Calibrating Complex Models using Emulation and History Matching

2023 IDM Annual Symposium

Andrew Iskauskas

Durham University
In collaboration with LSHTM

May 23, 2023

Complex Models of Real-World Phenomena

Complex computer models (or *simulators*) are used in a variety of fields, including

- Oil Industry (oil reservoir and geology models) [4]
- Climate Science (climate models of global warming) [11]
- Systems Biology (genetic and metabolic network models) [10]
- Cosmology (galaxy formation simulations) [9]
- Nuclear Physics (quantum many-body models of nuclei) [6]
- **Epidemiology** (which is why I'm here!) [1, 8]

Simulators are often computationally expensive: a full exploration of the parameter space using only the simulator is infeasible.

The Emulator

An *emulator* is a statistical approximation of a complex computer simulator [3].

Let $f(x)$ be an output from the simulator at a given parameter set $x \in \mathbb{R}^d$, corresponding to some real physical process y. Then we define a emulator for output $f(x)$ as

$$g(x) = \sum_i \beta_i h_i(x_A) + u(x_A) + w(x)$$

The $h_i(x_A)$ are a collection of basis functions in the *active variables* x_A, β_i the coefficients, $u(x_A)$ a weakly stationary process in the active variables, and $w(x)$ a 'nugget term'.

Bayes Linear Emulators

We need to provide specifications for the random quantities β, $u(x)$, $w(x)$, then update in light of data. We could use a *fully Bayesian approach*, putting distributions on each – but often we don't have enough information to. Even if we do, *to what extent do we trust those statements*?

Instead we take a *pragmatic* approach and require only the specification of second-order quantities: expectations, variances, and covariances. This is the *Bayes Linear* approach [5]. This is a much easier task.

Introduction	Emulation and History Matching	Application	Conclusion	References
○	○○○●○○	○○○○	○○	

Emulation

Example: 1d Model

Consider a model which can be described in closed form as

$$f(x) = 2x + 3x \sin\left(\frac{5\pi(x - 0.1)}{0.4}\right)$$

Prior specifications: assume a constant regression function $h(x) = 1$, $E[\beta] = 0.6$, $Var[\beta] = 0$, $E[u(x)] = 0$, and

$$\text{Cov}[u(x), u(x')] = \sigma^2 \exp\left\{-\frac{(x - x')^2}{\theta^2}\right\},$$

with $\sigma^2 = 7$ and $\theta = 0.3$.

Introduction	Emulation and History Matching	Application	Conclusion	References
O	000●000	0000	00	

Emulation

Example: 1d Model

Adjust this prior structure with reference to a collection of 10
points, spaced equally over the interval $[0, 0.5]$.

Introduction	Emulation and History Matching	Application	Conclusion	References
○	○○○●○○○	○○○○	○○	
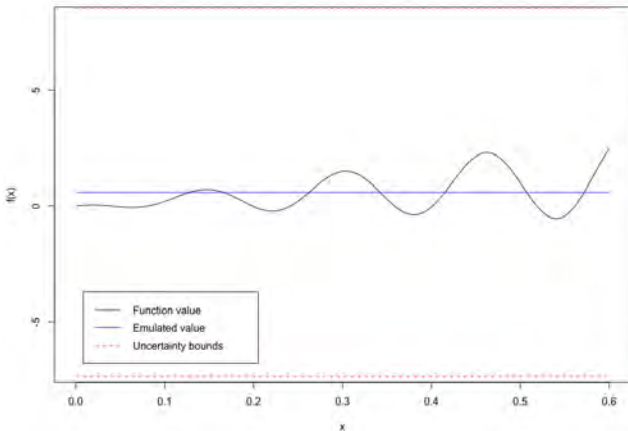
Emulation

Example: 1d Model

Adjust this prior structure with reference to a collection of 10 points, spaced equally over the interval $[0, 0.5]$.
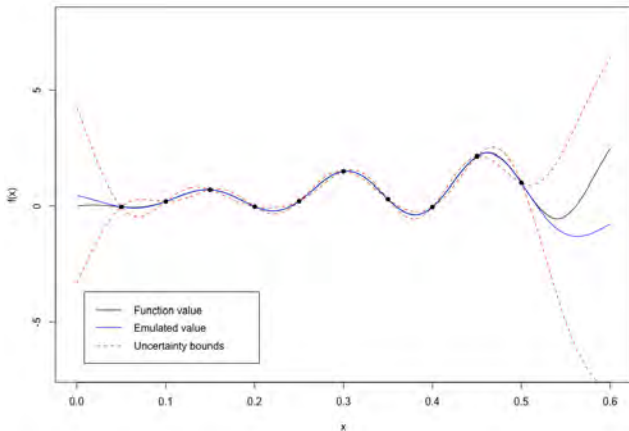
Emulation

Emulators are **fast to evaluate**, requiring only matrix multiplication. For complex models which can take anywhere from minutes to months to evaluate a limited ensemble of runs, an emulator can quickly investigate model behaviour across the entire parameter space.

Emulators **have uncertainty statements built-in**. Each prediction comes with a corresponding uncertainty, $\text{Var}_D[g(x)]$, which depends on the data provided to it and the proximity of the unseen points thereof.

History Matching

> Given observed data corresponding to a simulator output, what combinations of input parameters could give rise to output consistent with this observation?

History matching works on the principle of complementarity: a point x is considered unsuitable if **even accounting for the uncertainties in the system**, the prediction $E_D[g(x)]$ cannot be 'close' to the observed value z. Closeness is defined via an *implausibility* measure

$$I^2(x) = \frac{(E_D[g(x)] - z)^2}{\text{Var}_D[g(x)] + \text{Var}[e] + \text{Var}[\epsilon]}.$$

Emulation and HM: Summary

Emulators can **efficiently** and **robustly** predict simulator output at unseen points, given a small collection of known runs.

The corresponding emulator uncertainty is a **natural extension** to existing sources of uncertainty in our model.

History matching allows us to leverage the uncertainty structure to find **all** acceptable matches to data arising from our model.
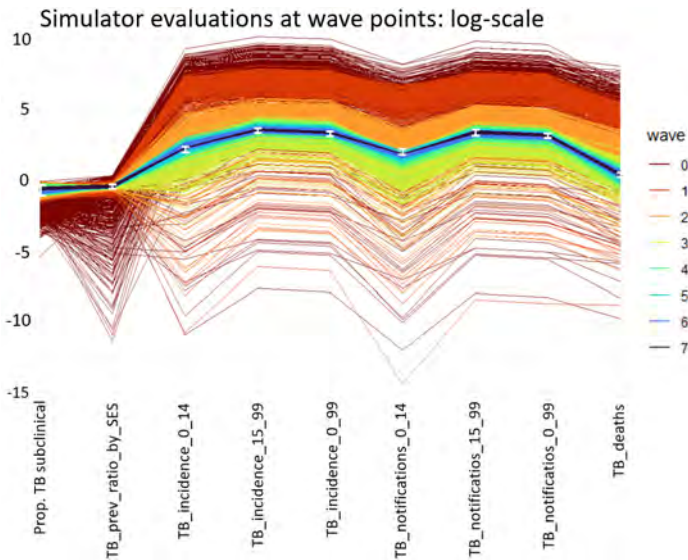
The hmer Package

The hmer package [7] was developed to make the tools of emulation and history matching accessible for modellers. It allows

- Careful prior specifications to be determined and emulators to be trained
- Diagnostics to be performed to assess suitability
- Appropriate choices of implausibility measure and design for further waves to be made.
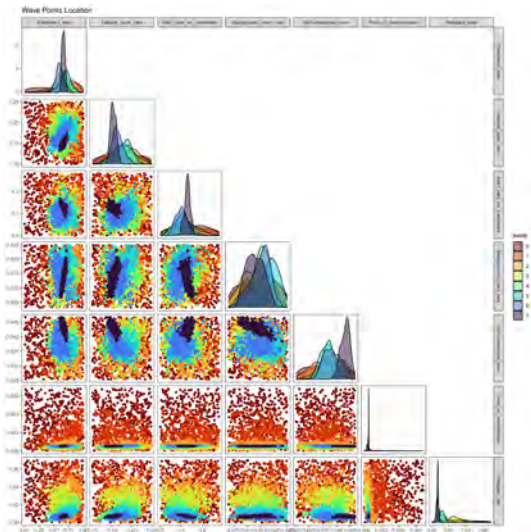
The TBvax model

- Deterministic hybrid model of tuberculosis and HIV, designed by colleagues at London School of Hygiene and Tropical Medicine [2, 8]
- Between 19 and 34 input parameters, and 9 to 30 outputs per country
- Run-time for a parameter set varied between seconds and minutes
- Modelling requested by the World Health Organisation in order to form the basis for policy decisions
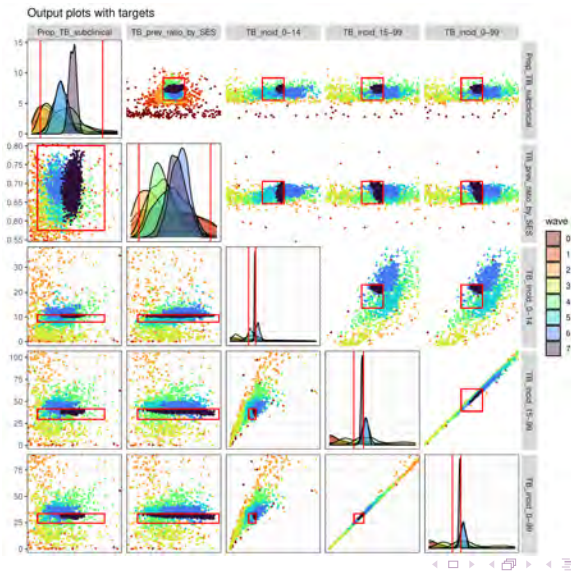- Expected to be used to match to observational data for 115 countries
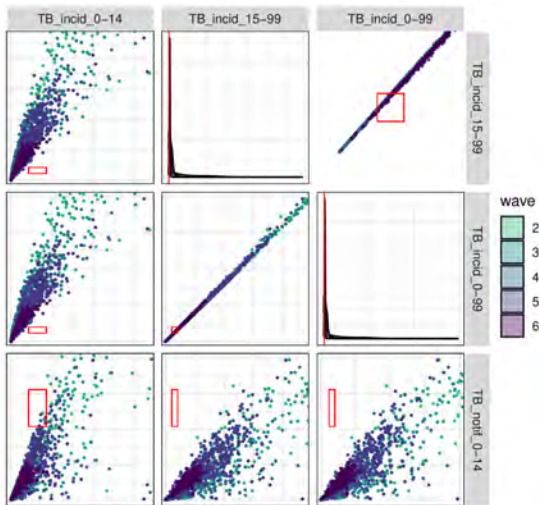
Results: Calibrated Countries

Results: Calibrated Countries

Introduction
o

Emulation and History Matching
oooooo

Application
oooo

Conclusion
oo

References

Results: Calibrated Countries

Results: Uncalibrated Countries

Active Development: Models Required!

The hmer package is in active development, with many possible extensions to the framework or improvements in visualisation of results.

Models (particularly of disease) have been, and will continue to be, powerful drivers of the direction of development. If you have a calibration problem, we'd like to know about it!

- CRAN Repository: cran.r-project.org/package=hmer
- Github Active Branch: github.com/andy-iskauskas/hmer
- Documentation, Tutorials, and Examples:
  https://hmer-package.github.io/website/

Summary

- Complex simulators are extremely valuable for understanding real-life processes, but can be slow to run and difficult to meaningfully and/or robustly analyse.

- Emulation allows us to quickly make predictions across a large-dimensional parameter space without recourse to the (slower) simulator.

- History Matching uses the induced uncertainty structure to determine the full set of parameter combinations that could give rise to the observed data.

- Flexible, low specificational burden, and extensible via hierarchical emulation, multistate emulation, known boundary emulation, . . .

- The hmer package is designed to make these tools more readily available for modellers.

References I

[1] Ioannis Andrianakis et al. "Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda". In: *PLoS computational biology* 11.1 (2015), e1003968.

[2] Rebecca A Clark et al. "The Impact of Alternative Delivery Strategies for Novel Tuberculosis Vaccines in Low- and Middle-income Countries: A Modelling Study". In: *The Lancet Global Health* 11.4 (2023), e546–e555.

[3] Peter S Craig et al. "Constructing Partial Prior Specifications for Models of Complex Physical Systems". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.1 (1998), pp. 37–53.

[4] Jonathan A Cumming and Michael Goldstein. "Bayes Linear Uncertainty Analysis for Oil Reservoirs Based on Multiscale Computer Experiments". In: *The Oxford handbook of applied Bayesian analysis* (2010), pp. 241–270.

[5] Michael Goldstein and David Wooff. *Bayes Linear Statistics: Theory and Methods*. Vol. 716. John Wiley & Sons, 2007.

[6] Baishan Hu et al. "Ab Initio Predictions Link the Neutron Skin and 208Pb to Nuclear Forces". In: *Nature Physics* 18.10 (2022), pp. 1196–1200.

[7] Andrew Iskauskas et al. "Emulation and History Matching using the hmer Package". In: *arXiv preprint arXiv:2209.05265* (2022).

References II

[8]   Danny Scarponi et al. "Demonstrating Multi-country Calibration of a
      Tuberculosis Model Using New History Matching and Emulation Package -
      hmer". In: *Epidemics* 43 (2023), p. 100678.

[9]   Ian Vernon, Michael Goldstein, and Richard Bower. "Galaxy Formation: A
      Bayesian Uncertainty Analysis". In: *Bayesian analysis* 5.4 (2010), pp. 619–669.

[10] Ian Vernon et al. "Bayesian Uncertainty Analysis for Complex Systems Biology
      Models: Emulation, Global Parameter Searches and Evaluation of Gene
      Functions". In: *BMC systems biology* 12.1 (2018), pp. 1–29.

[11] Daniel Williamson et al. "History Matching for Exploring and Reducing Climate
      Model Parameter Space using Observations and a Large Perturbed Physics
      Ensemble". In: *Climate dynamics* 41.7 (2013), pp. 1703–1729.