

# Replies to Reviewer5 bMGC

Thanks for the valuable questions.

**Q1.** *The experiment results cannot completely back up the major claim that the proposed deviation between global- and local- reconstructed time series is better at handling noise. In Fig. 4, the traditional reconstruction between the global reconstructed time series and the original time series achieves the best performance in the two datasets. Also, no false positive metric result is provided to support the claim that the proposed method could consistently achieve fewer false positives. The only quality analysis result for validation in Fig. 7 does not show the comparison between the proposed method and the traditional method.*

**Reply:**

$\hat{X}_G$  is different from traditional methods because its attention matrix mainly depends on remote points, owing to the KL divergence on the attention matrices between  $\hat{X}_L$  and  $\hat{X}_G$ .

$MSE(X, \hat{X}_G)$  can obtain better performance than  $MSE(\hat{X}_L, \hat{X}_G)$  when anomaly is as small as noise. Because  $\hat{X}_L$  and  $\hat{X}_G$  smooth the point by local and remote context, the small anomaly will be smoothed out by both of them and the  $MSE(\hat{X}_L, \hat{X}_G)$  will close to 0.

However,  $X$  retains the small anomaly, and therefore  $MSE(X, \hat{X}_G)$  can indicate these small anomalies. The small anomaly detection is very challenging and beyond our models' ability. We assume the noise is ubiquitous (white noise), the anomaly is sporadical and manifests in specific segments as described in Lines 89-90.

To verify MGRD's robustness against noise, we added noise to the test set with signal-to-noise ratios ranging from 10dB to 100dB. A higher signal-to-noise ratio indicates less noise, and "None" indicates no noise added. We experimentally compare the F1 scores using  $MSE(\hat{X}_L, \hat{X}_G)$  with  $MSE(X, \hat{X}_G)$  and report the results in the Figure a1. In the univariate dataset UCR and the multivariate dataset SWaT and PSM,  $MSE(X, \hat{X}_G)$  is weaker than  $MSE(\hat{X}_L, \hat{X}_G)$ . In the SWaT dataset, when the noise becomes 10 dB, the detection performance of  $MSE(X, \hat{X}_G)$  decreases sharply, while  $MSE(\hat{X}_L, \hat{X}_G)$  still maintains a more stable level. The above results indicate that using  $\hat{X}_L$  in the place of  $X$  can boost the robustness to noise.

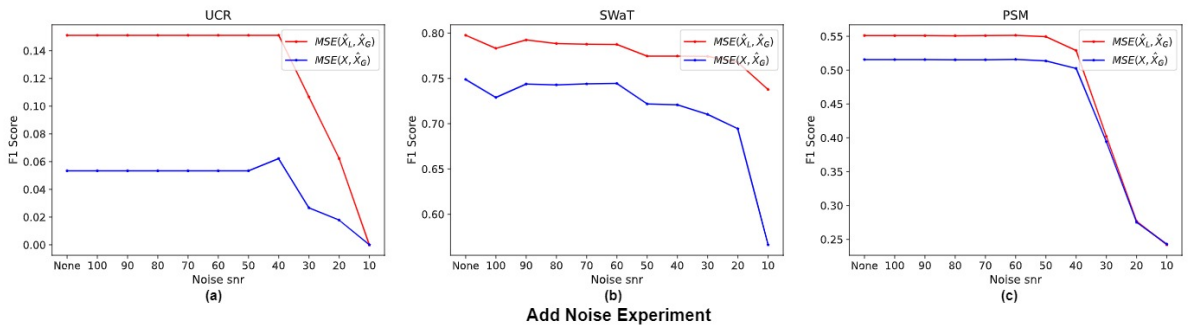


Figure a1. Robustness of MGRD on noise

To demonstrate if the MGRD can achieve fewer false positives, we report the FPR of  $MSE(\hat{X}_L, \hat{X}_G)$ ,  $MSE(X, \hat{X}_G)$ ,  $MSE(X, \hat{X}_L)$  on all datasets in Figure a2. The lower value indicates better performance. We can see that  $MSE(\hat{X}_L, \hat{X}_G)$  achieves the best on all datasets. Because both  $\hat{X}_L$  and  $\hat{X}_G$  can smooth noise, it can effectively reduce FPR.

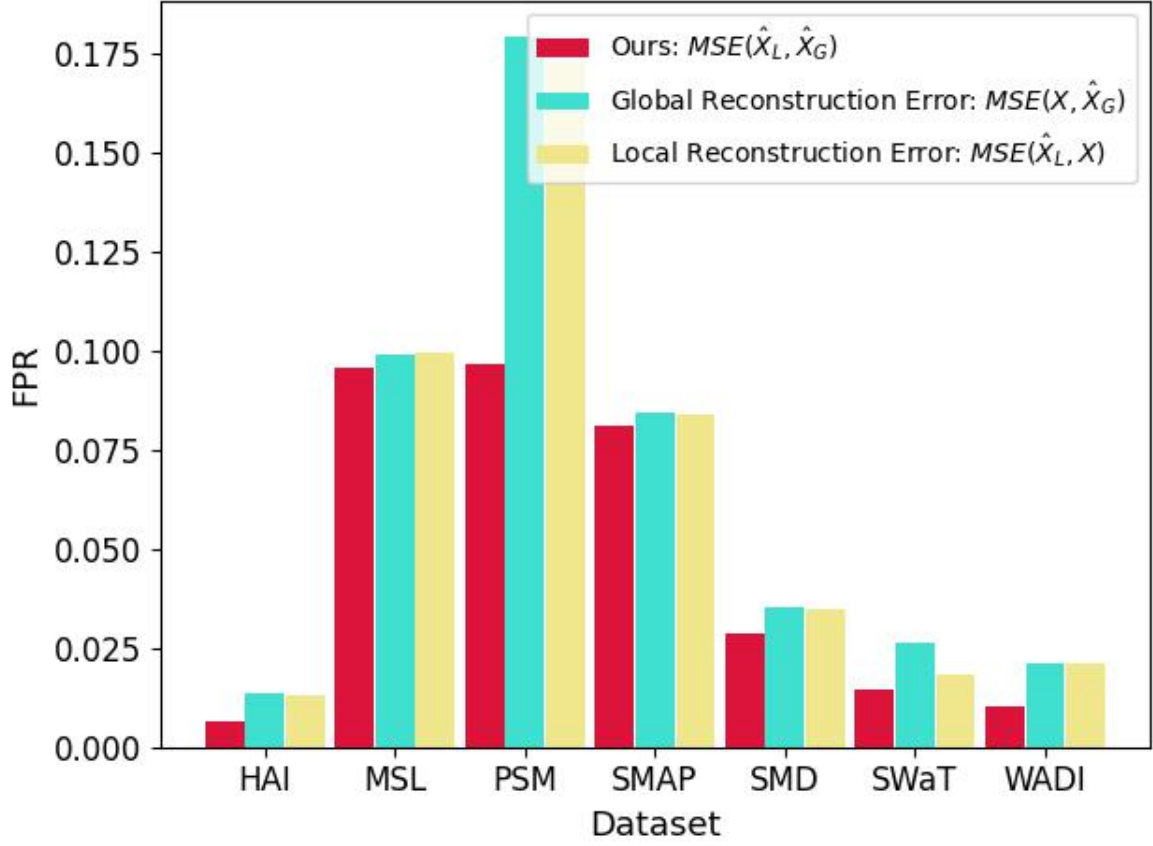


Figure a2. FPR on all datasets

We provide the Anomaly score of Anomaly Transformer(AT), TransAD, USAD, TimesNet, MSCRED, and CAE-M on PSM, SMD HAI in Figure a3, Figure a4, and Figure a5. We can see that our method can obtain better performance in the most anomalous areas.

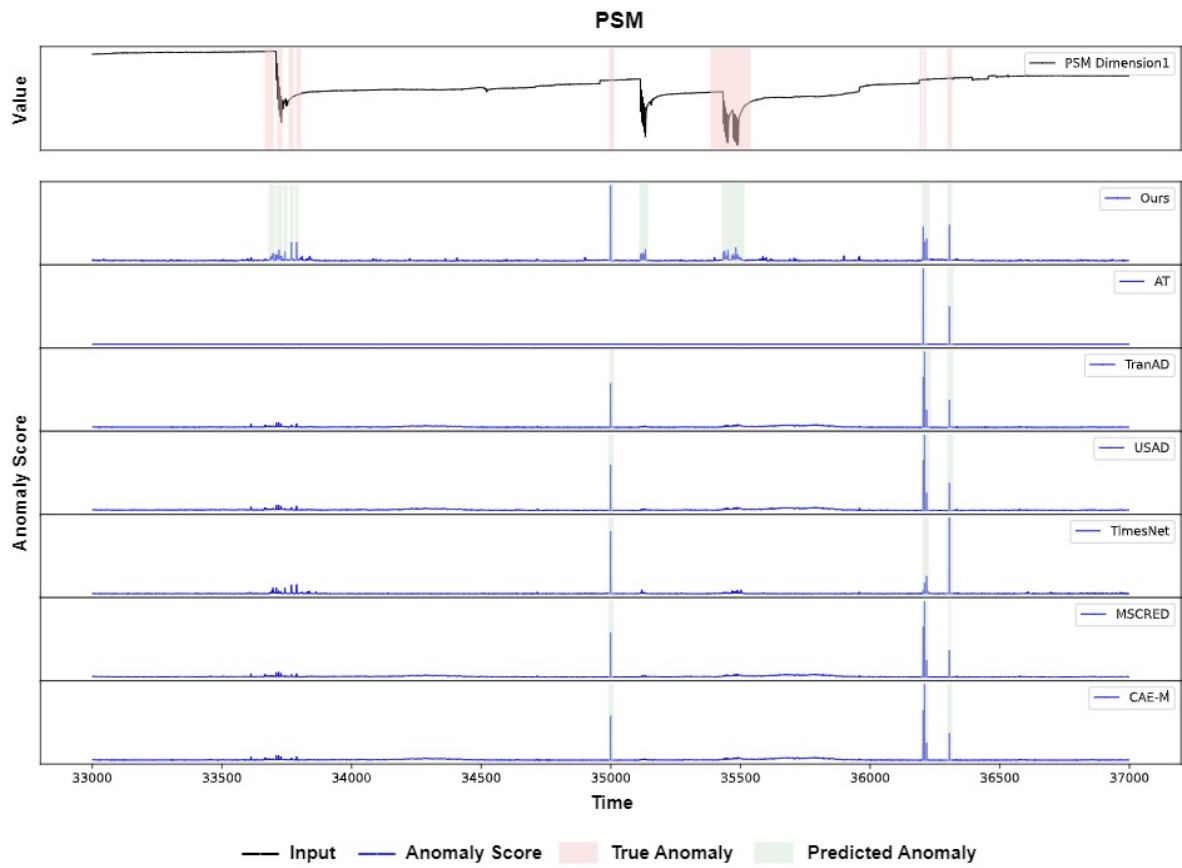


Figure a3. Anomaly Score Visualization on PSM

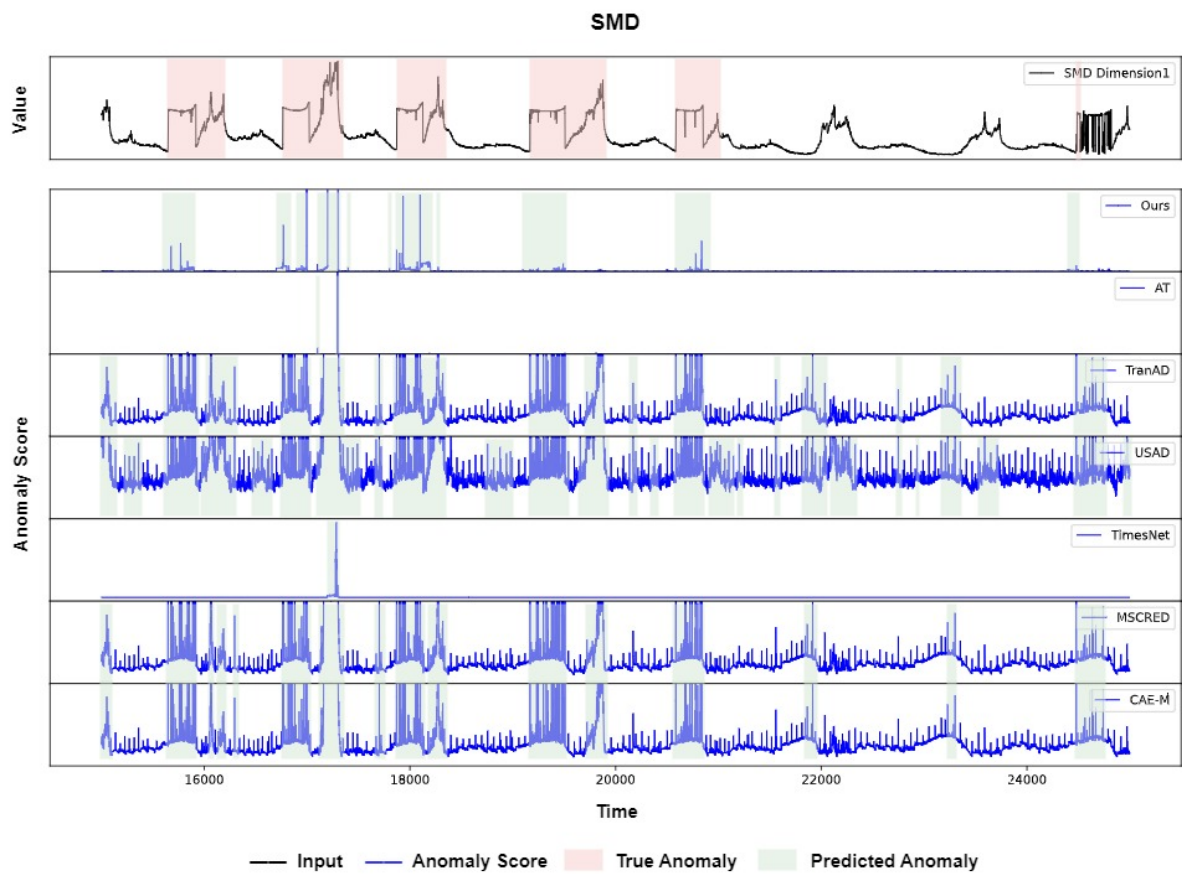


Figure a4. Anomaly Score Visualization on SMD

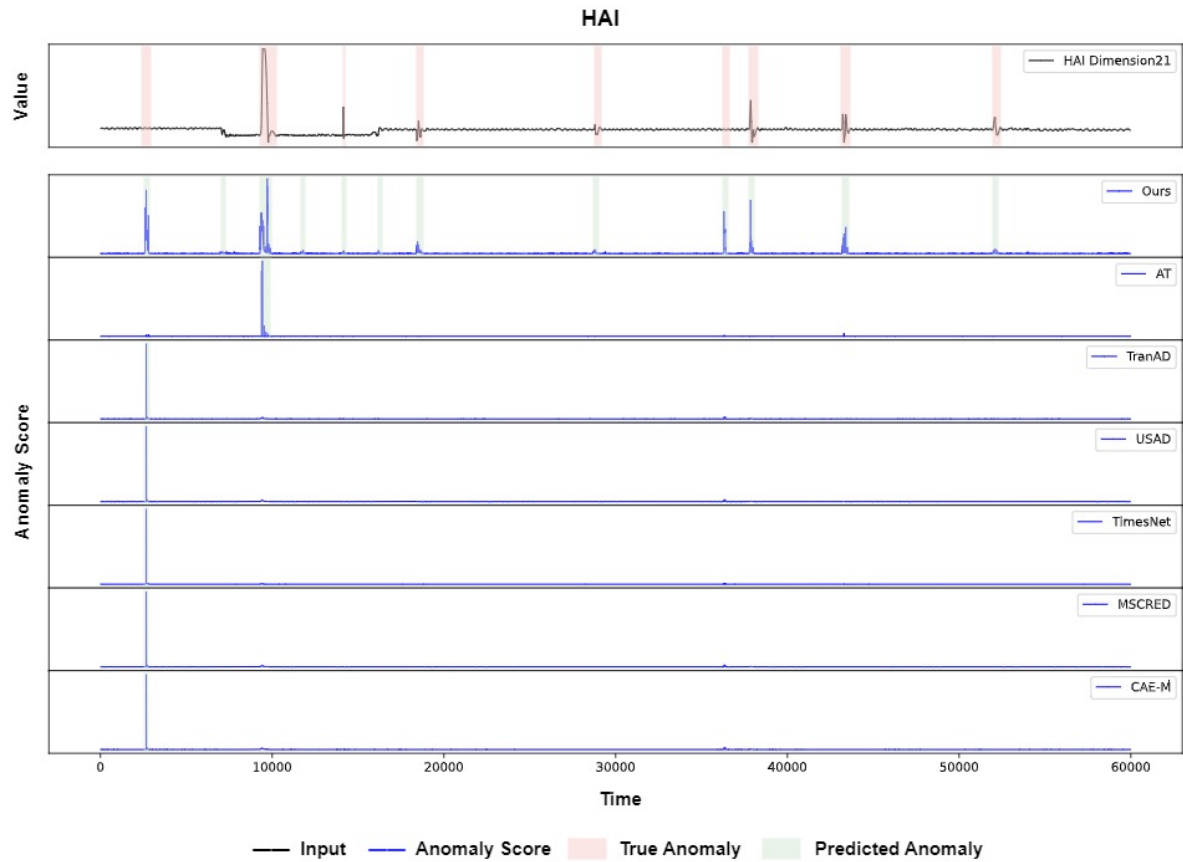


Figure a5. Anomaly Score Visualization on HAI

**Q2.** *There is no explicit definition for noise, which makes it hard to know what is the target the proposed method is really optimized.*

**Reply:** We assume that the noise is ubiquitous white noise, and the anomaly is sporadic and manifests in specific segments. Under these assumptions, noise can be smoothed out by local reconstruction and the anomaly can be smoothed by global reconstruction.

**Q3.** *Some compared methods' results are incomplete. There are 4 (out of 10) methods missing main results across multiple datasets. This decreases the confidence of the main comparison results.*

**Reply:** Our experiments were primarily conducted using a 16GB accelerator card. Adjusting the parameters to simplify the model resulted in significantly lower performance, which we deemed not representative of the model's capabilities. Consequently, these results were not included in the report. We believe that the results presented provide a fair evaluation of the model's performance. However, to address your concerns regarding the completeness of the comparison, we are open to removing the corresponding rows and columns from Table 2 to ensure clarity and maintain the integrity of our comparisons.