We thank all reviewers for their valuable comments and suggestions. To address the point raised by the first two reviewers regarding the comparison against recent DL methods, we evaluated our approach against DeepAR [1] and MQCNN [3], which we believe are a fair representation of the state-of-the-art in deep-learning-based forecasting. We also compared with DeepGLO [2] on two datasets provided by the authors.

dataset	exchange	solar	elec	traffic	taxi	wiki
DeepAR [1]	**0.007**	0.379	0.063	0.147	**0.332**	0.337
MQCNN [3]	0.013	0.482	0.078	0.177	0.657	0.277
GP-Copula (Ours)	0.008	**0.371**	**0.056**	**0.133**	0.360	**0.236**

dataset	electricity	traffic
DeepGLO [2]	0.109	0.221
TRMF [2]	0.105	0.210
GP-Copula (Ours)	**0.083**	**0.168**

Table 1: CRPS for additional baselines (left) and comparison with [2] when measuring WAPE (right).

We hope these additional experiments demonstrate that the proposed approach is not only competitive against classical statistical techniques, but also against state-of-the-art approaches. Also note that none of these competing approaches models correlations across time series in their forecasts (in fact DeepGLO only provides point forecasts). The code for running the benchmark will be released after publication to help the community to evaluate forecasting methods.

Reviewer #2 *"The paper does not seem to have enough original contribution. Authors mostly have adopted existing techniques (see references below) and algorithms and combined them together ..."* Thank you for the relevant references – we will add them to the paper. While we agree that the individual ingredients of our technique (Gaussian copula models, GPs with low-rank covariance matrices, RNN models for time series forecasting) have been proposed and studied before, we believe that the way they are combined in our approach is novel and non-trivial, and addresses what we believe to be a highly-relevant, practical problem, namely robust high-dimensional time series forecasting.

"The synthetic data are simple periodic data, expected that predicted line follow the synthetic much more closely." While the plot of Cov coefficient is a smooth cyclic periodic signal, the observed data is very noisy, making it hard to regress the signal: we added a plot of the raw series to better illustrate that the signal is very weak compared to the noise to show how this task is difficult.

"As main goal of the paper is to perform the superior forecasting, it will be fair that results will be compared to paper below" As mentioned above, DeepGLO [2] only produces point forecasts (not distributions) and does not deal with the high-dimensional covariance matrices that we tackle in this paper. Further, the referenced paper was published on arXiv only a week before the submission deadline. We have since contacted the authors and obtained the first two data sets used in their evaluation. The preliminary results above indicate that our method outperforms their approach in the simpler point forecasting setting.

Reviewer #3 *"In Table 1 ... more significant to validate that the model in this paper with the same settings can improve the performance of several state-of-the-art deep learning models."* We added a comparison with [1, 3, 2] to represent SOTA in deep-learning forecasting, see Tab. 1.

"The number of Monte Carlo sampling ... excessive sampling may increase the complexity of the entire model." Runtime complexity of prediction increases linearly with the number of samples (training is not affected as there is no sampling at that stage). We added an explanation in the manuscript and ran an experiment with different numbers of samples, characterizing the impact of this parameter on the model's performance.

Reviewer #4 *"How are the CRPS-sum error bars being computed?"* By rerunning each method with three different seed and reporting mean/std (this detail was inadvertently omitted in the submitted manuscript).

"More detailed experiments for certain aspects of the algorithm tuning different constant choices (e.g. rank, marginal discretization level, embedding vector size)." The rank hyperparameter is investigated in the appendix where we show that (as one would expect) using a larger rank decreases training error but increases test error due to overfiting. The sensitivity of the method to the other hyperparameters relative to properties of the data is an aspect that would be interesting to investigate further, but the extensive experiments required come at a significant hardware cost.

"CRPS is nice, but MSE, loglike, and visualizing temporal patterns in residuals ..." We added MSE and will add loglike when possible (as some models cannot compute it). Visualizing the pattern of residuals is done in the appendix; we will reference this more clearly in the main text and add the analysis for all datasets.

[1] David Salinas et al. Deepar: Probabilistic forecasting with autoregressive recurrent networks. 2017.

[2] Rajat Sen et al. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. 2019.

[3] Ruofeng Wen et al. A multi-horizon quantile recurrent forecaster. 2017.