Emulation of computer models when multi-fidelity data are available

Georgios Karagiannis

Group of Statistics
Department of Mathematical sciences,
Durham University

September, 2019

IDAS Launch Event

My research...

- Bayesian statistical modeling
  - Emulation of computer models
  - Calibration of computer models
- Bayesian computations
  - Markov chain Monte Carlo
  - Approximate Bayesian Computations
- Applications:
  - WRF climate model
  - ADCIRC storm surge model
  - PSU-3D ice model

Computer models, multifidelity, & Emulation

- Computer model (CM)
  - aim to reproduce the real system's behavior with high accuracy.

$$\text{for } x \in \mathcal{X} \quad \text{and} \quad y \in \mathbb{R} \qquad\qquad x \xrightarrow{S(x)} y$$

  - is software running on computers (or super-computers)
  - Expensive: only a limited number of simulations is performed.
- Emulators:
  - A cheap probabilistic approximation of the input-output mapping
  - run CM at different levels of fidelity, sophistication, or resolution $\{(y_t, \mathfrak{X}_t\}$.
  - Fit a Gaussian process regression

$$x \longmapsto y(x)$$

The ADCIRC model

- ADCIRC:
  - ADvanced CIRCulation storm surge model
  - it takes weeks to run
- Output y:
  - pick surge elevation for the landfalling hurricane
- Input x:
  - six parameters characterizing the storm
    - central pressure deficit of the storm (mb)
    - scale pressure radius in nautical miles
    - storm's forward speed (m/s)
    - storm's heading in degrees clockwise from north
    - Holland's B parameter (unitless)
    - landfall location in latitude and longitude
- Different fidelity levels:
  - ADCIRC
  - ADCIRC+SWAN (Simulating WAves Nearshore)

AR co-kriging (ARCK) model by K&O

- The output $\{y_t(\cdot)\}$ is modeled as

$$y_t(x) = \xi_{t-1}(\cdot) y_{t-1}(x) + \delta_t(x) \qquad \text{for } x \in \mathcal{X},\ t = 2, ..., S$$

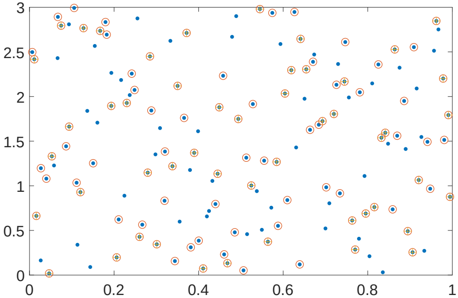
where

$$\begin{aligned} y_1(\cdot) &\sim \text{GP}(\mu_1(\cdot|\beta_1), c_1(\cdot,\cdot|\phi_1)); \\ \delta_t(\cdot) &\sim \text{GP}(\mu_t(\cdot|\beta_t), c_t(\cdot,\cdot|\phi_t)), \end{aligned}$$

- $\{\delta_t(\cdot)\}$ and $\{\xi_t\}$, account for 'missing' physical properties in $\mathscr{C}_{t-1}$ w.r.t. $\mathscr{C}_t$.

Challenge: Non-nested designs

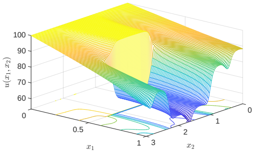
- Challenge: requires data-sets to be nested $\mathfrak{X}_1 \subseteq \mathfrak{X}_2 \subseteq \ldots$
- Create artificially nested data by imputing the data $\{y_t, \mathfrak{X}_t\}$ with missing data $\{\mathring{y}_t, \mathring{\mathfrak{X}}_t\}$

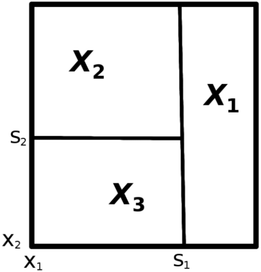


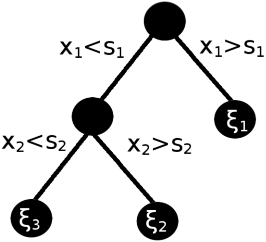
Challenge: Discontinuity/non-stationarity in the output

Introduce a Bayesian treed partitioning such as:

$$y_{k,t}(x) = \xi_{k,t-1}(\cdot) y_{k,t-1}(x) + \delta_{k,t}(x) \qquad \text{for } x \in \mathcal{X}_k$$

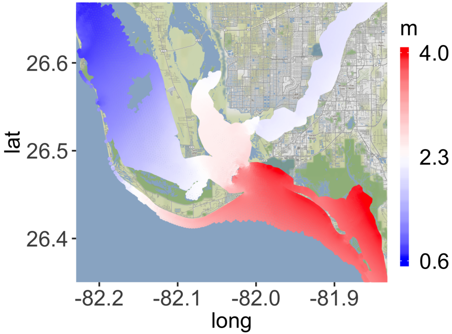






Challenge: High-dimensional output

Consider the output of ADCIRC model at certain input values



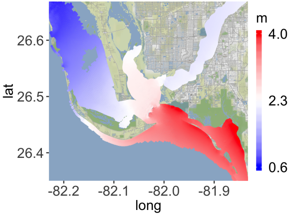
For the j-th coordinate, the following cokriging model is assumed:

$$y_{t,j}(\cdot) = \gamma_{t-1,j}(\cdot) y_{t-1,j}(\cdot) + \delta_{t,j}(\cdot)$$

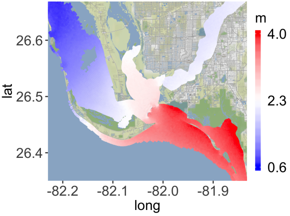
$$y_{1,j}(\cdot) \sim \text{GP}(\mu_{1,j}(\cdot|\beta_{1,j}), c_1(\cdot,\cdot|\phi_1)); \quad \delta_{t,j}(\cdot) \sim \text{GP}(\mu_{t,j}(\cdot|\beta_{t,j}), c_{t,j}(\cdot,\cdot|\phi_t)),$$

where ϕ_t are common.

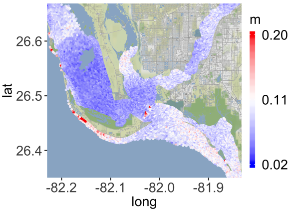
Emulating the ADCIRC model



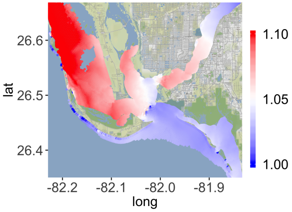
(a) exact PSE



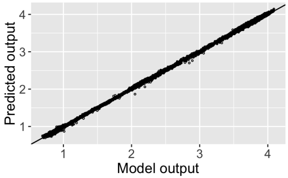
(b) predicted PSE



(c) sd. error



(d) scale discrepancy



(e) cross validation

My research...

- Bayesian statistical modeling
  - Emulation of computer models
  - Calibration of computer models
- Bayesian computations
  - Markov chain Monte Carlo
  - Approximate Bayesian Computations
- Applications:
  - WRF climate model
  - ADCIRC storm surge model
  - PSU-3D ice model