Deep Gaussian processes for estimation of failure probabilities in complex systems

Annie (Sauer) Booth
with Robert B. Gramacy and Ashwin Renganathan

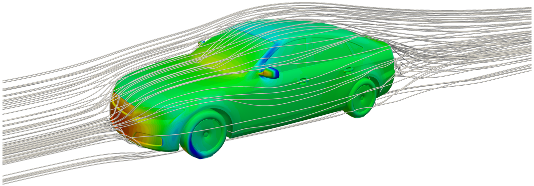
Virginia Tech Department of Statistics

2025

Problem: failure probability of complex systems

Consider an expensive "blackbox" computer experiment

$$f : \mathcal{X} \to \mathbb{R}$$



Failures occur when $f(\mathbf{x}) > t$.

Inputs are governed by a known distribution, $\mathbf{x} \sim p(\mathbf{x})$.

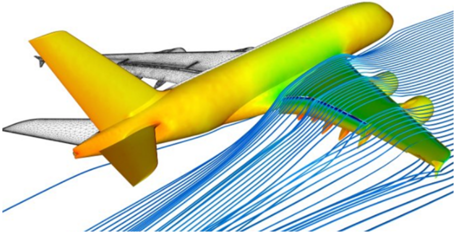
Our goal is to quantify the **failure probability**:

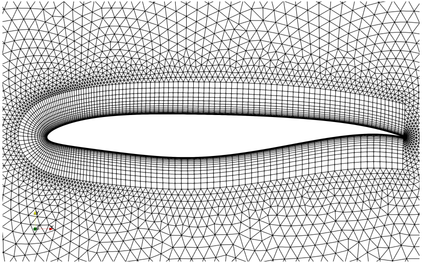
$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\, d\mathbf{x}}$$

Motivation: RAE-2822 transonic airfoil efficiency

Consider a simulation of air flow around an RAE-2822 transonic airfoil.

- Solved via SU2: https://su2code.github.io (Economon et al., 2016)





Motivation: RAE-2822 transonic airfoil efficiency

Consider a simulation of air flow around an RAE-2822 transonic airfoil.

- Solved via SU2: https://su2code.github.io (Economon et al., 2016)
- **7 inputs:** 4 wing shape parameters, angle, speed, Reynolds number
- **1 output:** lift/drag ratio (L/D)

L/D ratios less than 3 are considered failures.

Objective: to quantify the probability of system failure

$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\, d\mathbf{x}}$$

A straightforward solution? ... it's never that easy

Objective: to quantify the probability of system failure

$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\, d\mathbf{x}}$$

A true **Monte Carlo** approach estimates the integral as

$$\hat{\alpha}_{\mathrm{MC}} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\{f(\mathbf{x}_i)>t\}} \quad \text{for} \quad \mathbf{x}_i \overset{\text{iid}}{\sim} p(\mathbf{x})$$

A straightforward solution? ... it's never that easy

Objective: to quantify the probability of system failure

$$\boxed{\alpha = \int_{\mathbf{x} \in \mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x}) \, d\mathbf{x}}$$

A true **Monte Carlo** approach estimates the integral as

$$\hat{\alpha}_{\mathrm{MC}} = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{\{f(\mathbf{x}_i)>t\}} \quad \text{for} \quad \mathbf{x}_i \overset{\text{iid}}{\sim} p(\mathbf{x})$$

Challenge: the computer simulation is complex and **expensive**

A straightforward solution? ... it's never that easy

Objective: to quantify the probability of system failure

$$\boxed{\alpha = \int_{\mathbf{x} \in \mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x}) \, d\mathbf{x}}$$

A true **Monte Carlo** approach estimates the integral as

$$\hat{\alpha}_{\mathrm{MC}} = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{\{f(\mathbf{x}_i)>t\}} \quad \text{for} \quad \mathbf{x}_i \overset{\text{iid}}{\sim} p(\mathbf{x})$$

Challenge: the computer simulation is complex and **expensive**

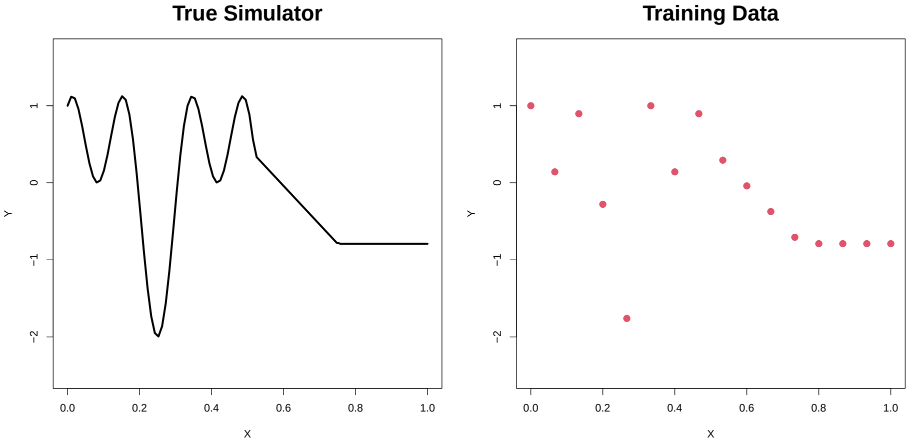
Solution (?): use a surrogate

$$\hat{\alpha}_{\mathrm{SURR}} = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}_{\{\hat{f}(\mathbf{x}_i)>t\}} \quad \text{for} \quad \mathbf{x}_i \overset{\text{iid}}{\sim} p(\mathbf{x})$$

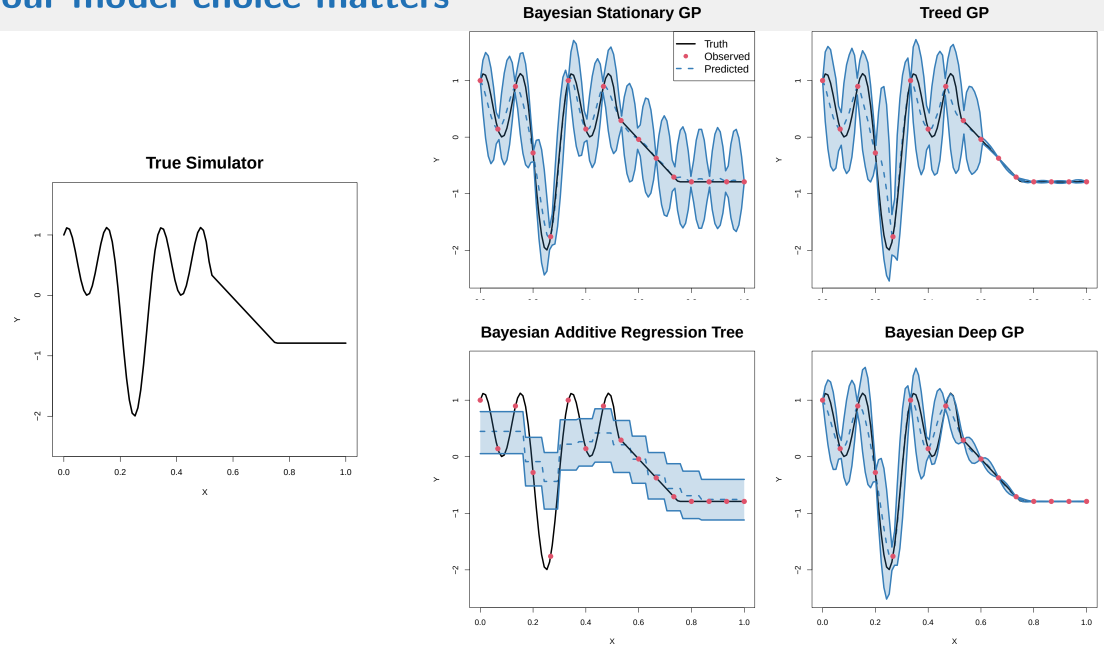
I probably don't need to tell you... but I will anyways

There are **two** essential pieces of a surrogate:

- A statistical model
- An experimental design

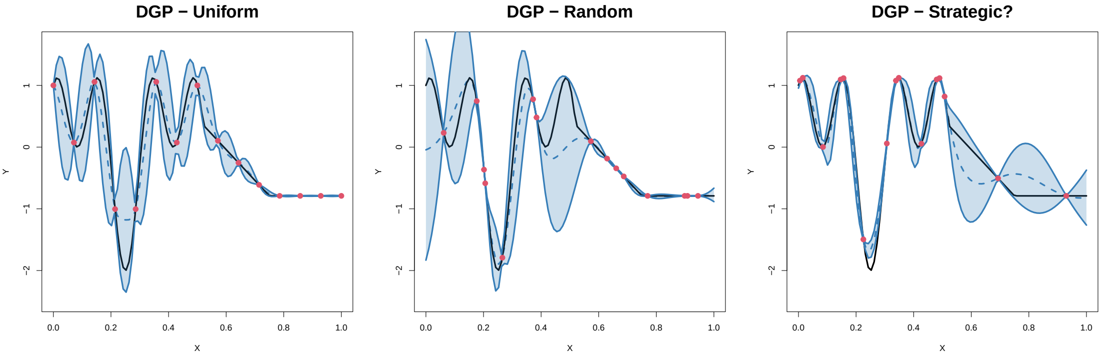


Your model choice matters



R packages (in order): `deepgp`, `tgp`, `BART`, `deepgp`

Your design choice matters



Motivation: RAE-2822 transonic airfoil efficiency

Consider a simulation of air flow around an RAE-2822 transonic airfoil.

- Solved via SU2: https://su2code.github.io (Economon et al., 2016)
- **7 inputs:** 4 wing shape parameters, angle, speed, Reynolds number
- **1 output:** lift/drag ratio (L/D)

L/D ratios less than 3 are considered failures.

Objective: to quantify the probability of system failure

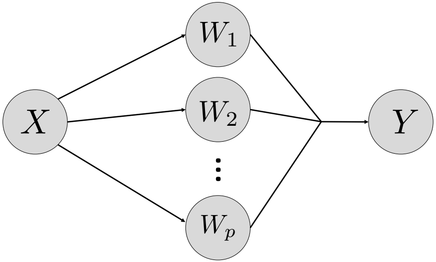
$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\, d\mathbf{x}}$$

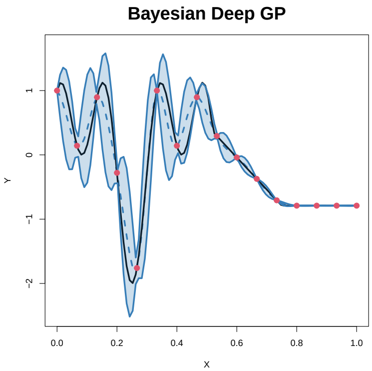
What model should we use? What design should we use?

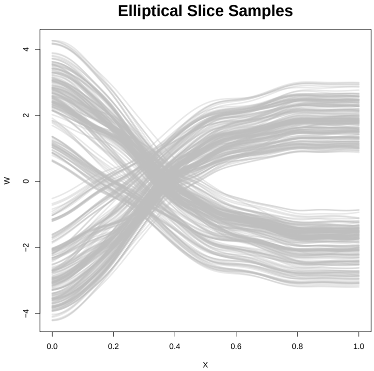
Ashwin thinks there will be **nonstationarity** in the L/D surface...

Deep GPs offer nonstationary flexibility while preserving UQ

- DGPs are functional compositions of GPs
- Latent layers act as warped inputs
- But they require MCMC sampling
  - Elliptical slice sampling
  - Murray et al. (2010)
- Wrapped in deepgp R-package



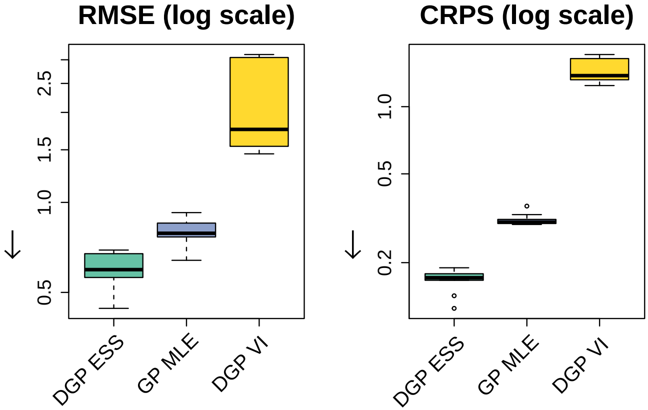




Motivation: RAE-2822 transonic airfoil efficiency

How well can we predict L/D?

- 500-point Latin hypercube sample for training
- 4,500-point hold-out testing set



Variational inference (VI) DGP using `gpytorch` (Gardnew et al., 2018)

Motivation: RAE-2822 transonic airfoil efficiency

Recall...

$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\, d\mathbf{x}} \qquad \hat{\alpha}_{\text{SURR}} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\{\hat{f}(\mathbf{x}_i)>t\}} \quad \text{for} \quad \mathbf{x}_i \overset{\text{iid}}{\sim} p(\mathbf{x})$$

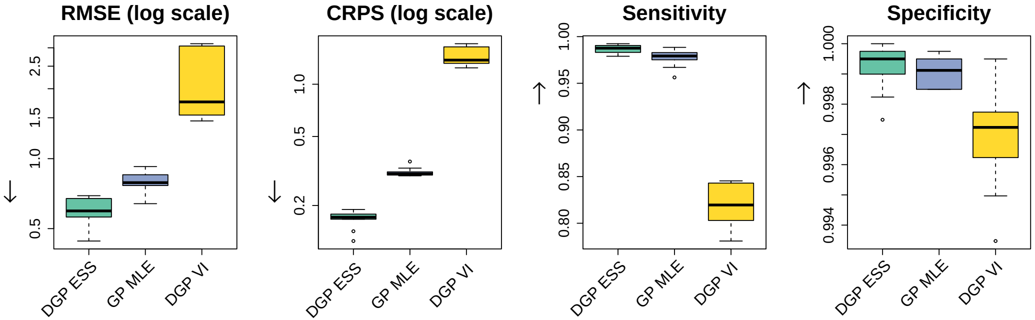
The primary purpose of our surrogate is to *predict passes/failures*...

Motivation: RAE-2822 transonic airfoil efficiency

Recall...

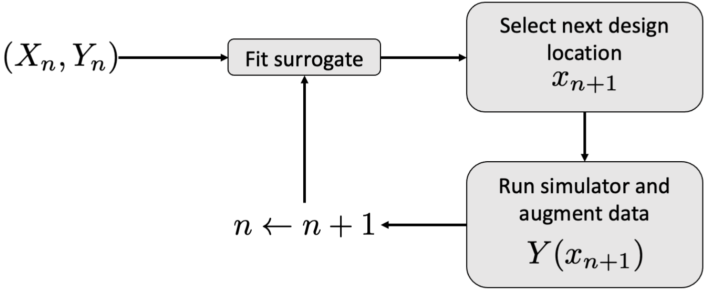
$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\ d\mathbf{x}} \qquad \hat{\alpha}_{\text{SURR}} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\{\hat{f}(\mathbf{x}_i)>t\}} \quad \text{for} \quad \mathbf{x}_i \overset{\text{iid}}{\sim} p(\mathbf{x})$$

The primary purpose of our surrogate is to *predict passes/failures*...



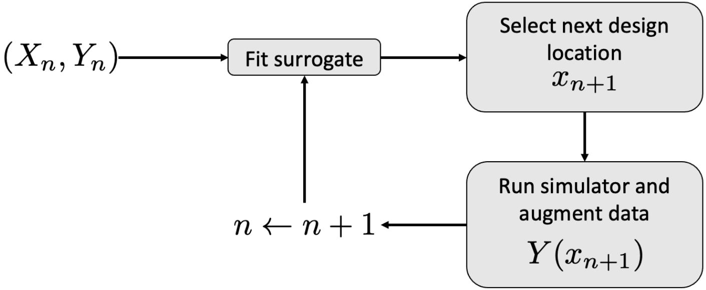
We can do better with a sequential design

Contour Location: strategic sequential design targeting a contour/level set



We can do better with a sequential design

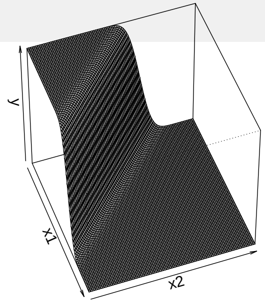
Contour Location: strategic sequential design targeting a contour/level set

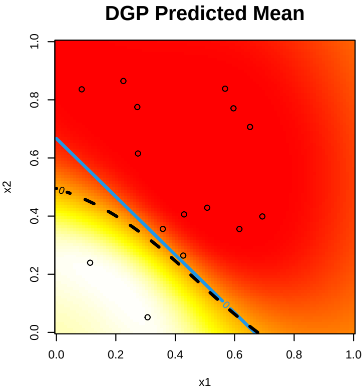


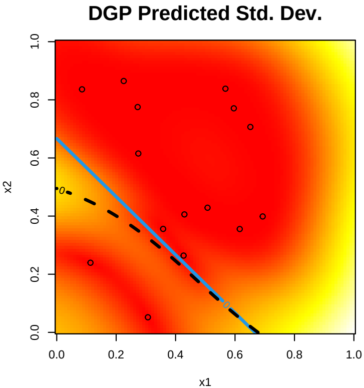
x_{n+1} is chosen by optimizing an **acquisition function**
We need an acquisition scheme that effectively targets the failure contour...

An illustrative example

- Plateau function in 2d
- Two-layer DGP surrogate fit to an initial random design of size $n = 15$

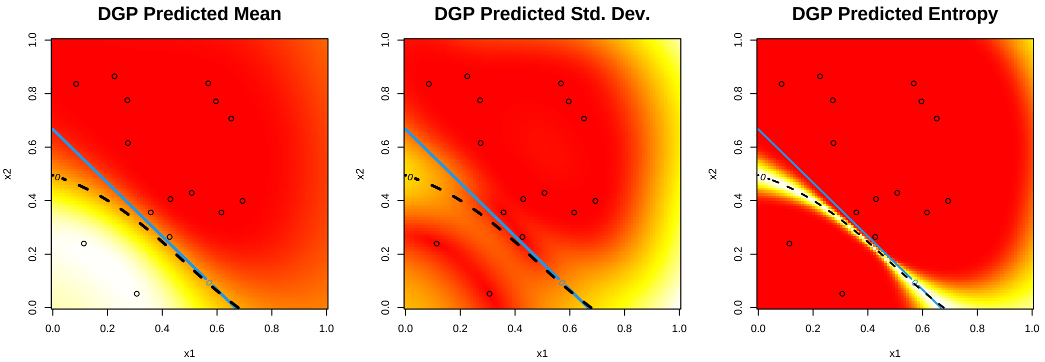






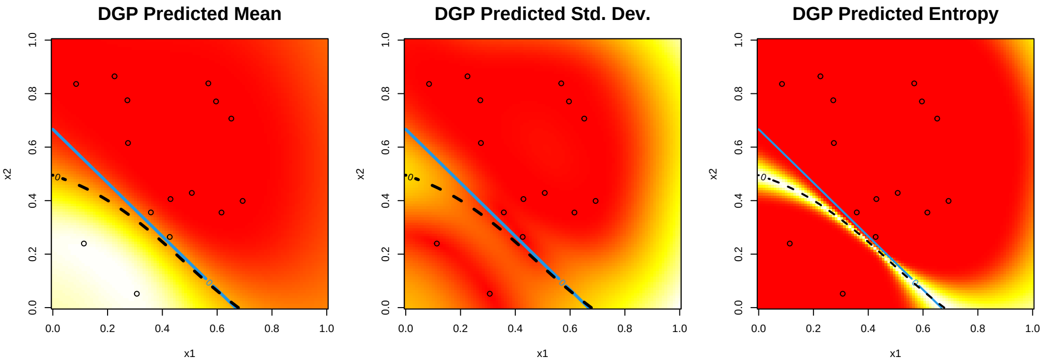
An illustrative example

State-of-the-art methods rely on numerical optimization of the entropy surface, or variations thereof (Bryan et al., 2005; Marques et al., 2018; Cole et al., 2022).



An illustrative example

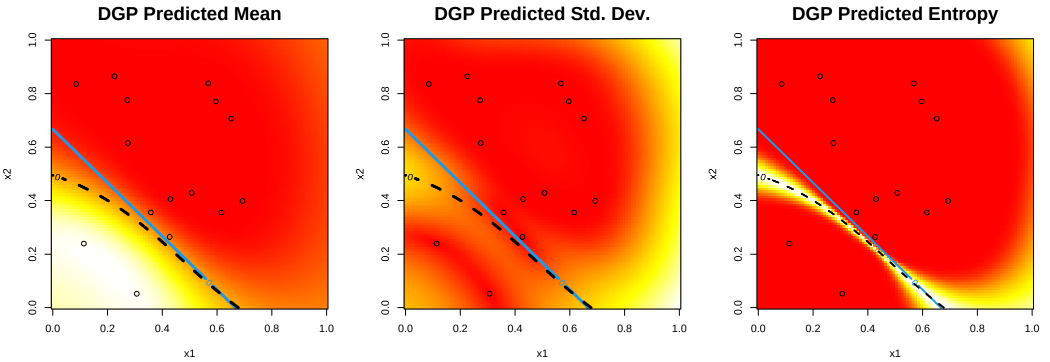
State-of-the-art methods rely on numerical optimization of the entropy surface, or variations thereof (Bryan et al., 2005; Marques et al., 2018; Cole et al., 2022).



- But numerical optimizations are hefty

An illustrative example

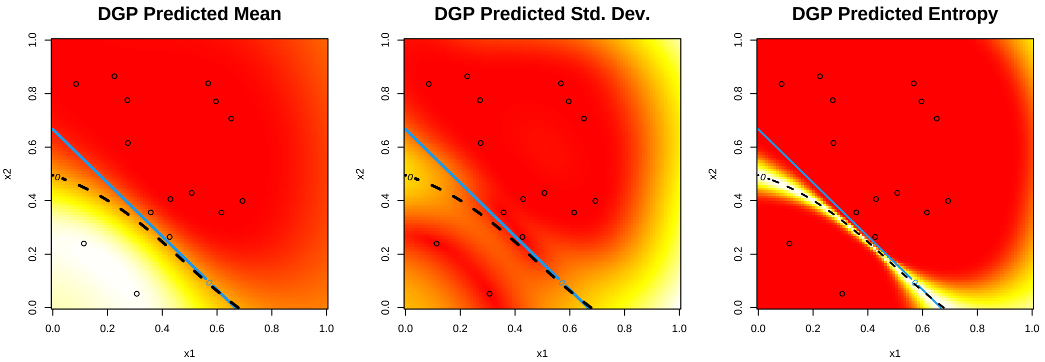
State-of-the-art methods rely on numerical optimization of the entropy surface, or variations thereof (Bryan et al., 2005; Marques et al., 2018; Cole et al., 2022).



- But numerical optimizations are hefty → **strategic candidates**

An illustrative example

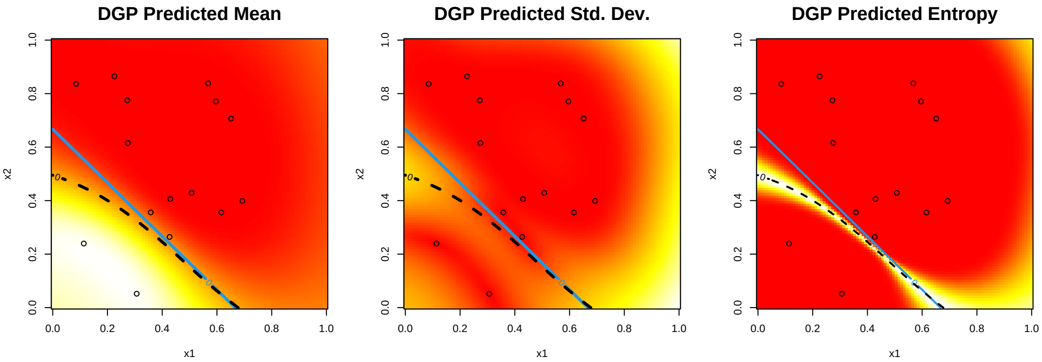
State-of-the-art methods rely on numerical optimization of the entropy surface, or variations thereof (Bryan et al., 2005; Marques et al., 2018; Cole et al., 2022).



- But numerical optimizations are hefty $\rightarrow$ **strategic candidates**
- And entropy does not encourage exploration

An illustrative example

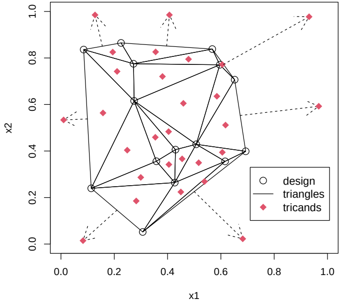
State-of-the-art methods rely on numerical optimization of the entropy surface, or variations thereof (Bryan et al., 2005; Marques et al., 2018; Cole et al., 2022).



- But numerical optimizations are hefty $\rightarrow$ **strategic candidates**
- And entropy does not encourage exploration $\rightarrow$ **Pareto front with uncertainty**

Candidates on the Pareto front of entropy and uncertainty

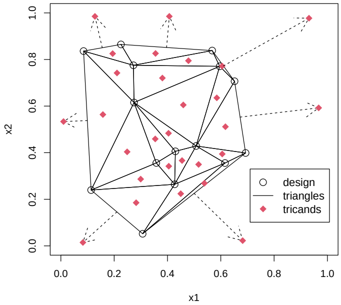
1. Propose **triangulation candidates**

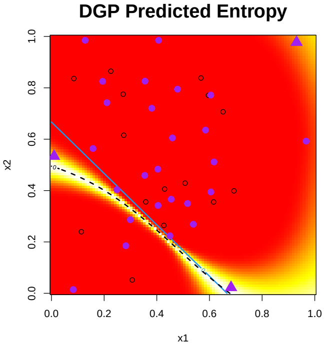


(Gramacy, Sauer, & Wycoff, 2022; Booth, Renganathan, & Gramacy, 2023)

Candidates on the Pareto front of entropy and uncertainty

1. Propose **triangulation candidates**
2. Calculate **entropy** and **uncertainty**

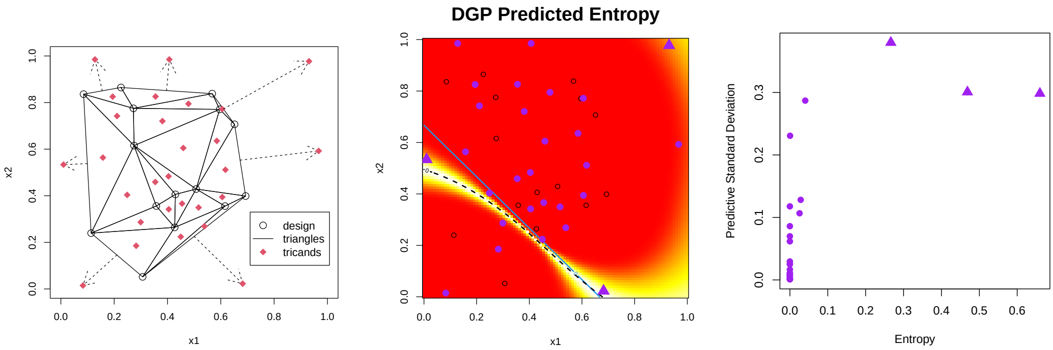




(Gramacy, Sauer, & Wycoff, 2022; Booth, Renganathan, & Gramacy, 2023)

Candidates on the Pareto front of entropy and uncertainty

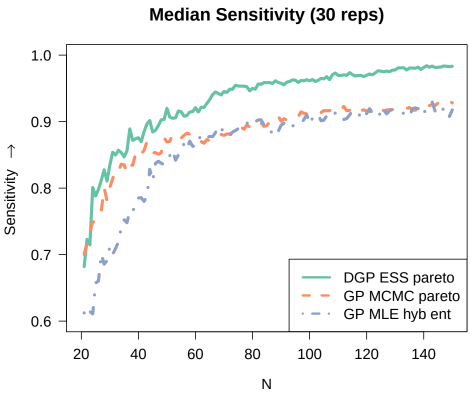
1. Propose **triangulation candidates**
2. Calculate **entropy** and **uncertainty**
3. Choose candidate on the Pareto front

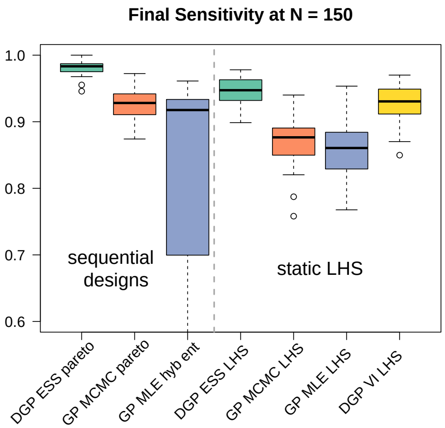


(Gramacy, Sauer, & Wycoff, 2022; Booth, Renganathan, & Gramacy, 2023)

The "secret sauce": DGP + tricands + Pareto front

Plateau function bumped up to $d = 5$

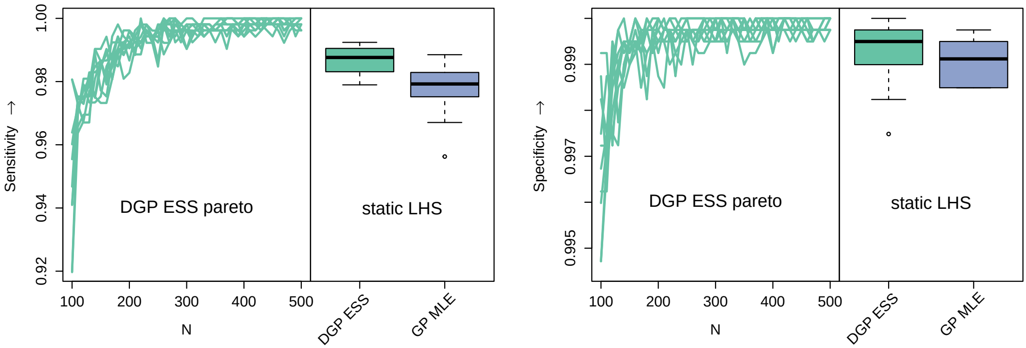




"hyb ent" uses strategic local optimization of entropy (Cole et al., 2022)

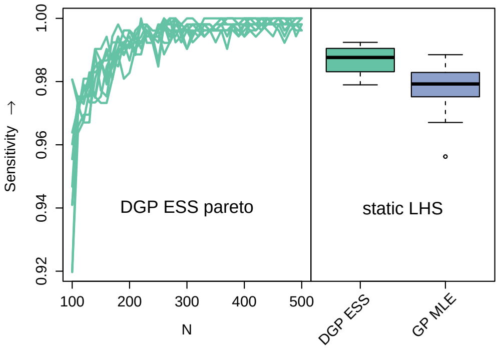
Motivation: RAE-2822 transonic airfoil efficiency

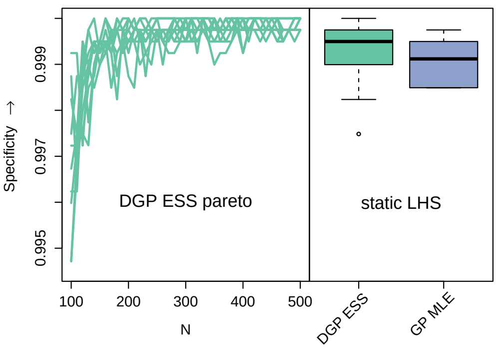
- 100-point initial LHS
- 400 acquisitions using DGP + tricands + Pareto
- Compared to static LHS of size 500



Motivation: RAE-2822 transonic airfoil efficiency

- 100-point initial LHS
- 400 acquisitions using DGP + tricands + Pareto
- Compared to static LHS of size 500





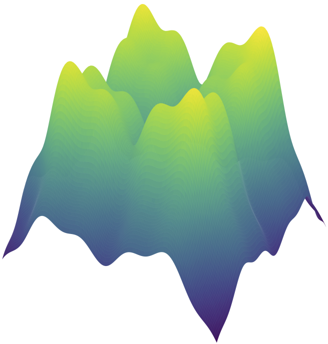
But what if $N = 500$ exhausts our *entire* budget...

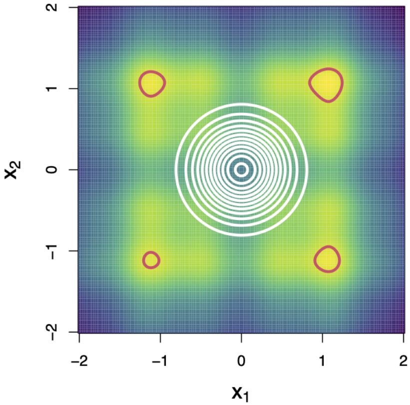
...should we spend it *all* on training the surrogate?

A new illustrative example

Recall...

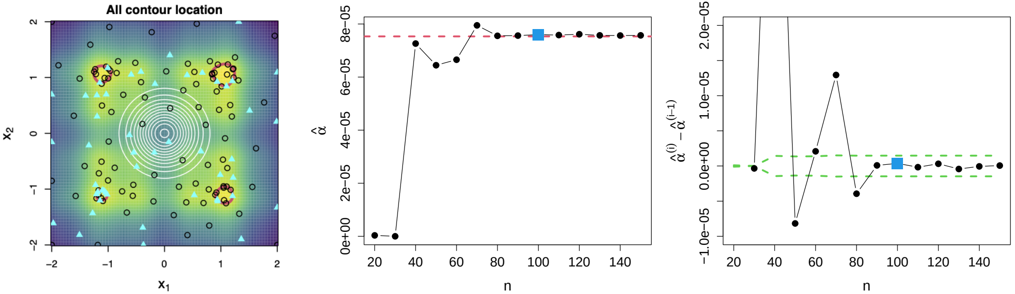
$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\, d\mathbf{x}} \qquad \hat{\alpha}_{\text{SURR}} = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\{\hat{f}(\mathbf{x}_i)>t\}} \quad \text{for} \quad \mathbf{x}_i \overset{\text{iid}}{\sim} p(\mathbf{x})$$





A new illustrative example

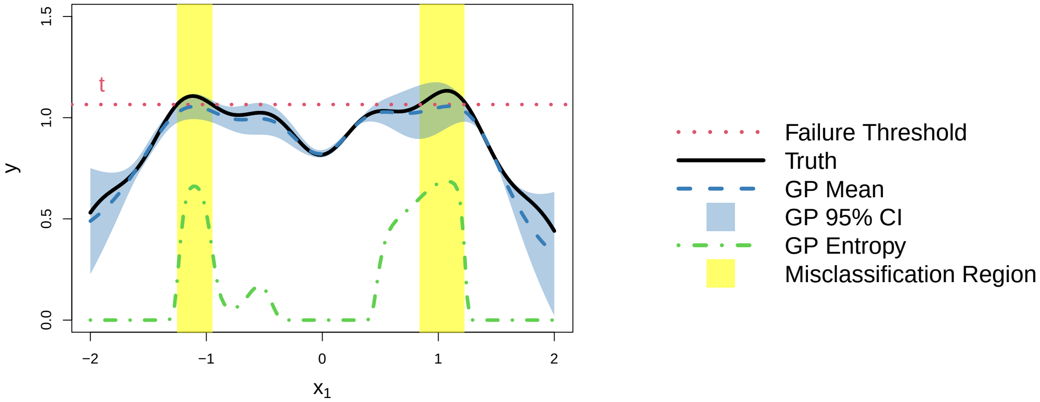
As we train our surrogate with contour location, the learning plateaus...



We propose a stopping rule based on the standard error of this estimate...
...then what do we do with our remaining budget?

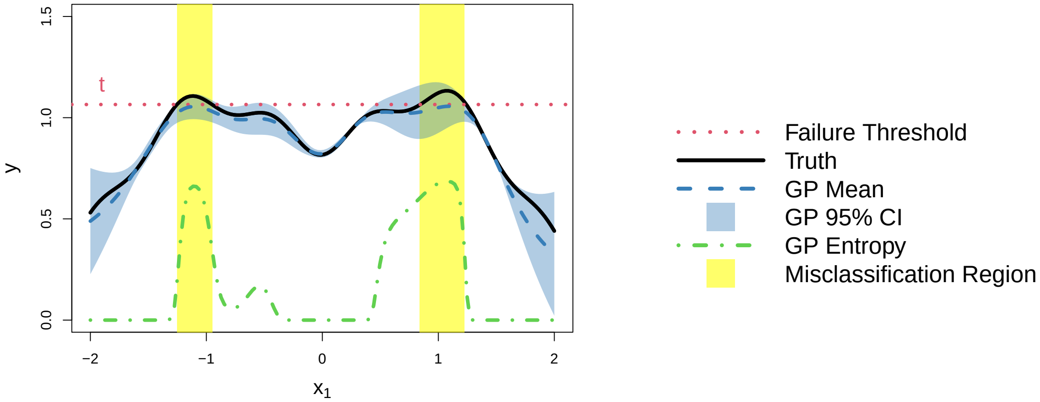
Two-stage design: what to do with our remaining budget?

Herbie function slice at $x_2 = 1.1$



Two-stage design: what to do with our remaining budget?

Herbie function slice at $x_2 = 1.1$

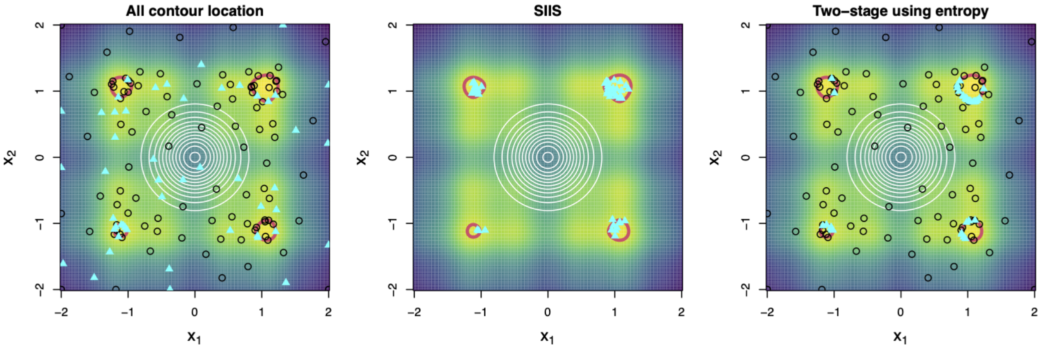


The misclassification region aligns with the peaks in entropy...

... we propose using the remaining budget on the highest entropy samples.

Two-stage design: CL + highest entropy samples

1. Perform contour location until learning has saturated
2. Use remaining budget on highest entropy samples (ensuring correct classification)

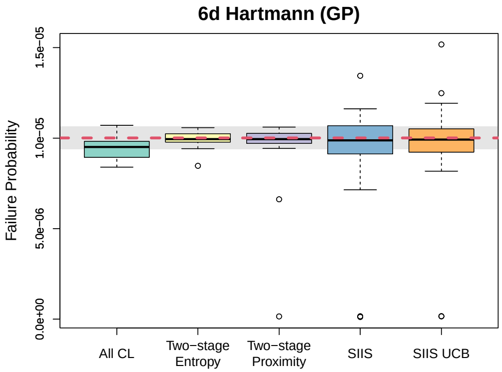


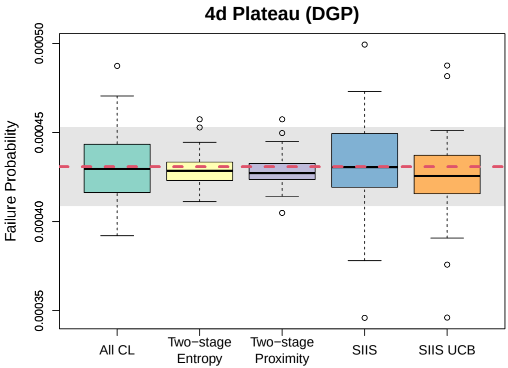
With the same surrogate and starting design, this outperforms

- Exhaustive contour location
- Surrogate-informed importance sampling (Peherstorfer et al., 2016)

Two-stage design outperforms exhaustive CL and IS

Function	True α	d	n	M	Surrogate	CL Scheme
Hartmann	1.001×10^{-5}	6	600	1.0×10^{8}	GP	Cole et al. (2023)
Plateau	8.530×10^{-4}	4	200	3.5×10^{7}	DGP	Booth et al. (2024)

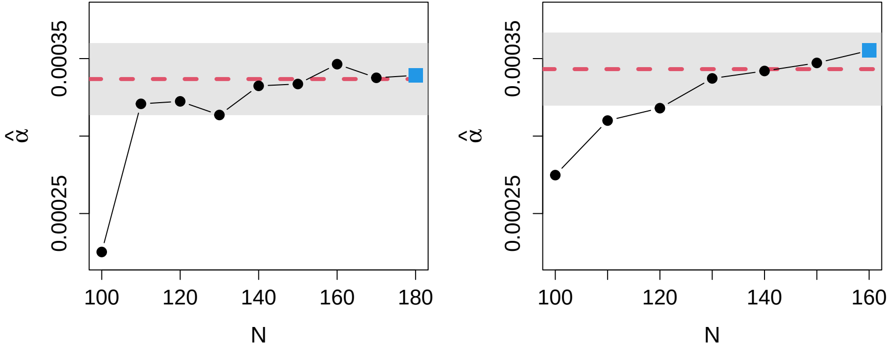




Motivation: RAE-2822 transonic airfoil simulation

$$\boxed{\alpha = \int_{\mathbf{x}\in\mathcal{X}} \mathbb{1}_{\{f(\mathbf{x})>t\}} p(\mathbf{x})\, d\mathbf{x}}$$

- Total budget of 500, $M = 2.5 \times 10^6$
- DGP with two-stage design (Pareto + tricands for contour location)
- Two repetitions with re-randomized starting designs
- Final estimate shown in red



Thanks!

Takeaways:

- Models matter, but they are intertwined with design
- A good surrogate with a strategic sequential design is hard to beat
- For failure probability estimation, a two-stage design is key
    1. Balance exploration and exploitation to learn the contour
    2. Greedily exploit to make sure uncertain Monte Carlo samples are classified correctly

Everything you saw today is supported by

- `deepgp` for R on CRAN (Booth, 2024)
- and two git repositories of examples:

`https://bitbucket.org/gramacylab/deepgp-ex/`
`https://bitbucket.org/boothlab/failprob/`