Variable-Dynamic Multivariate Time Series Forecasting for IoT Systems

This is the appendix to the paper entitled ‘Variable-Dynamic Multivariate Time Series Forecasting for IoT Systems’, submitted to IEEE Internet of Things Journal.

APPENDIX A
MORE DETAILS OF THE EMPLOYED BENCHMARKS

This section provides more details of the 10 employed benchmarks in the main text:

1) ETT (Electricity Transformer Temperature) [1]: This dataset owns two types of subsets: 1-hour-level datasets $\{\mathrm{ETTh}_1, \mathrm{ETTh}_2\}$ and 15-min-level datasets $\{\mathrm{ETTm}_1, \mathrm{ETTm}_2\}$. They are respectively composed of 2-year data from two electric stations in China.

2) ECL (Electricity Consuming Load) [2]: This dataset contains the electricity consumption (Kwh) of 321 clients lasting for 2 years. We follow the usage of [1] to transform it into a 1-hour-level dataset.

3) Traffic [3]: This dataset consists of the hourly road occupation rates in San Francisco Bay area freeways for 2 years.

4) Weather [4]: This dataset is a 10-minute-level dataset which describes the values of 21 meteorological indicators in Germany during 2020.

5) Solar [5]: This dataset describes the solar power productions of 137 different photovoltaic (PV) power plants every 10 minutes in Alabama State during 2006.

6) Air [6]: This 1-hour-level dataset is with respect to the air quality and provided by UCI Machine Learning Repository. It contains 12 responses that are collected by a gas multi-sensor device in an Italian city during 2004 and 2005. The missing values, which are tagged with -200, are replaced by the average of total normal values by us.

7) River [7]: This dataset is comprised of the hourly instances for the river flow observations of eight sites for about a year.

8) HomeC [8]: This datasest is a 1-minute-level dataset which contains the electricity consumption (Kwh) of house appliances and the weather conditions in a particular region. The data is sampled in the experiment so that the interval is 15 minutes to align with the settings of other minute-level datasets.

TABLE B.1
THE COMPARISON WITH VARIABLE-INDEPENDENT MODELS AND VARIABLE-SPECIFIC MODELS (INPUT LENGTH = 96)

Methods	Metrics	ETTh$_1$				ETTh$_2$				ECL				Traffic			
		192	336	720	Avg.	192	336	720	Avg.	192	336	720	Avg.	192	336	720	Avg.
VDformer	MSE	***0.412***	***0.434***	***0.461***	***0.436***	***0.321***	***0.358***	***0.412***	***0.364***	***0.146***	***0.162***	***0.198***	***0.169***	***0.397***	***0.412***	***0.451***	***0.420***
	MAE	***0.414***	***0.425***	***0.459***	***0.433***	***0.361***	***0.392***	***0.432***	***0.395***	***0.241***	***0.258***	***0.289***	***0.263***	***0.245***	***0.254***	***0.275***	***0.258***
Triformer	MSE	0.484	0.513	0.711	0.569	1.028	1.049	1.223	1.100	*0.191*	*0.209*	*0.253*	*0.218*	—	—	—	—
	MAE	0.486	0.489	0.638	0.538	0.708	0.745	0.836	0.763	0.289	0.308	0.344	0.314	—	—	—	—
PatchTST	MSE	*0.438*	*0.482*	*0.492*	*0.471*	*0.380*	*0.417*	*0.425*	*0.407*	0.199	0.214	0.256	0.223	*0.579*	*0.589*	*0.631*	*0.599*
	MAE	0.459	0.494	*0.519*	0.491	*0.395*	*0.428*	*0.443*	*0.422*	0.283	*0.299*	*0.331*	*0.305*	0.368	0.371	*0.391*	*0.377*
FiLM	MSE	0.474	0.510	0.557	0.514	0.384	0.425	0.443	0.417	0.199	0.217	0.279	0.232	0.600	0.610	0.694	0.635
	MAE	*0.452*	*0.468*	0.521	*0.480*	0.397	0.436	0.454	0.429	*0.278*	0.300	0.357	0.312	*0.362*	*0.367*	0.427	0.385
DLinear-i	MSE	0.492	0.534	0.553	0.526	0.872	1.050	1.267	1.063	0.219	0.232	0.268	0.240	0.750	0.756	0.799	0.768
	MAE	0.472	0.493	0.527	0.497	0.617	0.687	0.764	0.690	0.320	0.333	0.362	0.338	0.471	0.474	0.491	0.479

TABLE B.2
THE COMPARISON WITH MODELS THAT ANALYZED THE IMPORTANCE OF VARIABLE RELATIONSHIPS (INPUT LENGTH = 96)

Methods	Metrics	ETTh$_1$				ETTh$_2$				ECL				Traffic			
		192	336	720	Avg.	192	336	720	Avg.	192	336	720	Avg.	192	336	720	Avg.
VDformer	MSE	***0.412***	***0.434***	***0.461***	***0.436***	***0.321***	***0.358***	***0.412***	***0.364***	***0.146***	***0.162***	***0.198***	***0.169***	***0.397***	***0.412***	***0.451***	***0.420***
	MAE	***0.414***	***0.425***	***0.459***	***0.433***	***0.361***	***0.392***	***0.432***	***0.395***	***0.241***	***0.258***	***0.289***	***0.263***	***0.245***	***0.254***	***0.275***	***0.258***
LIFT	MSE	0.425	0.479	*0.510*	*0.472*	0.365	*0.435*	0.444	0.415	0.172	*0.186*	0.231	0.196	0.456	0.461	0.502	0.473
	MAE	0.432	0.449	*0.491*	0.457	0.395	*0.422*	0.438	0.418	0.260	*0.269*	0.308	0.279	*0.303*	*0.310*	*0.332*	*0.315*
Ada-MSHyper	MSE	0.513	0.581	0.633	0.576	0.461	0.520	0.503	0.495	0.204	0.227	0.277	0.236	0.521	0.567	0.628	0.572
	MAE	0.497	0.540	0.556	0.531	0.452	0.502	0.502	0.485	0.299	0.329	0.365	0.331	0.379	0.365	0.395	0.380
CCM	MSE	*0.420*	*0.474*	0.521	*0.472*	*0.361*	*0.435*	*0.440*	*0.412*	*0.165*	0.190	*0.219*	*0.191*	*0.442*	*0.456*	*0.487*	*0.462*
	MAE	*0.428*	*0.440*	0.496	*0.454*	*0.391*	0.427	*0.429*	*0.416*	*0.246*	*0.269*	*0.301*	*0.272*	0.309	*0.310*	0.343	0.321

TABLE B.3
THE COMPARISON WITH SPATIAL-TEMPORAL FORECASTING MODELS

Methods	Metrics	PeMSD3 192	PeMSD3 336	PeMSD3 720	PeMSD3 Avg.	PeMSD4 192	PeMSD4 336	PeMSD4 720	PeMSD4 Avg.	PeMSD7 192	PeMSD7 336	PeMSD7 720	PeMSD7 Avg.	PeMSD8 192	PeMSD8 336	PeMSD8 720	PeMSD8 Avg.
VDformer	MAE	***0.220***	***0.223***	***0.248***	***0.230***	***0.214***	***0.217***	***0.219***	***0.217***	***0.166***	***0.175***	***0.176***	***0.172***	***0.194***	***0.195***	***0.203***	***0.197***
	RMSE	***0.376***	***0.380***	***0.425***	***0.393***	***0.340***	***0.348***	***0.355***	***0.348***	***0.280***	***0.292***	***0.321***	***0.298***	***0.336***	***0.340***	***0.361***	***0.346***
	MAPE	***40.532***	***40.475***	***44.604***	***41.870***	***38.395***	***37.283***	***37.272***	***37.650***	***32.823***	***32.704***	***34.599***	***33.376***	***35.970***	***35.428***	***36.415***	***35.938***
SSTBAN	MAE	0.306	0.324	0.360	0.330	0.298	0.329	0.349	0.325	0.276	0.350	0.307	0.311	0.370	0.460	0.382	0.404
	RMSE	0.454	0.481	0.523	0.486	0.424	0.465	0.498	0.462	0.419	0.537	0.472	0.476	0.535	0.645	0.544	0.575
	MAPE	51.974	54.575	58.103	54.884	48.396	52.341	54.056	51.598	48.140	54.830	50.812	51.261	57.583	67.138	58.679	61.133
Cy2Mixer	MAE	*0.275*	*0.279*	*0.297*	*0.284*	*0.247*	*0.247*	*0.268*	*0.254*	*0.238*	*0.249*	*0.278*	*0.255*	*0.296*	*0.273*	*0.303*	*0.290*
	RMSE	*0.411*	*0.415*	*0.452*	*0.426*	*0.368*	*0.373*	*0.400*	*0.380*	*0.395*	*0.407*	*0.445*	*0.416*	*0.428*	*0.419*	*0.464*	*0.437*
	MAPE	48.861	*49.276*	*51.154*	*49.763*	*42.947*	*43.005*	*45.216*	*43.723*	*43.599*	*44.331*	*47.777*	*45.236*	*50.708*	*47.162*	*50.454*	*49.441*
Ada-MSHyper	MAE	0.287	0.298	0.364	0.316	0.314	0.332	0.393	0.346	0.302	0.317	0.360	0.326	0.348	0.362	0.418	0.376
	RMSE	0.458	0.475	0.552	0.495	0.475	0.498	0.577	0.517	0.455	0.477	0.532	0.488	0.543	0.559	0.613	0.571
	MAPE	*48.522*	49.827	57.305	51.885	49.969	51.829	58.069	53.289	50.212	52.242	56.497	52.984	54.547	56.260	62.096	57.635
FlashST	MAE	0.346	0.346	0.373	0.355	0.344	0.365	0.368	0.359	0.380	0.359	0.409	0.382	0.399	0.384	0.400	0.395
	RMSE	0.492	0.502	0.534	0.509	0.489	0.509	0.528	0.509	0.553	0.529	0.595	0.559	0.565	0.549	0.573	0.563
	MAPE	56.353	56.885	59.568	57.602	53.756	55.624	55.544	54.974	59.029	56.908	61.925	59.288	61.046	59.565	60.573	60.394

APPENDIX B
QUANTITATIVE RESULTS OF SOME OTHER MODELS

We provide the comparative results of VDformer and four additional models, including two variable-independent models (PatchTST (ICLR 2023) [9] and FiLM (NIPS 2022) [10]) and two variable-specific models (Triformer (IJCAI 2022) [11] and DLinear-I (AAAI 2023) [12]), in this section. $\{$ETTh$_1$, ETTh$_2$, ECL, Traffic$\}$ are chosen since they are the most prevalently used. '—' indicates that certain model is out of memory in such condition. It can be observed that VDformer outperforms all of these baselines.

Moreover, we adopt the other three existing works that have analyzed the importance of variable relationships, including LIFT [13] , Ada-MSHyper [14] and CCM [15] as the baselines for further comparison. Specifically, LIFT is combined with PatchTST, which is the leading model variant in the original work. As shown in Table B.2, the proposed VDformer model still outperforms these three baselines in the majority of cases.

Besides, an additional comparison experiment on PeMSD3, PeMSD4, PeMSD7 and PeMSD8 datasets [16], which are the most commonly used datasets for spatial-temporal forecasting, is used to compare the performance of VDformer with four SOTA spatial-temporal forecasting models, including SSTBAN [17], Cy2Mixer [18], Ada-MSHyper [14] and FlashST [19]. As shown in Table B.3, the experimental results, which are evaluated by MAE, RMSE and MAPE of the normalized ground truths and prediction results (Denormalized results are normally adopted in short-term spatial-temporal forecasting scenarios, whereas this work targets at long-term forecasting scenarios), illustrate that the proposed VDformer outperforms all these four baselines. Thus, the superior ability of VDformer to tackle dynamic variable correlations is further verified.

APPENDIX C
TUNING-AVAILABLE FORECASTING SCENARIO

In this scenario, we also tune the hyper-parameters of VDformer and compare its forecasting results with the original results of other baselines. ETTm$_2$, ECL, Traffic and Weather datasets are used since they are all adopted in the other baselines. We only tune the input sequence length, which is chosen within $\{96, 192, 336\}$, the stage number of encoder-decoder architecture, which is chosen within $\{1, 2, 3, 4, 5\}$ and the hidden dimension size, which is chosen within $\{8, 16, 32, 64, 128\}$. Only the MSE results are presented. The results are given in Table C.1. It can be observed that VDformer achieves a general MSE reduction by 2.9%/16.0%/5.6%/7.0%/8.6%/15.9%/15.8% when compared

TABLE C.1
MULTIVARIATE FORECASTING RESULTS WITH HYPER-PARAMETER TUNING

Methods	ETTm$_2$ 192	ETTm$_2$ 336	ETTm$_2$ 720	ETTm$_2$ Avg.	ECL 192	ECL 336	ECL 720	ECL Avg.	Traffic 192	Traffic 336	Traffic 720	Traffic Avg.	Weather 192	Weather 336	Weather 720	Weather Avg.
VDformer	***0.194***	***0.248***	***0.334***	***0.259***	***0.133***	***0.153***	***0.175***	***0.153***	***0.372***	***0.380***	***0.418***	***0.390***	***0.184***	*0.236*	***0.293***	***0.238***
ARM	*0.218*	*0.265*	0.357	*0.280*	*0.142*	*0.154*	*0.179*	*0.158*	*0.373*	*0.383*	*0.425*	*0.394*	*0.189*	***0.232***	*0.296*	*0.239*
iTransformer	0.250	0.311	0.412	0.324	0.162	0.178	0.225	0.188	0.417	0.433	0.467	0.439	0.221	0.278	0.358	0.286
ModernTCN	0.222	0.272	0.351	0.282	0.143	0.161	0.191	0.165	0.379	0.397	0.440	0.405	0.196	0.238	0.314	0.249
TSLANet	0.224	0.275	0.354	0.284	0.152	0.168	0.205	0.175	0.388	0.394	0.430	0.404	0.193	0.245	0.325	0.254
SparseTSF	0.218	0.272	*0.350*	0.280	0.151	0.166	0.205	0.174	0.398	0.411	0.448	0.419	0.215	0.260	0.318	0.264
TimeXer	0.237	0.296	0.392	0.308	0.157	0.176	0.211	0.181	0.448	0.473	0.516	0.479	0.204	0.261	0.340	0.268
CATS	0.248	0.304	0.402	0.318	0.163	0.180	0.219	0.187	0.436	0.453	0.484	0.458	0.208	0.264	0.342	0.271

TABLE D.1
THE RESULTS OF A CHALLENGING EXPERIMENT FOR VDFORMER

Methods		VDformer		ARM		iTransformer		ModernTCN		TSLANet		SparseTSF		TimeXer		CATS	
Input Length		96		192		192		192		192		192		192		192	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh_1$	192	***0.412***	***0.414***	0.563	0.533	0.444	0.437	0.446	0.437	0.446	0.439	0.427	*0.417*	0.442	0.444	*0.419*	0.422
	336	***0.434***	***0.425***	0.605	0.556	0.485	0.461	0.490	0.463	0.474	0.453	*0.449*	*0.430*	0.481	0.473	0.454	0.439
	720	*0.461*	*0.459*	0.625	0.577	0.492	0.489	0.670	0.564	0.508	0.494	***0.433***	***0.444***	0.529	0.514	0.477	0.475
	Avg.	***0.436***	*0.433*	0.597	0.555	0.474	0.462	0.535	0.488	0.476	0.462	***0.436***	***0.431***	0.484	0.477	*0.450*	0.445
$ETTh_2$	192	***0.321***	***0.361***	0.562	0.506	0.416	0.425	0.412	0.420	0.381	0.405	0.378	0.399	0.389	0.403	*0.367*	*0.399*
	336	***0.358***	***0.392***	0.699	0.573	0.420	0.433	0.397	0.425	0.403	0.423	*0.395*	*0.417*	0.426	0.433	0.405	0.428
	720	*0.412*	***0.432***	0.822	0.627	0.430	0.449	0.418	0.442	0.427	0.444	***0.410***	*0.433*	0.419	0.442	0.439	0.454
	Avg.	***0.364***	***0.395***	0.694	0.568	0.422	0.436	0.409	0.429	0.403	0.424	*0.395*	*0.417*	0.411	0.426	0.404	0.427
ECL	192	***0.146***	***0.241***	0.158	0.267	*0.153*	*0.247*	0.202	0.300	0.160	0.253	0.171	0.262	0.154	0.251	0.196	0.281
	336	***0.162***	***0.258***	0.170	0.277	*0.169*	*0.264*	0.208	0.309	0.178	0.271	0.187	0.278	0.173	0.268	0.214	0.297
	720	*0.198*	*0.289*	***0.180***	***0.287***	0.215	0.304	0.230	0.326	0.217	0.304	0.225	0.309	0.212	0.303	0.258	0.332
	Avg.	***0.169***	***0.263***	***0.169***	0.277	*0.179*	*0.271*	0.213	0.312	0.185	0.276	0.194	0.283	0.180	0.274	0.223	0.303
Traffic	192	***0.397***	***0.245***	0.753	0.391	*0.410*	*0.281*	0.460	0.313	0.435	0.301	0.463	0.302	0.424	0.286	0.634	0.389
	336	***0.412***	***0.254***	0.779	0.382	*0.428*	*0.291*	0.475	0.319	0.450	0.307	0.477	0.309	0.448	0.303	0.650	0.395
	720	***0.451***	***0.275***	1.068	0.383	*0.459*	0.311	0.503	0.337	0.479	0.325	0.505	0.329	0.479	*0.308*	0.675	0.405
	Avg.	***0.420***	***0.258***	0.867	0.386	*0.432*	*0.294*	0.479	0.323	0.455	0.311	0.482	0.313	0.450	0.299	0.653	0.397
Air	192	***0.823***	***0.665***	0.870	0.720	0.838	0.699	0.856	0.707	0.827	*0.690*	0.829	0.695	0.867	0.715	*0.823*	0.699
	336	***0.858***	***0.687***	0.941	0.759	0.906	0.735	0.963	0.759	0.914	0.734	*0.876*	*0.720*	0.951	0.753	0.895	0.734
	720	0.923	***0.721***	1.084	0.814	0.921	0.743	0.960	0.761	***0.909***	*0.736*	*0.913*	0.738	0.977	0.773	0.917	0.744
	Avg.	***0.868***	***0.691***	0.965	0.765	0.888	0.726	0.926	0.742	0.883	0.720	*0.873*	*0.718*	0.931	0.747	0.879	0.726
River	192	***0.058***	***0.136***	0.174	0.261	0.073	0.157	0.078	0.162	0.087	0.171	0.089	0.171	0.076	0.162	*0.069*	*0.153*
	336	***0.115***	***0.184***	0.300	0.341	0.170	0.232	0.175	0.234	*0.153*	*0.224*	0.170	0.231	0.207	0.256	0.166	0.229
	720	***0.306***	***0.309***	0.574	0.532	*0.345*	*0.334*	0.527	0.423	0.369	0.356	0.416	0.372	0.510	0.414	0.454	0.386
	Avg.	***0.160***	***0.210***	0.349	0.378	*0.196*	*0.241*	0.260	0.273	0.203	0.251	0.225	0.258	0.265	0.277	0.230	0.256

with ARM/iTransformer/ModernTCN/TSLANet/SparseTSF/TimeXer/CATS, which demonstrates that VDformer is capable of achieving better performance if the hyper-parameter tuning is allowed.

APPENDIX D
CHALLENGING FORECASTING SCENARIO

To highlight the forecasting capability of VDformer, we vouchsafe double input sequence length to the other baselines in the main text, which is supposed to enhance their forecasting performances, and render VDformer maintaining its original input sequence length. This experiment is conducted on six 1-hour-level datasets. Albeit under the disadvantageous condition, VDformer maintains the leading position under the majority of forecasting scenarios, as shown in Table D.1. This manifests the superior forecasting capability of VDformer.

APPENDIX E
EXTENSION CAPABILITIES OF PROPOSED METHODS

To show that the proposed methods in this work can also be widely adopted to other forecasting models. We separately combine PatchTST [9], TSMixer [20], ModernTCN [21], Crossformer [22], iTransformer [23] and SAMformer [24] with the EMD-based DyCVA. Crossformer is further combined with ELDI as it is the only baseline that owns decoder. The results are shown in Table E.1 and it can be observed that:

1) DyCVA enhances the forecasting performance of PatchTST, TSMixer and ModernTCN, which are respectively based on Transformer, MLP and CNN, a lot. The improved three models are even able to outperform more than half of the models in Table III of the main text except VDformer, demonstrating the success of DyCVA and its adaptivity to other models. As for the two models, iTransformer and SAMformer, that already achieve SOTA thanks to their conventional cross-variable attention modules, DyCVA further improves their performance, demonstrating the advantage of variable-dynamic forecasting approach. Moreover, the performance of Crossformer is greatly enhanced by DyCVA and ELDI, illustrating that ELDI is also available for the forecasting model with encoder-decoder architecture.
2) Among these baselines, PatchTST is variable-independent and the rest are variable-consistent. However, their performances are both enhanced when combined with DyCVA, which upgrades them to the variable-dynamic models. This verifies that the variable-dynamic forecasting strategy is genuinely better than the rest of forecasting strategies.

TABLE E.1
OTHER BASELINES COMBINED WITH OUR PROPOSED METHODS UNDER ECL

Methods	PatchTST		+DyCVA		TSMixer		+DyCVA		ModernTCN		+DyCVA	
Prediction Length	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
192	0.199	0.283	***0.189***	***0.271***	0.226	0.315	***0.196***	***0.276***	0.183	0.284	***0.179***	***0.264***
336	0.214	0.299	***0.204***	***0.286***	0.244	0.332	***0.203***	***0.284***	0.194	0.297	***0.186***	***0.273***
720	0.256	0.331	***0.220***	***0.302***	0.287	0.363	***0.224***	***0.309***	0.224	0.320	***0.219***	***0.306***
Avg.	0.223	0.305	***0.204***	***0.286***	0.252	0.337	***0.208***	***0.290***	0.200	0.300	***0.195***	***0.281***
Impr.			+8.4%	+6.1%			+17.7%	+13.9%			+2.9%	+6.4%
Methods	iTransformer		+DyCVA		SAMformer		+DyCVA		Crossformer		+DyCVA&ELDI	
Prediction Length	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
192	0.176	0.265	***0.166***	***0.247***	0.156	0.250	***0.149***	***0.235***	0.285	0.351	***0.171***	***0.257***
336	0.194	0.284	***0.194***	***0.284***	0.176	0.271	***0.176***	***0.271***	0.342	0.381	***0.180***	***0.268***
720	0.233	0.317	***0.233***	***0.317***	0.214	0.303	***0.214***	***0.303***	0.543	0.471	***0.210***	***0.296***
Avg.	0.201	0.289	***0.201***	***0.289***	0.182	0.274	***0.182***	***0.274***	0.390	0.401	***0.187***	***0.273***
Impr.			+5.4%	+6.7%			+4.7%	+5.9%			+52.1%	+31.8%

APPENDIX F
PARAMETER SENSITIVITY INVOLVING EMD

The parameter sensitivity experiment results of the range threshold r and SD threshold η are illustrated in Fig. F.1. Two datasets, Traffic and Air, are selected for the experiment. As indicated in Table III of the main text, the former one is more suitable for variable-consistent models and the later one is more suitable for variable-independent models. It can be observed from Fig. F.1 that the performance of VDformer is not sensitive to these two hyperparameters unless they are set too large so that EMD is unable to decompose enough IMFs.

APPENDIX G
EFFECTS OF LIMITATIONS OF EMD

As mentioned in the main text, EMD method inevitably possesses some weaknesses, including the end effects, being prone to noise, the overshoot and undershoot in spline fitting. Moreover, EEMD [25] and CEEMDAN [26], as two EMD derivates that greatly enhance the resistance to noise, fail to improve the performance of VDformer in the main text. Thus, this appendix section studies whether the other two limitations of EMD affect the forecasting ability of VDformer. Four variants are tested, including:

1) EMD+24: An additional sequence of length 24 is attached to the input sequence at the beginning when performing EMD. However, only the time span of original sequence is used for further analysis to reduce the end effect.

2) EMD+48: An additional sequence of length 48 is attached to the input sequence at the beginning when performing EMD. However, only the time span of original sequence is used for further analysis to reduce the end effect.

3) QSG: The cubic spline fitting method is replaced with quintic spline fitting method.

4) SSG: The cubic spline fitting method is replaced with seventh-order spline fitting method.

It can be observed in Table G.1, where the 'Avg.' results of six hour-level benchmarks are presented, that all these model variants affect the performance a little. This phenomenon demonstrates that the weaknesses of EMD are not significant in VDformer due to the qualitative usage of EMD, rather than the quantitative usages in previous works [27], [28], in VDformer.

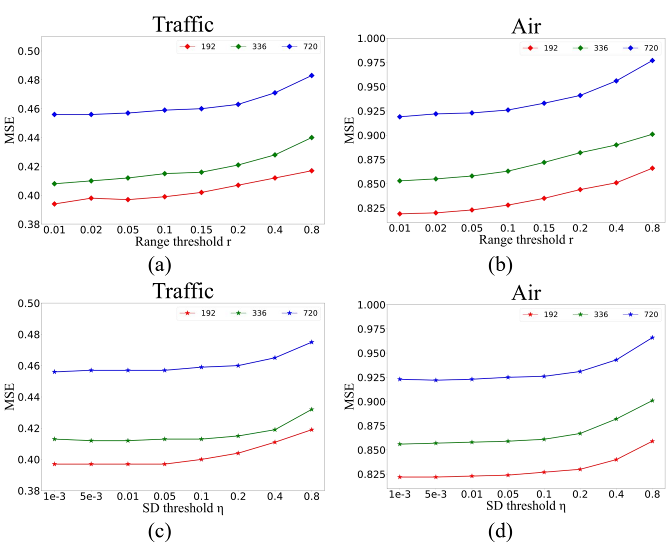

Fig. F.1. The MSE results of the VDformer with different r and η on two datasets. (a) and (b) illustrate the results with different r on Traffic and Air, respectively. (c) and (d) show the results with different η on Traffic and Air, respectively.

TABLE G.1
COMPARATIVE RESULTS INVOLVING EMD

Methods	VDformer		EMD+24		EMD+48		QSG		SSG	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh_1$	0.436	0.433	0.434	0.432	0.432	0.431	0.434	0.430	0.434	0.428
$ETTh_2$	0.364	0.395	0.363	0.392	0.361	0.391	0.363	0.392	0.360	0.390
ECL	0.169	0.263	0.167	0.261	0.164	0.259	0.166	0.259	0.165	0.259
Traffic	0.420	0.258	0.418	0.257	0.415	0.255	0.418	0.254	0.417	0.253
Air	0.868	0.691	0.867	0.690	0.864	0.687	0.863	0.686	0.866	0.689
River	0.160	0.210	0.158	0.208	0.155	0.207	0.155	0.205	0.158	0.209

TABLE H.1
ABLATION STUDY RESULTS INVOLVING REVIN

Methods	RevIN	ELDI	Metric	Weather 192	Weather 336	Weather 720	Weather Avg.	Traffic 192	Traffic 336	Traffic 720	Traffic Avg.
CATS	✔	✘	MSE	0.188	0.240	0.316	0.248	0.498	0.512	0.544	0.518
			MAE	0.240	0.282	0.336	0.286	0.297	0.303	0.320	0.307
	✘	✘	MSE	0.195	0.243	0.340	0.259	0.586	0.600	0.633	0.606
			MAE	0.259	0.295	0.368	0.307	0.304	0.310	0.327	0.314
TimeXer	✔	✘	MSE	0.204	0.257	0.341	0.267	0.477	0.495	0.530	0.501
			MAE	0.252	0.294	0.350	0.299	0.304	0.312	0.329	0.315
	✘	✘	MSE	0.242	0.315	0.407	0.322	0.629	0.643	0.680	0.650
			MAE	0.303	0.356	0.411	0.357	0.356	0.363	0.378	0.366
VDformer	✔	✔	MSE	0.190	0.241	0.313	0.248	0.397	0.412	0.451	0.420
			MAE	0.232	0.270	0.323	0.275	0.245	0.254	0.275	0.258
	✘	✔	MSE	0.205	0.261	0.327	0.264	0.409	0.430	0.479	0.439
			MAE	0.242	0.285	0.338	0.288	0.252	0.263	0.287	0.267
	✘	✘	MSE	0.230	0.289	0.363	0.294	0.512	0.547	0.594	0.551
			MAE	0.254	0.304	0.355	0.304	0.286	0.295	0.308	0.296

APPENDIX H
MORE ABLATION STUDIES

H.1 *Ablation Study on RevIN*

To validate that VDformer benefits much less from RevIN than other models, several model variants are evaluated in this appendix section. CATS and TimeXer, which are leading variable-independent and variable-consistent models in the main text respectively, are used for comparison. Weather and Traffic datasets are used for benchmarks. As shown in Table H.1, CATS and TimeXer suffer from significant performance degradation with RevIN removed. Although the performance of VDformer is worse with RevIN removed, the degradation rate is much lower. Moreover, if both ELDI and RevIN are removed, VDformer also suffers from heavy MSE increases, demonstrating the effectiveness of ELDI for tackling the nonstationarity.

H.2 *More Ablation Studies on ELDI*

To highlight the performance of ELDI, four ablation variants, which replace the ELDI with a small MLP (VDformer-M), the average of input sequence (VDformer-A), the last value of input sequence (VDformer-L) and the EDLI but using the dominant nodes obtained only by FFT (VDformer-F) respectively, are used for comparison. The experimental results in Table H.2 show that all four ablation variants perform worse relative to the original VDformer. Moreover, VDformer-A achieves the best performance among the four additional ablation variants, demonstrating that initializing the decoder with the general local input features is relatively better. Besides, VDformer-F counter-intuitively achieves worse performance relative to VDformer, which once more demonstrates that FFT is not effective for non-stationary data decomposition and reconstruction. A forecasting scenario, which is identical to the one in Fig. 9 (a1) of the main text, is visualized below to show the difference of VDformer and VDformer-F.

TABLE H.2
ABLATION STUDY RESULTS INVOLVING ELDI

Methods	Metric	Weather 192	Weather 336	Weather 720	Weather Avg.	Traffic 192	Traffic 336	Traffic 720	Traffic Avg.
VDformer	MSE	***0.190***	***0.241***	***0.313***	***0.248***	***0.397***	***0.412***	***0.451***	***0.420***
	MAE	***0.232***	***0.270***	***0.323***	***0.275***	***0.245***	***0.254***	***0.275***	***0.258***
VDformer-M	MSE	0.210	0.259	0.366	0.278	0.468	0.507	0.539	0.505
	MAE	0.276	0.302	0.382	0.320	0.290	0.314	0.332	0.312
VDformer-A	MSE	*0.196*	*0.253*	*0.341*	*0.263*	*0.443*	*0.478*	*0.531*	*0.484*
	MAE	*0.257*	*0.296*	*0.366*	*0.306*	*0.279*	*0.288*	*0.322*	*0.296*
VDformer-L	MSE	0.211	0.255	0.362	0.276	0.456	0.501	0.544	0.500
	MAE	0.276	0.294	0.381	0.317	0.281	0.321	0.336	0.313
VDformer-F	MSE	0.204	0.264	0.360	0.276	0.459	0.493	0.548	0.500
	MAE	0.266	0.305	0.379	0.317	0.285	0.299	0.337	0.307

APPENDIX I
ANALYSIS OF LASSO IN ELDI

Apart from LASSO, a linear regressor, a ridge regressor and a kernel ridge regressor with RBF kernel are tested for ELDI. LASSO regressors with different penalty factors α are also evaluated for ELDI for analyzing the sensitivity of α. The results in Table I.1 and I.2 show that ELDI is robust to various regressors, which means that the performance of VDformer changes slightly with different regressors. Moreover, the chosen LASSO regressor with α = 0.1 is the most suitable for ELDI according to is leading performance in Table I.1 and I.2.

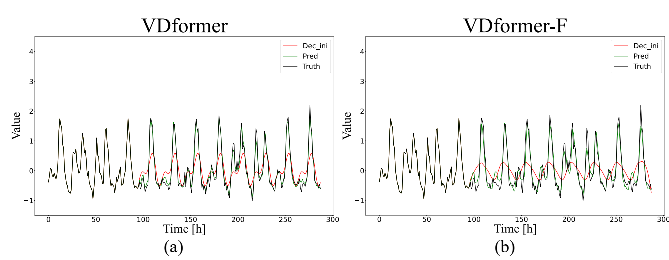

Fig. H.1. The visualization of the difference involving ELDI between VDformer and VDformer-F.

TABLE I.1
ABLATION STUDY RESULTS INVOLVING THE REGRESSOR USED IN ELDI

Variants	LASSO		Linear		Ridge		Kernel ridge	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh_1$	0.436	0.433	0.439	0.437	0.440	0.438	0.440	0.438
$ETTh_2$	0.364	0.395	0.365	0.398	0.369	0.401	0.365	0.397
ECL	0.169	0.263	0.173	0.268	0.171	0.266	0.173	0.268
Traffic	0.420	0.258	0.422	0.261	0.421	0.261	0.424	0.263
Air	0.868	0.691	0.869	0.694	0.872	0.696	0.869	0.693
River	0.160	0.210	0.164	0.215	0.161	0.213	0.165	0.216

TABLE I.2
PARAMETER SENSITIVITY INVOLVING THE PENALTY FACTOR OF LASSO

α	0.01		0.1		1		10		100	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh_1$	0.439	0.437	0.436	0.433	0.440	0.438	0.441	0.438	0.444	0.443
$ETTh_2$	0.366	0.398	0.364	0.395	0.366	0.398	0.368	0.400	0.373	0.405
ECL	0.173	0.268	0.169	0.263	0.172	0.267	0.170	0.266	0.176	0.271
Traffic	0.425	0.263	0.420	0.258	0.421	0.260	0.423	0.262	0.425	0.265
Air	0.870	0.695	0.868	0.691	0.871	0.695	0.869	0.693	0.877	0.702
River	0.161	0.212	0.160	0.210	0.163	0.214	0.162	0.213	0.165	0.217

APPENDIX J
ANALYSIS OF STABILITY

The standard deviations of multiple experimental results of VDformer and other baselines are compared in this section to measure their stability. The number of duplicate experiments for each setting is ten. As shown in Fig. J.1, VDformer not only achieves state-of-the-art forecasting accuracy, but also is robust enough to provide stable forecasting results. Specifically, even the worse forecasting performance of VDformer among ten duplicate experiments is better than the best forecasting performances of other baselines.

APPENDIX K
VISUALIZATION OF DYNAMIC VARIABLE CORRELATIONS

In addition, the dominant periodic ingredients of several variables from the Weather and Traffic datasets are visualized to show the dynamic variable correlations in real-world IoT systems. As shown in Fig. K.1, the dominant periodic ingredients of the variables, each of which is represented by the corresponding f_d, from the Weather dataset vary heavily and frequently, which indicates apparent dynamic variable correlations in the Weather dataset. Thus, the variable-independent models achieve relatively better performances than the variable-consistent models do when handling the Weather dataset in Table IV of the main text. The opposite phenomenon occurs for the variables of the Traffic dataset. Although the variable correlations and the dominant periodic ingredients rarely change, which leads to better performances of variable-consistent models than those of variable-independent models when handling the Traffic dataset in Table III of the main text, dynamic variable correlations still exist. Hence, variable-dynamic forecasting is always better than variable-consistent and variable-independent forecasting in these cases.

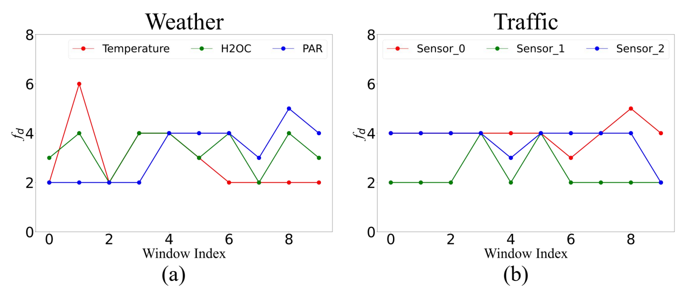

Fig. K.1. The frequencies of dominant periodic ingredients of several variables from Weather (a) and Traffic (b) datasets.

APPENDIX L
EFFICIENCY ANALYSIS

Here provides an efficiency analysis that evaluates the separate complexity of the components in VDformer. As shown in Table L.1: 1) The DyCVA module hardly imposes extra computation complexity to the conventional cross-variable attention (CVA) module and its improvement on model forecasting performance is significant. 2) The ELDI method is

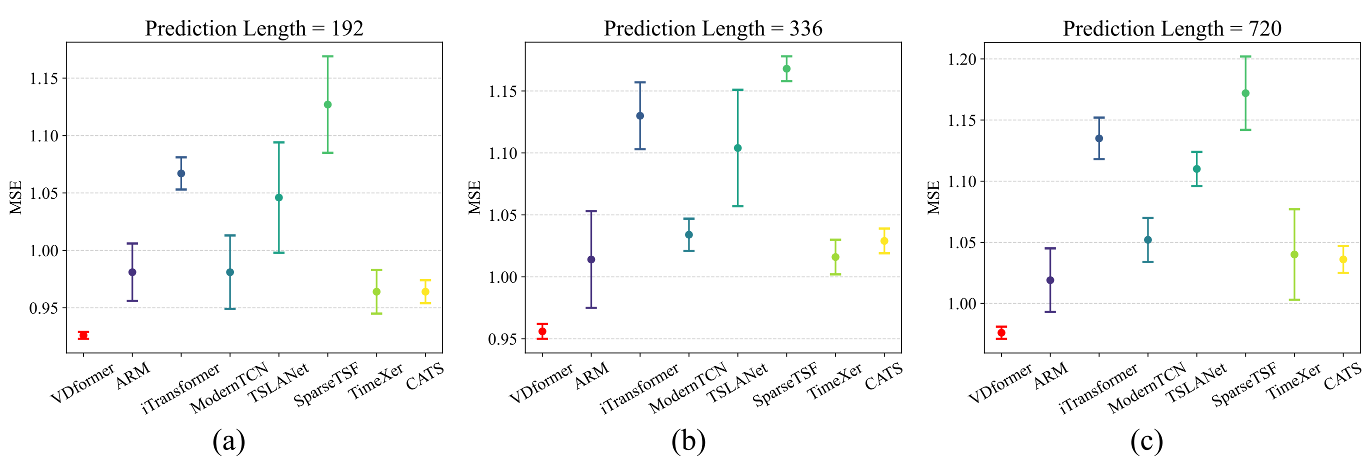

Fig. J.1. The stability comparison between VDformer and other baselines on HomeC dataset. MSE is selected as the evaluation metric.

TABLE L.1
THE SEPARATE COMPLEXITY OF THE COMPONENTS IN VDFORMER

VDformer	DyCVA	CVA	ELDI	Patch SA	Patch CA	FFN	Others	All
FLOPs (MB)	14.10	14.10	0.05	15.17	15.17	55.13	24.48	138.20
Parameters (MB)	0.19	0.19	0.00	0.19	0.19	0.74	0.20	1.70

nearly free in space complexity and cheap in time complexity, whereas it outperforms the other free initialization methods as shown in Appendix H.2 of the Supplementary Material.

Moreover, we would like answer a question: 'Though VDformer is not lightweight due to the complex feature extraction along the temporal and variable dimension, is it possible that a light version of VDformer is able to outperform the majority of Transformer-based forecasting models from the perspectives of both accuracy and efficiency?' To answer this question, a super lightweight VDformer, VDformer-small, wherein the temporal feature extraction modules are all removed and the decoder only possesses one stage for prediction, is used for the efficiency comparison with Crossformer [22], ARM [29], iTransformer [23] and TimeXer [30], which are all variable-consistent forecasting models focusing on both temporal and cross-variable feature extraction. Note that the encoder of VDformer-small only possesses an embedding module, three DyCVA modules and the merging operations. The results in Fig. L.1 show that VDformer-small has worse forecasting accuracy relative to the complete VDformer model, whereas its computation complexity is much lower and it still outperforms the other four forecasting models, demonstrating the effectiveness of the variable-dynamic forecasting approach and the proposed DyCVA and ELDI methods.

APPENDIX M
HOW DOES EMD CHANGE FFT?

This section presents a case study to evaluate the overlap of the dominant periodicities obtained via EMD+FFT with those obtained via FFT. As shown in Table M.1, the overlap percentages for different datasets are between 19.78% and 53.23% when the input sequence lengths are identical with the ones set for the quantitative experiment.

Moreover, we visualize how the dominant periodicities obtained via EMD+FFT and those obtained via FFT overlap in Fig. M.1. Traffic, whose the overlap rate is the highest, and HomeC, whose the overlap rate is the lowest, are adopted. It can be observed that the overlap region is neither successive nor regular in most of the cases, which demonstrates the existence of dynamic variable correlations and the significance to handle them.

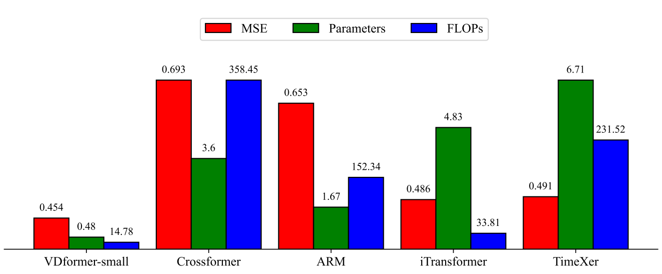

Fig. L.1. The complexity comparison of between VDformer-small and other four baselines. The red, green and blue bars are respectively the averaged MSE results, the number of learnable parameters (MB) and FLOPs (MB). The dataset is $ETTh_1$.

TABLE M.1
THE OVERLAP PERCENTAGES FOR ELEVEN DATASETS

Datasets	$ETTh_1$	$ETTh_2$	$ETTm_1$	$ETTm_2$	ECL	Traffic
Overlap	35.70%	35.86%	35.91%	35.42%	52.24%	53.23%
Datasets	Air	River	Weather	Solar	HomeC	
Overlap	26.58%	36.04%	33.18%	51.14%	19.78%	

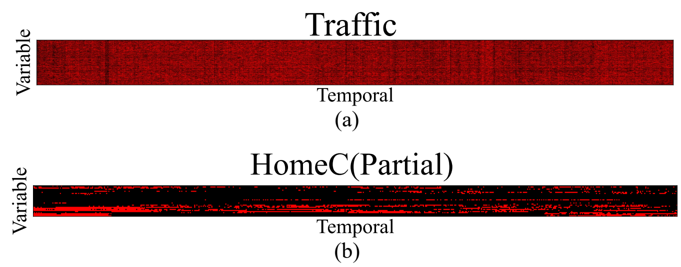

Fig. M.1. The visualization of the overlapping extents of the dominant periodicities obtained via EMD+FFT and those obtained via FFT on two datasets. (a) Traffic. (b) HomeC.

REFERENCES

[1] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, vol. 35, no. 12. AAAI Press, 2021, pp. 11 106–11 115.

[2] A. Trindade, "ElectricityLoadDiagrams20112014," UCI Machine Learning Repository, 2015, DOI: https://doi.org/10.24432/C58C86.

[3] (n.d.) Caltrans pems. [Online]. Available: http://pems.dot.ca.gov/

[4] (n.d.) Max-planck-institut fuer biogeochemie - wetterdaten. [Online]. Available: https://www.bgc-jena.mpg.de/wetter/

[5] (n.d.) Solar power data for integration studies. [Online]. Available: https://www.nrel.gov/grid/solar-power-data.html

[6] S. Vito, "Air Quality," UCI Machine Learning Repository, 2016, DOI: https://doi.org/10.24432/C59K5F.

[7] (2022) River flow-rf2. [Online]. Available: https://www.kaggle.com/datasets/samanemami/river-flowrf2

[8] (n.d.) Smart home dataset with weather information. [Online]. Available: https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information

[9] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=Jbdc0vTOcol

[10] T. Zhou, Z. Ma, xue wang, Q. Wen, L. Sun, T. Yao, W. Yin, and R. Jin, "FiLM: Frequency improved legendre memory model for long-term time series forecasting," in *Advances in Neural Information Processing Systems*, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022. [Online]. Available: https://openreview.net/forum?id=zTQdHSQUQWc

[11] R.-G. Cirstea, C. Guo, B. Yang, T. Kieu, X. Dong, and S. Pan, "Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, L. D. Raedt, Ed. International Joint Conferences on Artificial Intelligence Organization, 7 2022, pp. 1994–2001, main Track.

[12] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[13] L. Zhao and Y. Shen, "Rethinking channel dependence for multivariate time series forecasting: Learning from leading indicators," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=JiTVtCUOpS

[14] Z. Shang, L. Chen, B. Wu, and D. Cui, "Ada-mshyper: adaptive multi-scale hypergraph transformer for time series forecasting," *Advances in Neural Information Processing Systems*, vol. 37, pp. 33 310–33 337, 2024.

[15] J. Chen, J. E. Lenssen, A. Feng, W. Hu, M. Fey, L. Tassiulas, J. Leskovec, and R. Ying, "From similarity to superiority: Channel clustering for time series forecasting," in *Thirty-eighth Conference on Neural Information Processing Systems*, 2024. [Online]. Available: https://openreview.net/forum?id=MDgn9aazo0

[16] W. Duan, X. He, Z. Zhou, L. Thiele, and H. Rao, "Localised adaptive spatial-temporal graph neural network," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 448–458.

[17] S. Guo, Y. Lin, L. Gong, C. Wang, Z. Zhou, Z. Shen, Y. Huang, and H. Wan, "Self-supervised spatial-temporal bottleneck attentive network for efficient long-term traffic forecasting," in *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 2023, pp. 1585–1596.

[18] M. Lee, Y. Y. Choi, S. W. Park, S. Lee, J. Ko, and J. Hong, "Enhancing topological dependencies in spatio-temporal graphs with cycle message passing blocks," in *The Third Learning on Graphs Conference*, 2024.

[19] Z. Li, L. Xia, Y. Xu, and C. Huang, "Flashst: A simple and universal prompt-tuning framework for traffic prediction," in *International Conference on Machine Learning*, 2024.

[20] V. Ekambaram, A. Jati, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting," *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259187817

[21] L. donghao and wang xue, "ModernTCN: A modern pure convolution structure for general time series analysis," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=vpJMJerXHU

[22] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=vSVLM2j9eie

[23] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, "itransformer: Inverted transformers are effective for time series forecasting," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=JePfAI8fah

[24] R. Ilbert, A. Odonnat, V. Feofanov, A. Virmaux, G. Paolo, T. Palpanas, and I. Redko, "Unlocking the potential of transformers in time series forecasting with sharpness-aware minimization and channel-wise attention," *arXiv preprint arXiv:2402.10198*, 2024.

[25] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, pp. 1–41, 2009.

[26] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4144–4147.

[27] J. Yin, H. Wang, N. Wang, and X. Wang, "An adaptive real-time modular tidal level prediction mechanism based on emd and lipschitz quotients method," *Ocean Engineering*, vol. 289, p. 116297, 2023.

[28] N. Li, J. Dong, L. Liu, H. Li, and J. Yan, "A novel emd and causal convolutional network integrated with transformer for ultra short-term wind power forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 154, p. 109470, 2023.

[29] J. Lu, X. Han, and S. Yang, "ARM: Refining multivariate forecasting with adaptive temporal-contextual learning," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=JWpwDdVbaM

[30] Y. Wang, H. Wu, J. Dong, G. Qin, H. Zhang, Y. Liu, Y. Qiu, J. Wang, and M. Long, "Timexer: Empowering transformers for time series forecasting with exogenous variables," in *The Thirty-eighth Conference on Neural Information Processing Systems*, 2024. [Online]. Available: https://openreview.net/forum?id=INAeUQ04lT