Mini-Minimax Uncertainty Quantification for Emulators[*]

Jeffrey C. Regier[†] and Philip B. Stark[†]

Abstract. Consider approximating a "black box" function f by an *emulator* $\hat{f}$ based on n noiseless observations of f. Let w be a point in the domain of f. How big might the error $|\hat{f}(w) - f(w)|$ be? If f could be arbitrarily rough, this error could be arbitrarily large: we need some constraint on f besides the data. Suppose f is Lipschitz with a known constant. We find a lower bound on the number of observations required to ensure that for the best emulator $\hat{f}$ based on the n data, $|\hat{f}(w) - f(w)| \leq \epsilon$. But in general, we will not know whether f is Lipschitz, much less know its Lipschitz constant. Assume optimistically that f is Lipschitz-continuous with the smallest constant consistent with the n data. We find the maximum (over such regular f) of $|\hat{f}(w) - f(w)|$ for the best possible emulator $\hat{f}$; we call this the "mini-minimax uncertainty" at w. In reality, f might not be Lipschitz or—if it is—it might not attain its Lipschitz constant on the data. Hence, the mini-minimax uncertainty at w could be much smaller than $|\hat{f}(w) - f(w)|$. But if the mini-minimax uncertainty is large, then—even if f satisfies the optimistic regularity assumption—$|\hat{f}(w) - f(w)|$ could be large, no matter how cleverly we choose $\hat{f}$. For the Community Atmosphere Model, the maximum (over w) of the mini-minimax uncertainty based on a set of 1154 observations of f is no smaller than it would be for a single observation of f at the centroid of the 21-dimensional parameter space. We also find lower confidence bounds for quantiles of the mini-minimax uncertainty and its mean over the domain of f. For the Community Atmosphere Model, these lower confidence bounds are an appreciable fraction of the maximum. To know that the emulator estimates f accurately would require evidence that f is typically more regular than it is across the n sample values.

Key words. emulator, surrogate function, metamodel, minimax, Lipschitz, information-based complexity

AMS subject classifications. 68Q17, 65D05, 68U20, 62P12

DOI. 10.1137/130917909

1. Introduction. This paper studies the accuracy of emulators, also known as surrogate functions and metamodels. Emulators are important tools for approximating functions that have been observed only partially. Kriging, multivariate adaptive regression splines (MARS), projection pursuit regression, polynomial chaos (PC) expansions, Gaussian process (GP) models, and other Bayesian modeling techniques are common methods for constructing emulators [1, 2, 3]. We find error bounds for emulators in general—including the "best possible" method—rather than focusing on any particular emulation method.[1]

Emulators are frequently used to approximate expensive computer models, which are

[1]Software that computes the bounds described in this paper is freely available at https://github.com/jeff-regier/MiniMiniMaxUQ.

often deterministic functions.[2] Resources limit the number of times the computer model can be run, though typically an intractable number of inputs is possible—for instance if any input parameter is a floating point number. By fitting an emulator to the output of a tractable number of runs for different inputs, one can approximate the computer model inexpensively; the issue is the accuracy of that approximation.

Computer models known as *HEB* [4] may be particularly difficult to emulate: They depend on *H*igh-dimensional inputs; they are *E*xpensive to run; and they are effectively *B*lack boxes that are not amenable to closed-form, analytic study. Because such models have high-dimensional inputs, it takes prohibitively many runs to explore their domains: to attain a given sample density, the number grows exponentially in the dimension. Because the models are expensive, performing many runs is impractical or impossible. And because the models are black boxes, there are few (if any) constraints to ensure that the error in extrapolating from inputs actually tried to inputs not sampled is small.

HEB problems arise often in practice, for instance in the following:

- Climate models: [5] (21–28-dimensional domains; 1154 simulations; Kriging and MARS).
- Automobile crashes: [6] (15-dimensional domain; 55 simulations; polynomial response surfaces and artificial neural networks).
- Chemical reactions: [7] (30–50-dimensional domains; boosted surrogate models) and [8] (46-dimensional domain; seconds per simulation).
- Aircraft design: [9] (25-dimensional domain; 500 simulations; response surfaces and Kriging), [10] (22-dimensional domain; minutes per simulation; response surfaces and Kriging), and [11] (31-dimensional domain; 20 minutes to several days per simulation; Kriging).
- Electric circuits: [12] (60-dimensional domain; 216 simulations; Kriging).

How accurately can a function f be emulated from a given set of data? How many evaluations of f are required to guarantee that f can be emulated to a given level of accuracy?

Since f is a "black box," we do not know how rough it might be: extrapolating beyond the data could entail arbitrarily large errors. We assume that f is regular and find the resulting uncertainty in emulating f. If the regularity assumption fails, the uncertainty would be larger. We measure the regularity of f by its *absolute condition number* or *Lipschitz constant* K. Similar results could be derived for other measures of regularity, but Lipschitz bounds are particularly amenable to analysis.

The observations impose a lower bound $\hat{K}$ on K. Suppose, optimistically, that the true Lipschitz constant of f is equal to this lower bound. Then f might be any member of the set $\mathcal{F}_{\hat{K}}$ of functions that agree with the observations and have Lipschitz constant no greater than $\hat{K}$. If an emulator is guaranteed to do well no matter which member of $\mathcal{F}_{\hat{K}}$ f happens to be, then the uncertainty of that emulator is low. On the other hand, if there are elements of $\mathcal{F}_{\hat{K}}$

---

[2]They might not be entirely deterministic; for instance, they could involve Monte Carlo simulations. Moreover, in distributed parallel computations, numerical results can depend on the order in which subproblems happen to complete. These cases can be thought of as observing the function with noise. We do not address noise here; however, uncertainty in the observations makes accurate approximation more difficult. Because we focus on lower bounds on the difficulty of approximating the function accurately, our results generally remain lower bounds when the observations are not only incomplete but also noisy. To extend our methods to include noise would involve finding a lower confidence bound on the regularity of the function.

that an emulator cannot approximate well, the uncertainty is large.

Consider *all* emulators that can be computed from the observations alone, without additional knowledge of f; this collection includes emulators constructed using GP, PC, MARS, and all the other methods mentioned above. Viewed as a function of w in the domain of f, the minimax error among such emulation methods over the set $\mathcal{F}_{\hat{K}}$ of functions that agree with the observations and have Lipschitz constant no greater than $\hat{K}$ is the *mini-minimax uncertainty* $\mathcal{E}_{\hat{K}}(w)$ in the title of this paper.

The first "mini" refers to the regularity condition: since K is not smaller than $\hat{K}$, $\mathcal{E}_{\hat{K}}(w)$ is a lower bound on the minimax uncertainty for functions that are as regular as f. The second "mini" refers to emulators: this is the uncertainty for the best emulator—including all the standard ones. The "max" is over functions that agree with f at the observations and satisfy the optimistic regularity condition. That is, $\mathcal{E}_{\hat{K}}(w)$ is the smallest that the uncertainty at w could be, for the best emulator, over the set of functions that have the highest degree of regularity consistent with the observations and that agree with the observations. The maximum of $\mathcal{E}_{\hat{K}}(w)$ over w in the domain of f is an attainable lower bound on the maximum uncertainty of any emulator $\hat{f}$ of f.

If K were known, this would be a standard problem in information-based complexity [13, 14, 15]. We derive bounds on the uncertainty using the lower bound $\hat{K}$ computed from the observed variation of f. Section 3 derives a lower bound on the number of additional observations that might be necessary to learn f. Section 4 derives two lower bounds on the maximum uncertainty for approximating f from a fixed set of observations: a purely empirical bound and a bound expressed as a fraction of the unknown Lipschitz constant. The latter yields conditions under which emulating f by a constant function, equal to the value of f at the centroid of its domain, has smaller maximum uncertainty than any emulator based on the n actual observations.

Section 5 applies these bounds to two closed-form functions (a high-dimensional cone and the *borehole function* [16]) and to a black box function (the Community Atmosphere Model [5]). Section 6 extends the results for the maximum error to quantiles of the error and the mean of the error over the domain of f. Section 7 gives our conclusions.

2. Notation and problem formulation. We summarize key notation in Table 1. The function f is a fixed unknown real-valued function on $[0,1]^p$, the p-dimensional unit cube. The space of real-valued continuous functions on $[0,1]^p$ is $\mathcal{C}[0,1]^p$. The Roman letters i, j, p, q, and M_ϵ denote integers. Lowercase Greek letters denote real scalars, with the exception of μ, which denotes Lebesgue measure. Uppercase Roman letters such as X and D denote subsets of $[0,1]^p$; X is a fixed finite subset of $[0,1]^p$. Lowercase Roman letters from the end of the alphabet, such as v, w, x, y, and z, denote points in $[0,1]^p$. The lowercase Roman letters e, f, g, and h denote real-valued functions on (subsets of) $[0,1]^p$. The domain of a function g is $\mathrm{dom}(g)$. The restriction of a function g to $D \subset \mathrm{dom}(g)$ is denoted by $g|_D$. The observations from which f is to be emulated are $f|_X$; that is, we observe f on the set X. An emulator $\hat{f}$ is a real-valued function on $[0,1]^p$. Let $\|h\|_\infty \equiv \sup_{w \in \mathrm{dom}(h)} |h(w)|$, the infinity-norm of h. This paper studies how large $|\hat{f}(w) - f(w)|$ and $\|\hat{f} - f\|_\infty$ could be for the best $\hat{f}$ chosen on the basis of the data—without other information about f.

Table 1
Summary of key notation.

Symbol	Meaning
f	unknown function on $[0,1]^p$ to be emulated
$\hat{f}$	an emulator
X	finite subset of $[0,1]^p$ where f is observed
$g|_Y$	the restriction of the function g to the set $Y \subset [0,1]^p$
$f|_X$	*the data*: the restriction of f to X
K	Lipschitz constant of the function f
$\hat{K}$	smallest Lipschitz constant of any function that interpolates the data
$\mathcal{F}_{\kappa,Y}$	all functions that interpolate $f|_Y$ and have Lipschitz constant no larger than κ
$\mathcal{F}_\kappa$	$\mathcal{F}_{\kappa,X}$
$e_\kappa^+(w)$	maximum value at w among functions in $\mathcal{F}_\kappa$
$e_\kappa^-(w)$	minimum value at w among functions in $\mathcal{F}_\kappa$
$\hat{f}_\kappa(w)$	mean of $e_\kappa^+(w)$ and $e_\kappa^-(w)$; the minimax emulator at the point w over functions in $\mathcal{F}_\kappa$
$\mathcal{E}_{\kappa,Y}(w;\hat{f})$	*maximum uncertainty of $\hat{f}$ at w*: uncertainty of $\hat{f}$ at the point w over functions in $\mathcal{F}_{\kappa,Y}$
$\mathcal{E}_{\kappa,Y}(w)$	*minimax uncertainty at w*: uncertainty of the best possible emulator at the point w over functions in $\mathcal{F}_{\kappa,Y}$
$\mathcal{E}_{\kappa,Y}(\hat{f})$	*maximum uncertainty of $\hat{f}$*: maximum (over $w \in [0,1]^p$) uncertainty of $\hat{f}$ over functions in $\mathcal{F}_{\kappa,Y}$
$\mathcal{E}_{\kappa,Y}$	*minimax uncertainty*: maximum (over $w \in [0,1]^p$) uncertainty of the best possible emulator over functions in $\mathcal{F}_{\kappa,Y}$
$\mathcal{E}_\kappa(\cdots)$	when $Y = X$, we generally suppress X from the subscript, viz., $\mathcal{E}_\kappa(w;\hat{f})$, $\mathcal{E}_\kappa(w)$, $\mathcal{E}_\kappa(\hat{f})$, and $\mathcal{E}_\kappa$
M_ϵ	*minimum computational burden*: a lower bound on the number of additional observations needed to guarantee that the minimax uncertainty is no larger than ϵ

Let d be a metric on $\mathrm{dom}(g)$. The (best) Lipschitz constant for g is

$$(2.1) \qquad \mathrm{Lip}(g) \equiv \sup \left\{ \frac{g(v) - g(w)}{d(v,w)} : v, w \in \mathrm{dom}(g) \text{ and } v \neq w \right\}.$$

If $f \notin \mathcal{C}[0,1]^p$, then $\mathrm{Lip}(f) \equiv \infty$. Define

$$\mathcal{F}_\kappa(g) \equiv \{(h : [0,1]^p \to \Re) : \mathrm{Lip}(h) \leq \kappa \text{ and } h|_{\mathrm{dom}(g)} = g\}.$$

Then $\mathcal{F}_\infty(f|_X)$ is the space of (possibly discontinuous) functions that fit the n data. Some of our results involve values of f at points other than the points X at which f was observed; Y denotes a generic set of points in the domain of f. To simplify notation, we set

$$\mathcal{F}_{\kappa,Y} \equiv \mathcal{F}_\kappa(f|_Y).$$

When $Y = X$, we generally write $\mathcal{F}_\kappa$ in place of $\mathcal{F}_{\kappa,X}$.
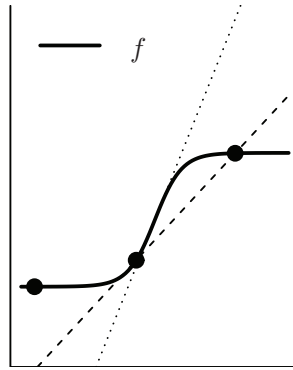
Figure 1. *Illustration of the difference between the true Lipschitz constant K and the empirical lower bound $\hat{K}$ for K. The dotted line is tangent to f where f attains its Lipschitz constant: it has slope K. The dashed line is the steepest line that intersects any pair of observations: it has slope $\hat{K} \leq K$.*

Definition. *The* uncertainty at w *of $\hat{f} : [0,1]^p \to \Re$ over the set of functions $\mathcal{F}_{\kappa,Y}$ is*

$$\mathcal{E}_{\kappa,Y}(w; \hat{f}) \equiv \sup_{g \in \mathcal{F}_{\kappa,Y}} |\hat{f}(w) - g(w)|.$$

The minimax uncertainty at w *over the set of functions $\mathcal{F}_{\kappa,Y}$ is*

$$\mathcal{E}_{\kappa,Y}(w) \equiv \inf_{\hat{f}:[0,1]^p \to \Re} \mathcal{E}_{\kappa,Y}(w; \hat{f}).$$

The maximum uncertainty *of $\hat{f} : [0,1]^p \to \Re$ over the set of functions $\mathcal{F}_{\kappa,Y}$ is*

$$\mathcal{E}_{\kappa,Y}(\hat{f}) \equiv \sup_{w \in [0,1]^p} \mathcal{E}_{\kappa,Y}(w; \hat{f}) = \sup_{g \in \mathcal{F}_{\kappa,Y}} \|\hat{f} - g\|_\infty.$$

The minimax maximum uncertainty *over the set of functions $\mathcal{F}_{\kappa,Y}$ is*

$$\mathcal{E}_{\kappa,Y} \equiv \inf_{\hat{f}:[0,1]^p \to \Re} \mathcal{E}_{\kappa,Y}(\hat{f}).$$

The emulator $\hat{f}$ approximates f within $\mathcal{E}_\infty(w; \hat{f})$ at the point w if f is in $\mathcal{F}_\infty$, the set of functions that agree with the observations. However, $\mathcal{E}_\infty(w; \hat{f})$ is infinite for every $\hat{f}$ unless $w \in X$, even if f is guaranteed to be continuous.[3] To guarantee that the uncertainty is finite requires stronger regularity than mere continuity.

Let $K \equiv \mathrm{Lip}(f)$ and $\hat{K} \equiv \mathrm{Lip}(f|_X)$. Because $X \subset [0,1]^p$, $\hat{K} \leq K$, as illustrated in Figure 1. (There and in subsequent figures, $p = 1$ and the bold black dots represent $f|_X$, the observations of f at $x \in X$.)

Define

$$e_\kappa^+(w) \equiv \min_{x \in X} [f(x) + \kappa d(x, w)]$$

---

[3]The set X is not dense in $[0,1]^p$, so for any $c > 0$, there exists some function $g \in \mathcal{F}_\infty(f|_X)$ satisfying $\|f - g\|_\infty > c$.
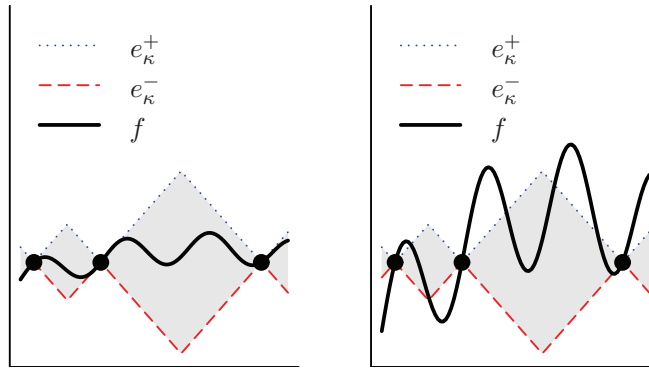
Figure 2. *Illustration of the upper and lower envelope functions e_κ^- and e_κ^+. In the left panel, $\kappa = K$; in the right, $\kappa < K$. If $\kappa \geq K$, then $e_\kappa^- \leq f \leq e_\kappa^+$, and, equivalently, $f \in \mathcal{F}_\kappa$.*
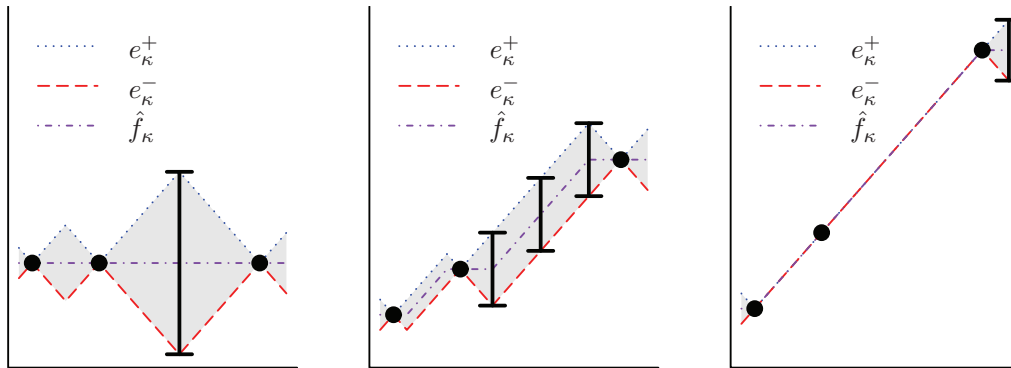

Figure 3. *Illustration of how the pointwise uncertainty depends on the observed variation of f: the uncertainty is smaller where the data require f to vary rapidly. The vertical distance between the blue and red curves is twice the uncertainty at the corresponding abscissa. The black error bars are at some points where the uncertainty is largest. The succession of panels shows that as the slope between observations approaches κ, $\mathcal{E}_\kappa(w)$ approaches 0 for points w between observations, and the maximum uncertainty decreases.*

and

$$e_\kappa^-(w) \equiv \max_{x \in X} \left[f(x) - \kappa d(x, w) \right].$$

The mean of the two is

$$\hat{f}_\kappa(w) \equiv \hat{f}_\kappa(w; X, \kappa) \equiv \frac{e_\kappa^-(w) + e_\kappa^+(w)}{2}.$$

Figures 2 and 3 illustrate these definitions. The proof of Proposition 1 shows that the function $\hat{f}_\kappa(w)$ is the minimax emulator for pointwise error over the class $\mathcal{F}_\kappa$ of functions that agree with the data and have Lipschitz constant no greater than κ. The minimax emulator $\hat{f}_\kappa(w)$ interpolates (rather than smooths) the data.

3. Bounds on the number of observations needed to approximate f well. In this section we construct a function $\bar{f}$ that agrees with the data $f|_X$, has Lipschitz constant $\hat{K}$ (the smallest Lipschitz constant consistent with the data), and yet would require a large number M_ϵ of additional observations $f|_Y$ to estimate f within ϵ on $[0,1]^p$.[4] The function $\bar{f}$ is not intended to be an emulator—it is a technical device. Since f could in fact be $\bar{f}$, this gives a lower bound on the number of additional observations that might be required to estimate f well, even if f is no rougher than the original data $f|_X$ reveal it to be.

Let $B(x,\delta)$ denote the open ball in $\mathbb{R}^p$ centered at x with radius δ. Since f has Lipschitz constant K, $f(y)$ is guaranteed to be within ϵ of $f(x)$ if $y \in B(x, \epsilon/K)$. But depending on f and X, it can happen that $\hat{f}_{\hat{K}}$ is guaranteed to be within $\pm\epsilon$ of every $g \in \mathcal{F}_K$ for parts of the domain not contained in $\cup_{x \in X} B(x, \epsilon/K)$. To see this, consider $p = 1$, $f(x) = x$, and let X be the two-element set $\{0, 1\}$. Then $K = \hat{K} = 1$. In this case, the observations $f|_X$ determine f exactly: the only function in $\mathcal{F}_K$ is f. In this example, for a function g to agree with the observations requires it to attain the Lipschitz constant K everywhere. A function cannot agree with the observations and "run away" from f very far.

More generally, if f varies on X, then for a function g to agree with f at the observations, g must vary too. That required variation "spends" some of g's Lipschitz constant, preventing g from running as far away from f as it could if f_X were constant. We now quantify this intuition to construct a function $\bar{f}$ that requires many additional observations to estimate well. The function $\bar{f}$ is constant "as much as possible" subject to the constraint that it interpolates the data and has Lipschitz constant $\hat{K}$. Since estimating $\bar{f}$ where it is constant is hard (as illustrated in Figure 3), the size of the set where $\bar{f}$ could be constant gives a lower bound on the number of additional observations that might be required.

Define $\bar{\gamma} \equiv \arg\min_{\gamma \in \mathbb{R}} \sum_{x \in X} |f(x) - \gamma|^p$. Computing $\bar{\gamma}$ is straightforward because the objective function is univariate and convex.[5] Let $X^+ \equiv \{x \in X : f(x) \geq \bar{\gamma}\}$, and let $X^- \equiv \{x \in X : f(x) < \bar{\gamma}\}$. Let

$$Q_+ \equiv \bigcup_{x \in X^+} \left\{ B\left(x, \frac{f(x) - \bar{\gamma}}{\hat{K}}\right) \bigcap [0,1]^p \right\}$$

and

$$Q_- \equiv \bigcup_{x \in X^-} \left\{ B\left(x, \frac{\bar{\gamma} - f(x)}{\hat{K}}\right) \bigcap [0,1]^p \right\}.$$

Then $Q_+ \cap Q_- = \emptyset$.[6]

---

[4]We do not discuss the choice of $\epsilon > 0$ in detail: scientific context should inform the choice. In examples below, we set ϵ to be an absolute tolerance, a fraction of $\hat{K}$, and a fraction of K. One might also consider relating ϵ to the "typical value" of f (e.g., the mean of f or of $f|_X$).

[5]Alternatively, we could set $\bar{\gamma} \equiv \frac{1}{\#X} \sum_{x \in X} f(x)$, where $\#X$ is the size of X. The resulting lower bound may not be as tight.

[6]Fix $x^+ \in X^+$ and $x^- \in X^-$. Then $|f(x^+) - f(x^-)|/d(x^+, x^-) \leq \hat{K}$. Equivalently, $d(x^+, x^-) \geq |f(x^+) - f(x^-)|/\hat{K}$. Let $B^+ = B(x^+, [f(x) - \bar{\gamma}]\hat{K})$ and $B^- = B(x^-, [\bar{\gamma} - f(x)]/\hat{K})$. Let a be the sum of the radii of B^+ and B^-. Then $a = (f(x^+) - \bar{\gamma})/\hat{K} + (\bar{\gamma} - f(x^-))/\hat{K} = (f(x^+) - f(x^-))/\hat{K}$, and $a \leq d(x^+, x^-)$. Therefore, $B^+ \cap B^- = \emptyset$. Because our selection of $x^+ \in X^+$ and $x^- \in X^-$ was arbitrary, $Q^+ \cap Q^- = \emptyset$.
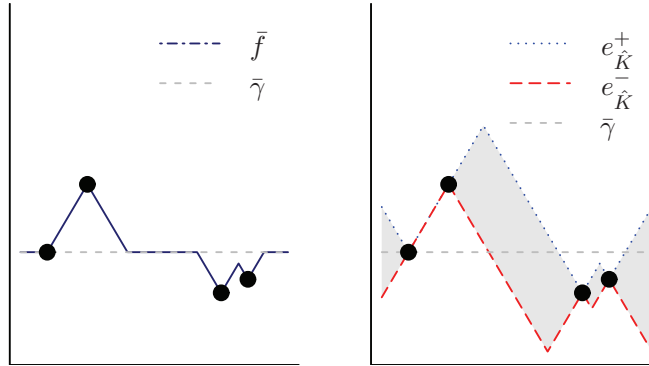
Figure 4. *A function that agrees with the data, has Lipschitz constant $\hat{K}$, and is hard to estimate because it is often constant. The function $\bar{f}$ (shown in the left panel) is comprised of segments of $e_{\hat{K}}^{+}$, $e_{\hat{K}}^{-}$ and the constant function $\bar{\gamma}$ (all shown in the right panel). It is constant over roughly half of the domain. No function between $e_{\hat{K}}^{-}$ and $e_{\hat{K}}^{+}$ (inclusive) is constant over a larger fraction of the domain.*

Define

$$\bar{f} : [0,1]^p \to \mathbb{R},$$

$$w \mapsto \begin{cases} e_{\hat{K}}^{-}(w), & w \in Q_+, \\ e_{\hat{K}}^{+}(w), & w \in Q_-, \\ \bar{\gamma} & \text{otherwise.} \end{cases}$$

Figure 4 illustrates this definition. If we know $f|_X$, we know $\bar{f}$. By construction, $\bar{f} \in \mathcal{F}_{\hat{K}} \subset \mathcal{F}_K$.

Let $\bar{Q} \equiv [0,1]^p \setminus (Q_+ \cup Q_-)$. Let μ be Lebesgue measure. By the union bound, because $\mu([0,1]^p) = 1$,

$$\mu(\bar{Q}) \geq 1 - \sum_{x \in X} \mu\left(B\left(x, |f(x) - \bar{\gamma}|/\hat{K}\right)\right).$$

Let $C_2 \equiv \frac{\pi^{p/2}}{\Gamma(p/2+1)}$ and $C_\infty \equiv 2^p$, where Γ is the gamma function. Then, for $q \in \{2, \infty\}$,

$$\mu(\bar{Q}) \geq 1 - C_q \sum_{x \in X} \left(|f(x) - \bar{\gamma}|/\hat{K}\right)^p.$$

If there is some $x \in X$ for which for all $g \in \mathcal{F}_{\hat{K},\{x\}}$, $|g(y) - f(x)| < \epsilon$ for all $y \in A \subset \bar{Q}$, then $\mu(A) \leq \mu(B(0, \epsilon/\hat{K}))$. Hence, because $\bar{f} \in \mathcal{F}_K$,

$$M_\epsilon \geq \left\lceil \frac{\mu(\bar{Q})}{\mu(B(0, \epsilon/\hat{K}))} \right\rceil$$

(3.1)
$$\geq \left\lceil \epsilon^{-p} \left[\frac{\hat{K}^p}{C_q} - \sum_{x \in X} |f(x) - \bar{\gamma}|^p \right] \right\rceil.$$

Section 5 shows that this lower bound, the *minimum computational burden*, can be extremely large for even modest problem dimensions p.

4. Bounds on the maximum uncertainty for a fixed experimental design. The previous section gave lower bounds on the number of additional observations of f required to attain a desired maximum uncertainty ϵ. This section gives two lower bounds on the maximum uncertainty $\mathcal{E}_K(\hat{f})$ for a fixed experimental design X: an absolute bound and a bound expressed as a fraction of K. The bound as a fraction of K can yield a strong negative result: when a statistic—calculable from the observations—exceeds a calculable threshold, the maximum uncertainty is not less than the maximum uncertainty of the best emulator based on a single observation at the centroid of the domain. If the goal is to minimize the maximum uncertainty, we could have just approximated f as constant and saved $\#X - 1$ observations.

4.1. Lower bounds. Consider the set $\mathcal{F}_\kappa$ of functions g that agree with the observations $f|_X$ and have Lipschitz constant no larger than κ. Consider all possible emulators $\hat{f}$. Proposition 1 states that the smallest (across emulators $\hat{f}$) maximum (across functions g) error at the point $w \in [0,1]^p$ is $[e_\kappa^+(w) - e_\kappa^-(w)]/2$, and the emulator $\hat{f}_\kappa(w)$ attains this bound at every w.

Proposition 1. *If $\kappa \geq \hat{K}$, then*

$$\mathcal{E}_\kappa(w) = \mathcal{E}_\kappa(w; \hat{f}_\kappa) = \frac{e_\kappa^+(w) - e_\kappa^-(w)}{2}.$$

Proofs, including the proof of Proposition 1, are in Appendix A.

Corollary 2. *For any emulator $\hat{f}$,*

(4.1) $$\mathcal{E}_K(w, \hat{f}) \geq \mathcal{E}_K(w) \geq \mathcal{E}_{\hat{K}}(w; \hat{f}_{\hat{K}}).$$

Corollary 2 follows from Proposition 1 and the fact that, since $\hat{K} \leq K$,

$$\mathcal{F}_{\hat{K}} \subset \mathcal{F}_K.$$

Corollary 2 is one of our principal results: $\mathcal{E}_{\hat{K}}$, a statistic calculable solely from the observations $f|_X$, is a lower bound on the maximum uncertainty for *any* emulator $\hat{f}$ based on the observations $f|_X$. Theorem 3 gives a stronger lower bound in terms of the unknown value of K.

Theorem 3. *For any $\lambda \in \Re^+$, if $\mathcal{E}_{\hat{K}} \geq \lambda\hat{K}$, then $\mathcal{E}_K(\hat{f}) \geq \lambda K$.*

4.2. Maximum uncertainty for an emulator based on one observation. In this section we work in ℓ_∞: $d(v,w) = \|v - w\|_\infty$. This simplifies the calculations and gives a particularly strong result.

Let $z \equiv (1/2, \ldots, 1/2)$, the centroid of $[0,1]^p$, and let $Z \equiv \{z\}$. Let $\hat{g} \in \mathcal{F}_{\infty,Z}$ be the constant function $\hat{g}(w) \equiv f(z)$ for all $w \in [0,1]^p$. The ℓ_∞ distance from z to any point on the boundary of $[0,1]^p$ is $1/2$, so

$$\mathcal{E}_{K,Z}(\hat{g}) = \frac{K}{2}.$$

That is, the maximum uncertainty of the emulator that is constant throughout $[0,1]^p$ and equal to the value of f at the centroid of the cube is $K/2$. Let $W \subset [0,1]^p$ be finite, and let $c \in \mathbb{R}$. Suppose f is constant on the set W and that W contains fewer than 2^p points. Let $\hat{h} \in \mathcal{F}_{\infty,W}$. By examining the corners of the domain, it follows that

$$\mathcal{E}_{K,W}(\hat{h}) \geq \frac{K}{2}.$$

Making 2^p observations of f is intractable for the Community Atmosphere Model and for many other applications. If f is nearly constant, the situation may still be hopeless.

How do we know whether $f|_X$ is too close to constant to benefit from observing it more than once but fewer than 2^p times?

Corollary 4. *If $\mathcal{E}_{\hat{K}} \geq \hat{K}/2$, then*

$$\mathcal{E}_K(\hat{f}) \geq \frac{K}{2} \geq \mathcal{E}_{K,Z}(\hat{g}).$$

That is, if $\mathcal{E}_{\hat{K}} \geq \hat{K}/2$, no emulator based on observing $f|_X$ has smaller maximum uncertainty than the constant emulator based on a single observation—f is too nearly constant. Corollary 4 follows directly from Theorem 3, taking $\lambda = \hat{K}/2$.

5. Applications. This section presents three examples of increasing complexity: two in which f is known analytically, and one in which f is *HEB* arising from a numerical model of climate. In this section the distance metric is $d(v, w) = \|v - w\|_\infty$, except where noted.

5.1. High-dimensional ℓ_∞ cone. Consider a emulating a function defined on the 21-dimensional hypercube $[0, 1]^{21}$; $z \equiv (0.5, \ldots, 0.5)$ denotes the center of that hypercube. Suppose

$$f(x) \equiv \|x - z\|_\infty.$$

We observe f at z and, for $i = 1, \ldots, 21$, at both points satisfying $x_i \in \{0, 1\}$ and $x_j = 0.5$ for $j \neq i$. (This is a "one-at-a-time" sampling design, where one component at a time is shifted from a typical value to a more extreme value.) These 43 points constitute X. Then

$$\hat{K} = K = 1.$$

Because every point $w \in [0, 1]^{21}$ is within 0.5 of $x \in X$ satisfying $f(x) = 0.5$,

$$e_{\hat{K}}^- \geq 0.$$

Because every point $w \in [0, 1]^{21}$ is within 0.5 of z, and $f(z) = 0$,

$$e_{\hat{K}}^+ \leq 0.5.$$

Hence, by Corollary 2,

$$\mathcal{E}_{\hat{K}} \leq 0.25.$$

Had we only observed f at z but fixed $\hat{K} = 1$ (or observed f at another point in addition to z and computed $\hat{K}$ from those two points),

$$\mathcal{E}_{\hat{K}} = 0.5.$$

In this example, despite the high dimension of dom(f), emulating f using a modest number of observations (43) has smaller maximum uncertainty than emulating f using just a single observation of f at z: a small number of observations may constrain a high-dimensional function globally. High-dimensional problems with small numbers of data do not necessarily

Table 2

Borehole function domain.

Variable	Range	Description
H_u	$[990, 1110]$	potentiometric head of upper aquifer (m)
H_ℓ	$[700, 820]$	potentiometric head of lower aquifer (m)
T_u	$[63070, 115600]$	transmissivity of upper aquifer (m^2/yr)
T_ℓ	$[63.1, 116]$	transmissivity of lower aquifer (m^2/yr)
r	$[100, 50000]$	radius of influence (m)
r_w	$[0.05, 0.15]$	radius of borehole (m)
L	$[1120, 1680]$	length of borehole (m)
K_w	$[9855, 12045]$	hydraulic conductivity of borehole (m/yr)

have large uncertainties, as "the curse of dimensionality" would suggest. The dimension matters, but so does f itself.

To connect our results to a common emulation method, we fit a GP to $f|_X$ by maximum likelihood using the R package `mlegp` [17]. For 100,000 points selected uniformly at random from $[0, 1]^p$, the mean error is 0.02, 2% of K. The maximum error at these 100,000 points is 0.23, but the error at $(0.6, \ldots, 0.6)$—which is not in the sample—is 0.38.[7] Because the error of $\hat{f}_{\hat{K}}$ is no greater than $\mathcal{E}_{\hat{K}} = 0.25$, for this f, the minimax emulator $\hat{f}_{\hat{K}}$ outperforms this GP emulator both in minimax uncertainty and in actual maximum error.

5.2. Borehole function. The commonly used test function

$$f_0(H_u, H_\ell, T_u, T_\ell, r, r_w, L, K_w) \equiv \frac{2\pi T_u (H_u - H_l)}{\log (r/r_w) \left(1 + \frac{2LT_u}{\log(r/r_w)r_w^2 K_w} + \frac{T_u}{T_\ell}\right)}$$

models water flow through a borehole [16]. Its input variables are described in Table 2, which also lists the ranges of those variables. The output is water flow rate in cubic meters per year. We rescale f_0 so that its inputs range over the 8-dimensional unit hypercube $[0, 1]^8$; the resulting function is denoted by f.

Reasoning about the functional form of f (Appendix B) shows that

$$944 \leq K \leq 1200.$$

Of course, if f really were a black box, such reasoning would be impossible. We estimated $\hat{K}$ from 1000 sample points selected in two different ways:

1. Select 1000 points by Latin hypercube sampling. This yields $\hat{K} = 367$.
2. Select 100 points by Latin hypercube sampling. For each of these points, draw an additional nine points a small distance (10^{-5}) from it in each coordinate in a random direction. This yields $\hat{K} = 576$.

We fix $\hat{K} = 576$ for the remainder of this example; note that this is roughly half the true value of K.

Now let X contain the following 273 points: all $2^8 = 256$ corners of $[0, 1]^8$, the center of the domain $(0.5, \ldots, 0.5)$, and, for $i = 1, \ldots, 8$, each of the two points satisfying $x_i \in \{0, 1\}$

---

[7]This point was found by searching the ray $c(1, \ldots, 1)$; there might be points with even larger errors.

and $x_j = 0.5$ for $j \neq i$. (The empirical Lipschitz constant of f on this set is less than 576.) By branch-and-bound we find

$$\mathcal{E}_{576} < 207,$$

which is less than $576/2$. Hence, by Corollary 2, the best emulator $\hat{f}_{576}$ based on $f|_X$ has lower maximum uncertainty than the best emulator based on $f|_{\{z\}}$ alone.

Holding X fixed, we now lower-bound M_ϵ, the minimum computational burden (section 3). Convex programming finds $\bar{\gamma} = 134.7$. The union bound implies that the proportion of the domain where f could be constant is $\mu\left(\bar{Q}\right) \geq 0.76$. Then for $\epsilon = 100$ (about 20% of $\hat{K}$ or 10% of K), $M_{100} \geq 3598$ additional observations might be needed. But for $\epsilon = 10$, $M_{10} \geq 3.59 \times 10^{11}$ additional observations might be required.

For comparison, we emulate f by a GP, again estimating the parameters using the R package `mlegp` [17] from the same set X of 273 points. For 100,000 points selected at random uniformly from $[0,1]^8$, the mean error is 37.3, approximately 3% of K. The maximum error at these points is 207.9, approximately 20% of K.

5.3. Climate modeling. The Uncertainty Quantification Initiative at Lawrence Livermore National Laboratory[8] provided results from 1154 climate simulations using the Community Atmosphere Model (CAM) with $p = 21$ parameters. Each parameter was scaled so that the interval $[0,1]$ contained all values considered physically reasonable. The output of interest was a scalar, the simulated global average upwelling longwave flux (FLUT) averaged over the third through twelfth years of the simulation (a 10-year average after a 2-year burn-in). Each such average is deterministic: repeating a run with the same input parameters should produce the same output. The simulator amounts to a function f that maps $[0,1]^p \to \mathbb{R}$. Running the simulator was computationally expensive; each run took several days on a supercomputer. The Lawrence Livermore National Laboratory team used several approaches to choose the points $X \subset [0,1]^p$ at which to run simulations, including Latin hypercube, one-at-a-time, and random-walk multiple-one-at-a-time [5]. The 1154 simulations include all points selected by any of those approaches.

For these observations, we find $\bar{\gamma} = 232.77$, $\hat{K} = 14.20$ for $q = 2$, and $\hat{K} = 34.68$ for $q = \infty$.

5.3.1. Computational burden. By (3.1),

$$M_\epsilon \geq \left\lceil \epsilon^{-21} \left[\frac{1.57 \times 10^{24}}{0.014} - 6.81 \times 10^{24} \right] \right\rceil > \epsilon^{-21} \times 10^{25}$$

for $q = 2$. For example, if ϵ is 1% of $\hat{K}$, then $M_\epsilon \geq 10^{43}$. Even if ϵ is 50% of $\hat{K}$, $M_\epsilon > 10^8$. For $q = \infty$,

$$M_\epsilon \geq \left\lceil \epsilon^{-21} \left[\frac{2.19 \times 10^{32}}{2^{21}} - 6.81 \times 10^{24} \right] \right\rceil > \epsilon^{-21} \times 10^{25}.$$

These lower bounds on the minimum computational burden are extreme for a wide range of values of ϵ: there are functions that fit the 1154 observations and are as regular as the

---

[8]This dataset was provided by the Institutional Science and Technology Office at Lawrence Livermore National Laboratory under the Uncertainty Quantification Strategic Initiative Laboratory-Directed Research and Development Project 10-SI-013.

observations allow but that cannot be approximated with useful uncertainty from any tractable number of observations. The function $\bar{f}$, which is simple to construct, attains these lower bounds on minimum computational burden. Note the contrast with the cone example, which was also 21-dimensional: the dimension of $\mathrm{dom}(f)$ does not by itself determine how hard it is to emulate f accurately.

5.3.2. Uncertainty. Is the maximum uncertainty of the best emulator based on observing f at the 1154 points in X lower than the maximum uncertainty of the constant emulator based on one observation of f at the centroid of $[0, 1]^p$? We cannot simply compute these two maximum uncertainties because K is unknown. But Corollary 4 applies if we can determine whether $\mathcal{E}_{\hat{K}} \geq \hat{K}/2$. Unfortunately, determining $\mathcal{E}_{\hat{K}}$ is difficult. In ℓ_∞, if $f|_X$ is constant, finding $\mathcal{E}_{\hat{K}}$ amounts to finding a maximal empty hypercube, a problem recently shown to be NP-hard in p [18]. It is generally no easier if f varies on X. Fortunately, it suffices to bound $\mathcal{E}_{\hat{K}}$. By working in ℓ_∞, we can bound $\mathcal{E}_{\hat{K}}$ above and below by considering just the corners of $[0, 1]^p$; we take $d(v, w) = \|v - w\|_\infty$ throughout this section.

Proposition 5. *Let* $\mathbf{0} \equiv (0, \ldots, 0)$, $\mathbf{1} \equiv (1, \ldots, 1)$, *and* $\tilde{d}(v) \equiv \max(d(v, \mathbf{0}), d(v, \mathbf{1}))$. *Then*

$$\mathcal{E}_{\hat{K}} \leq \frac{1}{2} \left\{ \min_{x \in X} \left[f(x) + \hat{K}\tilde{d}(x) \right] - \max_{x \in X} \left[f(x) - \hat{K}\tilde{d}(x) \right] \right\}.$$

Using this proposition, we calculate $\mathcal{E}_{\hat{K}} \leq 20.95$ for the CAM dataset. On the other hand, the maximum over all $[0, 1]^p$ is at least as large as the maximum over the corners of $[0, 1]^p$:

$$\mathcal{E}_{\hat{K}} \geq \max \left\{ \mathcal{E}_{\hat{K}}(w) : \forall w \in \{0, 1\}^p \right\}.$$

Perhaps surprisingly, this lower bound is essentially sharp for the CAM dataset. The domain $[0, 1]^p$ contains 2^p corners $\{r_i\}_{i=1}^{2^p}$. Divide $[0, 1]^p$ into 2^p hypercubes $\{R_i\}_{i=1}^{2^p}$ with edge-length $1/2$, disjoint interiors, each containing a different corner of $[0, 1]^p$ (e.g., one such hypercube is $[0, 1/2]^p$). Then the R_i are disjoint ℓ_∞-balls of radius $1/4$. Because X contains only 1154 points, the vast majority of $\{R_i\}_{i=1}^{2^p}$ do not contain any element of X. Because $\mathcal{E}_{\hat{K}}(w)$ tends to increase with distance from points in X, these unoccupied hypercubes are good regions to look for points with large values of $\mathcal{E}_{\hat{K}}(w)$. Within an unoccupied hypercube R_i, no point is farther in ℓ_∞ from any point in X than the corner r_i. So, the corners $\{r_i\}_1^{2^p}$ are good places to observe $\mathcal{E}_{\hat{K}}(w)$ to find a tight lower bound on $\mathcal{E}_{\hat{K}}$.

For the CAM dataset, one corner r_j attains $\mathcal{E}_{\hat{K}}(r_j) = 20.95$. Since this is also the numerical upper bound, $\mathcal{E}_{\hat{K}} = 20.95$.

Because $\mathcal{E}_{\hat{K}} \geq \hat{K}/2 = 17.34$, Theorem 3 says that $\mathcal{E}_K(\hat{f}) \geq K/2$ for any emulator $\hat{f}$. In other words, by the discussion in section 4.2, our maximum uncertainty would have been no greater had we just observed f once, at z, and predicted $\hat{f}(w) = f(z)$ for all $w \in [0, 1]^p$.

In some sense, this result is not surprising: if we had fixed $\hat{K}$ but replaced f with a constant function, and $\#X < 2^p$, then $\mathcal{E}_{\hat{K}} \geq \hat{K}/2$, with equality holding if and only if $z \in X$. By repeating the bounding procedures from the previous two sections with $\hat{K}/2 = 17.34$ fixed but f replaced with constant function c, we find $\mathcal{E}_{c, X, \hat{K}} = 26.95$. The increase in maximum uncertainty from 20.95 to 26.95 that results from replacing f with a constant shows that the observed variation in f reduces the maximum uncertainty considerably—although the maximum uncertainty remains quite large.

Table 3
Confidence bounds for quantiles and the mean of the uncertainty of the minimax emulator $\hat{f}_{\hat{K}}$ for CAM.

Norm	Units	95% lower confidence bound			Average
		Lower quartile	Median	Upper quartile	
Euclidean	$\hat{K}/2$	1.462	1.599	1.732	1.599
Supremum	$\hat{K}/2$	0.648	0.716	0.781	0.715
Euclidean	$\hat{\gamma}$	0.044	0.049	0.053	0.049
Supremum	$\hat{\gamma}$	0.048	0.053	0.058	0.053

Column 1: distance metric d used for the Lipschitz constant.
Columns 3–5: binomial lower 95% confidence bounds for quartiles of
the uncertainty, obtained by inverting binomial tests. Column 6: 95%
lower 95% confidence bound for the integral of the uncertainty over the
entire domain $[0,1]^p$, based on inverting z-tests. Columns 3–6 are ex-
pressed as a fraction of the quantity in column 2. Results are based on
10,000 uniform random samples from $[0,1]^p$.

To connect these theoretical results to common emulation methods, we fit a GP model [17]
and multivariate adaptive regression splines (MARS) [19] to the 110 CAM observations from
a Latin hypercube design, leaving 1043 observations for testing. On the test set, the mean
error of the Gaussian process model is 1.03 (3% of $\hat{K}$) and the maximum error is 6.73 (20% of
$\hat{K}$). For MARS, the mean error on the test set is 1.59 and the maximum error is 6.21. Since
the 1043 test points are all distant from many corners of $[0,1]^p$, the error of these methods
over $[0,1]^p$ might be far larger; it would take many more evaluations of f to tell. Absent such
data, there is no evidence that those methods have maximum error less than $\mathcal{E}_{\hat{K}} = 20.95$.

6. Extensions.

6.1. Distribution of the uncertainty. By drawing independent points $W \sim \text{Uniform}([0,1]^p)$
and evaluating $\mathcal{E}_{\hat{K}}(W)$, we construct lower confidence bounds for quantiles of the uncertainty
and the mean uncertainty over $[0,1]^p$. Table 3 shows the results for the CAM simulations
based on 10,000 random samples from $[0,1]^p$. Even the lower quartiles are a large fraction of
$\hat{K}$. For instance, at confidence level 95%, the uncertainty under the sup-norm metric exceeds
71.7% of $\hat{K}/2$ on at least 50% of the domain.

6.2. Uncertainty relative to typical values. We have focused on taking ϵ to be a frac-
tion of K or $\hat{K}$. When ϵ is chosen that way, sections 3 and 4 establish conditions under
which no emulator can be guaranteed to replicate the variation of f. Emulators are generally
constructed to capture the complexity of the model: tracking its variability. That suggests
approximating f to within a fraction of its variation, which is why we have calibrated ϵ to
$\hat{K}$. If the goal were to approximate f to within a fraction of its mean, and its mean is large
compared to its variation, approximating f globally by its sample mean might suffice. Then
it might make sense to set ϵ to be a fraction of a typical value of f, for instance, $\bar{\gamma}$ or the
sample mean

$$\hat{\gamma} = \frac{1}{\#X} \sum_{x \in X} f(x).$$

Table 4
Minimum computational burden for the CAM model.

Norm	$\epsilon/\hat{\gamma}$	Lower bound on M_ϵ
Euclidean	0.02	3.6×10^{12}
	0.04	1,720,354
	0.06	345
	0.08	1
Supremum	0.02	8.6×10^{10}
	0.04	413,595
	0.06	83
	0.08	1

The last two rows of Table 3 list confidence bounds for percentiles of the uncertainty as a fraction of $\hat{\gamma}$.

Similarly, for ϵ chosen suitably, inequality (3.1) gives a lower bound on M_ϵ for approximating f within a fraction of its typical value, rather than within a fraction of its observed variation. (Of course, the resulting bounds can be made arbitrarily small by adding a sufficiently large constant to f. One reason we think it is more interesting to calibrate ϵ as a fraction of K or $\hat{K}$ is that the results are invariant under affine transformations of f.)

For the CAM model, this lower bound on M_ϵ is trivial when ϵ is a large fraction of the typical value of f but grows rapidly as the fraction decreases (Table 4).

6.3. Other uses for e_κ^- and e_κ^+. We have primarily used e_κ^+ and e_κ^- to construct the minimax emulator and find its uncertainty. But if f is no less regular than it was observed to be, $e_{\hat{K}}^+$ is a pointwise upper bound on f and $e_{\hat{K}}^-$ is a pointwise lower bound on f. Moreover, if f is no less regular than the data require it to be, $\max_{w\in[0,1]^p} e_{\hat{K}}^+(w)$ is a global upper bound on f and $\min_{w\in[0,1]^p} e_{\hat{K}}^-(w)$ is a global lower bound on f.

Maximizing $e_{\hat{K}}^+$ or minimizing $e_{\hat{K}}^-$ exactly may not be tractable. For sup-norm, we can use the techniques from section 5.3 to bound these extrema from above and below: for the CAM model, those upper and lower bounds on $e_{\hat{K}}^+$ are equal, as they are for $e_{\hat{K}}^-$. The maximum of $e_{\hat{K}}^+$ is 253.78, and the minimum of $e_{\hat{K}}^-$ is 211.88.

7. Conclusions. We find a lower bound on the minimum (over emulators) maximum (over functions that agree with the data and are as regular as the data allow) error of emulators of a function f based on n observations. This "mini-minimax" uncertainty is optimistic because it assumes that f has the smallest Lipschitz constant consistent with the data. The mini-minimax uncertainty is an attainable bound on the error of the best emulator of f at w: for any emulator $\hat{f}$, there is a function g that is at least as regular as f, that agrees with f at the n observations, and for which $|\hat{f}(w) - g(w)|$ is at least this mini-minimax value.

In some problems, *every* emulator based on any tractable number of observations of f has large maximum uncertainty (and the uncertainty is large over much of the domain), even if f is as regular as the data allow. That is, there are functions g and h that agree perfectly with the observations, are as regular as the observations permit, and yet differ by a large amount

at some point in the domain of f.

We give sufficient conditions under which even the best possible emulator has large uncertainty. The conditions depend only on the observed values of f; they can be computed from the same observations used to train an emulator, at a cost that typically is small compared with the cost of generating those observations. The conditions are sufficient but not necessary because f could be less regular than any finite set of observations reveals it to be. It is not possible to give necessary conditions that depend only on the observed values of f; a priori bounds on the regularity of f would be needed.

The conditions seem likely to hold for many high-consequence applications. Indeed, we show quantitatively that the conditions hold for a large climate-modeling dataset. When the maximum uncertainty in approximating f everywhere by a constant—the value of f at the center of the domain—is no larger than the maximum uncertainty in approximating f from any tractable number of observations, emulators may not be useful. No emulator can then reliably model f as a function of its input $w \in [0,1]^p$.

Common techniques for assessing the accuracy of emulators (e.g., posterior variance or performance on hold-out data) understate the true uncertainty because they make strong assumptions about f that are based neither on the observations nor on known properties of f or because they focus on average error rather than worst-case error. However, as section 6 shows, even the average uncertainty and quartiles of the uncertainty for the CAM model are quite large.

The mini-minimax uncertainty is a one-sided tool: if this uncertainty is large, the data do not constrain f well, while if it is small, the data constrain f *only* if it is no less regular than the data collected so far show it must be. That said, if the mini-minimax uncertainty is uncomfortably large, there might be ways to reduce it. For instance, if the lower bound (3.1) on the computation burden required to reduce the uncertainty to a useful level ϵ is affordable, one might collect more data. Provided the new data do not increase $\hat{K}$ substantially, the mini-minimax uncertainty can be reduced at will. But when p is large, the lower bound is likely to be large because it grows exponentially with p. If observing f requires a real-world experiment, new technology might be required to make a useful number of additional observations affordable. When observing f involves running a simulator, collecting enough additional data to reduce the uncertainty to a reassuring value might require not only recruiting additional computational resources but also reducing the computational cost of each simulation—substantially.

In some cases, clever strategies can reduce the cost of computing f, at least to some known degree of approximation, but that is not always so. Cost reductions of orders of magnitude might require reducing the complexity of f. Reducing the dimension p of the domain of f is especially helpful because reducing p pays exponential dividends. But it requires scientific justification: In general, eliminating parameters from a model entails bias in the model with no a priori limit. It is hard to calibrate the tradeoff between fitting a model that is constrained by the data but is known or suspected to be overly simplistic—and therefore biased—and a model that has lower bias but cannot be estimated reliably from an affordable number of data. Subject-matter knowledge is key.

Without increasing the number of observations or revising the model, reducing the un-

certainty of emulators requires either more information about f[9] or changing the measure of uncertainty—changing the scientific question. Finally, approximating f pointwise is not usually the ultimate scientific goal. More important questions about f might be answered more directly.[10] These tactics are application-specific: the underlying science dictates the conditions that actually hold for f and the questions about f that matter.

Appendix A. Proofs. For real χ and ρ, define the interval

$$I(\chi, \rho) \equiv \begin{cases} [\chi - \rho, \chi + \rho], & \rho \geq 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

If I is an interval, $\mu(I)$ denotes its length; for instance, $\mu(I(\chi, \rho)) = \max(0, 2\rho)$.

Lemma 6. *Fix $\alpha \in [0,1]$, $\rho_1, \ldots, \rho_n \in [0, \infty)$, and $\chi_1, \ldots, \chi_n \in \mathbb{R}$. Let $I_1 \equiv \bigcap_{i=1}^n I(\chi_i, \rho_i)$ and $I_\alpha \equiv \bigcap_{i=1}^n I(\chi_i, \alpha\rho_i)$. Then $\alpha\mu(I_1) \geq \mu(I_\alpha)$.*

Proof. Because the intersection of intervals is itself an interval, there exist χ_0 and ρ_0 satisfying

$$I_\alpha = I(\chi_0, \rho_0).$$

Fix $i \in 1, \ldots, n$. Then

$$I(\chi_0, \rho_0) \subset I(\chi_i, \alpha\rho_i).$$

It follows that

$$\chi_0 - \rho_0 \geq \chi_i - \alpha\rho_i.$$

Then

$$\alpha \left(\rho_i - \frac{\rho_0}{\alpha} \right) \geq \chi_i - \chi_0.$$

Because $\alpha \leq 1$ and $\rho_i \geq 0$,

$$\rho_i - \frac{\rho_0}{\alpha} \geq \chi_i - \chi_0.$$

Finally,

$$\chi_0 - \frac{\rho_0}{\alpha} \geq \chi_i - \rho_i.$$

By symmetric reasoning we also have

$$\chi_0 + \frac{\rho_0}{\alpha} \leq \chi_i + \rho_i.$$

Therefore,

$$I \left(\chi_0, \frac{\rho_0}{\alpha} \right) \subset I(\chi_i, \rho_i).$$

---

[9]Common additional conditions include the following: parameters have only low-order interactions; the second derivative has an upper bound; the third derivative has a limited number of knots; the integral of the squared derivative of the model is bounded [20]. There are problems in which conditions like these may reflect actual knowledge about f. However, such conditions tend to be difficult to verify: simulation is perhaps most valuable when the underlying equations are not amenable to mathematical analysis.

[10]For example, for global optimization—finding maxima or minima—a form of adaptive sampling known as multistart methods yields good results [21].

Because i was arbitrary,

$$I\left(\chi_0, \frac{\rho_0}{\alpha}\right) \subset I_1.$$

Hence,

$$\mu\left(I_1\right) \geq \mu\left(I\left(\chi_0, \frac{\rho_0}{\alpha}\right)\right) = \frac{2\rho_0}{\alpha} = \frac{\mu\left(I_\alpha\right)}{\alpha}. \qquad \blacksquare$$

Lemma 6 is used in the proof of Theorem 3, below.

Proposition 1. *If $\kappa \geq \hat{K}$, then*

$$\mathcal{E}_\kappa(w) = \mathcal{E}_\kappa(w; \hat{f}_\kappa) = \frac{e_\kappa^+(w) - e_\kappa^-(w)}{2}.$$

Proof.

Step 1: e_κ^+ *and* e_κ^- *are Lipschitz continuous with constant* κ. For $v, w \in [0,1]^p$, there exist $x, y \in X$ satisfying

$$e_\kappa^+(v) = f(x) + \kappa d(x, v) \quad \text{and} \quad e_\kappa^+(w) = f(y) + \kappa d(y, w).$$

Suppose without loss of generality that $e_\kappa^+(v) \geq e_\kappa^+(w)$. By construction, $e_\kappa^+(v) \leq f(y) + \kappa d(y, v)$. Hence

$$\begin{aligned}
0 \leq e_\kappa^+(v) - e_\kappa^+(w) &\leq f(y) + \kappa d(y, v) - e_\kappa^+(w) \\
&= f(y) + \kappa d(y, v) - f(y) - \kappa d(y, w) \\
&\leq \kappa(d(y, v) - d(y, w)) \\
&\leq \kappa d(v, y)
\end{aligned}$$

by the triangle inequality. Hence e_κ^+ has Lipschitz constant κ. An analogous argument shows that e_κ^- also has Lipschitz constant κ.

Step 2: e_κ^+ *and* e_κ^- *agree with* f *on* X. (Hence, $\hat{f}_\kappa = (e_\kappa^+ + e_\kappa^-)/2$ agrees with f on X.) We have

$$\kappa \geq \hat{K} \equiv \max_{x,y \in X: x \neq y} \frac{|f(x) - f(y)|}{d(x, y)},$$

and hence $|f(x) - f(y)| \leq \kappa d(x, y)$ for all $x, y \in X$. Thus

$$\min_{x \in X}[f(x) + \kappa d(x, y)] = \min\left\{f(y), \min_{x \in X, x \neq y}[f(x) + \kappa d(x, y)]\right\} = f(y).$$

Similarly, $\max_{x \in X}[f(x) - \kappa d(x, y)] = f(y)$ for $y \in X$. Hence, $e_\kappa^+(y) = e_\kappa^-(y) = f(y)$ for $y \in X$. Since, as shown in Step 1, e_κ^+ and e_κ^- are Lipschitz with constant κ, e_κ^+ and $e_\kappa^- \in \mathcal{F}_\kappa$.

Step 3: e_κ^- *is the pointwise infimum of* $\mathcal{F}_\kappa$, *and* e_κ^+ *is the pointwise supremum of* $\mathcal{F}_\kappa$. Suppose to the contrary that there exist $w \in [0,1]^p$, $x \in X$, and $g \in \mathcal{F}_\kappa$ for which

$$g(w) > f(x) + \kappa d(x, w).$$

Recall that $g \in \mathcal{F}_\kappa$ implies that $g(x) = f(x)$ for all $x \in X$. Hence

$$g(w) - g(x) > f(x) + \kappa d(x, w) - f(x) = \kappa d(x, w).$$

That is, g has a Lipschitz constant greater than κ, a contradiction. Hence, $e_\kappa^+(w) = \sup\{g(w) : g \in \mathcal{F}_\kappa\}$ for all $w \in [0,1]^p$. The same argument, mutatis mutandi, shows that

$$e_\kappa^-(w) = \inf\{g(w) : g \in \mathcal{F}_\kappa\} \quad \forall w \in [0,1]^p.$$

Step 4: The maximum uncertainty of $\hat{f}_\kappa$ at w, $\mathcal{E}_\kappa(w; \hat{f}_\kappa)$, equals $[e_\kappa^+(w) - e_\kappa^-(w)]/2$.

$$\mathcal{E}_\kappa(w; \hat{f}_\kappa) \equiv \sup_{g \in \mathcal{F}_\kappa(w)} \left| \hat{f}_\kappa(w) - g(w) \right|$$

$$= \max \left\{ \sup_{g \in \mathcal{F}_\kappa(w)} g(w) - \hat{f}_\kappa(w), \hat{f}_\kappa(w) - \inf_{g \in \mathcal{F}_\kappa(w)} g(w) \right\}$$

(A.1)
$$= \max \left\{ e_\kappa^+(w) - \hat{f}_\kappa(w), \hat{f}_\kappa(w) - e_\kappa^-(w) \right\}$$

$$= \max \left\{ e_\kappa^+(w) - \frac{e_\kappa^+(w) + e_\kappa^-(w)}{2}, \frac{e_\kappa^+(w) + e_\kappa^-(w)}{2} - e_\kappa^-(w) \right\}$$

$$= \frac{e_\kappa^+(w) - e_\kappa^-(w)}{2}.$$

Equality (A.1) follows from Step 3.

Step 5: The minimax uncertainty at w, $\mathcal{E}_\kappa(w)$, equals $[e_\kappa^+(w) - e_\kappa^-(w)]/2$. Suppose $\hat{f}(w) > \hat{f}_\kappa(w)$. Then

$$|\hat{f}(w) - e_\kappa^-(w)| > \frac{e_\kappa^+(w) - e_\kappa^-(w)}{2} = \mathcal{E}_\kappa(w; \hat{f}_\kappa).$$

Suppose $\hat{f}(w) < \hat{f}_\kappa(w)$. Then

$$|\hat{f}(w) - e_\kappa^+(w)| > \frac{e_\kappa^+(w) - e_\kappa^-(w)}{2} = \mathcal{E}_\kappa(w; \hat{f}_\kappa).$$

Hence, $\hat{f}_\kappa(w)$ is minimax, and $\mathcal{E}_\kappa(w) = \mathcal{E}_\kappa(w, \hat{f}_\kappa) = [e_\kappa^+(w) - e_\kappa^-(w)]/2$. ∎

Lemma 7. *For $\kappa \geq 0$,*

$$\mathcal{E}_\kappa(w) = \frac{1}{2}\mu\left(\bigcap_{x \in X} I\left(f(x), \kappa d(x, w)\right)\right).$$

Proof.

$$\mathcal{E}_\kappa(w) = \frac{1}{2}\left[e_\kappa^+(w) - e_\kappa^-(w)\right]$$

$$= \frac{1}{2}\left\{\min_{x \in X}\{f(x) + \kappa d(x, w)\} - \max_{x \in X}\{f(x) - \kappa d(x, w)\}\right\}$$

$$= \frac{1}{2}\mu\left(\left[\max_{x \in X}\{f(x) - \kappa d(x, w)\}, \min_{x \in X}\{f(x) + \kappa d(x, w)\}\right]\right)$$

$$= \frac{1}{2}\mu\left(\bigcap_{x \in X} I\left(f(x), \kappa d(x, w)\right)\right).$$

The first equality follows from Proposition 1. ∎

Theorem 3. *For any* $\lambda \in \Re^+$, *if* $\mathcal{E}_{\hat{K}} \geq \lambda \hat{K}$, *then* $\mathcal{E}_K(\hat{f}) \geq \lambda K$.

Proof. Let $w^\star \equiv \arg\max_w \mathcal{E}_{\hat{K}}(w)$. Then

$$\mathcal{E}_K(\hat{f}) \geq \mathcal{E}_K(\hat{f}_K)$$
$$= \mathcal{E}_K(w^\star)$$

(A.2)
$$\geq \frac{K}{\hat{K}} \cdot \mathcal{E}_{\hat{K}}(w^\star)$$

(A.3)
$$\geq \frac{K}{\hat{K}} \cdot \lambda \hat{K}$$
$$= \lambda K.$$

Inequality (A.3) follows from (A.2) by hypothesis. Inequality (A.2) is a consequence of Lemma 6: Let $\alpha = \hat{K}/K \leq 1$. For $i = 1, \ldots, \#X$, let $\rho_i = f(x_i)$ and $\chi_i = Kd(x, w)$. Then, by Lemma 7, $\mu(I_1)/2 = \mathcal{E}_K$ and $\mu(I_\alpha)/2 = \mathcal{E}_{\hat{K}}$. ∎

Proposition 5. *Let* $\mathbf{0} \equiv (0, \ldots, 0)$, $\mathbf{1} \equiv (1, \ldots, 1)$, *and* $\tilde{d}(v) \equiv \max\left(d(v, \mathbf{0}), d(v, \mathbf{1})\right)$. *Then*

$$\mathcal{E}_{\hat{K}} \leq \frac{1}{2} \left\{ \min_{x \in X} \left[f(x) + \hat{K}\tilde{d}(x) \right] - \max_{x \in X} \left[f(x) - \hat{K}\tilde{d}(x) \right] \right\}.$$

Proof. Fix $w \in [0, 1]^p$. Let $w_{(i)}$ denote the ith component of w. Then

$$d(v, w) = \max_{i \in \{1, \ldots, p\}} \left| v_{(i)} - w_{(i)} \right|$$
$$\leq \max_{i \in \{1, \ldots, p\}} \max_{\delta \in \{0, 1\}} \left| v_{(i)} - \delta \right|$$
$$= \max_{i \in \{1, \ldots, p\}} \max_{y \in \{\mathbf{0}, \mathbf{1}\}} \left| v_{(i)} - y_{(i)} \right|$$
$$= \max_{y \in \{\mathbf{0}, \mathbf{1}\}} \max_{i \in \{1, \ldots, p\}} \left| v_{(i)} - y_{(i)} \right|$$
$$= \max_{y \in \{\mathbf{0}, \mathbf{1}\}} d(v, y)$$
$$= \max(d(v, \mathbf{0}), d(v, \mathbf{1})).$$

Hence,

(A.4)
$$\mathcal{E}_{\hat{K}}(w) = \frac{1}{2} \mu \left(\bigcap_{x \in X} I\left(f(x), \hat{K}d(x, w) \right) \right)$$

(A.5)
$$\leq \frac{1}{2} \mu \left(\bigcap_{x \in X} I\left(f(x), \hat{K}\tilde{d}(x) \right) \right)$$
$$= \frac{1}{2} \left\{ \min_{x \in X} \left[f(x) + \hat{K}\tilde{d}(x) \right] - \max_{x \in X} \left[f(x) - \hat{K}\tilde{d}(x) \right] \right\},$$

where (A.4) follows from Lemma 7. Because the right-hand side of this inequality does not depend on w, the proposition follows by taking suprema. ∎

Appendix B. The Lipschitz constant K for the borehole function. The borehole function is

$$f_0(H_u, H_\ell, T_u, T_\ell, r, r_w, L, K_w) = \frac{2\pi T_u\,(H_u - H_l)}{\log\,(r/r_w)\left(1 + \frac{2LT_u}{\log(r/r_w)r_w^2 K_w} + \frac{T_u}{T_\ell}\right)}.$$

The variables are restricted to the ranges in Table 2. We rescale f_0 so that its inputs range over the 8-dimensional unit hypercube $[0,1]^8$; the resulting function is denoted by f.

In ℓ_∞, because f is differentiable and $\mathrm{dom}(f)$ is convex,

$$K = \sup_{w \in \mathrm{dom}(f)} \|Df(w)\|_\infty = \sup_{w \in \mathrm{dom}(f)} \sum_{i=1}^{8} \left| \frac{\partial}{\partial w_i} f(w) \right|.$$

Let

$$H = 2\pi(H_u - H_\ell),$$
$$R = \log(r/r_w),$$
$$M = 2L/K_w,$$
$$t = T_\ell^{-1} + T_u^{-1},$$

and

$$S = M + R r_w^2 t.$$

Now

$$f_0(H_u, H_\ell, T_u, T_\ell, r, r_w, L, K_w) = \frac{H r_w^2}{S}.$$

We bound each partial derivative of f using the ranges of the input variables:

$$\left| \frac{\partial f_0}{\partial H_\ell} \right| = \left| \frac{\partial f_0}{\partial H_u} \right| = \frac{2\pi r_w^2}{S} \le 0.76 \implies \left| \frac{\partial f}{\partial H_\ell} \right| = \left| \frac{\partial f}{\partial H_u} \right| \le 91.2,$$

$$\left| \frac{\partial f_0}{\partial T_u} \right| = \frac{H R r_w^4}{S^2 T_u^2} \le 0.01 \implies \left| \frac{\partial f}{\partial T_u} \right| \le 0.01,$$

$$\left| \frac{\partial f_0}{\partial T_l} \right| = \frac{H R r_w^4}{S^2 T_\ell^2} \le 0.13 \implies \left| \frac{\partial f}{\partial T_l} \right| \le 6.8,$$

$$\left| \frac{\partial f_0}{\partial r} \right| = \frac{H r_w^4 t}{S^2 r} \le 0.01 \implies \left| \frac{\partial f}{\partial r} \right| \le 290.8,$$

$$\left| \frac{\partial f_0}{\partial r_w} \right| = \frac{H r_w^3 t}{S^2} + \frac{2H}{S(1/r_w + R r_w t/M)} \le 4050.2 \implies \left| \frac{\partial f}{\partial r_w} \right| \le 405.0,$$

$$\left| \frac{\partial f_0}{\partial L} \right| = \frac{2 H r_w^2}{S^2 K_w} \le 0.34 \implies \left| \frac{\partial f}{\partial L} \right| \le 190.4,$$

$$\left| \frac{\partial f_0}{\partial K_w} \right| = \frac{2 L H r_w^2}{S^2 K_w^2} \le 0.06 \implies \left| \frac{\partial f}{\partial K_w} \right| \le 123.7.$$

Summing these upper bounds for the partial derivatives of f yields

$$\sup_{w \in \mathrm{dom}(f)} \|Df(w)\|_\infty < 1200.$$

Moreover, for $w_0 = (1100, 700, 115547, 116, 100, 0.15, 1120, 12045)$,

$$\|Df(w_0)\|_\infty = 944.$$

Hence, for the rescaled borehole function f,

$$944 \le K \le 1200.$$

REFERENCES

[1] J. SACKS, W.J. WELCH, T.J. MITCHELL, AND H.P. WYNN, *Design and analysis of computer experiments*, Statist. Sci., 4 (1989), pp. 409–435.

[2] E.N. BEN-ARI AND D.M. STEINBERG, *Modeling data from computer experiments: An empirical comparison of Kriging with MARS and projection pursuit regression*, Qual. Eng., 19 (2007), pp. 327–338.

[3] R.G. GHANEM, A. DOOSTAN, AND J. RED-HORSE, *A probabilistic construction of model validation*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 2585–2595.

[4] S. SHAN AND G.G. WANG, *Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions*, Struct. Multidiscip. Optim., 41 (2010), pp. 219–241.

[5] C. COVEY, S. BRANDON, P.T. BREMER, D. DOMYANCIS, X. GARAIZAR, G. JOHANNESSON, R. KLEIN, S.A. KLEIN, D.D. LUCAS, J. TANNAHILL, AND Y. ZHANG, *A New Ensemble of Perturbed-Input-Parameter Simulations by the Community Atmosphere Model*, Technical report, Lawrence Livermore National Laboratory, Livermore, CA, 2011.

[6] D. ASPENBERG, J. JERGEUS, AND L. NILSSON, *Robust optimization of front members in a full frontal car impact*, Eng. Optim., 45 (2013), pp. 245–264.

[7] M. HOLENA, D. LINKE, AND U. RODEMERCK, *Generator approach to evolutionary optimization of catalysts and its integration with surrogate modeling*, Catalysis Today, 159 (2011), pp. 84–95.

[8] J.A. SHORTER, P.C. IP, AND H.A. RABITZ, *An efficient chemical kinetics solver using high dimensional model representation*, J. Phys. Chem. A, 103 (1999), pp. 7192–7198.

[9] A. SRIVASTAVA, K. HACKER, K. LEWIS, AND T.W. SIMPSON, *A method for using legacy data for metamodel-based design of large-scale systems*, Struct. Multidiscip. Optim., 28 (2004), pp. 146–155.

[10] P.N. KOCH, T.W. SIMPSON, AND J.K. ALLEN, *Statistical approximations for multidisciplinary design optimization: The problem of size*, J. Aircraft, 36 (1999), pp. 275–286.

[11] A.J. BOOKER, J.E. DENNIS, P.D. FRANK, D.B. SERAFINI, V. TORCZON, AND M.W. TROSSET, *A rigorous framework for optimization of expensive functions by surrogates*, Struct. Multidiscip. Optim., 17 (1999), pp. 1–13.

[12] R.A. BATES, R.J. BUCK, E. RICCOMAGNO, AND H.P. WYNN, *Experimental design and observation for large systems*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 77–111.

[13] E.W. PACKEL, *Do linear problems have linear optimal algorithms?*, SIAM Rev., 30 (1988), pp. 388–403.

[14] J. TRAUB AND H. WOŹNIAKOWSKI, *A General Theory of Optimal Algorithms*, Academic Press, New York, 1980.

[15] J.F. TRAUB, G.W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, *Information-Based Complexity*, Academic Press, Boston, 1988.

[16] S. SURJANOVIC AND D. BINGHAM, *Virtual Library of Simulation Experiments: Test Functions and Datasets*, http://www.sfu.ca/~ssurjano/emulat.html (accessed March 3, 2014).

[17] G.M. DANCIK, *mlegp: Maximum Likelihood Estimates of Gaussian Processes*, R package version 3.1.4, http://cran.r-project.org/package=mlegp (2013).

[18] J. BACKER AND J.M. KEIL, *The mono- and bichromatic empty rectangle and square problems in all dimensions*, in LATIN 2010: Theoretical Informatics, Springer-Verlag, Berlin, 2010, pp. 14–25.

[19] T. HASTIE AND R. TIBSHIRANI, *mda: Mixture and Flexible Discriminant Analysis*, R package version 0.4.4, http://cran.r-project.org/package=mda (2013).

[20] M. LAMBONI, B. IOOSS, A.L. POPELIN, AND F. GAMBOA, *Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests*, Math. Comput. Simulation, 87 (2013), pp. 45–54.

[21] F.J. HICKERNELL, *A simple multistart algorithm for global optimization*, OR Transactions, 1 (1997), pp. 1–12.