Metrics

Saverio Dieni

November 2025

1 Results

Dataset / Metric	NsDiff	NsRatd-v1 w/o ref	NsRatd-v1 with ref	NsRatd-v2 w/o ref	NsRatd-v2 with ref
ETTm1					
MSE	0.4468	**0.4442**	*0.4460*	0.4920	0.4583
MAE	0.4333	*0.4307*	**0.4307**	0.4495	0.4413
CRPS	0.3402	0.3321	**0.3297**	0.3485	*0.3320*
QICE	2.4697	2.1753	*1.9338*	1.9374	**1.2383**
ETTm2					
MSE	0.3499	0.3488	*0.3471*	0.3482	**0.3470**
MAE	0.3876	0.3874	**0.3863**	*0.3868*	0.3868
CRPS	0.2912	*0.2909*	**0.2903**	0.2920	0.2925
QICE	2.6343	*2.5926*	**2.5649**	2.5935	2.7174
ETTh1					
MSE	0.6606	0.6672	0.6666	*0.6593*	**0.6284**
MAE	0.5638	0.5649	0.5646	*0.5624*	**0.5586**
CRPS	0.4149	0.4157	0.4153	*0.4107*	**0.4105**
QICE	*0.9897*	1.1159	1.0770	**0.9840**	1.1681
ETTh2					
MSE	**0.5062**	0.5346	0.5341	*0.5332*	0.5342
MAE	**0.4948**	0.5048	0.5044	*0.5042*	0.5050
CRPS	**0.3679**	0.3739	0.3728	0.3723	*0.3721*
QICE	1.7390	1.3623	1.2758	**1.1942**	*1.2521*
ILI					
MSE	0.2102	0.2101	0.2079	*0.2002*	**0.1862**
MAE	0.3079	0.3086	0.3048	*0.2935*	**0.2855**
CRPS	0.2991	0.2997	0.2971	*0.2730*	**0.2677**
QICE	7.5231	7.5151	7.5161	*7.2827*	**7.2796**
EXG					
MSE	*0.0204*	**0.0204**	0.0205	0.0205	0.0206
MAE	0.1049	0.1049	*0.1048*	**0.1047**	0.1048
CRPS	**0.0776**	0.0777	0.0776	*0.0776*	0.0777
QICE	*2.5381*	2.5942	2.5478	**2.5277**	2.5551
Linear Synthetic					
MSE	*0.9236*	0.9252	1.0350	**0.9151**	1.0229
MAE	*0.7007*	0.7013	0.7411	**0.6973**	0.7362
CRPS	*0.4998*	0.5019	0.5323	**0.4968**	0.5328
QICE	**0.7572**	1.1359	1.3589	*0.7218*	2.3495
Quadratic Synthetic					
MSE	*1.0707*	1.0752	1.1603	**1.0680**	1.2198
MAE	*0.6606*	0.6625	0.6905	**0.6597**	0.6990
CRPS	*0.4711*	0.4759	0.4943	**0.4709**	0.5115
QICE	*0.8848*	2.1357	**0.4244**	1.1605	1.3291

Table 1: Comparison of models on ETT, ILI, EXG, Linear and Quadratic Synthetic datasets. Best results bold, second best italics.

Model	Best MSE	Best MAE	Best CRPS	Best QICE	Total Best
NsDiff	1	1	2	1	**5**
NsRatd-v1 w/o ref	2	0	0	0	**2**
NsRatd-v1 with ref	0	2	2	2	**6**
NsRatd-v2 w/o ref	2	3	3	2	**10**
NsRatd-v2 with ref	3	2	2	2	**9**

Table 2: Number of best results (lower is better) per model.

Model	MSE Top-2	MAE Top-2	CRPS Top-2	QICE Top-2	Total Top-2
NsDiff	4	3	4	4	**15**
NsRatd-v1 w/o ref	2	1	1	1	**5**
NsRatd-v1 with ref	2	3	2	3	**10**
NsRatd-v2 w/o ref	5	7	5	5	**22**
NsRatd-v2 with ref	3	2	4	3	**12**

Table 3: Top-2 results (first or second best) per model.

1.1 Results on the real-world datasets

Dataset / Metric	NsDiff	NsRatd-v1 w/o ref	NsRatd-v1 with ref	NsRatd-v2 w/o ref	NsRatd-v2 with ref
ETTm1					
MSE	0.4468	**0.4442**	*0.4460*	0.4920	0.4583
MAE	0.4333	*0.4307*	**0.4307**	0.4495	0.4413
CRPS	0.3402	0.3321	**0.3297**	0.3485	*0.3320*
QICE	2.4697	2.1753	*1.9338*	1.9374	**1.2383**
ETTm2					
MSE	0.3499	0.3488	*0.3471*	0.3482	**0.3470**
MAE	0.3876	0.3874	**0.3863**	*0.3868*	0.3868
CRPS	0.2912	*0.2909*	**0.2903**	0.2920	0.2925
QICE	2.6343	*2.5926*	**2.5649**	2.5935	2.7174
ETTh1					
MSE	0.6606	0.6672	0.6666	*0.6593*	**0.6284**
MAE	0.5638	0.5649	0.5646	*0.5624*	**0.5586**
CRPS	0.4149	0.4157	0.4153	*0.4107*	**0.4105**
QICE	*0.9897*	1.1159	1.0770	**0.9840**	1.1681
ETTh2					
MSE	**0.5062**	0.5346	0.5341	*0.5332*	0.5342
MAE	**0.4948**	0.5048	0.5044	*0.5042*	0.5050
CRPS	**0.3679**	0.3739	0.3728	0.3723	*0.3721*
QICE	1.7390	1.3623	1.2758	**1.1942**	*1.2521*
ILI					
MSE	0.2102	0.2101	0.2079	*0.2002*	**0.1862**
MAE	0.3079	0.3086	0.3048	*0.2935*	**0.2855**
CRPS	0.2991	0.2997	0.2971	*0.2730*	**0.2677**
QICE	7.5231	7.5151	7.5161	*7.2827*	**7.2796**
EXG					
MSE	*0.0204*	**0.0204**	0.0205	0.0205	0.0206
MAE	0.1049	0.1049	*0.1048*	**0.1047**	0.1048
CRPS	**0.0776**	0.0777	0.0776	*0.0776*	0.0777
QICE	*2.5381*	2.5942	2.5478	**2.5277**	2.5551

Table 4: Comparison of models on six real-world datasets. Best results bold, second best italics.

Model	Best MSE	Best MAE	Best CRPS	Best QICE	Total Best
NsDiff	1	1	2	0	**4**
NsRatd-v1 w/o ref	2	0	0	0	**2**
NsRatd-v1 with ref	0	2	2	1	**5**
NsRatd-v2 w/o ref	0	1	0	3	**4**
NsRatd-v2 with ref	3	2	2	2	**9**

Table 5: Number of best results (lower is better) per model.

Given the complete metric table, I counted how many times each model achieved the best value for each metric. I then produced two summaries: one considering only the best (top-1) results, and another counting both the best and second-best(top-2) results. Looking at the top-1 summary, the second version of NsRatd achieved the highest number of best metrics, especially when a reference is used, although NsDiff also showed good performance. The top-2 summary indicates that NsRatd-v2 achieves superior performance compared to NsRatd-v1. When considering only the real-world datasets, the advantage of NsRatd-v2, especially in the reference-based variant, becomes even more pronounced.

Model	MSE Top-2	MAE Top-2	CRPS Top-2	QICE Top-2	Total Top-2
NsDiff	2	1	2	2	**7**
NsRatd-v1 w/o ref	2	1	1	1	**5**
NsRatd-v1 with ref	2	3	2	2	**9**
NsRatd-v2 w/o ref	3	5	3	4	**15**
NsRatd-v2 with ref	3	2	4	3	**12**

Table 6: Top-2 results (first or second best) per model.

2 Results with a masked Input

Dataset / Metric	NsRatd-v1 w/o ref	NsRatd-v1 with ref	NsRatd-v2 w/o ref	NsRatd-v2 with ref
ETTm1				
MSE	0.8115	0.8131	*0.6156*	**0.5780**
MAE	0.6400	0.6396	*0.5406*	**0.5320**
CRPS	0.4818	0.4802	*0.4063*	**0.3921**
QICE	1.9580	1.8689	*1.1800*	**0.7987**
ETTm2				
MSE	0.7465	0.7455	**0.7440**	*0.7449*
MAE	0.6244	*0.6241*	0.6245	**0.6239**
CRPS	**0.4476**	*0.4480*	0.4485	0.4489
QICE	5.2870	5.3079	**4.7904**	*5.2561*
ETTh1				
MSE	0.8186	0.8177	*0.8009*	**0.7496**
MAE	0.6544	0.6541	*0.6504*	**0.6373**
CRPS	0.4729	0.4725	*0.4681*	**0.4661**
QICE	**0.9118**	*0.9342*	1.1839	1.9962
ETTh2				
MSE	**0.6615**	0.6631	0.6609	*0.6634*
MAE	**0.5721**	0.5724	*0.5713*	0.5740
CRPS	*0.4191*	0.4191	0.4185	**0.4265**
QICE	*1.9684*	1.9847	2.0698	**2.6071**
ILI				
MSE	0.4439	0.4328	*0.4256*	**0.4028**
MAE	0.4787	0.4671	*0.4648*	**0.4506**
CRPS	0.3759	0.3703	*0.3510*	**0.3439**
QICE	5.0345	4.9894	**4.2482**	*4.3229*
EXG				
MSE	0.4928	**0.4922**	0.4929	*0.4928*
MAE	0.5025	**0.5021**	0.5023	*0.5022*
CRPS	**0.3805**	*0.3808*	0.3809	0.3809
QICE	**4.9703**	5.0579	*5.0563*	5.0593
Linear Synthetic				
MSE	*0.9393*	1.0281	**0.9299**	1.0139
MAE	*0.7093*	0.7412	**0.7055**	0.7369
CRPS	*0.5051*	0.5362	**0.5016**	0.5377
QICE	*0.6184*	2.0087	**0.4192**	3.0205
Quadratic Synthetic				
MSE	*1.0763*	1.1500	**1.0709**	1.1836
MAE	*0.6640*	0.6856	**0.6612**	0.6896
CRPS	*0.4744*	0.4924	**0.4712**	0.5075
QICE	1.4214	*1.0978*	**0.5048**	1.8914

Table 7: Comparison of models on ETT, ILI, EXG, Linear and Quadratic Synthetic datasets. The results are computed with a masked input, with 20% of the series masked. Best results bold, second best italics.

Model	Best MSE	Best MAE	Best CRPS	Best QICE	Total Best
NsRatd-v1 w/o ref	1	1	2	2	**6**
NsRatd-v1 with ref	1	1	0	0	**2**
NsRatd-v2 w/o ref	3	2	2	4	**11**
NsRatd-v2 with ref	3	4	4	2	**13**

Table 8: Number of best results (lower is better) per model.

Model	MSE Top-2	MAE Top-2	CRPS Top-2	QICE Top-2	Total Top-2
NsRatd-v1 w/o ref	3	3	5	5	**16**
NsRatd-v1 with ref	1	2	2	2	**7**
NsRatd-v2 w/o ref	5	6	5	6	**22**
NsRatd-v2 with ref	6	5	4	4	**19**

Table 9: Top-2 results (first or second best) per model.

2.1 Results on the real-world datasets, with a masked input

Dataset / Metric	NsRatd-v1 w/o ref	NsRatd-v1 with ref	NsRatd-v2 w/o ref	NsRatd-v2 with ref
		ETTm1		
MSE	0.8115	0.8131	*0.6156*	**0.5780**
MAE	0.6400	0.6396	*0.5406*	**0.5320**
CRPS	0.4818	0.4802	*0.4063*	**0.3921**
QICE	1.9580	1.8689	*1.1800*	**0.7987**
		ETTm2		
MSE	0.7465	0.7455	**0.7440**	*0.7449*
MAE	0.6244	*0.6241*	0.6245	**0.6239**
CRPS	**0.4476**	*0.4480*	0.4485	0.4489
QICE	5.2870	5.3079	**4.7904**	*5.2561*
		ETTh1		
MSE	0.8186	0.8177	*0.8009*	**0.7496**
MAE	0.6544	0.6541	*0.6504*	**0.6373**
CRPS	0.4729	0.4725	*0.4681*	**0.4661**
QICE	**0.9118**	*0.9342*	1.1839	1.9962
		ETTh2		
MSE	**0.6615**	0.6631	0.6609	*0.6634*
MAE	**0.5721**	0.5724	*0.5713*	0.5740
CRPS	*0.4191*	0.4191	0.4185	**0.4265**
QICE	*1.9684*	1.9847	2.0698	**2.6071**
		ILI		
MSE	0.4439	0.4328	*0.4256*	**0.4028**
MAE	0.4787	0.4671	*0.4648*	**0.4506**
CRPS	0.3759	0.3703	*0.3510*	**0.3439**
QICE	5.0345	4.9894	**4.2482**	*4.3229*
		EXG		
MSE	0.4928	**0.4922**	0.4929	*0.4928*
MAE	0.5025	**0.5021**	0.5023	*0.5022*
CRPS	**0.3805**	*0.3808*	0.3809	0.3809
QICE	**4.9703**	5.0579	*5.0563*	5.0593

Table 10: Comparison of models on six real-world datasets. The results are computed with a masked input, with 20% of the series masked. Best results bold, second best italics.

Model	Best MSE	Best MAE	Best CRPS	Best QICE	Total Best
NsRatd-v1 w/o ref	1	1	2	2	**6**
NsRatd-v1 with ref	1	1	0	0	**2**
NsRatd-v2 w/o ref	1	0	0	2	**3**
NsRatd-v2 with ref	3	4	4	2	**13**

Table 11: Number of best results (lower is better) per model.

In this section, I analyze the impact of providing a reference when the input series is heavily degraded. I simulated this condition by masking 20% of the input. Even under this setting, NsRatd-v2 continues to outperform NsRatd-v1. The use of a reference is particularly beneficial for the real-world datasets, where NsRatd-v2 shows a marked improvement, as illustrated in Table 11.

Model	MSE Top-2	MAE Top-2	CRPS Top-2	QICE Top-2	Total Top-2
NsRatd-v1 w/o ref	1	1	3	4	**9**
NsRatd-v1 with ref	1	2	2	1	**6**
NsRatd-v2 w/o ref	3	4	3	4	**14**
NsRatd-v2 with ref	6	5	4	4	**19**

Table 12: Top-2 results (first or second best) per model.