A MOMOGP STRUCTURE CONSTRUCTION

Table 1 compares the results obtained from MOMoGPs constructed either using conditional independence tests or using random splitting. We construct MOMoGPs using conditional independence tests for the splitting of the output space, we employed the randomized conditional correlation test [Strobl et al., 2019]. In all experiments, we used a p-value of 0.5. We see that the use of a conditional independence test for the structure construction results in an overall improvement of the performance of MOMoGPs with respect to the RMSE and the MAE.

B MULTI-OUTPUT REGRESSION BENCHMARK RESULTS

We train and test GP, DSMGP, sumGP, and MOMoGP on all data sets (except for usFlight) five times with different random seeds. For usFlight, the above models are trained twice. Table 2 shows both average and standard deviation of the multiple runs. MOGP and MOSVGP have only been trained once for all data sets, thus, their results are not compared in Table 2.

References

Table 2: Mean and standard deviation of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Negative Log Predictive Density (NLPD) of state-of-the-art approaches and MOMoGP (our work) on benchmark data sets. Smaller values are better. Best result is indicated in **bold** and comparison of MOMoGP to DSMGP is indicated using arrows ↑ / ↓.