Supplement to "Analysis Methods for Computer Experiments: How to Assess and What Counts?"

Hao Chen, Jason L. Loeppky, Jerome Sacks and William J. Welch

University of British Columbia and National Institute of Statistical Sciences

1. TEST FUNCTIONS (FAST CODES)

1.1 Borehole function

The output y is generated by

$$y = \frac{2\pi T_u (H_u - H_l)}{\log(r/r_w)\left(1 + \frac{2LT_u}{\log(r/r_w)r_w^2 K_w} + T_u/T_l\right)},$$

where the 8 input variables and their respective ranges of interest are as in Table 1.

Variable	Description (units)	Range
r_w	radius of borehole (m)	$[0.05, 0.15]$
r	radius of influence (m)	$[100, 5000]$
T_u	transmissivity of upper aquifer (m^2/yr)	$[63070, 115600]$
H_u	potentiometric head of upper aquifer (m)	$[990, 1110]$
T_l	transmissivity of lower aquifer (m^2 / yr)	$[63.1, 116]$
H_l	potentiometric head of lower aquifer (m)	$[700, 820]$
L	length of borehole (m)	$[1120, 1680]$
K_w	hydraulic conductivity of borehole (m/yr)	$[9855, 12045]$

TABLE 1
Borehole function input variables, units, and ranges. All ranges are converted to $[0, 1]$ for statistical modeling.

Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: hao.chen@stat.ubc.ca; will@stat.ubc.ca). Statistics, University of British Columbia, Kelowna, BC V1V 1V7, Canada (e-mail: jason.loeppky@ubc.ca). National Institute of Statistical Sciences, Research Triangle Park, NC 27709, USA (e-mail: sacks@niss.org).

Variable	Description	Range
u_1	rate constant	$[2 \times 10^6, 2 \times 10^7]$
u_2	rate constant	5×10^{-3} (fixed)
u_3	rate constant	1×10^{-3} (fixed)
u_4	rate constant	2×10^{-3} (fixed)
u_5	rate constant	8 (fixed)
u_6	rate constant	$[3 \times 10^{-5}, 3 \times 10^{-4}]$
u_7	rate constant	$[0.3, 3]$
u_8	rate constant	1 (fixed)
x	initial concentration	$[1.0 \times 10^{-9}, 1.0 \times 10^{-6}]$

TABLE 2

G-protein code input variables and ranges. All variables are transformed to log scales on $[0,1]$ for statistical modeling.

1.2 G-protein code

The differential equations generating the output y of the G-protein system dynamics are

$$
\begin{aligned}
\dot{\eta}_1 &= -u_1\eta_1 x + u_2\eta_2 - u_3\eta_1 + u_5, \\
\dot{\eta}_2 &= u_1\eta_1 x - u_2\eta_2 - u_4\eta_2, \\
\dot{\eta}_3 &= -u_6\eta_2\eta_3 + u_8(G_{\text{tot}} - \eta_3 - \eta_4)(G_{\text{tot}} - \eta_3), \\
\dot{\eta}_4 &= u_6\eta_2\eta_3 - u_7\eta_4, \\
y &= (G_{\text{tot}} - \eta_3)/G_{\text{tot}},
\end{aligned}
$$

where $\eta_1, \ldots, \eta_4$ are concentrations of 4 chemical species, $\dot{\eta}_1 \equiv \frac{\partial \eta_1}{\partial t}$, etc., and $G_{\text{tot}} = 10000$ is the (fixed) total concentration of G-protein complex after 30 seconds.

The input variables in this system are described in Table 2. Only $d = 4$ inputs are varied: we model y as a function of $\log(x)$, $\log(u_1)$, $\log(u_6)$, $\log(u_7)$.

1.3 PTW code

The Preston-Tonks-Wallace (PTW) model describes the plastic deformation of various metals over a range of strain rates and temperatures (Preston, Tonks, and Wallace, 2003). For our purposes the model contains $d = 11$ input variables (parameters), where three of these are physically based (temperature, strain rate, strain), and the remaining 8 inputs can be used to tune the model to match data from physical experiments. Additional information on the model and the calibration problem can be found in Fugate et al. (2005). The inputs are all scaled to the unit interval $[0, 1]$.

2. DATA

2.1 Volcano code

The data for the 32 runs of the volcano code are in Table 3.

2.2 Arctic sea ice code

The data are available in the supplementary files `ice-x.txt` and `ice-y.txt`.

3. FURTHER RESULTS FOR RMSE

Here we give further results for normalized holdout RMSE of prediction, $e_{\text{rmse,ho}}$.

Run	Volume	BasalAngle	Height	logHeight	sqrtHeight
1	9.64	16.48	43.72	1.65	6.61
2	10.20	15.08	151.16	2.18	12.29
3	9.83	13.95	68.12	1.84	8.25
4	10.02	17.61	105.60	2.03	10.28
5	10.58	16.20	341.39	2.53	18.48
6	10.77	15.36	540.20	2.73	23.24
7	10.39	14.23	195.45	2.29	13.98
8	10.95	17.33	675.34	2.83	25.99
9	9.55	12.83	38.58	1.60	6.21
10	10.11	18.73	123.66	2.10	11.12
11	9.73	19.86	58.17	1.77	7.63
12	9.92	11.70	81.92	1.92	9.05
13	10.48	13.11	278.09	2.45	16.68
14	10.67	18.45	476.99	2.68	21.84
15	10.30	19.58	168.18	2.23	12.97
16	10.86	11.98	606.20	2.78	24.62
17	9.36	14.52	21.93	1.36	4.68
18	8.80	15.92	8.64	0.98	2.94
19	9.17	17.05	17.63	1.27	4.20
20	8.98	13.39	13.93	1.17	3.73
21	8.42	14.80	0.45	0.16	0.67
22	8.23	15.64	0.00	0.00	0.00
23	8.61	16.77	2.39	0.53	1.55
24	8.05	13.67	0.06	0.02	0.24
25	9.45	18.17	27.26	1.45	5.22
26	8.89	12.27	14.68	1.20	3.83
27	9.27	11.14	30.43	1.50	5.52
28	9.08	19.30	12.95	1.14	3.60
29	8.52	17.89	0.08	0.03	0.28
30	8.33	12.55	1.83	0.45	1.35
31	8.70	11.42	9.83	1.03	3.14
32	8.14	19.02	0.00	0.00	0.00

TABLE 3

Volcano code: inputs (Volume and BasalAngle) and output (Height, logHeight, or sqrtHeight)
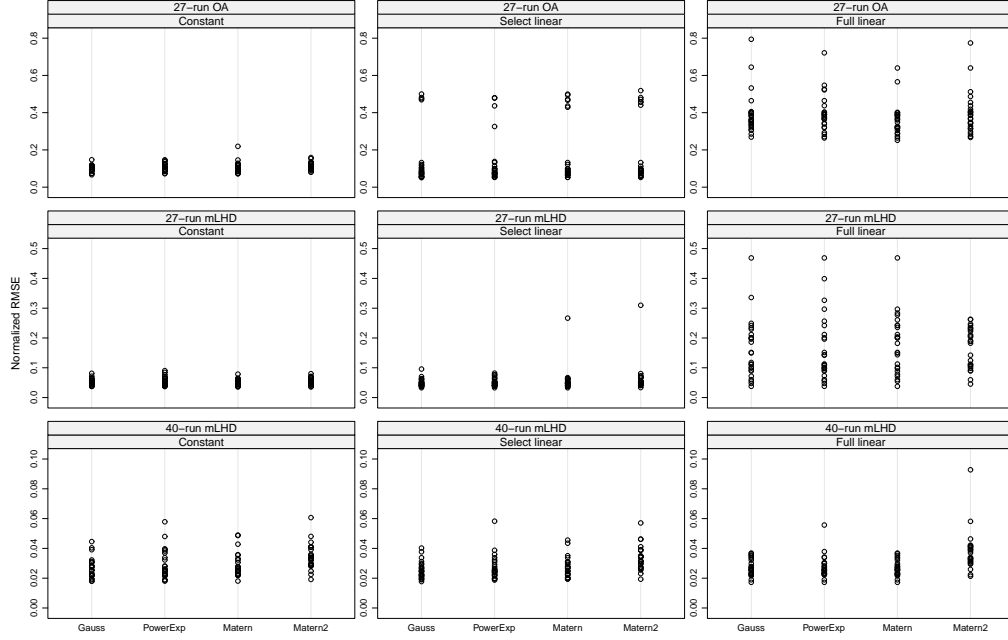
.

Fig 1. *Borehole function: Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, for GaSP with all combinations of three regression models and four correlation functions. Every model has a fitted nugget term. There are three base designs: a 27-run OA (top row); a 27-run mLHD (middle row); and a 40-run mLHD (bottom row). For each base design, 25 random permutations between columns give the 25 values of $e_{\mathrm{rmse,ho}}$ in a dot plot.*

3.1 Borehole function

Figure 1 shows normalized RMSE results for models with a fitted nugget. Comparison with Figure 2 in the main article for no-nugget models shows little difference, except that fitting a nugget term gives a small increase in the frequency of relatively poor results for models with a full-linear regression.

3.2 PTW code

Figure 2 gives results for $e_{\mathrm{rmse,ho}}$ for three regression models and four correlation functions. It is produced using the methods in the main paper's Section 3. Figure 3 gives $e_{\mathrm{rmse,ho}}$ results for the Bayes-GEM-SA and CGP methods in the main paper's Sections 5.1 and 5.2. Neither of these figures show practical differences from the various modeling strategies.

3.3 Nilson-Kuusk Model

Table 4 gives results for 150 training runs and 100 runs for the hold-out test set. The roles of the two data sets are switched relative to the main paper's Table 1. Again, Gauss is inferior, and PowerExp and Matérn are the best performers.

3.4 Arctic sea ice

There are 13 input variables and 157 runs of the code available. Repeat experiments were generated by sampling $n = 10d = 130$ runs, leaving 27 holdout observations.

Figure 4 gives results for $e_{\mathrm{rmse,ho}}$ for three regression models and four correlation functions. The four correlation functions give similar results, but the
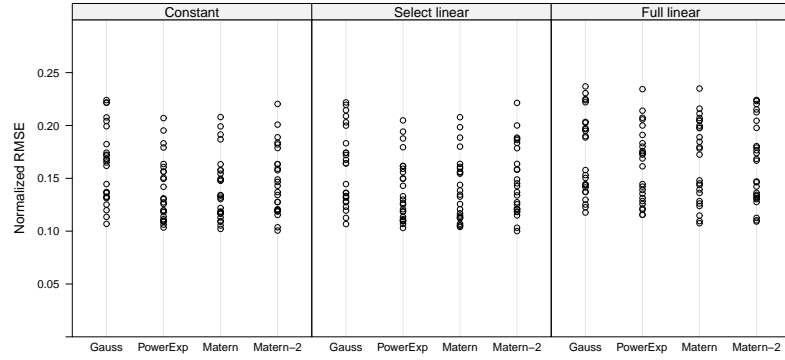
FIG 2. *PTW code: Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, for GaSP with all combinations of three regression models and four correlation functions. The base design is a 110-run mLHD; 25 random permutations between columns give the 25 values of $e_{\mathrm{rmse,ho}}$ in a dot plot.*
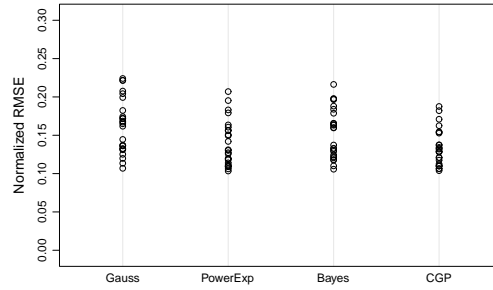

FIG 3. *PTW code: Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. The latter two methods also have constant regression models. The base design is a 110-run mLHD; 25 random permutations between columns give the 25 values of $e_{\mathrm{rmse,ho}}$ in a dot plot.*

	$e_{\mathrm{rmse,ho}}$			
Regression	Correlation function			
model	Gauss	PowerExp	Matérn-2	Matérn
Constant	0.111	0.080	0.087	0.078
Select linear	0.113	0.080	0.091	0.079
Full linear	0.109	0.079	0.090	0.076
Quartic	0.104	0.079	0.088	0.078

TABLE 4

Nilson-Kuusk model: Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, for four regression models and four correlation functions. The experimental data are from a 150-run LHD, and the hold-out set is from a 100-run LHD.
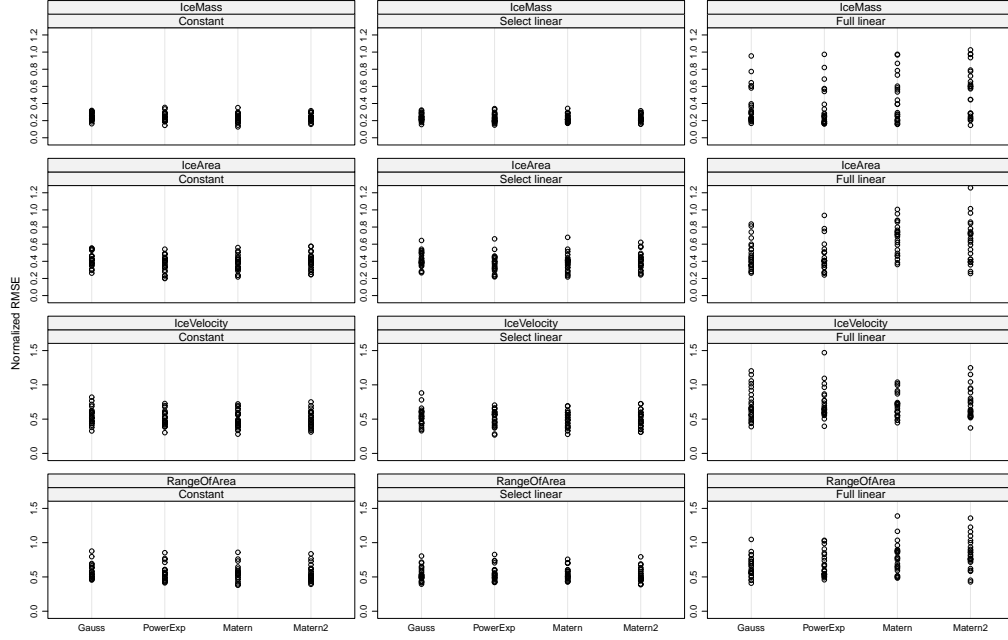
FIG 4. *Arctic sea ice code: Normalized holdout RMSE of prediction, $e_{\text{rmse,ho}}$, for GaSP with all combinations of three regression models and four correlation functions. The 157 runs available are randomly split 25 times into 130 runs for fitting and 27 hold-out runs to give the 25 values of $e_{\text{rmse,ho}}$ in a dot plot. Results are given for four output variables: ice mass, ice area, ice velocity, and range of area.*

full-linear regression model is inferior for all four output variables.

Figure 5 compares GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. The four strategies give similar results.

Figure 6 gives $e_{\text{rmse,ho}}$ results for models with estimated nugget terms. Comparison with the results for the same models without a fitted nugget in Figure 4 shows no practical differences.

3.5 2-d example

Figure 7 gives results for the function in (5.1) from three sets of repeat experiments. The designs leading to Figure 13 in the main paper have x_1, x_2, or both x_1 and x_2 reflected within columns about the centers of their ranges to generate equivalent designs and new training data. The resulting $e_{\text{rmse,ho}}$ values are plotted in Figure 7.

4. RESULTS FOR MAXIMUM ABSOLUTE ERROR

Here we give results for normalized holdout maximum absolute error of prediction, $e_{\text{max,ho}}$, defined as

$$(4.1) \qquad e_{\text{max,ho}} = \frac{\max\limits_{i=1,...,N} \left| \hat{y}(\mathbf{x}_{\text{ho}}^{(i)}) - y(\mathbf{x}_{\text{ho}}^{(i)}) \right|}{\max\limits_{i=1,...,N} \left| \bar{y} - y(\mathbf{x}_{\text{ho}}^{(i)}) \right|}.$$

Again, the measure of error is relative to the performance of the trivial predictor, $\bar{y}$.
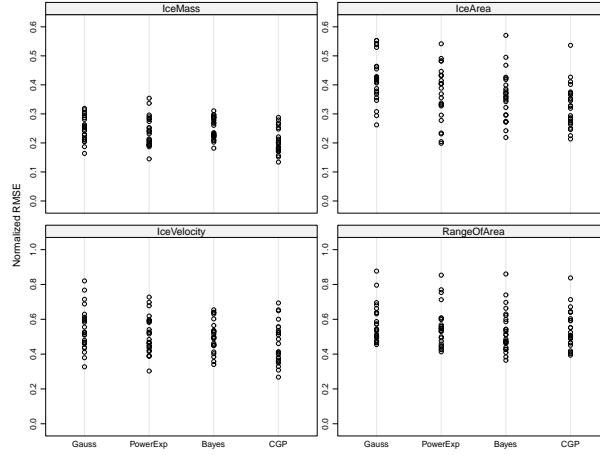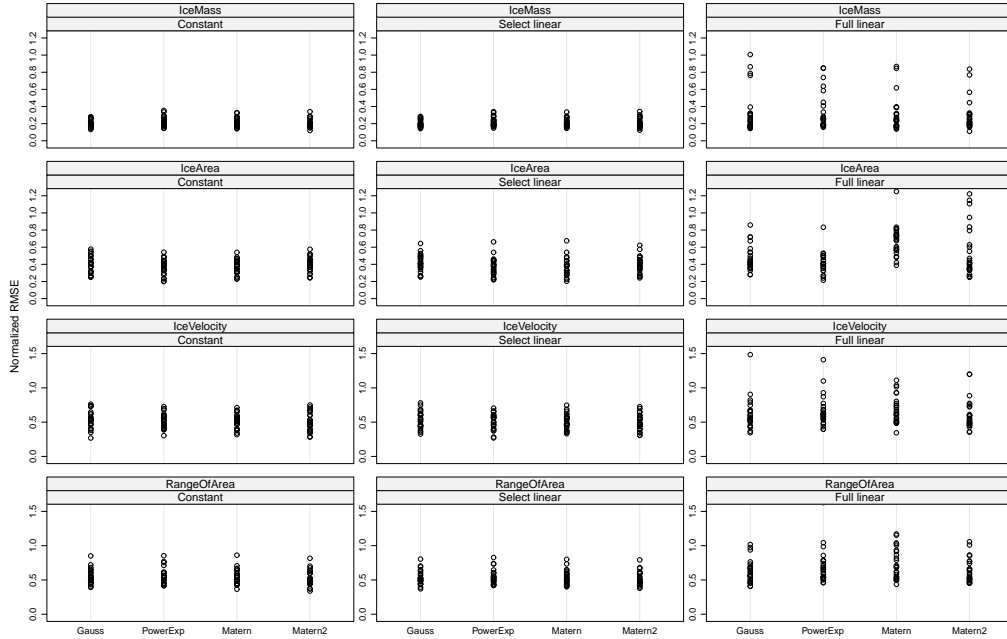
FIG 5. *Arctic sea ice code: Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. The latter two methods also have constant regression models. The 157 runs available are randomly split 25 times into 130 runs for fitting and 27 hold-out runs to give the 25 values of $e_{\mathrm{rmse,ho}}$ in a dot plot.*
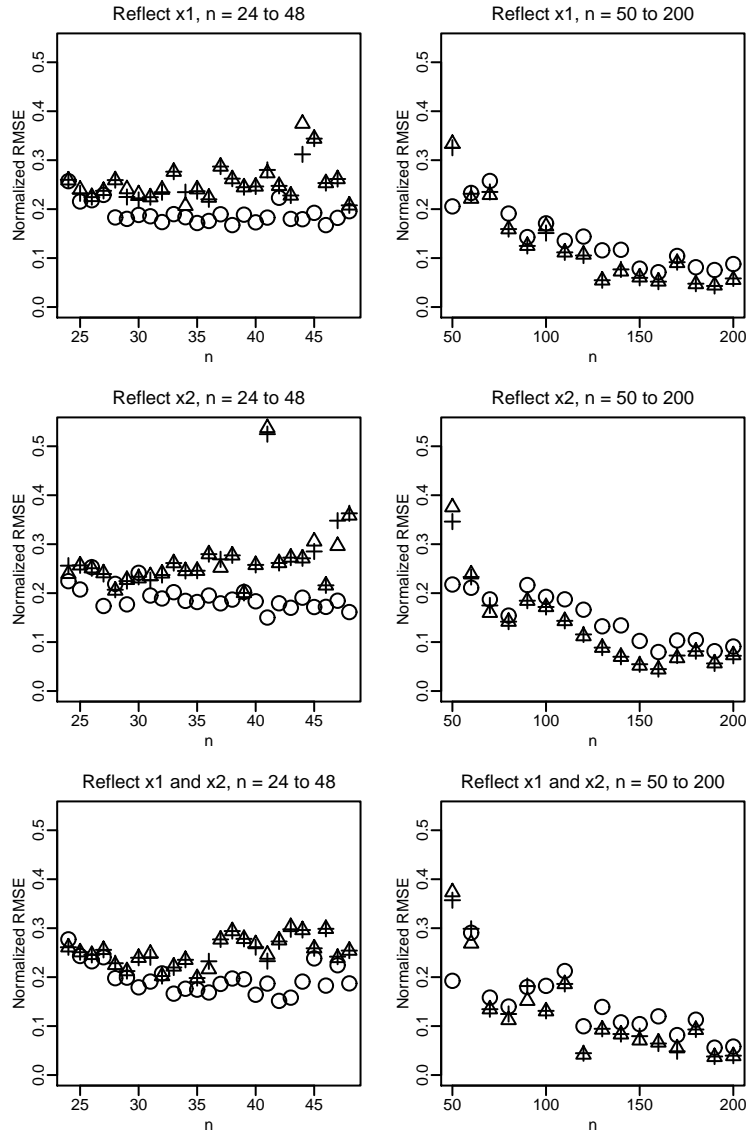

FIG 6. *Arctic sea ice code: Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, for GaSP with all combinations of three regression models and four correlation functions. Every model has a fitted nugget term. The 157 runs available are randomly split 25 times into 130 runs for fitting and 27 hold-out runs to give the 25 values of $e_{\mathrm{rmse,ho}}$ in a dot plot.*

FIG 7. *2-d function: Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, versus n for CGP ($\circ$), GaSP(Const, PowerExp) ($\triangle$) and GaSP(Const, PowerExp) with nugget ($+$). For each value of n the base mLHD has x_1 reflected around the center of its range (top row), x_2 reflected (middle row), or both x_1 and x_2 reflected (bottom row).*
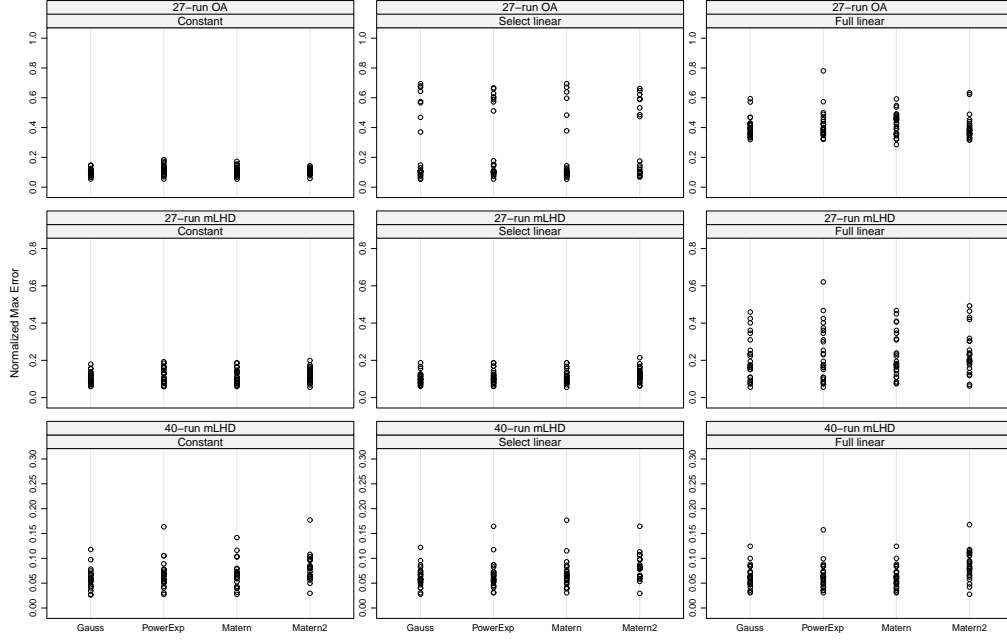
FIG 8. *Borehole function: Normalized holdout maximum absolute error of prediction, $e_{\max,ho}$, for GaSP with all combinations of three regression models and four correlation functions. There are three base designs: a 27-run OA (top row); a 27-run mLHD (middle row); and a 40-run mLHD (bottom row). For each base design, 25 random permutations between columns give the 25 values of $e_{\max,ho}$ in a dot plot.*

4.1 Borehole function

Figure 8 gives results for $e_{\max,ho}$ for three regression models and four correlation functions. Figure 9 compares GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. These two figures reporting $e_{\max,ho}$ results are analogous to the main paper's Figures 2 and 8 for $e_{rmse,ho}$; the same patterns emerge.

Figure 10 gives $e_{\max,ho}$ results for models with estimated nugget terms. Compared with Figure 8, inclusion of a nugget makes little difference except for a small increase in frequency of poor results with select-linear and full-linear regression models.

4.2 G-protein

Figure 11 gives results for $e_{\max,ho}$ for three regression models and four correlation functions. Figure 12 compares GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. These two figures reporting $e_{\max,ho}$ results are analogous to the main paper's Figures 4 and 9 for $e_{rmse,ho}$. The same conclusions emerge: there is little practical difference between the various modeling strategies.

4.3 PTW code

Figure 13 gives results for $e_{\max,ho}$ for three regression models and four correlation functions, and Figure 14 compares GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. These two figures reporting $e_{\max,ho}$ results are analogous to Figures 2 and 3 in Section 3.2 of this supplement relating to
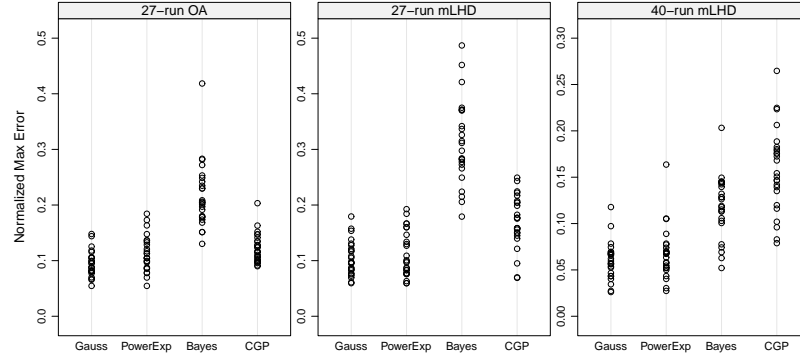
FIG 9. *Borehole function: Normalized holdout maximum absolute error of prediction, $e_{\max,\text{ho}}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. The latter two methods also have constant regression models.*
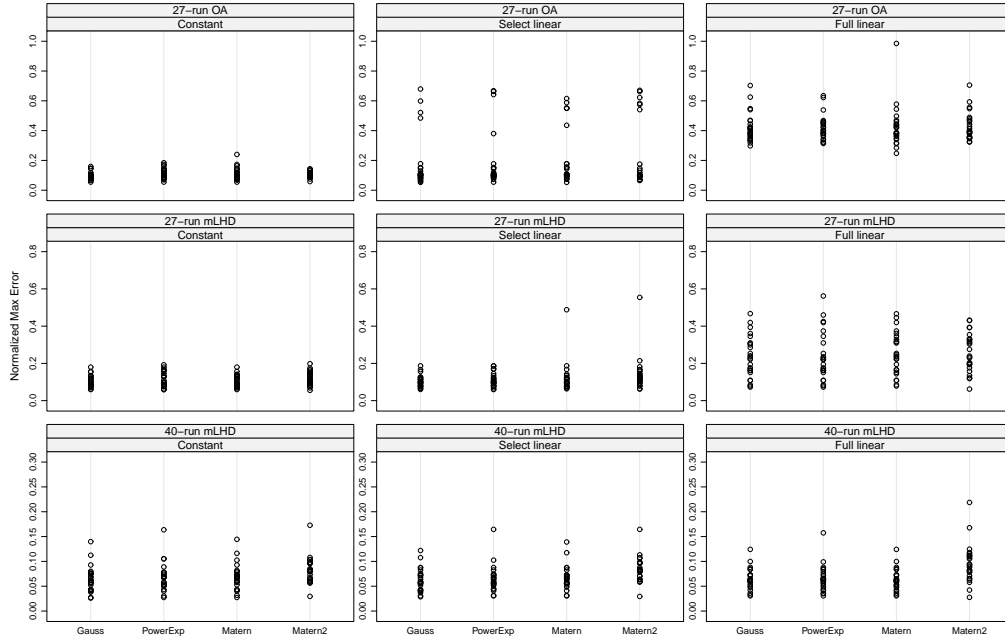

FIG 10. *Borehole function: Normalized holdout maximum absolute error of prediction, $e_{\max,\text{ho}}$, for GaSP with all combinations of three regression models and four correlation functions. Every model has a fitted nugget term. There are three base designs: a 27-run OA (top row); a 27-run mLHD (middle row); and a 40-run mLHD (bottom row). For each base design, 25 random permutations between columns give the 25 values of $e_{\max,\text{ho}}$ in a dot plot.*
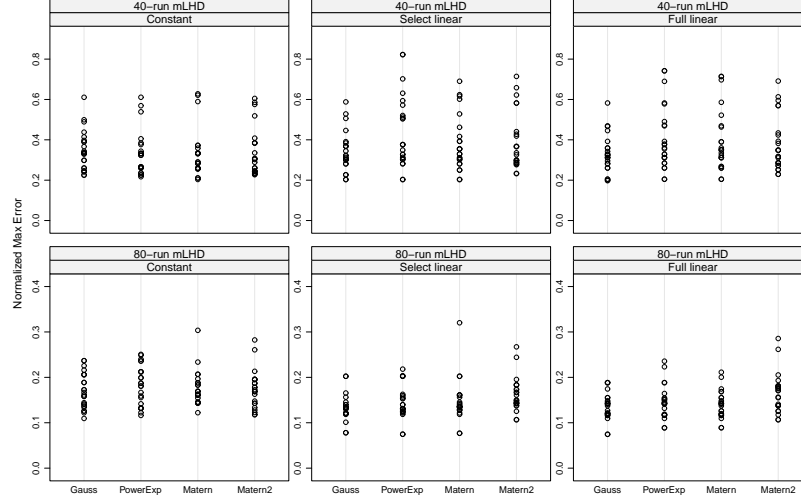
FIG 11. *G-protein code: Normalized holdout maximum absolute error of prediction, $e_{\max,\mathrm{ho}}$, for all combinations of three regression models and four correlation functions. There are two base designs: a 40-run mLHD (top row); and an 80-run mLHD (bottom row). For each base design, all 24 permutations between columns give the 24 values of $e_{\max,\mathrm{ho}}$ in each dot plot.*
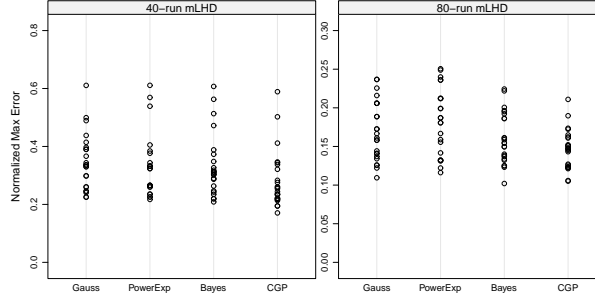

FIG 12. *G-protein code: Normalized holdout maximum error of prediction, $e_{\max,\mathrm{ho}}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. The latter two methods also have constant regression models.*
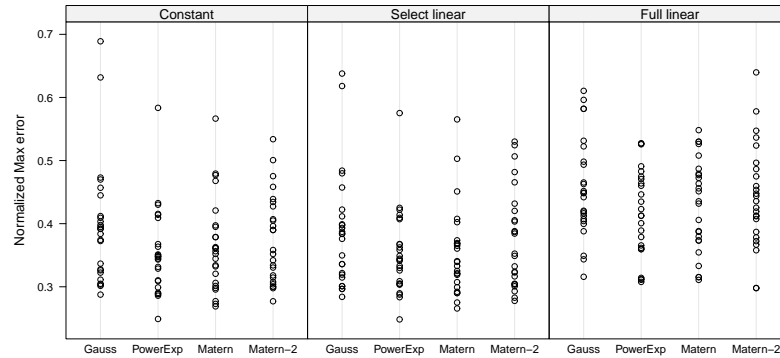

FIG 13. *PTW code: Normalized holdout maximum absolute error of prediction, $e_{\max,\mathrm{ho}}$, for GaSP with all combinations of three regression models and four correlation functions. The base design is a 110-run mLHD; 25 random permutations between columns give the 25 values of $e_{\max,\mathrm{ho}}$ in a dot plot.*
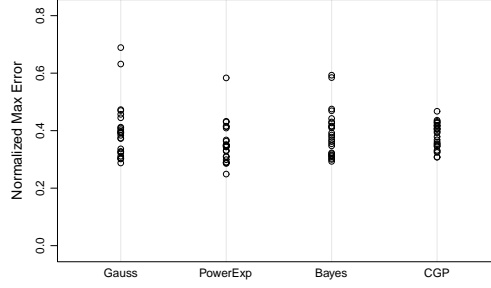
FIG 14. *PTW code: Normalized holdout maximum absolute error of prediction, $e_{\max,ho}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. The latter two methods also have constant regression models. The base design is a 110-run mLHD; 25 random permutations between columns give the 25 values of $e_{\max,ho}$ in a dot plot.*

| Regression model | $e_{\max,ho}$ Correlation function | | | |
	Gauss	PowerExp	Matérn-2	Matérn
Constant	0.35	0.27	0.28	0.28
Select linear	0.35	0.26	0.27	0.24
Full linear	0.30	0.27	0.28	0.28
Quartic	0.29	0.28	0.29	0.31

TABLE 5

Nilson-Kuusk model: Normalized holdout maximum absolute error of prediction, $e_{\max,ho}$, for four regression models and four correlation functions. The experimental data are from a 100-run LHD, and the hold-out set is from a 150-run LHD.

$e_{\mathrm{rmse,ho}}$. The $e_{\max,ho}$ criterion leads to similar conclusions: there are no practical differences in performance.

4.4 Nilson-Kuusk Model

The $e_{\max,ho}$ results in Tables 5 and 6 are for 100 training runs and 150 training runs, respectively. They are analogous to Table 1 of the main paper and supplementary Table 4, which relate to $e_{\mathrm{rmse,ho}}$. Again, Gauss is inferior. The differences in performance between the other correlation functions are small, but PowerExp is the best performer 7 out of 8 times.

Figure 15 gives results for $e_{\max,ho}$ for three regression models and four correlation functions, and Figure 16 compares GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. These two figures reporting $e_{\max,ho}$ re-

| Regression model | $e_{\max,ho}$ Correlation function | | | |
	Gauss	PowerExp	Matérn-2	Matérn
Constant	0.21	0.16	0.18	0.17
Select linear	0.25	0.16	0.19	0.18
Full linear	0.23	0.16	0.21	0.17
Quartic	0.23	0.16	0.21	0.17

TABLE 6

Nilson-Kuusk model: Normalized holdout maximum absolute error of prediction, $e_{\max,ho}$, for four regression models and four correlation functions. The experimental data are from a 150-run LHD, and the hold-out set is from a 100-run LHD.
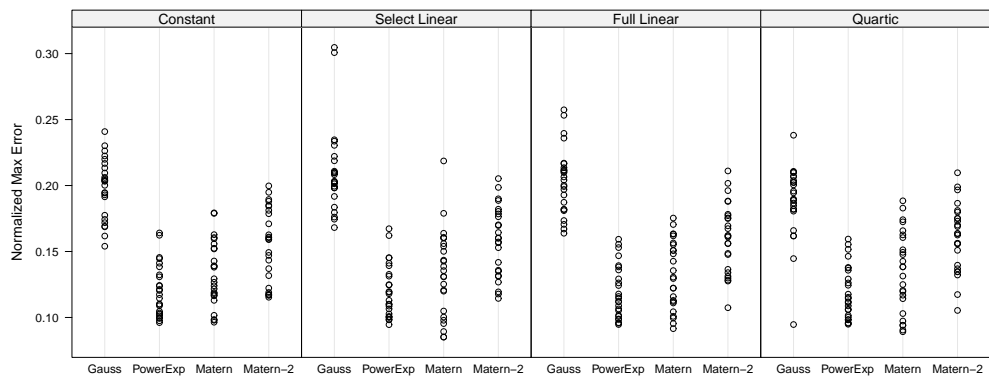
FIG 15. *Nilson-Kuusk code: Normalized holdout maximum absolute error of prediction, $e_{\max,\mathrm{ho}}$, for four regression models and four correlation functions. Twenty-five designs are created from a 150-run LHD base plus 50 random points from a 100-run LHD. The remaining 50 points in the 100-run LHD form the holdout set for each repeat.*
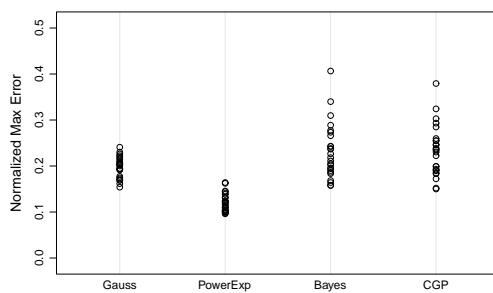

FIG 16. *Nilson-Kuusk model: Normalized holdout maximum absolute error of prediction, $e_{\max,\mathrm{ho}}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP.*
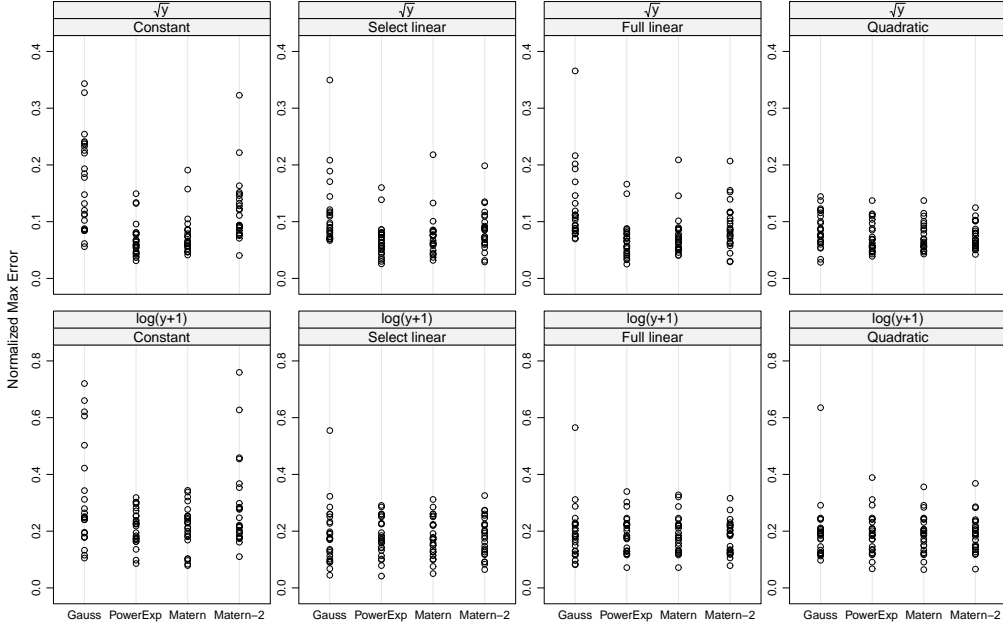
Fig 17. *Volcano model: Normalized holdout maximum absolute error, $e_{\mathrm{max,ho}}$, for three regression models and two correlation functions.*

sults are analogous to the main paper's Figures 5 and 10, which relate to $e_{\mathrm{rmse,ho}}$. Figures 15 and 16 lead to the same conclusions: R=PowerExp performs best, R=Gauss is inferior, there is no advantage from any of the non-constant regression functions, and neither Bayes-GEM-SA nor CGP perform well here.

4.5 Volcano code

Figure 17 gives results for $e_{\mathrm{max,ho}}$ for three regression models and four correlation functions, and Figure 18 compares GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. These two figures reporting $e_{\mathrm{max,ho}}$ results are analogous to the main paper's Figures 7 and 11, which relate to $e_{\mathrm{rmse,ho}}$. Figure 17 shows that the regression model makes little difference if combined with PowerExp, which always performs well here. For the $\sqrt{y}$ response, the quadratic model makes the performances of the other correlation functions similar to that
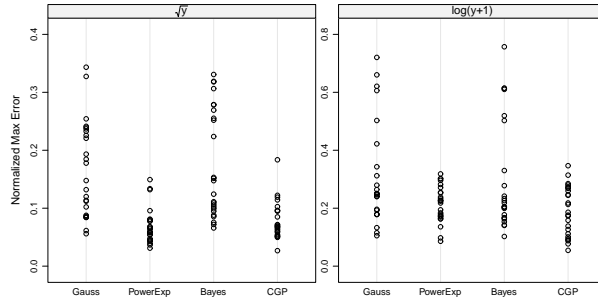

Fig 18. *Volcano model: Normalized holdout maximum absolute error of prediction, $e_{\mathrm{max,ho}}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP.*
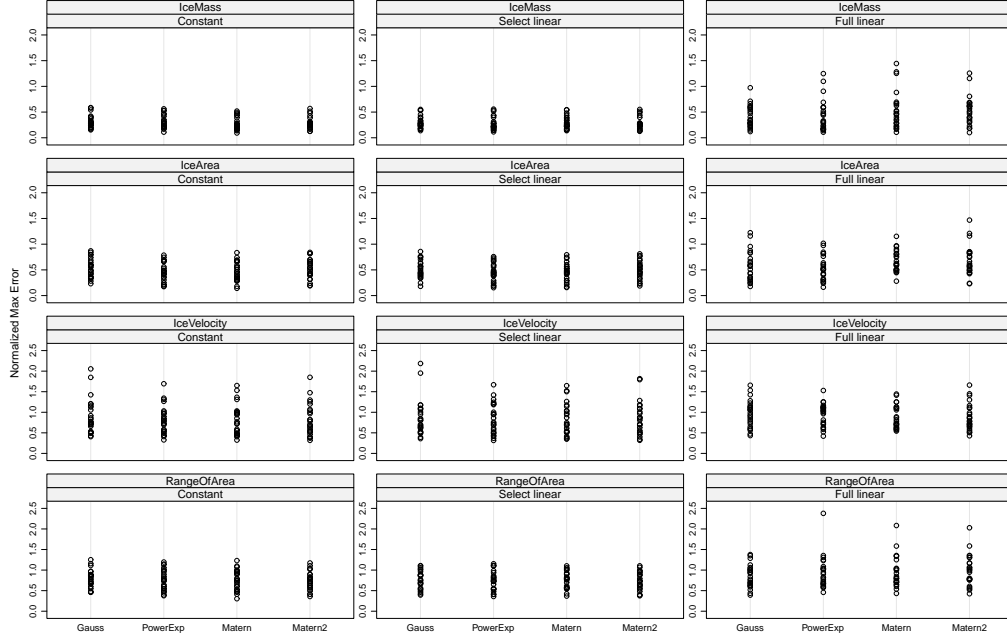
FIG 19. *Arctic sea ice code: Normalized holdout maximum error of prediction, $e_{\mathrm{max,ho}}$, for GaSP with all combinations of three regression models and four correlation functions. The 157 runs available are randomly split 25 times into 130 runs for fitting and 27 hold-out runs to give the 25 values of $e_{\mathrm{max,ho}}$ in a dot plot.*

of PowerExp, but there is no advantage over GaSP(Const, PowerExp). Figure 18 again shows for this application that R=Gauss and Bayes-GEM-SA are inferior, while PowerExp and CGP perform about the same.

4.6 Arctic sea ice

Figure 19 gives results for $e_{\mathrm{max,ho}}$ for three regression models and four correlation functions, and Figure 20 compares GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. These two figures reporting $e_{\mathrm{max,ho}}$ results are analogous to supplementary Figures 4 and 5 in Section 3.4, which relate to $e_{\mathrm{rmse,ho}}$. Figure 19 shows there is little practical advantage to be gained from a non-constant regression model. The small improvement seen for the response IceVelocity is in the context of $e_{\mathrm{max,ho}}$ often greater than 1 for all models, i.e., no predictive ability. Figure 20 shows that GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP give similar results.

Figure 21 gives $e_{\mathrm{max,ho}}$ results for models with estimated nugget terms. Comparison with the results for the same models without nugget terms in Figure 19 shows no practical improvement.

4.7 Friedman function

Figure 22 compares prediction accuracy of models with and without a nugget via the $e_{\mathrm{max,ho}}$ criterion. As in Figure 14 in the main article, no advantage from fitting a nugget is apparent.
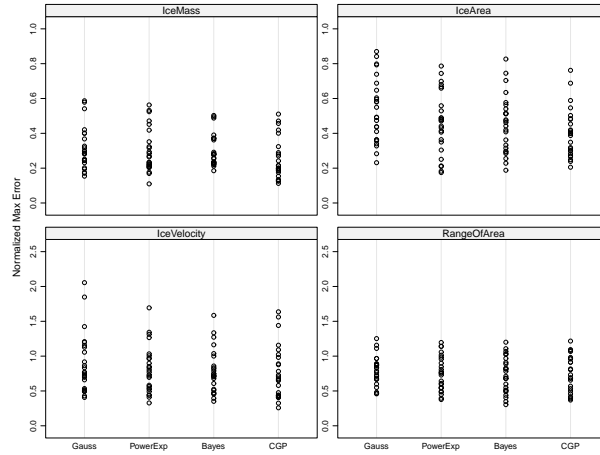
FIG 20. *Arctic sea ice code: Normalized holdout maximum error of prediction, $e_{\max,\mathrm{ho}}$, for GaSP(Const, Gauss), GaSP(Const, PowerExp), Bayes-GEM-SA, and CGP. The latter two methods also have constant regression models. The 157 runs available are randomly split 25 times into 130 runs for fitting and 27 hold-out runs to give the 25 values of $e_{\max,\mathrm{ho}}$ in a dot plot.*

FIG 21. *Arctic sea ice code: Normalized holdout maximum error of prediction, $e_{\max,\mathrm{ho}}$, for GaSP with all combinations of three regression models and four correlation functions. Every model has a fitted nugget term. The 157 runs available are randomly split 25 times into 130 runs for fitting and 27 hold-out runs to give the 25 values of $e_{\max,\mathrm{ho}}$ in a dot plot.*
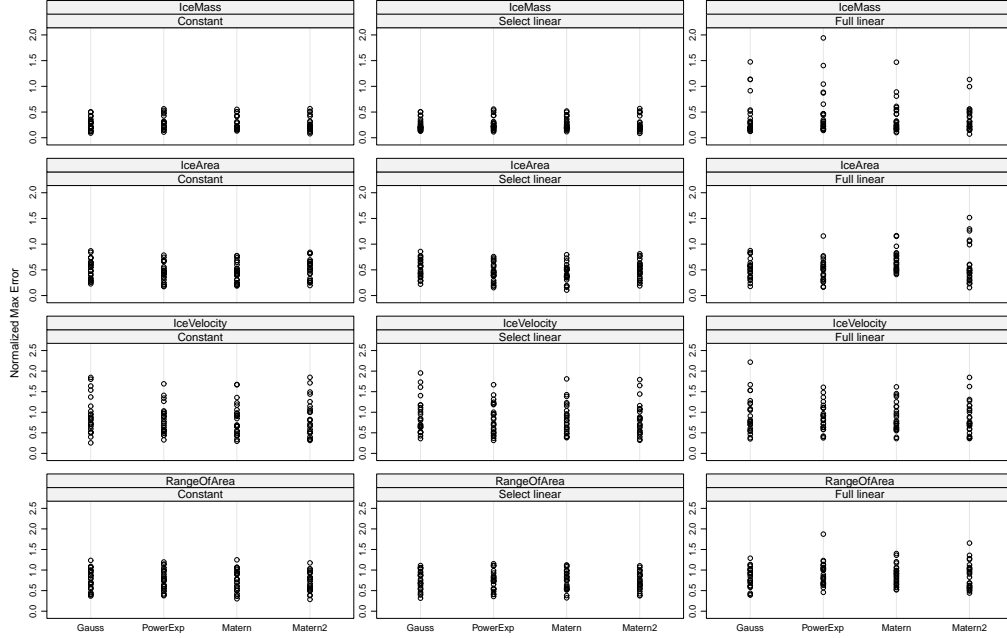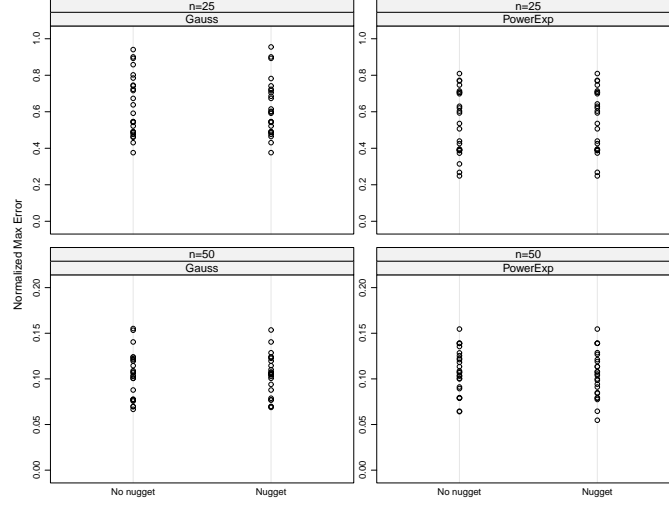
FIG 22. *Friedman function: Normalized holdout maximum error of prediction, $e_{\max,\text{ho}}$, for GaSP(Const, Gauss) and GaSP(Const, PowerExp) models with no nugget term versus the same models with a nugget. There are two base designs: a 25-run mLHD (top row); and a 50-run mLHD (bottom row). For each base design, 25 random permutations between columns give the 25 values of $e_{\max,\text{ho}}$ in a dot plot.*

5. UNCERTAINTY OF PREDICTION

The results in Figure 23 compare the distribution of ACP values with and without a nugget term for the Friedman function in equation (5.2) of the main article. With $n = 25$, substantial under-coverage occurs frequently; fitting a nugget term makes little difference. For $n = 50$ there is a modest improvement in the distribution towards the nominal 95% coverage when a nugget is fitted. Substantial under-coverage is still frequent, however.

6. REGRESSION TERMS

How the inclusion of regression terms performs in extreme settings can be seen in the following simulations. Functions were generated with 5-dimensional input **x** from a GaSP with regression component $\mu = 10(x_1 + x_2 + x_3 + x_4 + x_5)$, variance 1, and one of four GPs with Gaussian correlation functions differing in their true parameter values. The vector of correlation parameter values was either: (1) $\boldsymbol{\theta}_A = (6.7232, 2.4992, 0.6752, 0.0992, 0.0032)$, which sum to 10; (2) $\boldsymbol{\theta}_B = (3.3616, 1.2496, 0.3376, 0.0496, 0.0016)$, half the values of $\boldsymbol{\theta}_A$; (3) $\boldsymbol{\theta}$ has constant elements with value 2; and (4) $\boldsymbol{\theta}$ has constant elements with value 1.

With sample size $n = 50$ from a mLHD and 500 random test points, 25 sample functions from the 4 GPs were taken and normalized RMSEs calculated. The average $e_{\text{rmse,ho}}$ for the four simulation scenarios for Const versus FL fits are in Table 7.

What can be drawn from these numbers is that the strong trend composed with less correlated (rougher) functions gives fitting the linear regression little practical advantage. There can be a *relative* advantage for FL when $e_{\text{rmse,ho}}$ is very small anyway for both methods.

The trend here is very extreme: a total trend of magnitude 50 across the input
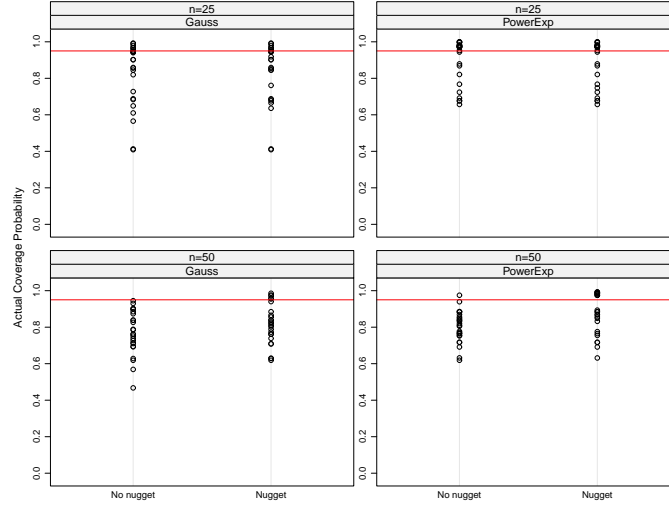
FIG 23. *Friedman function: ACP of nominal 95% confidence or credibility intervals for GaSP(Const, Gauss) and GaSP(Const, PowerExp) models, each without or with a nugget term. There are two base designs: a 25-run mLHD (top row); and a 50-run mLHD (bottom row). For each base design, 25 random permutations between columns give the 25 values of ACP in a dot plot.*

Correlation	Average $e_{\mathrm{rmse,ho}}$ when fitting	
parameters	GaSP(Const, PowerExp)	GaSP(FL, PowerExp)
$\boldsymbol{\theta}_A$	0.04	0.02
$\boldsymbol{\theta}_B$	0.02	0.01
$\boldsymbol{\theta = 2}$	0.08	0.06
$\boldsymbol{\theta = 1}$	0.04	0.02

TABLE 7

Normalized holdout RMSE of prediction, $e_{\mathrm{rmse,ho}}$, from fitting constant versus full-linear regression models. The data are generated using one of four vectors of true values of the correlation parameters.

space of all 5 variables, relative to a GP with variance 1. Other simulations with less extreme trend show even less difference between the results for the Const and FL models.

REFERENCES

Fugate, M., Williams, B., Higdon, D., Hanson, K. M., Gattiker, J., Chen, S.-R., and Unal, C. (2005), "Hierarchical Bayesian analysis and the Preston-Tonks-Wallace model," Tech. Rep. LA-UR-05-3935, Los Alamos National Laboratory, Los Alamos, NM.

Preston, D. L., Tonks, D. L., and Wallace, D. C. (2003), "Model of Plastic Deformation for Extreme Loading Conditions," *Journal of Applied Physics*, 93, 211–220.