# Mini-Minimax Uncertainty Quantification for Emulators

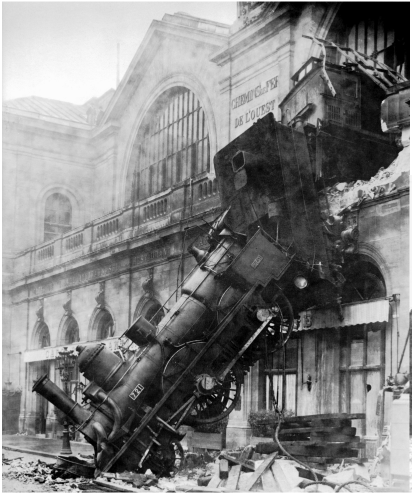
http://arxiv.org/abs/1303.3079

Philip B. Stark and Jeffrey C. Regier

Department of Statistics
University of California, Berkeley

2nd ISNPS Conference
Cadiz, Spain
13 June 2014

# Why Uncertainty Quantification Matters



??

James Bashford / AP

# Why Uncertainty Quantification Matters

NASA

Reuters / Japan TSB

## Emulators, Surrogate functions, Metamodels

- Can evaluate $f$ w/o noise.
- $f$ expensive to evaluate—experiment or big computation
- $f$ typically "black-box"
- Want "cheap" approximation of $f$ based on affordable number of samples.
- Emulators are essentially interpolators:
    - Kriging
    - Gaussian process models (GP)
    - Polynomial Chaos Expansions
    - Multivariate Adaptive Regression Splines (MARS)
    - Projection Pursuit Regression
    - Neural networks
    - etc.

## Noiseless non-parametric function estimation

- Estimate $f$ on domain $\mathrm{dom}(f)$ from $\{f(x_1), \ldots, f(x_n)\}$
- $f$ infinite-dimensional.
- $\mathrm{dom}(f)$ typically has dimension 5–100.
- Observe only $f|_X$, where $X = \{x_1, \ldots, x_n\}$. No noise.
- Estimating $f$ is grossly underdetermined problem (worse with noise).
- Usual context: A question that requires knowing $f(x)$ for $x \notin X$

## Common context

Part of larger problem in uncertainty quantification (UQ)

- Real-world phenomenon
- Physics description of phenomenon
- Theoretical simplification/approximation of the physics
- $f$ is the numerical solution of the approximation
- Emulation of the numerical solution of the approximation $\hat{f}$
- Calibration to noisy data
- "Inference"

High-consequence decisions are made on the basis of $\hat{f}$.
How well does $\hat{f}$ approximate $f$? The real world?

# Common strategies to estimate accuracy

## Bayesian Emulators (GP, Kriging, . . . )

- Use the posterior distribution (Tebaldi & Smith 2005)
- Posterior depends on prior and likelihood, but inputs are generally fixed parameters, not random.

## Others

- Using holdout data (Fang et al. 2006)
- Relevant only if the error at the held-out data is representative of the error everywhere.
  Data not usually IID; values of $f$ not IID.

Required conditions generally unverifiable or demonstrably false.

## Need constraints to say anything

- In rare cases, "physics" provides constraints, but generally, uncertainty estimates are driven by *assumptions* about $f$.
- Absent some regularity, no reliable way to extrapolate data to values of $f$ at unobserved inputs: completely uncertain.
- Stronger assumptions $\rightarrow$ smaller apparent uncertainties.
- What's the most optimistic assumption the data don't contradict?

### (Best) Lipschitz constant

Given a metric $d$ on $\mathrm{dom}(g)$, *best Lipschitz constant* $K$ *for* $g$ is

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

## How bad *must* the uncertainty be?

- Data $f|_X$ impose a lower bound on $K(f)$ (but no upper bound): Data *require* some lack of regularity.
- Intentional optimism: assume $f$ is as regular as possible while fitting the data
- Is there any $\hat{f}$ guaranteed to be close to $f$—no matter what $f$ is—provided $f$ fits the data and is that regular?

## Minimax formulation: Information-Based Complexity (IBC)

- $\mathcal{F}_{\kappa,Y}$: functions $g$ s.t. $\text{Lip}(g) \leq \kappa$ and $g|_Y = f|_Y$.
- *uncertainty at $w$ of $\hat{f}$ over $\mathcal{F}_{\kappa,Y}$:*

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

- *minimax uncertainty at $w$:*

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

- *maximum uncertainty of $\hat{f}$:*

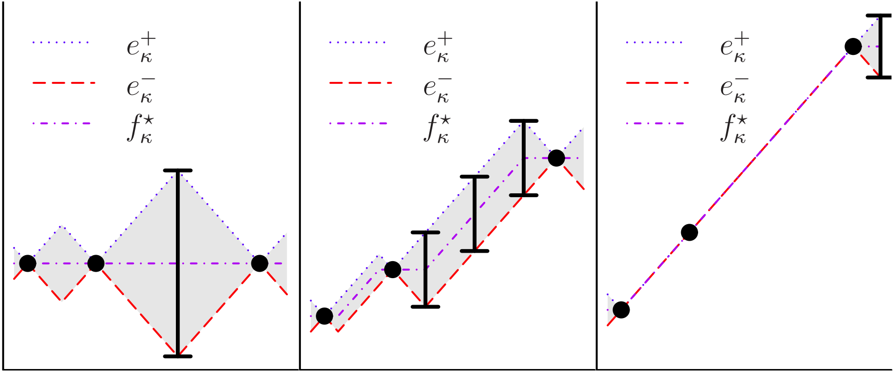
$$\mathcal{E}_{\kappa,Y}(\hat{f}) \equiv \sup_{w \in [0,1]^p} \mathcal{E}_{\kappa,Y}(w;\hat{f}) = \sup_{g \in \mathcal{F}_{\kappa,Y}} \|\hat{f} - g\|_\infty.$$

- *minimax uncertainty:*

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

## Pointwise minimax emulator and its uncertainty

- $e^+_\kappa(w) \equiv \min_{x \in X} [f(x) + \kappa d(x, w)]$
- $e^-_\kappa(w) \equiv \max_{x \in X} [f(x) - \kappa d(x, w)]$.
- $\mathcal{E}_{\kappa,X}(w) = e^\star_\kappa(w) \equiv \frac{e^-_\kappa(w) - e^+_\kappa(w)}{2}$ (theorem).
- If $\mathrm{Lip}(f) = \kappa$, $\hat{f}_\kappa(w) \equiv \frac{e^-_\kappa(w) + e^+_\kappa(w)}{2}$ is minimax (theorem).
- $e^*_\kappa$, $\hat{f}_\kappa(w)$ are computable from $f|_X$.

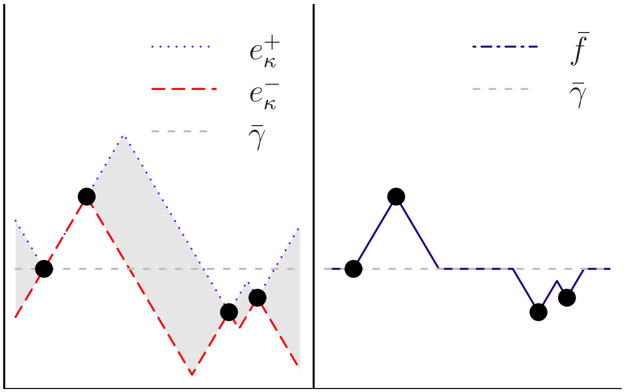



Black error bars are double $\sup_w e_\kappa^\star(w)$. As the slope between observations approaches $\kappa$, $e^\star(w)$ approaches 0 for points $w$ between observations, and $\sup_w e_\kappa^\star(w)$ decreases

## Lower bounds on computational burden

- Construct $\bar{f}$ that agrees with $f|_X$, has $\mathrm{Lip}(\bar{f}) = \hat{K}$, and requires $M_\epsilon$ additional observations $f|_Y$ to approximate within $\epsilon$ on $[0,1]^p$.
- Since $f$ *could be* $\bar{f}$, this gives a lower bound on the number of additional observations that might be required to approximate $f$ well, even if $f$ is not rougher than original data $f|_X$ require it to be.
- $\bar{f}$ is constant "as much as possible" while fitting the data and having $\mathrm{Lip}(\bar{f}) \leq \hat{K}$
- $\bar{\gamma} \equiv \arg\min_{\gamma \in \mathbb{R}} \sum_{x \in X} |f(x) - \gamma|^p$

# $\bar{f}$ is constant "as much as possible"



## Potential computational burden

- $C_q$: volume of $p$-dimensional unit ball in $q$ norm: $C_2 \equiv \frac{\pi^{p/2}}{\Gamma(p/2+1)}$ and $C_\infty \equiv 2^p$.
- $M_\epsilon$: observations potentially required to emulate $f$ within $\epsilon$.

$$M_\epsilon \quad \geq \quad \left\lceil \epsilon^{-p} \left[ \frac{\hat{K}^p}{C_q} - \sum_{x \in X} |f(x) - \bar{\gamma}|^p \right] \right\rceil . \tag{1}$$

## Uncertainty Quantification Strategic Initiative–LLNL

- Uncertainty Quantification Strategic Initiative at LLNL: 1154 climate simulations using the Community Atmosphere Model (CAM).
- $p = 21$ parameters scaled so that $[0, 1]$ has all plausible values.
- $f$ is global average upwelling longwave flux (FLUT) approximately 50 years in the future.
- Each run took several days on a supercomputer.
- Several approaches to choose $X \subset [0, 1]^p$: Latin hypercube, one-at-a-time, and random-walk multiple-one-at-a-time.
- 1154 simulations total.

## CAM calculations

- $\bar{\gamma} = 232.77$
- For $q = 2$, $\hat{K} = 14.20$:
  $M \geq \left\lceil \epsilon^{-21} \left[ \frac{1.57\times10^{24}}{0.0038} - 6.81 \times 10^{24} \right] \right\rceil > \epsilon^{-21} \times 10^{26}$
  If $\epsilon$ is 1% of $\hat{K}$, then $M \geq 10^{43}$.
  Even if $\epsilon$ is 50% of $\hat{K}$, $M > 10^{8}$.
- For $q = \infty$, $\hat{K} = 34.68$:
  $M \geq \left\lceil \epsilon^{-21} \left[ \frac{2.19\times10^{32}}{2^{21}} - 6.81 \times 10^{25} \right] \right\rceil > \epsilon^{-21} \times 10^{25}$

## More isn't necessarily better

If $\mathcal{E}_{\hat{K}} \geq \hat{K}/2$, then

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

No $\hat{f}$ based on $f|_X$ has smaller maximum potential error than the constant emulator based on one observation at the centroid $z$ of $[0,1]^p$

## Implications for CAM

- $\sup e^{\star}_{\hat{K}} = 20.95 \geq 17.34 = \hat{K}/2$
- Hence, $\mathcal{E}_K(\hat{f}) \geq K/2$ for every emulator $\hat{f}$.
- Maximum potential error would have been no greater had we just observed $f$ at $z$ and emulated by $\hat{f}(w) = f(z)$ for all $w \in [0,1]^p$.

## Extensions

- Can estimate the measure of $\{w : e^*_\kappa(w) \geq \epsilon > 0\}$ by sampling.
- Draw points $w \in [0, 1]^p$ at random; evaluate $e^*$ at each $w$—cheap.
- Yields binomial lower confidence bounds for the fraction of $[0, 1]^p$ where uncertainty is large, and confidence bounds for quantiles of the potential error.

# CAM: bounds on percentiles of error

| | | 95% lower confidence bound | | | |
|---|---|---|---|---|---|
| norm | units | lower quartile | median | upper quartile | average |
| 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 |

Error of minimax emulator $f^{\star}_{\hat{K}}$ of CAM model from 1154 LLNL observations. Col 1: metric $d$ used to define $K$. Cols 3–5: binomial lower confidence bounds for quartiles of the pointwise error, obtained by inverting binomial tests.

Col 6: 95% lower confidence bound for integral of the pointwise error over $[0,1]^p$, based on inverting a $z$-test.

Cols 3–6 are expressed as a fraction of the quantity in col 2. Based on 10,000 random samples.

# Computational Burden for “typical value”

| norm | $\epsilon$ | lower bound on $M$ |
|---|---|---|
| Euclidean | $0.02\hat{\gamma}$ | $3.6 \times 10^{12}$ |
| | $0.04\hat{\gamma}$ | 1,720,354 |
| | $0.06\hat{\gamma}$ | 345 |
| | $0.08\hat{\gamma}$ | 1 |
| supremum | $0.02\hat{\gamma}$ | $8.6 \times 10^{10}$ |
| | $0.04\hat{\gamma}$ | 413,595 |
| | $0.06\hat{\gamma}$ | 83 |
| | $0.08\hat{\gamma}$ | 1 |

## Conclusions

- In some problems, *every* emulator based on any tractable number of observations of $f$ has large potential error over much of its domain, even if $f$ is no less regular than the data *require*.
- Can find sufficient conditions under which all emulators are have large minimax error over much of their domain, even if $f$ is no less regular than the data *require*.
- Conditions depend only on the data; can be computed from the same data used to train emulator, at small incremental cost.
- Conditions hold for some problems of societal interest.

## Directions

- Reducing uncertainty in HEB problems requires knowing more about $f$ or changing the question.
- Both tactics application-specific: the science dictates what constraints $f$ satisfies and the senses in which it is useful to approximate $f$.
- Not clear that simulation and emulators help address the most important questions.
- Approximating $f$ pointwise rarely ultimate goal; most properties of $f$ are nuisance parameters.
- Important questions about $f$ might be answered more directly.
- Heroic simulations and emulators may be distractions.