Stochastic Differential Equations with Variational Wishart Diffusions
Supplementary Material

Martin Jørgensen [1]   **Marc Peter Deisenroth** [2]   **Hugh Salimbeni** [3]

1. RMSE plot and further results

A plot of the test-set RMSE, relative to SGP, is provided in Figure 1. Further, we have included the actual results for reproducability; these can be seen in Table 1.

Dataset		bike	boston	concrete	kin8nm
Dimension		14	13	8	8
Test-set log-likelihood	No noise	-0.4169 ± 0.0235	-1442.5453 ± 512.4655	-25.4477 ± 12.9521	1.1458 ± 0.0696
	diffGP	-0.6842 ± 0.1143	-2.7586 ± 0.7301	-3.0401 ± 0.1460	1.3046 ± 0.0302
	diffWGP (rho = 5)	-0.6569 ± 0.0381	-3.1014 ± 0.8984	-3.0164 ± 0.1352	1.3123 ± 0.0305
	diffWGP (rho = 10)	-0.7171 ± 0.1621	-2.9540 ± 0.8506	-3.0141 ± 0.1559	1.3021 ± 0.0304
Test-set RMSE	No noise	0.0706 ± 0.0305	4.7823 ± 1.0100	7.5674 ± 1.3031	0.0688 ± 0.0026
	diffGP	0.1845 ± 0.0507	2.9628 ± 0.6991	4.9913 ± 0.6657	0.0646 ± 0.0021
	diffWGP (rho = 5)	0.1737 ± 0.0254	3.1713 ± 0.8296	4.9086 ± 0.6442	0.0642 ± 0.0023
	diffWGP (rho = 10)	0.1935 ± 0.0919	3.0800 ± 0.8009	4.8708 ± 0.7079	0.0649 ± 0.0024
Dataset		naval	power	protein	wine_white
Dimension		26	4	9	11
Test-set log-likelihood	No noise	NA	-2.8269 ± 0.0618	-2.7678 ± 0.0187	-7.2889 ± 0.9605
	diffGP	7.2256 ± 0.0949	-2.7753 ± 0.0402	-2.8599 ± 0.0109	-1.0616 ± 0.0448
	diffWGP (rho = 5)	8.2736 ± 0.0677	-2.7736 ± 0.0413	-2.8342 ± 0.0114	-1.0603 ± 0.0443
	diffWGP (rho = 10)	8.2696 ± 0.0575	NA	-2.8480 ± 0.0103	-1.0620 ± 0.0448
Test-set RMSE	No noise	NA	3.8966 ± 0.1674	3.8029 ± 0.0552	0.9298 ± 0.0543
	diffGP	$< 0.0001 \pm < 0.0001$	3.8497 ± 0.1496	4.1717 ± 0.0479	0.7001 ± 0.0295
	diffWGP (rho = 5)	$< 0.0001 \pm < 0.0001$	3.8492 ± 0.1502	4.0677 ± 0.0484	0.6980 ± 0.0296
	diffWGP (rho = 10)	$< 0.0001 \pm < 0.0001$	NA	4.1267 ± 0.0441	0.7003 ± 0.0296

Table 1. Test set log-likelihood and RMSE on 8 UCI benchmark datasets. Mean and standard deviations over 20 splits.
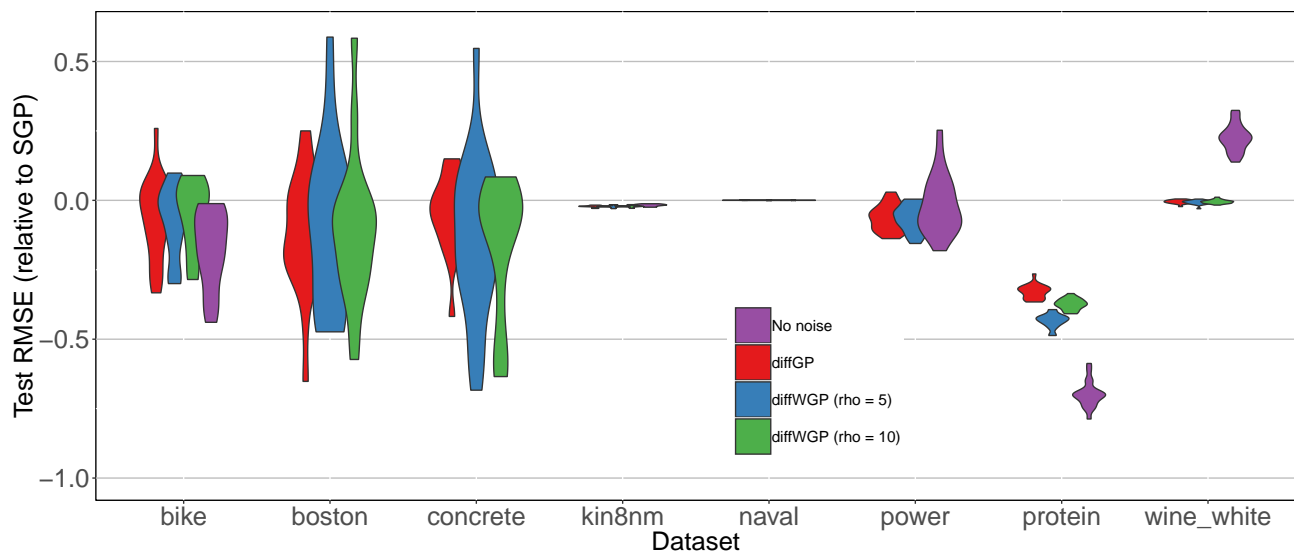
Figure 1. Test-set RMSE values on eight UCI regression datasets. The violin plots show the test-set RMSE difference of baseline diffusion models with respect to the SGP baseline. Values *less* than 0 indicate an improvement over SGP. We remark that this Figure has been cut for readability—this explains why occasionally purple violins are missing.