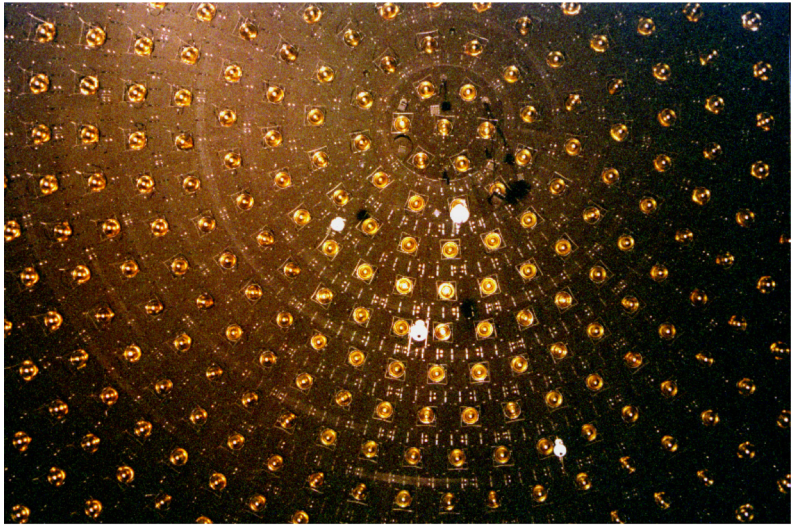
Detecting Bias:

How Rare is Rare?

Theo Faridani, University of Oregon - Department of Physics

Background: Rare Events

- What is an event?
  - A distinct point in time
  - An associated tuple of information
- What makes an event “rare”?
  - Hard to detect but actually common
  - Intrinsically rare
- Reality is convolved with Observing Profile



CCBias: A First-Use Tool

- Three main features
  - Given information about events, generates sample data
    - Convolves that with a given Observing Profile
    - Relatively general, given our definition
  - Given information about events, optimizes Observing Profile parameters
    - Requires a scoring function
  - Given observed data, estimates what bias is present
    - Requires certain functional properties

Use Case: LSST

- Primary goals include detecting events whose brightness vary significantly over time
    - Long observing times
- Telescope time is finite: want the most out of it

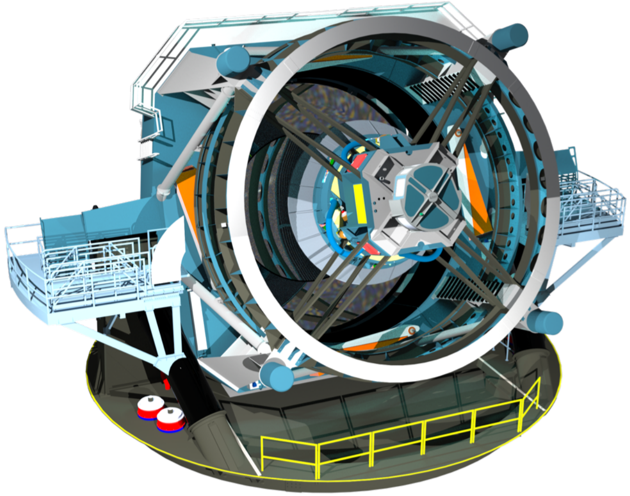
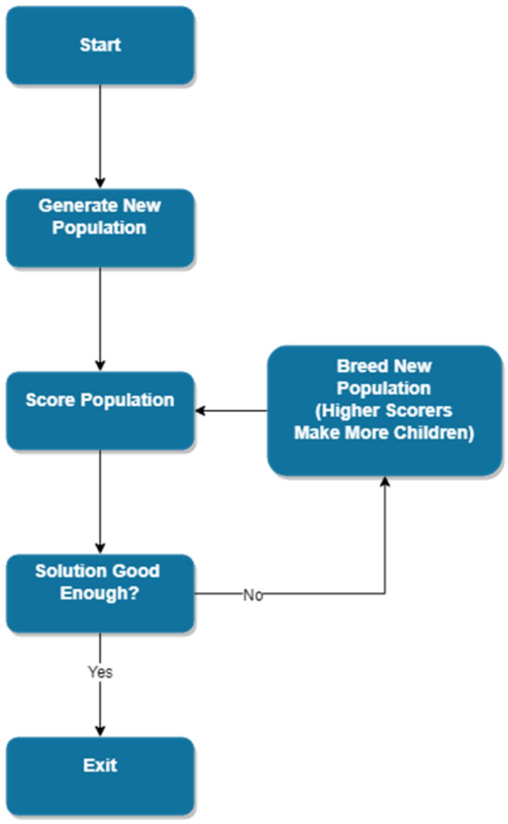


Image Credit: LSST

Survey Strategy Optimization

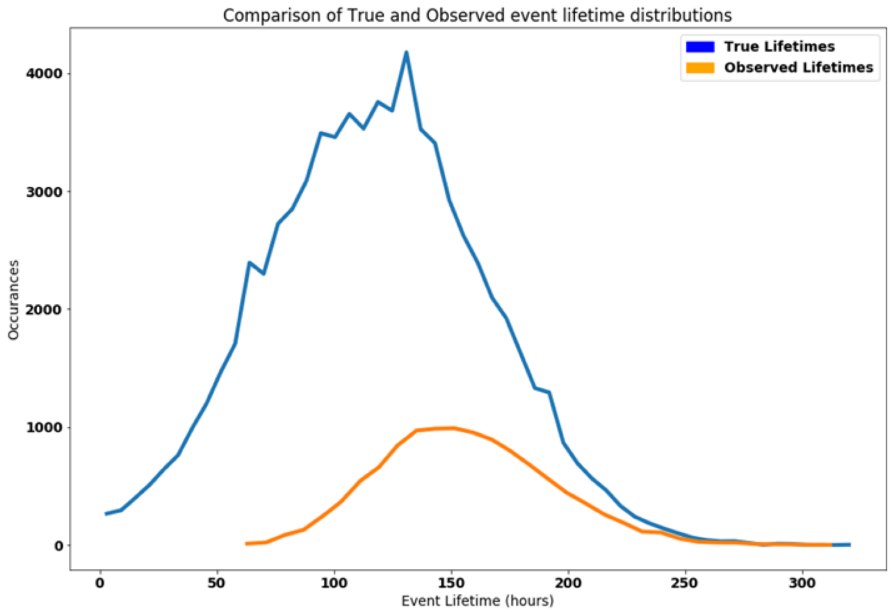
- Motivation: get something for free if we can
- Computationally Cheap
    - Means we can be wasteful in our algorithm choice
- CCBias uses a run-of-the-mill genetic algorithm



Recast the Bias Problem

- Say you have some observed data
  - Want true distributions of event properties
- Have the ability to simulate data given intrinsic event properties, i.e. model parameters

- Define a metric on the space of output data
  - Define a space of parameters
  - Solve for the set of event properties in that space to minimize the distance Between output data and observed data



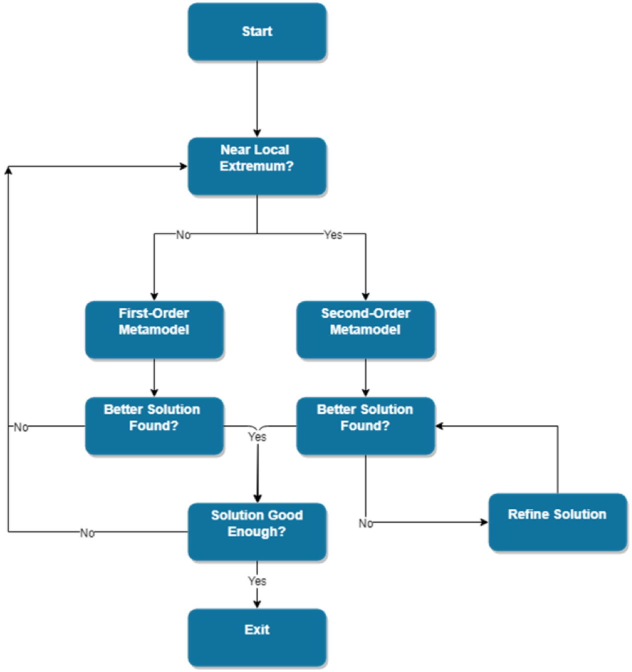
Detecting Bias as an Optimization Problem

- Let $d(\mathbf{x}_1,\mathbf{x}_2)$ be the metric on the data-space, $\mathbf{X}$, and $\mathbf{x}_0 \in \mathbf{X}$ the observed data
- Let $\mathbf{Y}$ be the space of simulation parameters and define $G: \mathbf{Y} \rightarrow \mathbf{X}$ the simulation

- For a vector of parameters $\mathbf{y} \in \mathbf{Y}$, then $d(G(\mathbf{y}), \mathbf{x}_0)$ is the quantity we want to minimize as a function of $\mathbf{y}$

Detecting Bias/The STRONG-S Algorithm[1]

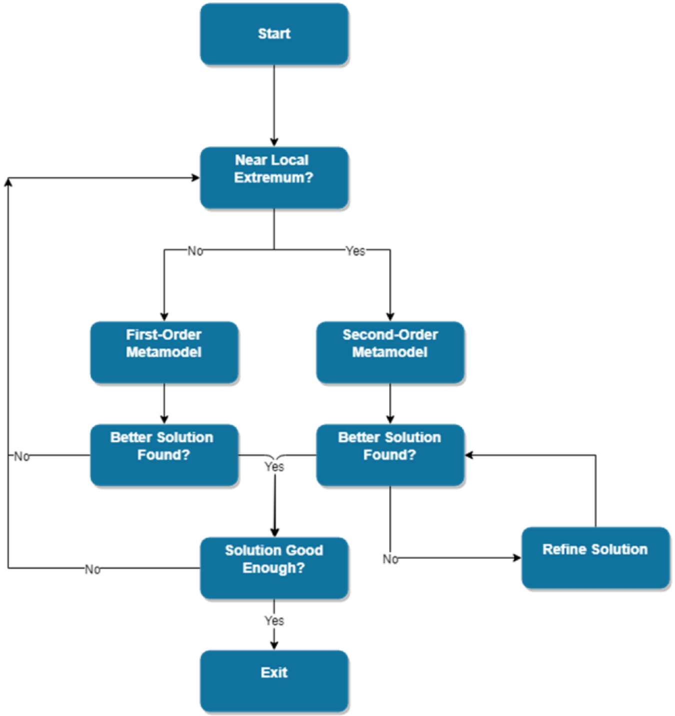
- Real use cases are complex
  - Many parameters
  - Long simulation time
- Naïve optimization algorithms too inefficient



[1] Kuo-Hao Chang , Ming-Kai Li & Hong Wan (2014) Combining STRONG with screening designs for large-scale simulation optimization, IIE Transactions, 46:4, 357-373, DOI: 10.1080/0740817X.2013.812268

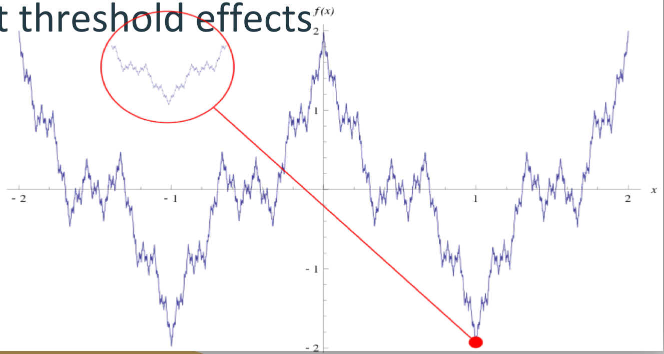
STRONG-S

- Metamodel-based algorithm
  - Based on Response Surface Models
- Automatically screens out unimportant variables
- Chooses to use first- or second-order models based on success of previous iterations



Convergence?

- Convergence of STRONG-S can be proved given two assumptions
  - *Assumption 1*: d(G(**y**),$\mathbf{x}_0$) is twice-differentiable and bounded
  - *Assumption 2*: Given enough data points, the metamodels (first/second order regressions) model `mean`(G(**y**)) to arbitrarily high accuracy
- Reasonable unless the modeled system has significant threshold effects
  - Weierstrass function distributions are forbidden



Acknowledgements

- Thanks to the Oregon NASA Space Grant Consortium for allowing me to talk today
- Special thanks to Dr Greg Bothun, Jonah Rose, and Peter Lovett for their ideas and support
- This work is supported in part through funding from NASA / Oregon Space Grant Consortium, grant NNX15AJ14H.