A Supplamentary of Main Results

We report the results in terms of MSE and MAE over three re-runs by:

1. First average predictions over re-runs, then estimate MSE and MAE, reported in Table 1 and 2.
2. First estimate MSE and MAE for each re-runs and report the average over re-runs, reported in Table 3 and 4.

The results with respect to the first method are reported in Table 1 and 2, where the results with respect to the second method are reported in Table 3 and 4.

For the significance analysis, we report:

1. paired-t-test over MSE/MAE at each point for both evaluation methods discussed above.
2. unpaired standard-error over the average of MSE and MAE of re-runs for the the second evaluation method discussed above.

We specifically report the paired-t-test, because when comparing the predictions of two methods over time, it is important to consider that the mean performance of the methods may be a mix of several populations with different levels of difficulty. Some time periods may be easier to predict than others, which can influence the overall mean performance of the methods. The paired t-test helps handle the case where some predictions are easy and some are difficult by comparing the differences (deltas) between the paired observations from the two methods. The test assesses whether the difference in performance between the two methods is consistent across all observations, regardless of the variation in difficulty between them. However, we show that even if we are using a standard-error overlap test, we find significant differences in many comparisons.

1.1 Results and Discussion

Table 1 and 2 summarize the evaluation results of autoformer and informer model's treatmeants on the three datasets, results are reported as MSE and MAE over averaged predictions across three independent restarts. All forecasting models are evaluated on their ability to predict for 24, 48, 72, and 96 future time steps. Our proposed GP-based corruption-resilient forecasting model outperforms other treatments of autoformer and informer models in **83**% of the cases. Compared to using a corruption/denoising approach during training (D-C-Input) and corruption/denoising with isotropic noise, our model consistently performs better or equal, supporting our hypothesis on benefits of 1) our end-to-end predict-corrupt-denoise model and 2) our proposed GP corruption model.

Next we report the results as average of MSE and MAE over three independent re-runs instead of average over predictions in Table 3 and 4. Here, in addition to paired t-test over average of MSE and MAE of each time step, we report the unpaired standard error to show that adopting a very conservative stance, our proposed model significantly outperforms other treatments for the majority of cases. Our proposed GP-based corruption-resilient forecasting model outperforms other treatments of autoformer and informer models in **71**%/**63**% and **83**%/**71**% of all cases with respect to paired t-test and unpaired standard error respectively.

Due to space limitations, we did not include these results and the corresponding significant analysis in the main manuscript.

Table 1: Overall results of the quantitative evaluation of corruption-resilient forecasting models in terms of MSE and MAE. We compare all forecasting models on all three datasets with different number of forecasting steps. A lower MSE and MAE indicates a better model. In **83**% of the cases our predict-corrupt-denoise approach with GPs improves other denoising baselines and the original forecasting methods. The results are reported as MSE and MAE of **average over predictions** of three independent re-runs. The ▼ denotes significant deterioration using a paired-t-test at $p \leq 0.05$ compared to the best model.

Dataset	Horizon	Metric	Autoformer					
			No-D	D-GP (Ours)	D-Pred	Res-Boos	D-Iso	D-C-Input
Traffic	24	MSE	0.405▼	**0.388**	0.400▼	0.415▼	0.412▼	0.430▼
		MAE	0.340▼	**0.333**	0.345▼	0.345▼	0.334▼	0.349▼
	48	MSE	0.416▼	**0.387**	0.417▼	0.422▼	0.422▼	0.410▼
		MAE	0.368▼	**0.328**	0.351▼	0.350▼	0.343▼	0.361▼
	72	MSE	0.394▼	**0.380**	0.398▼	0.403▼	0.383▼	0.404▼
		MAE	**0.356**	0.358▼	0.361▼	0.365▼	**0.356**	0.379▼
	96	MSE	0.411▼	**0.376**	0.405▼	0.416▼	0.403▼	0.422▼
		MAE	0.366▼	**0.333**	0.362▼	0.376▼	0.359▼	0.379▼
Electricity	24	MSE	0.171▼	**0.165**	0.178▼	0.167▼	0.173▼	0.170▼
		MSE	0.258▼	**0.249**	0.272▼	0.258▼	0.265▼	0.263▼
	48	MSE	0.207▼	**0.188**	0.209▼	0.195▼	0.204▼	0.200▼
		MAE	0.301▼	**0.275**	0.306▼	0.285▼	0.292▼	0.288▼
	72	MSE	0.198▼	0.209▼	**0.197**	0.277▼	0.203▼	0.212▼
		MAE	0.297▼	0.303▼	**0.295**	0.329▼	0.303▼	0.305▼
	96	MSE	0.286▼	**0.211**	0.227▼	0.233▼	0.231▼	0.219▼
		MAE	0.372▼	**0.304**	0.324▼	0.325▼	0.325▼	0.318▼
Solar	24	MSE	0.473▼	**0.446**	0.449▼	0.480▼	0.474▼	0.457▼
		MAE	0.566▼	**0.548**	0.549▼	0.569▼	0.561▼	0.555▼
	48	MSE	0.574▼	**0.546**	0.605▼	0.588▼	0.603▼	0.598▼
		MAE	0.638▼	**0.612**	0.656▼	0.650▼	0.656▼	0.655▼
	72	MSE	0.698▼	0.671▼	0.691▼	**0.661**	0.667▼	0.670▼
		MAE	0.729▼	**0.702**	0.724▼	**0.702**	**0.702**	0.709▼
	96	MSE	0.730▼	**0.713**	0.732▼	0.739▼	0.739▼	0.733▼
		MAE	0.747▼	**0.725**	0.746▼	0.739▼	0.754▼	0.745▼

Table 2: Overall results of the quantitative evaluation of corruption-resilient forecasting models in terms of MSE and MAE. We compare all forecasting models on all three datasets with different number of forecasting steps. A lower MSE and MAE indicates a better model. In **83**% of the cases our predict-corrupt-denoise approach with GPs improves other denoising baselines and the original forecasting methods. The results are reported as MSE and MAE of **average over predictions** of three independent re-runs. The ▼ denotes significant deterioration using a paired-t-test at $p \leq 0.05$ compared to the best model.

Dataset	Horizon	Metric	Informer					
			No-D	D-GP (Ours)	D-Pred	Res-Boos	D-Iso	D-C-Input
Traffic	24	MSE	0.421▼	**0.398**	0.406▼	0.435▼	0.415▼	0.473▼
		MAE	**0.329**	0.335▼	0.337▼	0.331▼	0.342▼	0.379▼
	48	MSE	0.434▼	**0.399**	0.420▼	0.438▼	0.421▼	0.447▼
		MAE	0.359▼	**0.329**	0.347▼	0.350▼	0.352▼	0.372▼
	72	MSE	0.436▼	**0.380**	0.392▼	0.407▼	0.395▼	0.421▼
		MAE	0.377▼	**0.345**	0.348▼	0.353▼	**0.353**	0.375▼
	96	MSE	0.402▼	0.397▼	**0.394**	0.412▼	0.402▼	0.414▼
		MAE	0.354▼	0.350▼	**0.348**	0.379▼	0.362▼	0.361▼
Electricity	24	MSE	0.222▼	**0.193**	0.204▼	0.225▼	0.212▼	0.230▼
		MAE	0.300▼	**0.290**	0.295▼	0.302▼	0.298▼	0.318▼
	48	MSE	0.262▼	**0.222**	0.241▼	0.261▼	0.229▼	0.256▼
		MAE	0.349▼	**0.311**	0.333▼	0.343▼	0.325▼	0.343▼
	72	MSE	0.280▼	**0.238**	**0.263**	0.262▼	0.253▼	0.268▼
		MAE	0.371▼	**0.336**	0.362▼	0.359▼	0.359▼	0.367▼
	96	MSE	0.289▼	**0.242**	0.279▼	0.283▼	0.275▼	0.275▼
		MAE	0.378▼	**0.342**	0.384▼	0.375▼	0.379▼	0.370▼
Solar	24	MSE	0.524▼	**0.455**	0.457▼	0.498▼	0.465▼	0.512▼
		MAE	0.597▼	0.553▼	**0.551**	0.573▼	0.563▼	0.596▼
	48	MSE	0.629▼	**0.556**	0.590▼	0.623▼	0.570▼	0.629▼
		MAE	0.681▼	**0.624**	0.649▼	0.675▼	0.635▼	0.681▼
	72	MSE	0.729▼	**0.643**	0.708▼	0.748▼	0.707▼	0.726▼
		MAE	0.752▼	**0.690**	0.736▼	0.763▼	0.735▼	0.735▼
	96	MSE	0.770▼	**0.708**	0.739▼	0.781▼	0.766▼	0.777▼
		MAE	0.772▼	**0.727**	0.753▼	0.777▼	0.764▼	0.766▼

Table 3: Overall results of the quantitative evaluation of corruption-resilient forecasting models in terms of MSE and MAE. We compare all forecasting models on all three datasets with different number of forecasting steps. A lower MSE and MAE indicates a better model. The results are reported as **average of MSE and MAE** over three independent re-runs. The ▼ denotes significant deterioration using a paired-t-test at $p \leq 0.05$ compared to the best model. We also report the unpaired standard-error over the MSE and MAE of different re-runs. The results indicated with underline are significantly better with respect to paired t-test and unpaired standard-error. In **71**% and **63**% of all cases our predict-corrupt-denoise approach with GPs significantly improves other denoising baselines and the original forecasting methods with respect to paired-test and unpaired standard error.

Dataset	Horizon	Autoformer					
		No-D MSE MAE	D-Pred MSE MAE	Res-Boos MSE MAE	D-Iso MSE MAE	D-C-Input MSE MAE	D-GP (Ours) MSE MAE
Traffic	24	0.428▼ ±0.006 0.358▼ ±0.007	0.424▼ ±0.005 0.364▼ ±0.009	0.436▼ ±0.006 0.362▼ ±0.005	0.434▼ ±0.003 0.358▼ ±0.007	0.454▼ ±0.015 0.367▼ ±0.005	**0.409**±0.006 **0.350**±0.010
	48	0.441▼ ±0.004 0.388▼ ±0.006	0.438▼ ±0.007 0.371▼ ±0.005	0.440▼ ±0.002 0.368▼ ±0.002	0.442▼ ±0.001 0.367▼ ±0.006	0.436▼ ±0.007 0.390▼ ±0.016	**0.436**±0.001 **0.347**±0.001
	72	0.417▼ ±0.002 **0.375**± 0.003	0.421▼ ±0.004 0.383▼ ±0.002	0.428▼ ±0.004 0.387▼ ±0.005	0.412▼ ±0.002 0.381▼ ±0.005	0.426▼ ± 0.006 0.396▼ ± 0.001	**0.405**±0.001 0.379▼ ±0.013
	96	0.436▼ ±0.004 0.386▼ ±0.004	0.425▼ ±0.001 0.382▼ ±0.004	0.438▼ ±0.002 0.397▼ ±0.005	0.430▼ ±0.002 0.385▼ ±0.004	0.444▼ ±0.003 0.397▼ ±0.005	**0.394**±0.003 **0.352**±0.000
Electricity	24	0.184▼ ±0.003 0.273▼ ±0.008	0.186▼ ±0.001 0.283▼ ±0.002	0.176▼ ±0.001 0.271▼ ±0.003	0.182▼ ±0.001 0.276▼ ±0.002	0.184▼ ±0.007 0.280▼ ±0.002	**0.174**±0.001 **0.259**±0.002
	48	0.219▼ ±0.007 0.314▼ ±0.008	0.217▼ ±0.002 0.315▼ ±0.002	0.206▼ ±0.002 0.300▼ ±0.003	0.215▼ ± 0.003 0.306▼ ±0.002	0.210▼ ±0.003 0.301▼ ±0.002	**0.197**±0.003 **0.285**±0.004
	72	**0.209**±0.004 **0.311**±0.006	0.213▼ ±0.010 0.312▼ ±0.009	0.245▼ ±0.022 0.346▼ ±0.021	0.215▼ ±0.004 0.318▼ ±0.004	0.225▼ ±0.002 0.322▼ ±0.003	0.220▼ ±0.001 0.315▼ ±0.004
	96	0.308▼ ±0.014 0.393▼ ±0.010	0.238▼ ±0.008 0.337▼ ±0.005	0.245▼ ±0.011 0.341▼ ±0.005	0.243▼ ±0.002 0.340▼ ±0.002	0.231▼ ±0.004 0.332▼ ±0.005	**0.220**±0.001 **0.315**±0.001
Solar	24	0.522▼ ±0.003 0.589▼ ±0.002	**0.505**±0.015 **0.575**±0.008	0.528▼ ±0.006 0.593▼ ±0.005	0.523▼ ±0.006 0.585▼ ±0.008	0.519▼ ±0.004 0.585▼ ±0.006	0.508▼ ±0.009 0.577▼ ±0.009
	48	0.618▼ ±0.003 0.658▼ ±0.003	0.640▼ ±0.001 0.672▼ ±0.002	0.629▼ ±0.005 0.668▼ ±0.003	0.641▼ ±0.005 0.673▼ ±0.003	0.633▼ ±0.003 0.671▼ ±0.003	**0.605**±0.005 **0.639**±0.003
	72	0.732▼ ± 0.023 0.742▼ ±0.017	0.720▼ ±0.012 0.735▼ ±0.008	0.708▼ ±0.010 0.722▼ ±0.010	0.710▼ ±0.002 0.718▼ ±0.001	**0.700**±0.006 **0.712**±0.004	0.715▼ ±.003 0.719▼ ±0.001
	96	0.765▼ ±0.009 0.760▼ ±0.009	0.766▼ ±0.010 0.758▼ ±0.006	0.762▼ ±0.009 0.755▼ ±0.007	0.772▼ ±0.007 0.766▼ ±0.005	0.767▼ ±0.006 0.758▼ ±0.006	**0.756**±0.004 **0.741**±0.000

Table 4: Overall results of the quantitative evaluation of corruption-resilient forecasting models in terms of MSE and MAE. We compare all forecasting models on all three datasets with different number of forecasting steps. A lower MSE and MAE indicates a better model. The results are reported as **average of MSE and MAE** over three independent re-runs. The ▼ denotes significant deterioration using a paired-t-test at $p \leq 0.05$ compared to the best model. We also report the unpaired standard-error over the MSE and MAE of different re-runs. The results indicated with underline are significantly better with respect to paired t-test and unpaired standard-error. In **83**% and **71**% of all cases our predict-corrupt-denoise approach with GPs significantly improves other denoising baselines and the original forecasting methods with respect to paired-test and unpaired standard error.

Dataset	Horizon	Informer					
		No-D	D-Pred	Res-Boos	D-Iso	D-C-Input	D-GP (Ours)
		MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
Traffic	24	0.432▼ ±0.005	0.422▼ ±0.002	0.451▼ ±0.005	0.436▼ ±0.002	0.482▼ ±0.003	**0.418** ±0.005
		0.340 ±0.006	0.346▼ ±0.003	0.347▼ ±0.003	0.363▼ ±0.004	0.390▼ ±0.003	0.357▼ ±0.007
	48	0.455▼ ±0.014	0.433▼ ±0.003	0.458▼ ±0.011	0.442▼ ±0.007	0.472▼ ±0.014	**0.413** ±0.004
		0.380▼ ±0.013	0.364▼ ±0.003	0.372▼ ±0.009	0.372▼ ±0.005	0.394▼ ± 0.006	**0.346** ±0.001
	72	0.462▼ ±0.015	0.409▼ ±0.001	0.431▼ ±0.009	0.412▼ ±0.001	0.452▼ ±0.011	**0.398** ±0.001
		0.394▼ ± 0.010	**0.364** ±0.001	0.372▼ ±0.005	0.373▼ ±0.004	0.396▼ ±0.006	**0.364** ±0.003
	96	0.420▼ ±0.002	**0.409** ±0.003	0.435▼ ±0.007	0.415▼ ±0.006	0.438▼ ±0.015	0.411▼ ±0.003
		0.373▼ ±0.006	**0.364** ±0.008	0.396▼ ±0.005	0.376▼ ±0.007	0.380▼ ±0.011	**0.364** ±0.004
Electricity	24	0.242▼ ±0.006	0.214▼ ±0.005	0.241▼ ±0.009	0.232▼ ±0.001	0.243▼ ±0.007	**0.204** ±0.003
		0.319▼ ±0.005	0.306▼ ±0.004	0.314▼ ±0.009	0.314▼ ±0.003	0.327▼ ±0.003	**0.301** ±0.008
	48	0.289▼ ±0.013	0.262▼ ±0.007	0.274▼ ±0.014	0.252▼ ±0.003	0.283▼ ±0.004	**0.240** ±0.003
		0.369▼ ±0.009	0.349▼ ±0.004	0.353▼ ±0.009	0.343▼ ±0.002	0.364▼ ±0.005	**0.327** ±0.002
	72	0.315▼ ±0.006	0.290▼ ±0.013	0.278▼ ±0.008	0.280▼ ±0.006	0.298▼ ±0.008	**0.261** ±0.001
		0.397▼ ±0.002	0.384▼ ±0.008	0.374▼ ±0.006	0.382▼ ±0.008	0.389▼ ±0.008	**0.352** ±0.003
	96	0.309▼ ±0.017	0.299▼ ±0.004	0.303▼ ±0.000	0.295▼ ±0.005	0.299▼ ±0.007	**0.263** ±0.000
		0.392▼ ±0.011	0.397▼ ±0.006	0.391▼ ±0.001	0.392▼ ±0.005	0.389▼ ±0.007	**0.358** ±0.004
Solar	24	0.560▼ ±0.002	**0.501** ±0.006	0.539▼ ±0.010	**0.501** ±0.006	0.565▼ ±0.012	0.510▼ ±0.007
		0.612▼ ±0.002	**0.573** ±0.001	0.594▼ ±0.008	0.581▼ ±0.003	0.619▼ ±0.007	0.580▼ ±0.005
	48	0.693▼ ±0.021	0.654▼ ±0.016	0.670▼ ±0.016	0.635▼ ±0.007	0.694▼ ±0.023	**0.615** ±0.011
		0.705▼ ±0.013	0.675▼ ±0.011	0.694▼ ±0.012	0.663▼ ±0.005	0.705▼ ±0.012	**0.650** ±0.007
	72	0.804▼ ± 0.024	0.769▼ ±0.014	0.805▼ ±0.010	0.771▼ ±0.026	0.775▼ ±0.006	**0.714** ±0.022
		0.777▼ ±0.017	0.758▼ ±0.011	0.783▼ ±0.020	0.759▼ ±0.019	0.753▼ ±0.004	**0.717** ±0.013
	96	0.838▼ ±0.017	0.803▼ ±0.010	0.833▼ ±0.017	0.809▼ ±0.006	0.817▼ ±0.000	**0.762** ±0.010
		0.795▼ ±0.012	0.775▼ ±0.005	0.795▼ ±0.014	0.779▼ ±0.004	0.779▼ ±0.002	**0.748** ±0.006