80.30	68.89	76.16
		Ti-MAE [8]	69.99	70.25	63.44	70.89	68.64
		TF-C [25]	81.71	72.65	81.59	76.83	78.20
		SimMTM	**97.56**	**98.33**	**98.04**	**98.14**	**98.02**

Table 18: In- and cross-domain settings of classification **based on the unified 1-D ResNet encoder**. For in-domain setting, we pre-train and fine-tune on the same dataset: Epilepsy. For cross-domain setting, we pre-train on SleepEEG and fine-tune on different domain datasets: Epilepsy, FD-B, Gesture, and EMG.

Scenarios		Models	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Avg (%)
In-Domain	Epilepsy ↓ Epilepsy	Random init.	89.83	92.13	74.47	79.59	84.00
		TS2vec [23]	92.33	94.53	81.11	86.33	88.58
		CoST[20]	92.35	94.73	81.16	85.92	88.54
		LaST [17]	-	-	-	-	-
		TST [24]	80.89	90.38	51.73	48.01	67.75
		Ti-MAE [8]	80.34	90.16	50.33	45.20	66.51
		TF-C [25]	93.96	94.87	85.82	89.46	91.03
		SimMTM	**94.75**	**95.60**	**89.93**	**91.41**	**92.92**
Cross-Domain	SleepEEG ↓ Epilepsy	Random init.	89.83	92.13	74.47	79.59	84.00
		TS2vec [23]	94.46	91.99	90.28	91.10	91.95
		CoST[20]	93.66	91.39	88.08	89.60	90.68
		LaST [17]	-	-	-	-	-
		TST [24]	82.89	86.15	79.02	80.44	82.13
		Ti-MAE [8]	73.45	72.56	65.34	77.20	72.14
		TF-C [25]	94.95	**94.56**	89.08	91.49	92.52
		SimMTM	**95.49**	93.36	**92.28**	**92.81**	**93.49**
	SleepEEG ↓ FD-B	Random init.	47.36	48.29	52.35	49.11	49.28
		TS2vec [23]	60.74	59.60	64.27	61.07	61.42
		CoST[20]	54.82	51.92	63.30	54.34	56.09
		LaST [17]	-	-	-	-	-
		TST [24]	65.57	70.05	67.57	64.41	66.90
		Ti-MAE [8]	67.98	62.83	64.45	63.36	64.66
		TF-C [25]	69.38	**75.59**	72.02	74.87	72.97
		SimMTM	**69.40**	74.18	**76.41**	**75.11**	**73.78**
	SleepEEG ↓ Gesture	Random Init.	42.19	47.51	49.63	48.86	47.05
		TS2vec [23]	73.33	70.88	73.33	71.56	72.27
		CoST[20]	73.33	74.37	73.33	71.16	73.04
		LaST [17]	-	-	-	-	-
		TST [24]	75.12	76.05	67.74	73.24	73.04
		Ti-MAE [8]	75.54	69.32	72.42	69.32	71.65
		TF-C [25]	76.42	77.31	74.29	75.72	75.94
		SimMTM	**80.00**	**79.03**	**80.00**	**78.67**	**79.43**
	SleepEEG ↓ EMG	Random init.	77.80	59.09	66.67	62.38	66.49
		TS2vec [23]	80.92	69.63	67.65	67.90	71.52
		CoST[20]	73.17	70.47	69.84	70.00	70.87
		LaST [17]	-	-	-	-	-
		TST [24]	75.89	74.67	80.66	78.48	77.43
		Ti-MAE [8]	63.52	67.77	70.55	58.32	65.04
		TF-C [25]	81.71	72.65	81.59	76.83	78.20
		SimMTM	**97.56**	**98.33**	**98.04**	**98.14**	**98.02**

Table 19: Full ablation studies for in-domain and cross-domain settings of classification. Under the *Avg* metric, the standard deviations of SimMTM are within 0.2% for Epilepsy, within 0.5% for FD-B, within 0.6% for Gesture, and within 0.1% for EMG.

Scenarios		Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Avg (%)
Epilepsy ↓ Epilepsy	Random init.	89.83	92.13	74.47	79.59	84.00
	W/o $\mathcal{L}_{\text{reconstruction}}$	93.80	**96.11**	86.11	89.45	91.37
	W/o $\mathcal{L}_{\text{constraint}}$	90.99	92.81	79.86	84.13	86.95
	SimMTM	**94.75**	95.60	**89.93**	**91.41**	**92.92**
SleepEEG ↓ Epilepsy	Random init.	89.83	92.13	74.47	79.59	84.00
	W/o $\mathcal{L}_{\text{reconstruction}}$	94.54	**93.87**	88.46	90.84	91.93
	W/o $\mathcal{L}_{\text{constraint}}$	91.73	90.57	82.21	85.53	87.51
	SimMTM	**95.49**	93.36	**92.28**	**92.81**	**93.49**
SleepEEG ↓ FD-B	Random init.	47.36	48.29	52.35	49.11	49.28
	W/o $\mathcal{L}_{\text{reconstruction}}$	66.11	67.97	74.70	70.01	69.70
	W/o $\mathcal{L}_{\text{constraint}}$	53.71	69.48	62.67	50.86	59.18
	SimMTM	**69.40**	**74.18**	**76.41**	**75.11**	**73.78**
SleepEEG ↓ Gesture	Random init.	42.19	47.51	49.63	48.86	47.05
	W/o $\mathcal{L}_{\text{reconstruction}}$	78.50	79.01	78.50	77.17	78.30
	W/o $\mathcal{L}_{\text{constraint}}$	76.67	74.91	76.67	74.80	75.76
	SimMTM	**80.00**	**79.03**	**80.00**	**78.67**	**79.43**
SleepEEG ↓ EMG	Random init.	77.80	59.09	66.67	62.38	66.49
	W/o $\mathcal{L}_{\text{reconstruction}}$	90.24	94.20	78.04	81.53	86.00
	W/o $\mathcal{L}_{\text{constraint}}$	85.37	89.97	69.62	70.74	78.93
	SimMTM	**97.56**	**98.33**	**98.04**	**98.14**	**98.02**

References

[1] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 2001.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*, 2018.

[3] Tianyu Gao, Adam Fisch, and Danqi Chen. Making pre-trained language models better few-shot learners. In *IJCNLP*, 2020.

[4] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, 2022.

[5] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[6] Bob Kemp, Aeilko H Zwinderman, Bert Tuk, Hilbert AC Kamphuisen, and Josefien JL Oberye. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg. *TBME*, 2000.

[7] Christian Lessmeier, James Kuria Kimotho, Detmar Zimmer, and Walter Sextro. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *PHM*, 2016.

[8] Zhe Li, Zhongwen Rao, Lujia Pan, Pengyun Wang, and Zenglin Xu. Ti-mae: Self-supervised masked time series autoencoders. *arXiv preprint arXiv:2301.08871*, 2023.

[9] Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. In *PMC*, 2009.

[10] Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *NeurIPS*, 2022.

[11] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *ICLR*, 2023.

[12] Adam Paszke, S. Gross, Francisco Massa, A. Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Z. Lin, N. Gimelshein, L. Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.

[13] PeMS. Traffic Dataset. `http://pems.dot.ca.gov/`.

[14] PhysioToolkit PhysioBank. Physionet: components of a new research resource for complex physiologic signals. *Circulation*, 2000.

[15] UCI. UCI Electricity Load Time Series Dataset. `https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014`.

[16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

[17] Zhiyuan Wang, Xovee Xu, Weifeng Zhang, Goce Trajcevski, Ting Zhong, and Fan Zhou. Learning latent seasonal-trend representations for time series forecasting. In *NeurIPS*, 2022.

[18] Wetterstation. Weather Dataset. `https://www.bgc-jena.mpg.de/wetter/`.

[19] Alexander Wettig, Tianyu Gao, Zexuan Zhong, and Danqi Chen. Should you mask 15% in masked language modeling? In *EACL*, 2023.

[20] Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. In *ICLR*, 2022.

[21] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *NeurIPS*, 2021.

[22] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: A simple framework for masked image modeling. In *CVPR*, 2022.

[23] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. TS2Vec: Towards Universal Representation of Time Series. In *AAAI*, 2022.

[24] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. A transformer-based framework for multivariate time series representation learning. In *SIGKDD*, 2021.

[25] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. In *NeurIPS*, 2022.

[26] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *AAAI*, 2021.