

## A DATASET STATISTICS AND EXPERIMENTAL SETUPS

**Dataset statistics.** We assess the performance of SOR-Mamba across 13 datasets, with the dataset statistics detailed in Table A.1, where  $C$  and  $T$  denote the number of channels and timesteps, respectively.

**Experimental setups.** We follow the same data processing steps and train-validation-test split protocol as used in S-Mamba (Wang et al., 2024), maintaining a chronological order in the separation of training, validation, and test sets, using a 6:2:2 ratio for the Solar-Energy, ETT, and PEMS datasets, and a 7:1:2 ratio for the other datasets. The results are shown in Table A.1, where  $N, L$ , and  $H$  represent the dataset size, the size of the lookback window, and the size of the forecast horizon, respectively. For all datasets and all models,  $L$  is uniformly set to 96. We do not tune any hyperparameters and adhere to those used in S-Mamba, except for  $\lambda$ , which is related to the proposed regularization, and is tuned using a grid search over  $[0.001, 0.01, 0.1]$ .

Dataset	Statistics		Experimental Setups		
	$C$	$T$	$(N_{\text{train}}, N_{\text{val}}, N_{\text{test}})$	$L$	$H$
ETTh1 (Zhou et al., 2021)	7	17420	(8545, 2881, 2881)	96	{96, 192, 336, 720}
ETTh2 (Zhou et al., 2021)		17420	(8545, 2881, 2881)		
ETTm1 (Zhou et al., 2021)		69680	(34465, 11521, 11521)		
ETTm2 (Zhou et al., 2021)		69680	(34465, 11521, 11521)		
Exchange (Wu et al., 2021)	8	7588	(5120, 665, 1422)		
Weather (Wu et al., 2021)	21	52696	(36792, 5271, 10540)		
ECL (Wu et al., 2021)	321	26304	(18317, 2633, 5261)		
Traffic (Wu et al., 2021)	862	17544	(12185, 1757, 3509)		
Solar-Energy (Lai et al., 2018)	137	52560	(36601, 5161, 10417)		
PEMS03 (Liu et al., 2022)	358	26209	(15617, 5135, 5135)		{12, 24, 48, 96}
PEMS04 (Liu et al., 2022)	307	15992	(10172, 3375, 3375)		
PEMS07 (Liu et al., 2022)	883	28224	(16911, 5622, 5622)		
PEMS08 (Liu et al., 2022)	170	17856	(10690, 3548, 3548)		

Table A.1: Datasets for TS forecasting.

## B BASELINE METHODS

- S-Mamba (Wang et al., 2024): S-Mamba utilizes the bidirectional Mamba to capture channel dependencies in TS by scanning the channels from both directions.
- PatchTST (Nie et al., 2023): PatchTST segments TS into patches and feeds them into a Transformer in a channel independent manner.
- iTransformer (Liu et al., 2024a): iTransformer reverses the conventional role of the Transformer in the TS domain by treating each channel rather than patches as a token, thereby emphasizing channel dependencies over temporal dependencies.
- Crossformer (Zhang & Yan, 2023): Crossformer employs a cross-attention mechanism to capture both temporal and channel dependencies in TS.
- TimesNet (Wu et al., 2023): TimesNet captures both intraperiod and interperiod variations in 2D space using a parameter-efficient inception block.
- RLinear (Li et al., 2023): RLinear is a simple linear model that integrates reversible normalization and channel independence.
- DLinear (Zeng et al., 2023): DLinear is a simple linear model with channel independent architecture, that employs TS decomposition.

## C S-MAMBA VS. SOR-MAMBA

Figure C.1 visualizes the comparison between S-Mamba (Wang et al., 2024), which employs the bidirectional Mamba to capture CD, and our method, SOR-Mamba, which uses a single unidirectional Mamba with regularization to capture CD.

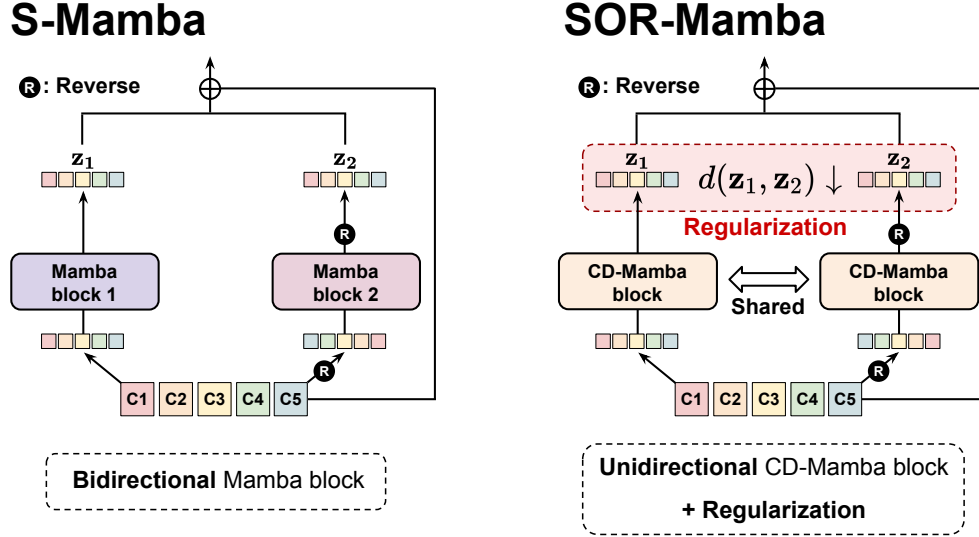


Figure C.1: Comparison of S-Mamba and SOR-Mamba.

## D REMOVAL OF 1D-CONVOLUTION

The original Mamba block (Gu & Dao, 2023) integrates the H3 block (Fu et al., 2023) with a gated MLP, where the H3 block uses a 1D-conv before the SSM layer to capture local information within nearby tokens, as illustrated in Figure D.1. However, since channels in TS do not have an inherent sequential order, we eliminate the 1D-conv from the Mamba block, resulting in the proposed CD-Mamba block. Figure D.2 shows the overall architecture of the proposed CD-Mamba block, where the 1D-conv before the selective SSM is removed from the original Mamba block (Gu & Dao, 2023).

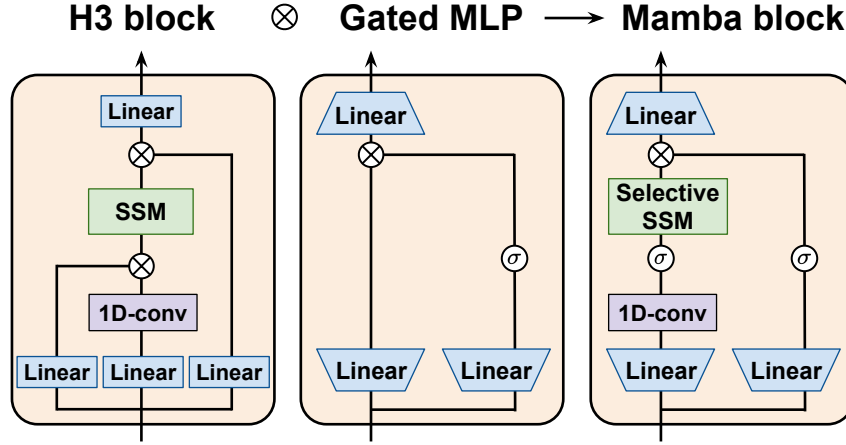


Figure D.1: **Architecture of the original Mamba block.** The original Mamba block contains 1D-conv before the SSM layer to capture local information within nearby tokens.

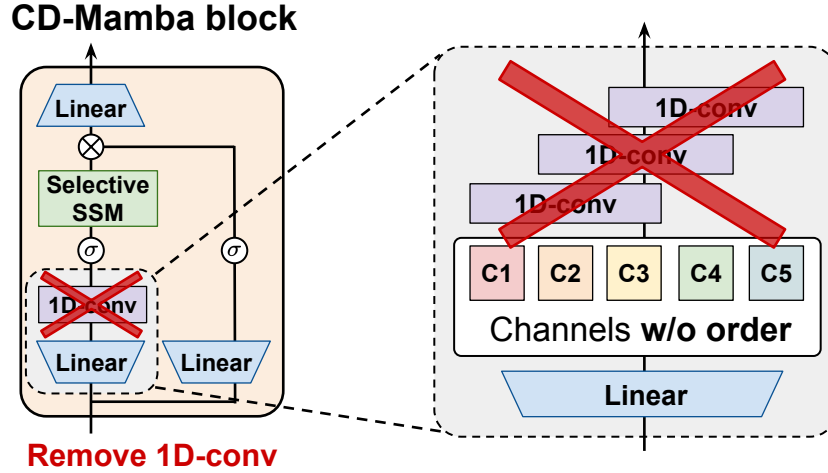


Figure D.2: **Architecture of the CD-Mamba block.** 1D-conv before the selective SSM is removed from the original Mamba block, as the channels do not have a sequential order.

## E FULL RESULTS OF TIME SERIES FORECASTING

Table E.1 shows the full results of TS forecasting tasks across four different horizons, highlighting the effectiveness of our method.

Models	SOR-Mamba				S-Mamba		iTransformer		RLinear		PatchTST		Crossformer		TiDE		TimesNet		DLinear		
	FT		SL		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	<u>377</u>	<u>398</u>	<u>385</u>	<u>398</u>	<u>385</u>	404	387	405	386	<u>395</u>	414	419	423	448	479	464	384	402	386	400
	192	<u>428</u>	<u>429</u>	<u>435</u>	<u>428</u>	<u>445</u>	441	441	436	437	<u>424</u>	460	445	471	474	525	492	436	429	437	432
	336	<u>464</u>	<u>448</u>	<u>474</u>	<u>448</u>	491	462	487	458	479	<u>446</u>	501	466	570	546	565	515	491	469	481	459
	720	<u>464</u>	<u>469</u>	<u>478</u>	<u>471</u>	506	497	509	494	481	<u>470</u>	500	488	653	621	594	558	521	500	519	516
	Avg.	<u>433</u>	<u>436</u>	<u>442</u>	438	457	452	457	449	446	<u>434</u>	469	454	529	522	541	507	458	450	456	452
ETTh2	96	<u>292</u>	<u>348</u>	299	<u>348</u>	297	349	301	350	<u>288</u>	<u>338</u>	302	<u>348</u>	745	584	400	440	340	374	333	387
	192	<u>372</u>	<u>397</u>	375	399	378	399	381	399	<u>374</u>	<u>390</u>	388	400	877	656	528	509	402	414	477	476
	336	<u>415</u>	<u>431</u>	<u>423</u>	435	425	435	427	434	<u>415</u>	<u>426</u>	426	433	1.043	731	643	571	452	452	594	541
	720	<u>423</u>	<u>445</u>	431	446	432	448	430	446	<u>420</u>	<u>440</u>	431	446	1.104	763	874	679	462	468	831	657
	Avg.	<u>376</u>	<u>405</u>	382	407	383	408	384	407	<u>374</u>	<u>398</u>	387	407	942	684	611	550	414	427	559	515
ETTm1	96	<u>324</u>	<u>362</u>	<u>326</u>	<u>367</u>	<u>326</u>	368	342	377	355	376	329	<u>367</u>	404	426	364	387	338	375	345	372
	192	<u>369</u>	<u>385</u>	375	<u>387</u>	378	393	383	396	391	392	<u>367</u>	<u>385</u>	450	451	398	404	374	<u>387</u>	380	389
	336	<u>402</u>	<u>408</u>	408	<u>408</u>	410	414	418	418	424	415	<u>399</u>	<u>410</u>	532	515	428	425	410	411	413	413
	720	<u>467</u>	<u>444</u>	472	<u>444</u>	474	451	487	456	487	<u>450</u>	<u>454</u>	<u>439</u>	666	589	487	461	478	450	474	453
	Avg.	<u>391</u>	<u>400</u>	396	<u>401</u>	398	407	408	412	414	407	<u>387</u>	<u>400</u>	513	496	419	419	400	406	403	407
ETTm2	96	<u>179</u>	<u>261</u>	181	265	182	266	186	272	182	265	<u>175</u>	<u>259</u>	287	366	207	305	187	267	193	292
	192	<u>241</u>	<u>304</u>	<u>246</u>	307	252	313	254	314	<u>246</u>	<u>304</u>	<u>241</u>	<u>302</u>	414	492	290	364	249	309	284	362
	336	<u>302</u>	<u>342</u>	306	345	313	349	317	353	307	<u>342</u>	<u>305</u>	<u>343</u>	597	542	377	422	321	351	369	427
	720	<u>401</u>	<u>400</u>	403	401	416	409	412	407	<u>398</u>	<u>402</u>	<u>400</u>	1.730	1.042	558	524	408	403	554	522	
	Avg.	<u>281</u>	<u>327</u>	<u>284</u>	329	290	333	293	337	286	<u>327</u>	<u>281</u>	<u>326</u>	757	610	358	404	291	333	350	401
PEMS03	12	<u>066</u>	<u>170</u>	<u>066</u>	<u>170</u>	<u>066</u>	<u>171</u>	071	174	126	236	099	216	090	203	178	305	085	192	122	243
	24	<u>088</u>	<u>197</u>	<u>090</u>	<u>200</u>	<u>088</u>	<u>197</u>	097	208	246	334	142	259	121	240	257	371	118	223	201	317
	48	<u>134</u>	<u>245</u>	167	280	165	277	<u>161</u>	<u>272</u>	551	529	211	319	202	317	379	463	155	260	333	425
	96	<u>193</u>	<u>297</u>	225	318	<u>213</u>	<u>313</u>	240	338	1.057	787	269	370	262	367	490	539	228	317	457	515
	Avg.	<u>121</u>	<u>227</u>	137	242	<u>133</u>	<u>240</u>	142	248	495	472	180	291	169	281	326	419	147	248	278	375
PEMS04	12	<u>074</u>	<u>175</u>	077	180	<u>073</u>	<u>177</u>	081	188	138	252	105	224	098	218	219	340	087	195	148	272
	24	<u>086</u>	<u>192</u>	091	<u>197</u>	<u>084</u>	<u>192</u>	099	211	258	348	153	275	131	256	292	398	103	215	224	340
	48	<u>106</u>	<u>214</u>	115	221	101	<u>213</u>	133	246	572	544	229	339	205	326	409	478	136	250	355	437
	96	<u>129</u>	<u>233</u>	143	248	<u>125</u>	<u>236</u>	172	283	1.137	820	291	389	402	457	492	532	190	303	452	504
	Avg.	<u>099</u>	<u>203</u>	107	212	<u>096</u>	<u>205</u>	121	232	526	491	195	307	209	314	353	437	129	241	295	388
PEMS07	12	<u>059</u>	<u>155</u>	<u>060</u>	<u>156</u>	<u>060</u>	157	067	165	118	235	095	207	094	200	173	304	082	181	115	242
	24	<u>076</u>	<u>174</u>	<u>082</u>	<u>182</u>	<u>082</u>	184	088	190	242	341	150	262	139	247	271	383	101	204	210	329
	48	<u>098</u>	<u>199</u>	107	209	<u>100</u>	<u>204</u>	113	218	562	541	253	340	311	369	446	495	134	238	398	458
	96	<u>117</u>	<u>218</u>	117	<u>218</u>	<u>117</u>	<u>218</u>	172	283	1.096	795	346	404	396	442	628	577	181	279	594	553
	Avg.	<u>088</u>	<u>186</u>	091	<u>191</u>	<u>090</u>	<u>191</u>	102	205	504	478	211	303	235	315	380	440	124	225	329	395
PEMS08	12	<u>078</u>	<u>178</u>	<u>076</u>	<u>176</u>	<u>076</u>	<u>178</u>	088	193	133	247	168	232	165	214	227	343	112	212	154	276
	24	<u>103</u>	<u>205</u>	109	<u>212</u>	110	216	138	243	249	343	224	281	215	260	318	409	141	238	248	353
	48	<u>159</u>	<u>250</u>	<u>172</u>	264	173	<u>254</u>	334	353	569	544	321	354	315	355	497	510	198	283	440	470
	96	<u>229</u>	<u>295</u>	290	334	<u>271</u>	<u>321</u>	458	436	1.166	814	408	417	377	397	721	592	320	351	674	565
	Avg.	<u>142</u>	<u>232</u>	162	247	<u>157</u>	<u>242</u>	254	306	529	487	280	321	268	307	441	464	193	271	379	416
Exchange	96	<u>085</u>	<u>204</u>	<u>085</u>	<u>205</u>	<u>086</u>	206	<u>086</u>	206	093	217	088	<u>205</u>	256	367	094	218	107	234	088	218
	192	179	<u>301</u>	179	<u>301</u>	181	303	177	<u>299</u>	184	307	<u>176</u>	<u>299</u>	470	509	184	307	226	344	<u>176</u>	315
	336	329	<u>415</u>	331	417	331	417	338	422	351	432	<u>301</u>	<u>397</u>	1.268	883	349	431	367	448	<u>313</u>	427
	720	<u>838</u>	<u>690</u>	860	698	858	<u>599</u>	847	691	886	714	901	714	1.767	1.068	852	698	964	746	<u>839</u>	695
	Avg.	<u>358</u>	<u>402</u>	363	405	364	407	368	409	378	417	367	<u>404</u>	940	707	370	413	416	443	<u>354</u>	414
Weather	96	174	<u>212</u>	175	215	<u>165</u>	<u>209</u>	174	215	192	232	177	218	<u>158</u>	230	202	261	172	220	196	255
	192	221	<u>255</u>	221	<u>255</u>	<u>215</u>	<u>255</u>	224	<u>258</u>	240	271	225	259	<u>206</u>	277	242	298	219	261	237	296
	336	<u>277</u>	<u>295</u>	<u>277</u>	<u>296</u>	<u>273</u>	<u>296</u>	281	298	292	307	278	297	<u>273</u>	335	287	335	280	306	283	335
	720	<u>353</u>	<u>348</u>	355	<u>348</u>	<u>353</u>	<u>349</u>	359	351	364	353	<u>354</u>	<u>348</u>	398	418	351	386	365	359	<u>345</u>	381
	Avg.	<u>256</u>	<u>277</u>	257	<u>278</u>	<u>252</u>	<u>277</u>	260	281	272	291	259	281	259	315	271	320	259	287	265	317
Solar	96	<u>194</u>	<u>229</u>	207	246	207	246	201	<u>234</u>	322	339	234	286	310	331	312	399	250	292	290	378
	192	<u>228</u>	<u>256</u>	239	270	240	272	<u>238</u>	<u>261</u>	359	356	267	310	734	725	339	416	296	318	320	398
	336	<u>247</u>	<u>276</u>	260	287	262	290	<u>248</u>	<u>273</u>	397	369	290	315	750	735	368	430	319	330	353	415
	720	<u>251</u>	<u>275</u>	264	<u>291</u>	267	293	<u>249</u>	<u>275</u>	397	356	289	317	769	765	370	425	338	337	356	413
	Avg.	<u>230</u>	<u>259</u>	242	274	244	275	<u>234</u>	<u>261</u>	369	356	270	307	641	639	347	417	301	319	330	401
ECL	96	<u>139</u>	<u>235</u>	<u>139</u>	<u>233</u>	<u>139</u>	237	<u>148</u>	240	201	281	181	270	219	314	237	329	168	272	197	282
	192	<u>160</u>	<u>254</u>	<u>158</u>	<u>249</u>	165	261	167	258	201	283	188	274	231	322	236	330	184	289	196	285
	336	<u>176</u>	<u>271</u>	<u>177</u>	<u>271</u>	177	274	179	<u>272</u>	215	298	204	293	246	337	249	344	198	300	209	301
	720	<u>198</u>	<u>292&lt;/</u>																		

## F ABLATION STUDY

To demonstrate the effectiveness of our method, we conduct an ablation study using four ETT datasets (Zhou et al., 2021) to assess the impact of the following components, where the results are shown in Table F.1. The results indicate that incorporating all components yields the best performance, and adding the regularization term enhances the performance even with the bidirectional Mamba.

Method	Mamba		Reg.	CCM	ETTh1	ETTh2	ETTm1	ETTm2	Avg.
	#	w/o conv.							
S-Mamba	Bi	-	-	-	.457	.383	.398	.290	.382
-	Bi	✓	-	-	.441	.383	.396	.285	.376
-	Bi	-	✓	-	.452	.382	.394	.286	.378
-	Bi	✓	✓	-	.443	<u>.381</u>	.393	.285	<u>.376</u>
-	Bi	✓	✓	✓	<u>.435</u>	<u>.376</u>	<u>.390</u>	<u>.281</u>	<u>.370</u>
-	Uni	-	-	-	.455	.383	.403	.289	.383
-	Uni	✓	-	-	.442	.382	.400	.285	.377
-	Uni	-	✓	-	.449	.382	.396	.285	.378
-	Uni	✓	✓	-	.442	.382	.396	<u>.284</u>	<u>.376</u>
SOR-Mamba	Uni	✓	✓	✓	<u>.433</u>	<u>.376</u>	<u>.391</u>	<u>.281</u>	<u>.370</u>

Table F.1: Ablation studies with four ETT datasets.

## G CHANNEL ORDERS FOR TWO VIEWS

Figure G.1 illustrates the four candidates for generating two embedding vectors,  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , for regularization, based on whether the channel order is fixed or randomly permuted in each iteration. Results in Table 12 indicate that fixing the order during training yields the best performance, with performance degrading as the order becomes random, especially with many channels, though it remains robust with fewer channels. We argue that a fixed order is preferable due to the instability introduced by randomness during training, as shown in Figure G.1, which displays the training loss for two datasets (Zhou et al., 2021; Liu et al., 2022) with varying numbers of channels. The figure indicates that a random order causes instability, particularly with the regularization loss.

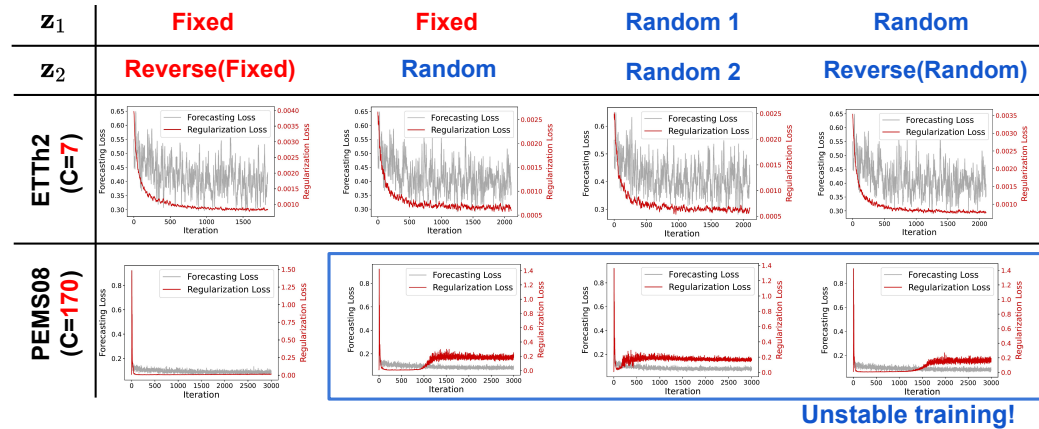


Figure G.1: Fixed vs. random order for generating two views,  $\mathbf{z}_1$  and  $\mathbf{z}_2$ .

## H ROBUSTNESS TO CHANNEL ORDER

To demonstrate that the proposed method effectively addresses the sequential order bias, we evaluate performance variations by permuting the channel order with five datasets (Zhou et al., 2021; Wu et al., 2021). Table H.1 shows the results, which indicate a small standard deviation across all horizons.

$H$	ETTh1	ETTh2	ETTm1	ETTm2	Exchange
96	.377 $\pm$ .0003	.292 $\pm$ .0011	.324 $\pm$ .0005	.179 $\pm$ .0003	.085 $\pm$ .0001
192	.428 $\pm$ .0002	.372 $\pm$ .0000	.369 $\pm$ .0005	.241 $\pm$ .0002	.179 $\pm$ .0001
336	.464 $\pm$ .0002	.415 $\pm$ .0002	.402 $\pm$ .0003	.302 $\pm$ .0001	.329 $\pm$ .0002
720	.464 $\pm$ .0004	.423 $\pm$ .0001	.467 $\pm$ .0009	.401 $\pm$ .0001	.838 $\pm$ .0014
Avg.	.434 $\pm$ .0002	.423 $\pm$ .0003	.391 $\pm$ .0001	.281 $\pm$ .0001	.358 $\pm$ .0003

Table H.1: Robustness to channel order.

## I ROBUSTNESS TO HYPERPARAMETER $\lambda$

Table I.1 shows the average MSE across four different horizons for the four ETT datasets (Zhou et al., 2021), using various values of  $\lambda$  that control the contribution of the regularization term. The results demonstrate the effectiveness of the regularization and its robustness to  $\lambda$ .

Dataset	SOR-Mamba					S-Mamba
	w/o Reg.	w/ Reg.				
	0	0.001	0.01	0.1	0.2	
ETTh1	<a href="#">.439</a>	<b>.433</b>	<b>.433</b>	<b>.433</b>	<b>.433</b>	.457
ETTh2	<a href="#">.382</a>	<b>.376</b>	<b>.376</b>	<b>.376</b>	<b>.376</b>	.383
ETTm1	.403	<b>.391</b>	<b>.391</b>	<b>.391</b>	<b>.391</b>	<a href="#">.398</a>
ETTm2	<a href="#">.285</a>	<b>.281</b>	<b>.281</b>	<b>.281</b>	<b>.281</b>	.290

Table I.1: Robustness to choice of  $\lambda$  for regularization.

## J ROBUSTNESS TO DISTANCE METRIC

To assess whether SOR-Mamba is sensitive to the choice of distance metric  $d$  for the regularization term and CCM when comparing the two matrices, we compare various metrics, including (negative) cosine similarity,  $\ell_1$  loss, and  $\ell_2$  loss. Tables J.1 and J.2 show the average MSE across four different horizons for the distance metric used in the regularization term and CCM, respectively, demonstrating that the performance is robust to the choice of distance metric, where we choose  $\ell_2$  loss throughout the experiment for both metrics.

Dataset	SOR-Mamba-SL			S-Mamba
	Cosine	$\ell_1$ Loss	$\ell_2$ Loss	
ETTh1	<b>.442</b>	<b>.442</b>	<b>.442</b>	<u>.457</u>
ETTh2	<b>.382</b>	<b>.382</b>	<b>.382</b>	<u>.383</u>
ETTm1	<b>.396</b>	<b>.396</b>	<b>.396</b>	<u>.398</u>
ETTm2	<b>.284</b>	<b>.284</b>	<b>.284</b>	<u>.290</u>
PEMS03	.145	.147	<u>.137</u>	<b>.133</b>
PEMS04	<u>.105</u>	<u>.105</u>	.107	<b>.096</b>
PEMS07	<u>.091</u>	<u>.091</u>	<u>.091</u>	<b>.090</b>
PEMS08	.162	<u>.159</u>	.162	<b>.157</b>
Exchange	.365	.365	<b>.363</b>	<u>.364</u>
Weather	<u>.256</u>	.257	.257	<b>.252</b>
Solar	<b>.242</b>	<b>.242</b>	<b>.242</b>	<u>.244</u>
ECL	<b>.167</b>	<u>.168</u>	.169	.174
Traffic	<u>.414</u>	<u>.414</u>	<b>.412</b>	.417
Average	<b>.265</b>	<b>.265</b>	<b>.265</b>	<u>.266</u>

Table J.1: Robustness to  $d$  for regularization.

Dataset	SOR-Mamba-SSL		S-Mamba
	$\ell_1$ Loss	$\ell_2$ Loss	
ETTh1	<u>.434</u>	<b>.433</b>	.457
ETTh2	<u>.379</u>	<b>.376</b>	.383
ETTm1	<b>.391</b>	<b>.391</b>	<u>.398</u>
ETTm2	<b>.281</b>	<b>.281</b>	<u>.290</u>
PEMS03	<b>.121</b>	<b>.121</b>	<u>.133</u>
PEMS04	<u>.099</u>	<u>.099</u>	<b>.096</b>
PEMS07	<u>.089</u>	<b>.088</b>	.090
PEMS08	<b>.140</b>	<u>.142</u>	.157
Exchange	<b>.358</b>	<b>.358</b>	<u>.364</u>
Weather	<u>.256</u>	<u>.256</u>	<b>.252</b>
Solar	<u>.232</u>	<b>.230</b>	.244
ECL	<b>.167</b>	<u>.168</u>	.174
Traffic	<b>.402</b>	<b>.402</b>	<u>.417</u>
Average	<u>.258</u>	<b>.257</b>	.266

Table J.2: Robustness to  $d$  for CCM.

## K COMPARISON OF GPU MEMORY USAGE

Figure K.1 visualizes GPU memory usage by dataset and method, demonstrating that our method is more efficient than both S-Mamba (Wang et al., 2024) and iTransformer (Liu et al., 2024a). Specifically, Mamba-based methods are more efficient than Transformer-based methods when  $C$  is large, as Mamba has nearly-linear complexity, whereas Transformers have quadratic complexity.

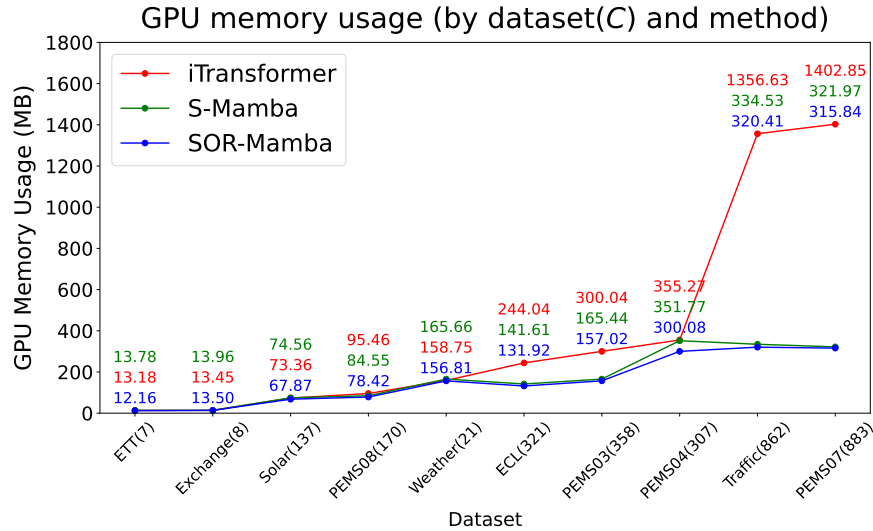


Figure K.1: Comparison of GPU memory usage.

## L STATISTICS OF RESULTS OVER MULTIPLE RUNS

To assess the consistency of SOR-Mamba’s performance, we present the statistics from results using five different random seeds. We calculate the mean and standard deviation for both MSE and MAE, detailed in Tables L.1, L.2, and L.3. which reveal that our method maintains consistent performance in both self-supervised and supervised settings.

Models		Ours			
		FT		SL	
Metric		MSE	MAE	MSE	MAE
ETTh1	96	.377 $\pm$ .001	.398 $\pm$ .001	.385 $\pm$ .000	.398 $\pm$ .000
	192	.428 $\pm$ .001	.429 $\pm$ .000	.432 $\pm$ .001	.428 $\pm$ .000
	336	.464 $\pm$ .001	.448 $\pm$ .001	.476 $\pm$ .000	.448 $\pm$ .000
	720	.464 $\pm$ .001	.469 $\pm$ .006	.476 $\pm$ .003	.476 $\pm$ .002
	Avg.	.433 $\pm$ .000	.436 $\pm$ .002	.442 $\pm$ .001	.438 $\pm$ .000
ETTh2	96	.292 $\pm$ .004	.348 $\pm$ .003	.299 $\pm$ .001	.348 $\pm$ .001
	192	.372 $\pm$ .001	.397 $\pm$ .001	.375 $\pm$ .001	.399 $\pm$ .001
	336	.415 $\pm$ .001	.431 $\pm$ .000	.423 $\pm$ .000	.435 $\pm$ .000
	720	.423 $\pm$ .001	.445 $\pm$ .001	.431 $\pm$ .002	.446 $\pm$ .001
	Avg.	.376 $\pm$ .001	.405 $\pm$ .001	.382 $\pm$ .001	.407 $\pm$ .000
ETTm1	96	.324 $\pm$ .002	.362 $\pm$ .002	.324 $\pm$ .004	.367 $\pm$ .003
	192	.369 $\pm$ .002	.385 $\pm$ .001	.375 $\pm$ .002	.387 $\pm$ .001
	336	.402 $\pm$ .002	.408 $\pm$ .001	.408 $\pm$ .000	.408 $\pm$ .000
	720	.467 $\pm$ .002	.444 $\pm$ .001	.472 $\pm$ .001	.444 $\pm$ .001
	Avg.	.391 $\pm$ .001	.400 $\pm$ .001	.396 $\pm$ .001	.401 $\pm$ .001
ETTm2	96	.179 $\pm$ .001	.261 $\pm$ .001	.181 $\pm$ .000	.265 $\pm$ .000
	192	.241 $\pm$ .000	.304 $\pm$ .000	.246 $\pm$ .001	.307 $\pm$ .001
	336	.302 $\pm$ .002	.342 $\pm$ .002	.306 $\pm$ .001	.345 $\pm$ .000
	720	.401 $\pm$ .002	.400 $\pm$ .002	.403 $\pm$ .002	.401 $\pm$ .001
	Avg.	.281 $\pm$ .001	.327 $\pm$ .000	.284 $\pm$ .001	.329 $\pm$ .000

Table L.1: Results of TS forecasting over five runs - 1) ETT datasets.

+



Models		Ours			
		FT		SL	
Metric		MSE	MAE	MSE	MAE
PEMS03	12	.066 $\pm$ .001	.170 $\pm$ .001	.066 $\pm$ .001	.170 $\pm$ .001
	24	.088 $\pm$ .001	.197 $\pm$ .001	.090 $\pm$ .001	.200 $\pm$ .001
	48	.134 $\pm$ .002	.245 $\pm$ .003	.167 $\pm$ .001	.280 $\pm$ .001
	96	.193 $\pm$ .005	.297 $\pm$ .006	.225 $\pm$ .003	.318 $\pm$ .002
	Avg.	.121 $\pm$ .002	.227 $\pm$ .002	.137 $\pm$ .001	.242 $\pm$ .001
PEMS04	12	.074 $\pm$ .002	.175 $\pm$ .003	.077 $\pm$ .000	.180 $\pm$ .000
	24	.086 $\pm$ .003	.192 $\pm$ .005	.091 $\pm$ .001	.197 $\pm$ .001
	48	.106 $\pm$ .001	.214 $\pm$ .005	.115 $\pm$ .002	.221 $\pm$ .003
	96	.129 $\pm$ .003	.233 $\pm$ .004	.143 $\pm$ .002	.248 $\pm$ .002
	Avg.	.099 $\pm$ .001	.203 $\pm$ .002	.107 $\pm$ .001	.212 $\pm$ .001
PEMS07	12	.059 $\pm$ .001	.155 $\pm$ .001	.060 $\pm$ .000	.156 $\pm$ .000
	24	.076 $\pm$ .005	.174 $\pm$ .004	.082 $\pm$ .000	.182 $\pm$ .000
	48	.098 $\pm$ .001	.199 $\pm$ .001	.107 $\pm$ .001	.209 $\pm$ .000
	96	.117 $\pm$ .003	.218 $\pm$ .003	.117 $\pm$ .001	.218 $\pm$ .001
	Avg.	.088 $\pm$ .001	.186 $\pm$ .001	.091 $\pm$ .000	.191 $\pm$ .000
PEMS08	12	.078 $\pm$ .000	.178 $\pm$ .000	.076 $\pm$ .001	.176 $\pm$ .000
	24	.103 $\pm$ .001	.205 $\pm$ .002	.109 $\pm$ .001	.212 $\pm$ .001
	48	.159 $\pm$ .001	.250 $\pm$ .001	.172 $\pm$ .003	.264 $\pm$ .003
	96	.229 $\pm$ .001	.295 $\pm$ .002	.290 $\pm$ .002	.334 $\pm$ .002
	Avg.	.142 $\pm$ .000	.232 $\pm$ .001	.162 $\pm$ .001	.247 $\pm$ .001

Table L.2: Results of TS forecasting over five runs - 2) PEMS datasets.

Models		Ours			
		FT		SL	
Metric		MSE	MAE	MSE	MAE
Exchange	96	.085 $\pm$ .001	.204 $\pm$ .002	.085 $\pm$ .001	.205 $\pm$ .001
	192	.179 $\pm$ .000	.301 $\pm$ .000	.179 $\pm$ .002	.301 $\pm$ .001
	336	.329 $\pm$ .001	.415 $\pm$ .001	.331 $\pm$ .000	.417 $\pm$ .000
	720	.838 $\pm$ .005	.690 $\pm$ .002	.860 $\pm$ .001	.698 $\pm$ .001
	Avg.	.358 $\pm$ .001	.402 $\pm$ .001	.363 $\pm$ .001	.405 $\pm$ .001
Weather	96	.174 $\pm$ .000	.212 $\pm$ .000	.175 $\pm$ .001	.215 $\pm$ .000
	192	.221 $\pm$ .000	.255 $\pm$ .000	.221 $\pm$ .000	.255 $\pm$ .000
	336	.277 $\pm$ .000	.295 $\pm$ .001	.277 $\pm$ .001	.296 $\pm$ .001
	720	.353 $\pm$ .001	.348 $\pm$ .001	.355 $\pm$ .000	.348 $\pm$ .000
	Avg.	.256 $\pm$ .000	.277 $\pm$ .000	.257 $\pm$ .000	.278 $\pm$ .000
Solar	96	.194 $\pm$ .005	.229 $\pm$ .004	.207 $\pm$ .000	.246 $\pm$ .001
	192	.228 $\pm$ .002	.256 $\pm$ .003	.239 $\pm$ .001	.270 $\pm$ .001
	336	.247 $\pm$ .006	.276 $\pm$ .005	.260 $\pm$ .001	.287 $\pm$ .001
	720	.251 $\pm$ .003	.275 $\pm$ .003	.264 $\pm$ .001	.291 $\pm$ .001
	Avg.	.230 $\pm$ .002	.259 $\pm$ .002	.242 $\pm$ .000	.274 $\pm$ .000
ECL	96	.139 $\pm$ .001	.235 $\pm$ .002	.139 $\pm$ .001	.233 $\pm$ .001
	192	.160 $\pm$ .002	.254 $\pm$ .002	.158 $\pm$ .001	.249 $\pm$ .001
	336	.176 $\pm$ .003	.271 $\pm$ .003	.177 $\pm$ .001	.271 $\pm$ .001
	720	.198 $\pm$ .003	.292 $\pm$ .006	.201 $\pm$ .003	.293 $\pm$ .002
	Avg.	.168 $\pm$ .001	.264 $\pm$ .001	.169 $\pm$ .001	.262 $\pm$ .001
Traffic	96	.378 $\pm$ .000	.258 $\pm$ .000	.378 $\pm$ .000	.259 $\pm$ .000
	192	.393 $\pm$ .001	.267 $\pm$ .001	.399 $\pm$ .000	.270 $\pm$ .000
	336	.399 $\pm$ .001	.276 $\pm$ .002	.416 $\pm$ .001	.279 $\pm$ .000
	720	.437 $\pm$ .001	.289 $\pm$ .002	.456 $\pm$ .001	.297 $\pm$ .001
	Avg.	.402 $\pm$ .000	.273 $\pm$ .001	.412 $\pm$ .000	.276 $\pm$ .000

Table L.3: Results of TS forecasting over five runs - 3) Other datasets.