Comments On Dave Higdon's "Simulation-Aided Inference In Cosmology"

Carlo Graziani

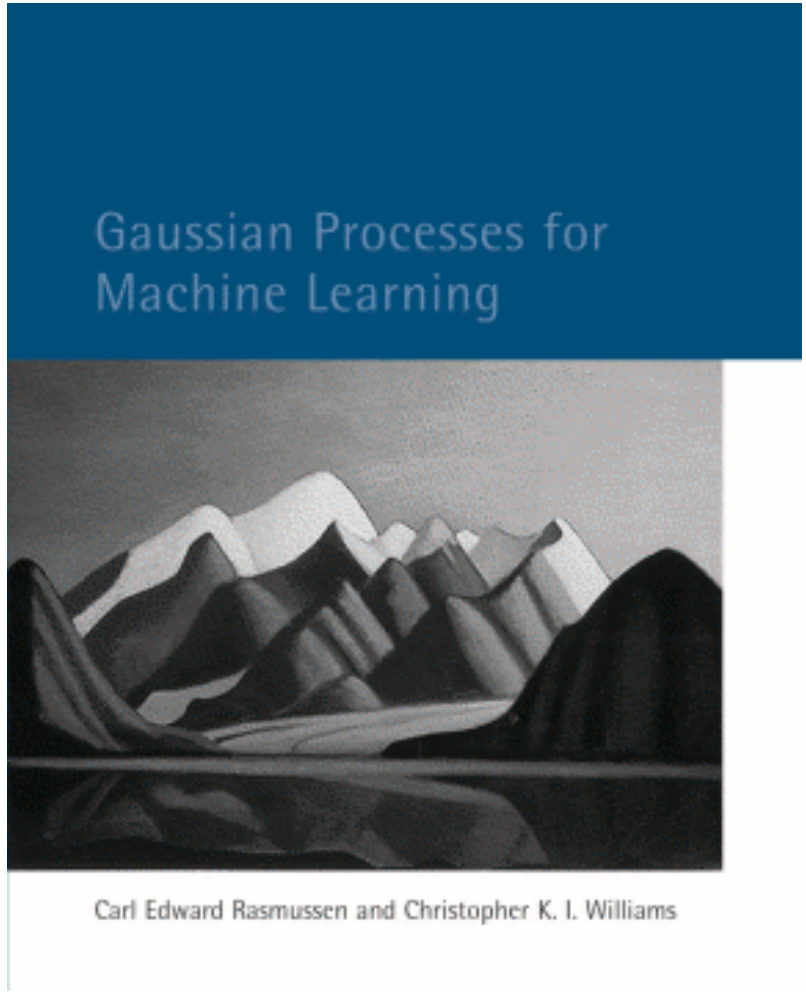
Flash Center For Computational Physics
The University of Chicago

14 June 2011

Computer Models, Meet Data

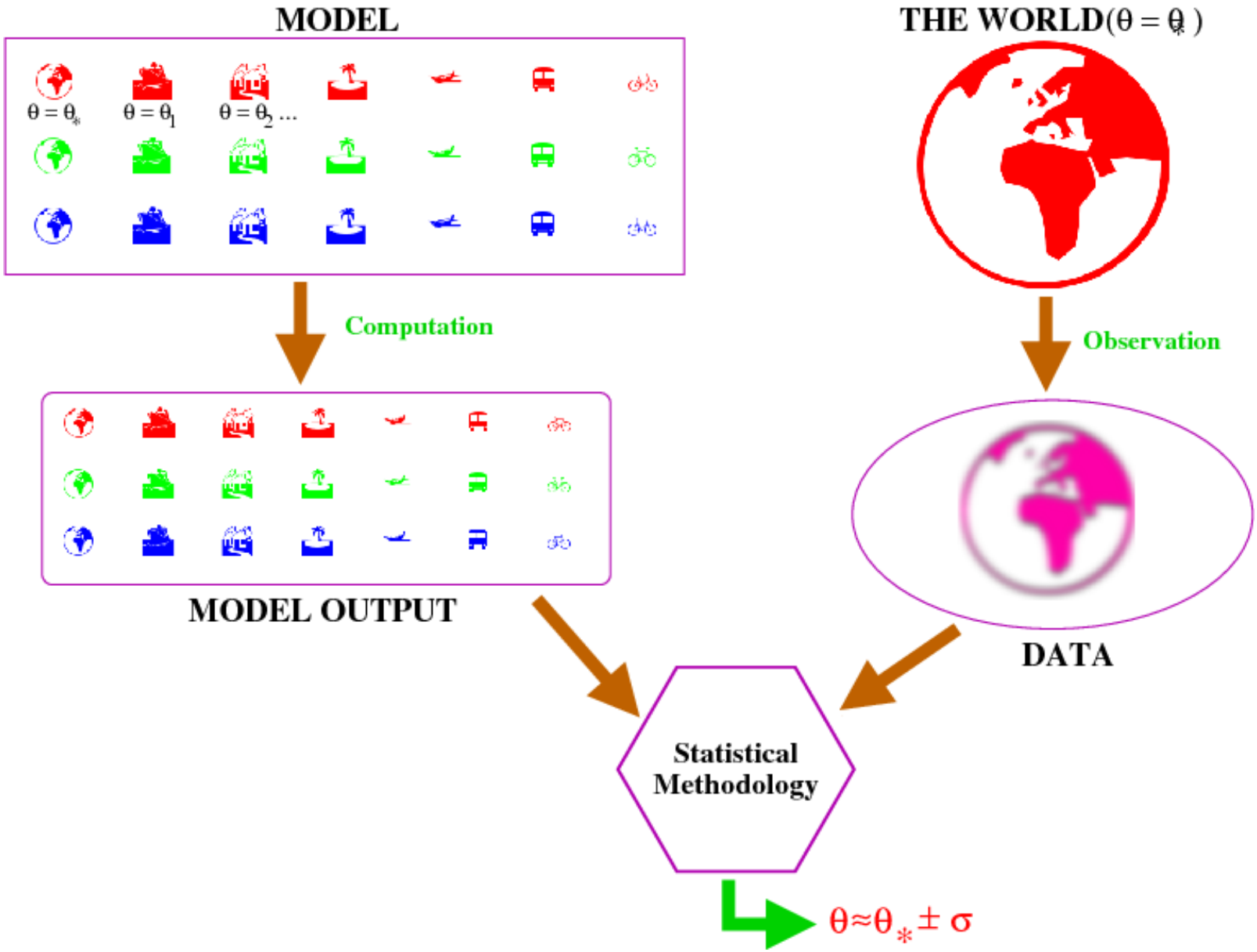
- The comparison of computer model output to data raises new and urgent issues at the forefront of scientific computing.
  - Our group simulates Type Ia SN expolosions with taxing physics fidelity (hydrodynamics, nuclear flame propagation, detonation physics, sub-grid turbulence modeling, radiation trasport...) at taxing resolution. Each sim costs $\sim 10^6$ CPU-hours on ANL's *Intrepid* BG/P system, over about 100 hours. Our models have quite a few important parameters. How are we supposed to do parameter studies?
- This kind of computing will soon be coming to many other fields in astrophysics, as University computational clusters grow in capability.
- We're all going to need something like what Dave is selling.

GP Theory For Busy Scientists

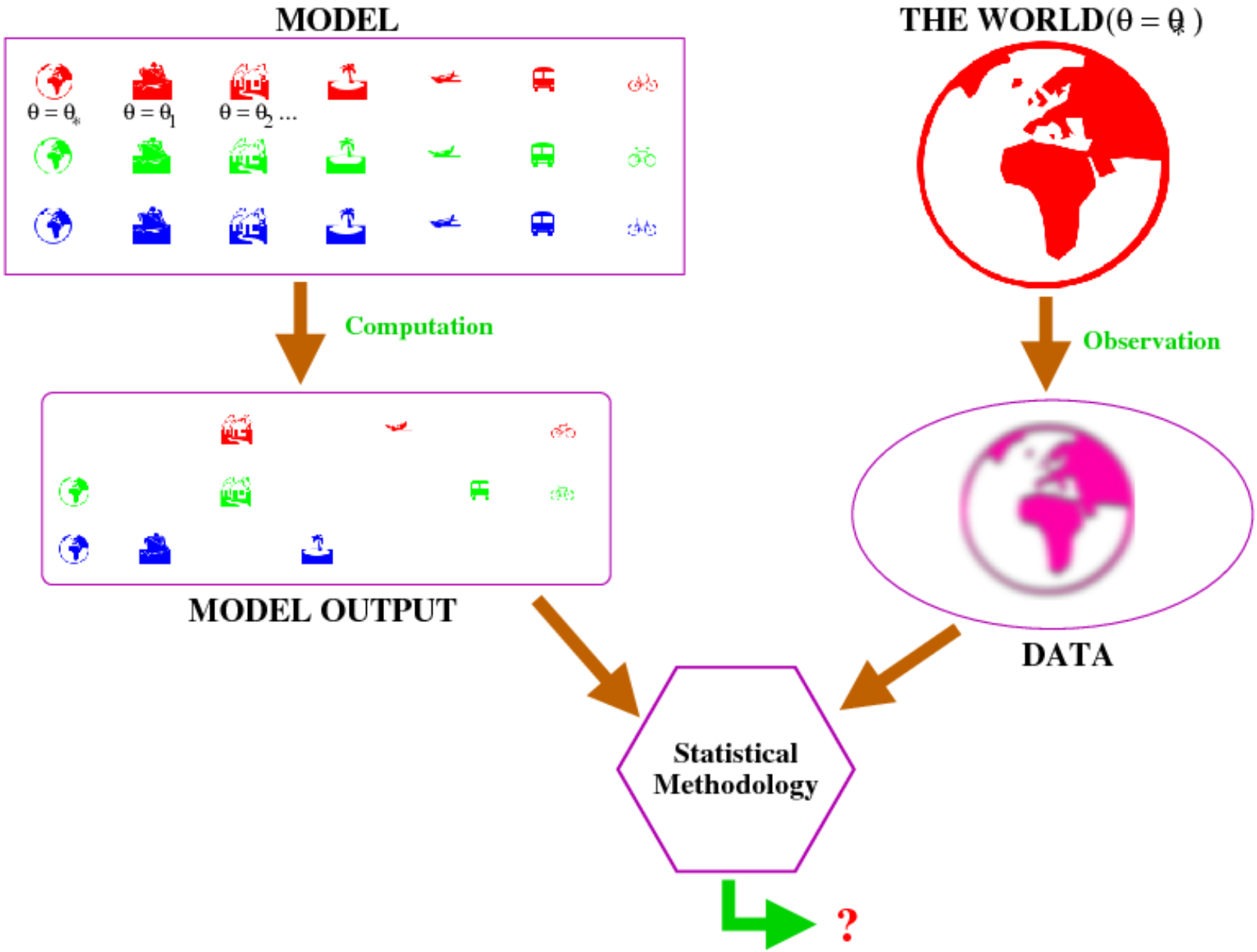


MIT Press. Also available on-line, at
`http://www.gaussianprocess.org/gpml/`

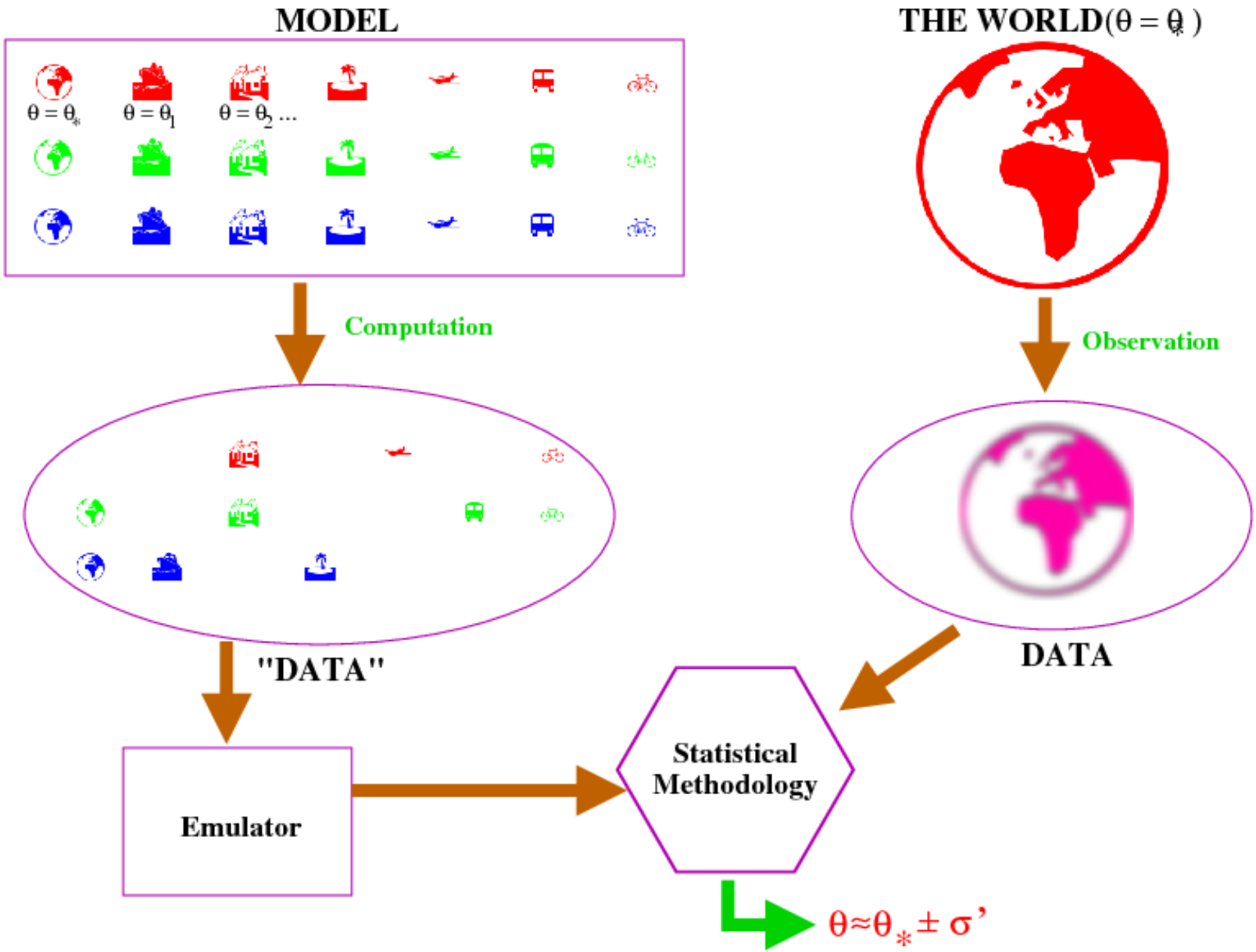
Sparsely Sampled Model Output and Emulators



Sparsely Sampled Model Output and Emulators



Sparsely Sampled Model Output and Emulators



Joint Fitting of Measurement Data and Computer "Data"

$$\pi(\theta, \eta(\cdot)|y, \eta^*) \propto L(y|\eta(\theta)) \times L(\eta^*|\eta(\cdot)) \times \pi(\eta(\cdot)) \times \pi(\theta)$$

- θ: Model parameters;
- $\eta(\cdot)$: Response surface;
- y: Field measurements;
- η^*: Simulation "data";
- $\pi(\eta(\cdot))$: GP prior
- $L(\eta^*|\eta(\cdot))$: GP Likelihood of simulation "data" (AKA "training")

"Opportunities" I: The Dimensionality Curse Just Got Worse

- Classic GP emulation à la Kennedy-O'Hagan (2001) requires the solution of linear problems of dimensionality $N_d \times N_d$, where $N_d \equiv N_{sim} \times N_{output}$. This scales as $\mathcal{O}(N_d^3)$.
- Radical data reduction is required in order for this approach to be feasible — hence Dave's Principal Components Analysis (PCA), which crushes the problem down to $N_{components}$ GPs, each constrained by N_{sim} model runs. This still scales as $\mathcal{O}(N_{components} \times N_{sim}^3)$.

Other Possible Curse-Abatement Strategies

- Are there other data reduction approaches that can beat PCA?
  - It would be interesting to check the performance of, e.g. SCA in this application. However...
  - Even reducing data dimension to $N_{components} << N_{output}$ doesn't solve the $\mathcal{O}(N_{sim}^3)$ problem...
- The Squared-Exponential covariance is not even a little bit sparse. What about other choices?
  - Tapering?
  - Relevance Vector Machine (RVM) approaches?
  - Reduced-rank approximations?
- Approximate inversion (e.g. Conjugate Gradient)?
  - Determinants are tricky...

"Opportunities" II: Covariance Choice

$$
\begin{aligned}
\eta(\theta, k) &= \sum_{j=1}^{p_\eta} w_j(\theta)\phi_j(k) \\
w_j(\cdot) &\sim GP\left(0, \lambda_{w_j}^{-1} C_j(\cdot,\cdot)\right) \\
C_j(\theta, \theta') &= \prod_{l=1}^{p} \exp\left[-\beta_{jl}(\theta_l - \theta_l')^2\right]
\end{aligned}
$$

- This "squared-exponential" covariance choice is convenient and popular, but it has some drawbacks.
  - Principal axes of covariance aligned with parameter coordinate axes;
  - Amount of variability of response surface is the same along all axes ($= \lambda_{w_j}^{1/2}$);
  - Translationally-invariant ("stationary") in parameter space.
- All these are generalizable at small cost.

"Opportunities" III: Model Fidelity

- One usually has the option of running one's computational model at lower — and cheaper — levels of model fidelity.
- For example, in SNIa simulations:
  - Lower mesh resolution;
  - 2-D instead of 3-D;
  - Cheaper physics/surrogate models.
- This opens the possibility of establishing a *Fidelity Hierarchy*, wherein we probe model parameter space using abundant sprays of cheaper, lower-fidelity models runs. Discrepancies from the high-fidelity model are calibrated using less-frequent high-fidelity runs at carefully-chosen parameter settings.
- Some results exist (Kennedy & O'Hagan 2000, Qian & Wu 2008, Cumming & Goldstein 2009). Some generalization desirable, e.g. non-stationary kernels, fidelity levels not strictly ordered by informativeness.

"Opportunities" IV: Adaptive Numerical Experimental Design

- At what parameter settings should simulations be performed?
- An *a priori* design, such as latin hypercube, is unlikely to capture important features in a complex-structure response surface.
- An *adaptive*, iterative scheme, which learns from previous emulator/data comparison to predict new, useful simulation parameter settings seems crucial, especially as the parameter space gets large.
- The scheme must resolve a tension — the *Exploration-Exploitation Tradeoff* — between two important objectives: Undestanding the response surface everywhere, and locating regions where it most resembles the experimental data.

Adaptive Numerical Experimental Design, Cont'd

- Efforts to-date have been focused on exploration — good characterization of response surface to be emulated, ignoring the data (e.g. Gramacy & Lee 2009, Cumming & Goldstein 2009).
- For consideration of the full exploration/exploitation tension one has to look to the literature on *physical* experimental design.
- A particularly promising development: Loredo & Chernoff (2004) show that physical measurements currently "in the can" can be used to calculate the expected information gain — negative Shannon entropy, measured in bits of information — from a future measurement with selected experimental parameters, and so choose those parameters so as to maximize that information.

Adaptive Numerical Experimental Design, Cont'd

This development translates immediately over from physical to numerical experimental design. Moreover, GP structure of emulator results in easily-computable information measures.

$$\begin{aligned}\text{Info}(\boldsymbol{\theta}_+) &= H\left[\pi(\mathbf{y}_+|\boldsymbol{\theta}_+, \mathbf{Y}, \boldsymbol{\Theta})\right] \\ &\quad - \int d\boldsymbol{\theta}_T \, \pi(\boldsymbol{\theta}_T|\mathbf{y}_{meas}, \mathbf{Y}, \boldsymbol{\Theta}) H\left[\pi(\mathbf{y}_+|\boldsymbol{\theta}_+, \mathbf{Y}, \mathbf{y}, \boldsymbol{\Theta}, \boldsymbol{\theta}_T)\right] .\end{aligned}$$

- $H[\pi(\cdot)]$ is entropy of PDF π; $\boldsymbol{\theta}_+$ is proposed new parameter point; $(\mathbf{Y}, \boldsymbol{\Theta})$ is the current numerical design; $\mathbf{y}_{meas}$ are the observations.
- The first term embodies exploration (by itself, it yields Maxent sampling). The second term embodies exploitation, rewarding smaller predictive uncertainty near best-fit parameter point.

Conclusions

- Gaussian Process Emulators are likely to be a very common feature of computational modeling of data in the coming decade.
- Several challenges must be met if big data is really to meet big models in a GP emulator.
  - Radical data reduction (SCA?)
  - Low-rank/sparse covariance kernels
  - Approximate inverses
- Multi-fidelity simulation hierarchies can expand our ability to probe the parameter space.
- Adaptive numerical experimental design can help choose simulation parameter settings as efficiently as possible.