Calibration of stochastic models using history matching and emulation.

Ioannis Andrianakis

London School of Hygiene and Tropical Medicine

GPUQSS 2016, Sheffield, 15 September 2016

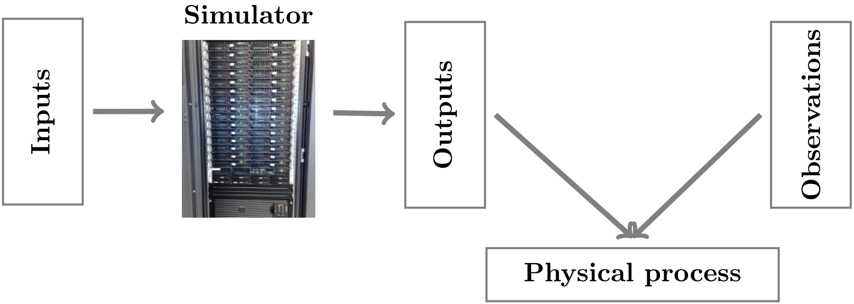
Joint work with: I.Vernon, N.McCreesh, TJ McKinley, J.Oakley, R.Nsubuga, M.Goldstein and R.White

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Outline

- Motivation.
- Emulation.
- History matching.
- Results.

Calibration of computer models



Calibration objective:

To find a set of input values so that the simulator represents best the physical process as this is described by observations.

The 'Mukwano' simulator

- A dynamic, stochastic, individual based model that simulates heterosexual sexual partnerships and HIV transmission.
- 22 inputs inc. contact rates, concurrency parameters, relationship duration, 2 sexual activity groups (high/low), 2 concurrency groups (high/low), 3 discrete behaviour periods.
- 18 outputs inc. population size, HIV prevalence, prevalence of men and women in long/short duration partnerships with one or more partners.
- Run time varies from 10 mins to >3 hours for 1 simulator run.
- Calibration data provided by a general population cohort in Uganda.

The 'Mukwano' simulator

- A dynamic, stochastic, individual based model that simulates heterosexual sexual partnerships and HIV transmission.
- 22 inputs inc. contact rates, concurrency parameters, relationship duration, 2 sexual activity groups (high/low), 2 concurrency groups (high/low), 3 discrete behaviour periods.
- 18 outputs inc. population size, HIV prevalence, prevalence of men and women in long/short duration partnerships with one or more partners.
- Run time varies from 10 mins to >3 hours for 1 simulator run.
- Calibration data provided by a general population cohort in Uganda.

The 'Mukwano' simulator

- A dynamic, stochastic, individual based model that simulates heterosexual sexual partnerships and HIV transmission.
- 22 inputs inc. contact rates, concurrency parameters, relationship duration, 2 sexual activity groups (high/low), 2 concurrency groups (high/low), 3 discrete behaviour periods.
- 18 outputs inc. population size, HIV prevalence, prevalence of men and women in long/short duration partnerships with one or more partners.
- Run time varies from 10 mins to >3 hours for 1 simulator run.
- Calibration data provided by a general population cohort in Uganda.

The 'Mukwano' simulator

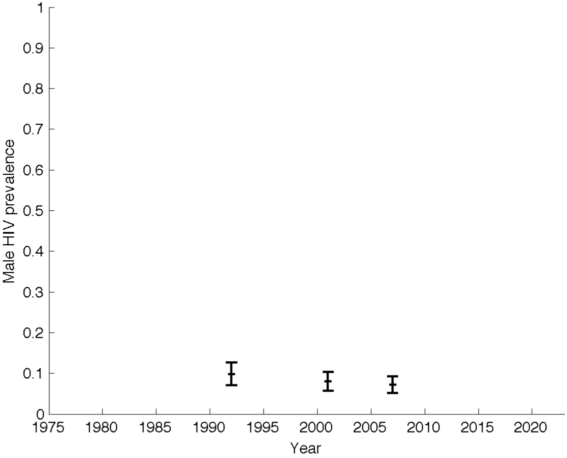
- A dynamic, stochastic, individual based model that simulates heterosexual sexual partnerships and HIV transmission.
- 22 inputs inc. contact rates, concurrency parameters, relationship duration, 2 sexual activity groups (high/low), 2 concurrency groups (high/low), 3 discrete behaviour periods.
- 18 outputs inc. population size, HIV prevalence, prevalence of men and women in long/short duration partnerships with one or more partners.
- Run time varies from 10 mins to >3 hours for 1 simulator run.
- Calibration data provided by a general population cohort in Uganda.

The 'Mukwano' simulator

- A dynamic, stochastic, individual based model that simulates heterosexual sexual partnerships and HIV transmission.
- 22 inputs inc. contact rates, concurrency parameters, relationship duration, 2 sexual activity groups (high/low), 2 concurrency groups (high/low), 3 discrete behaviour periods.
- 18 outputs inc. population size, HIV prevalence, prevalence of men and women in long/short duration partnerships with one or more partners.
- Run time varies from 10 mins to >3 hours for 1 simulator run.
- Calibration data provided by a general population cohort in Uganda.

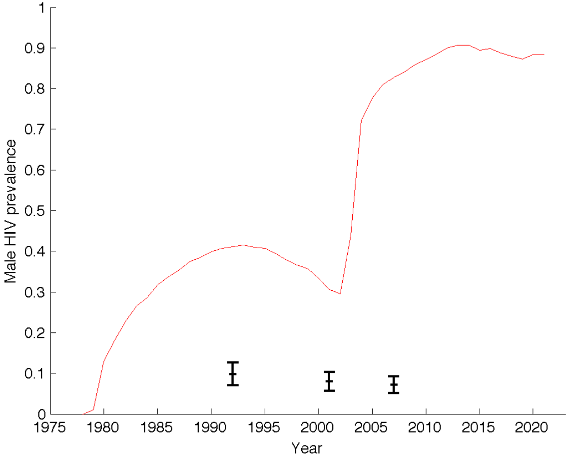
A manual approach

Suppose we want to match male HIV prevalences at 3 points in time.



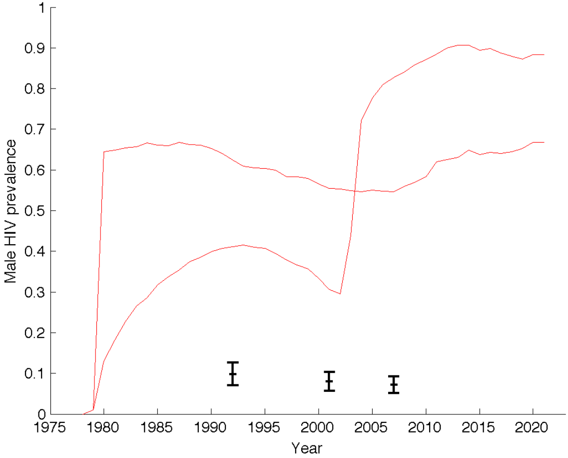
A manual approach

We choose a set of inputs run the model and...



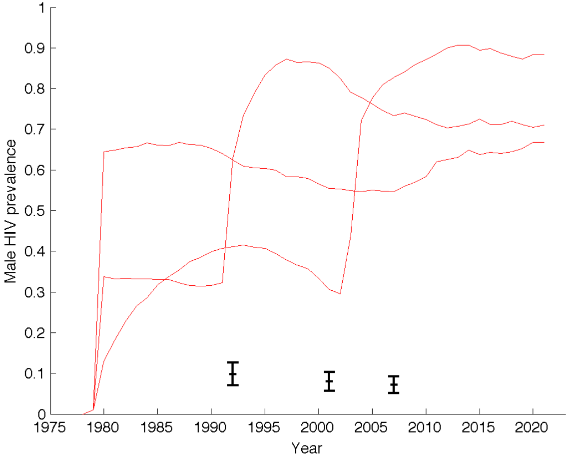
A manual approach

...we try again...



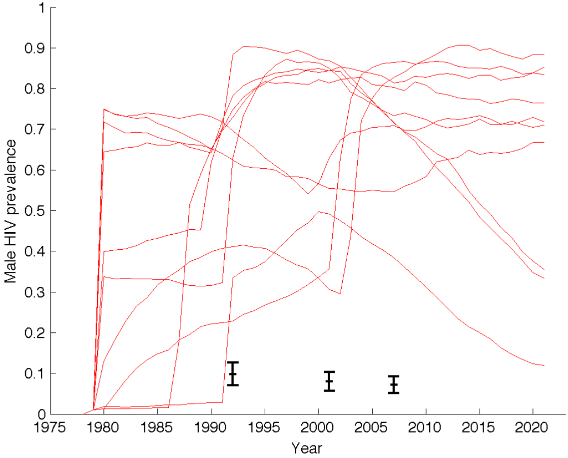
A manual approach

...and again...



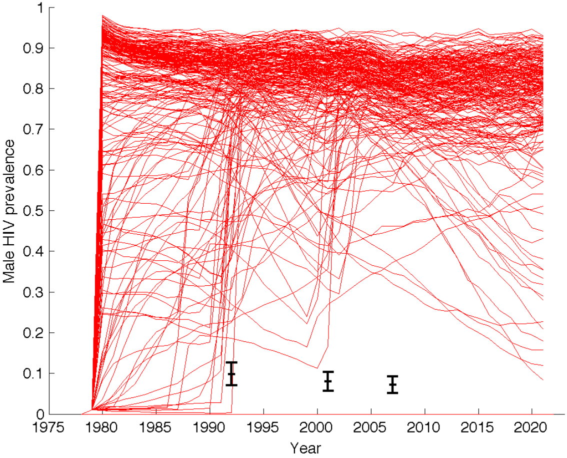
A manual approach

...after 10 runs...



A manual approach

...after 250 runs.



History matching

- Rather than looking for the best input values, history matching identifies and discards those unlikely to provide a match to the empirical data.

- The *implausible* input space is discarded in iterations known as waves.

- Not all inputs/outputs need to be considered at once.

- The simulator is often ‘better behaved’ in smaller areas of input space.

- History matching relies on *emulators* for computational efficiency.

History matching

- Rather than looking for the best input values, history matching identifies and discards those unlikely to provide a match to the empirical data.
- The *implausible* input space is discarded in iterations known as waves.
- Not all inputs/outputs need to be considered at once.
- The simulator is often ‘better behaved’ in smaller areas of input space.
- History matching relies on *emulators* for computational efficiency.

History matching

- Rather than looking for the best input values, history matching identifies and discards those unlikely to provide a match to the empirical data.
- The *implausible* input space is discarded in iterations known as waves.
- Not all inputs/outputs need to be considered at once.
- The simulator is often 'better behaved' in smaller areas of input space.
- History matching relies on *emulators* for computational efficiency.

History matching

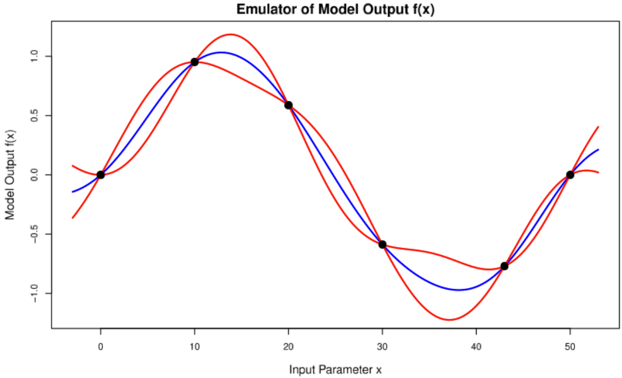
- Rather than looking for the best input values, history matching identifies and discards those unlikely to provide a match to the empirical data.
- The *implausible* input space is discarded in iterations known as waves.
- Not all inputs/outputs need to be considered at once.
- The simulator is often 'better behaved' in smaller areas of input space.
- History matching relies on *emulators* for computational efficiency.

History matching

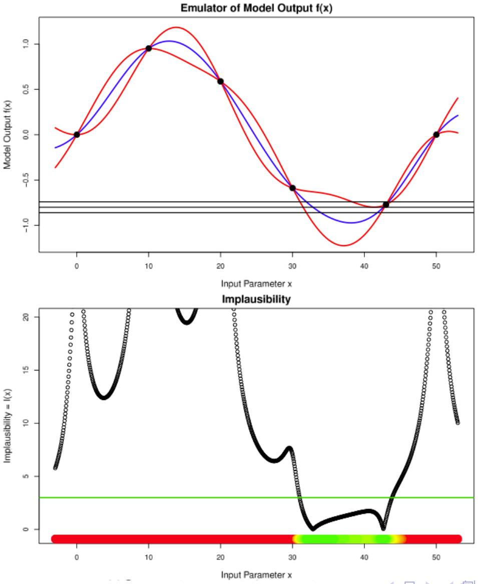
- Rather than looking for the best input values, history matching identifies and discards those unlikely to provide a match to the empirical data.
- The *implausible* input space is discarded in iterations known as waves.
- Not all inputs/outputs need to be considered at once.
- The simulator is often 'better behaved' in smaller areas of input space.
- History matching relies on *emulators* for computational efficiency.

An emulator example

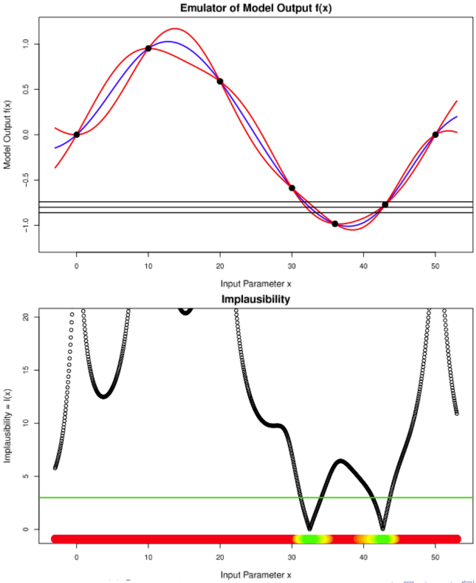
The emulator gives a posterior distribution for the model output, conditioned on the model runs we have seen so far.



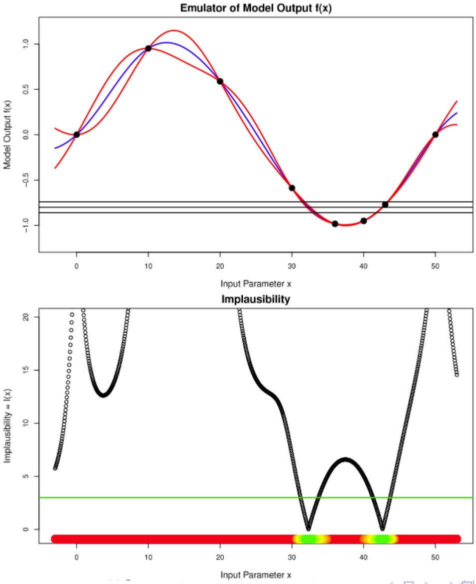
History matching - wave 1



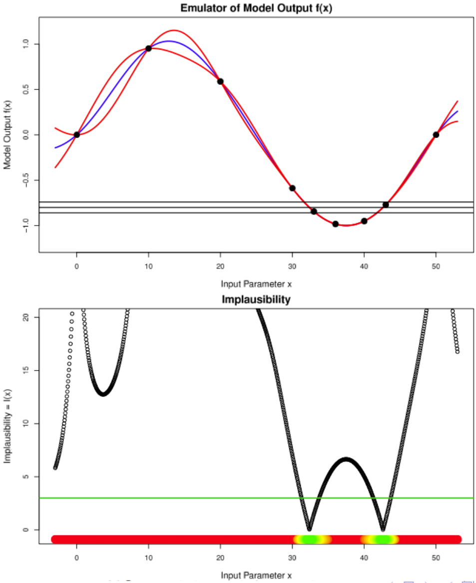
History matching - wave 2



History matching - wave 3



History matching - wave 4



Challenges

- Large number of inputs and outputs.
- Unavailable model likelihood.
- Long simulator running times.
- Stochastic model.

Challenges

- Large number of inputs and outputs.
- Unavailable model likelihood.
- Long simulator running times.
- Stochastic model.

Challenges

- Large number of inputs and outputs.
- Unavailable model likelihood.
- Long simulator running times.
- Stochastic model.

Challenges

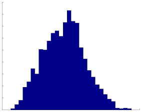
- Large number of inputs and outputs.
- Unavailable model likelihood.
- Long simulator running times.
- Stochastic model.

Outline

- Motivation.
- Emulation.
- History matching.
- Results.

Stochastic model

For an input $\mathbf{x}$, the simulator's output is a draw from an unknown distribution, with mean $g(\mathbf{x})$ and variance $s(\mathbf{x})$.



$$g(\mathbf{x}) \equiv \mathrm{E}[f(\mathbf{x})]$$

$$s(\mathbf{x}) \equiv \mathrm{Var}[f(\mathbf{x})]$$

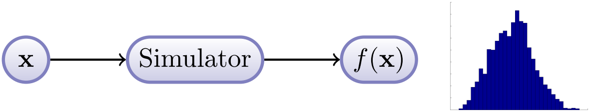
We can write this as:

$$f(\mathbf{x}) = g(\mathbf{x}) + \epsilon(\mathbf{x})$$

where $\epsilon(\mathbf{x})$ is a zero mean r.v. with variance $s(\mathbf{x})$.

Stochastic model

For an input $\mathbf{x}$, the simulator's output is a draw from an unknown distribution, with mean $g(\mathbf{x})$ and variance $s(\mathbf{x})$.



$$g(\mathbf{x}) \equiv \mathrm{E}[f(\mathbf{x})]$$

$$s(\mathbf{x}) \equiv \mathrm{Var}[f(\mathbf{x})]$$

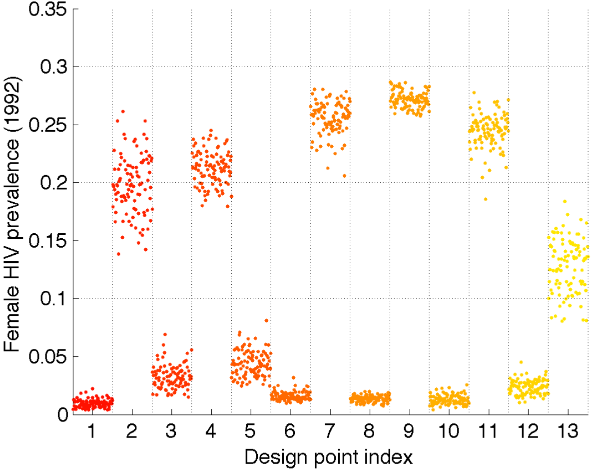
We can write this as:

$$f(\mathbf{x}) = g(\mathbf{x}) + \epsilon(\mathbf{x})$$

where $\epsilon(\mathbf{x})$ is a zero mean r.v. with variance $s(\mathbf{x})$.

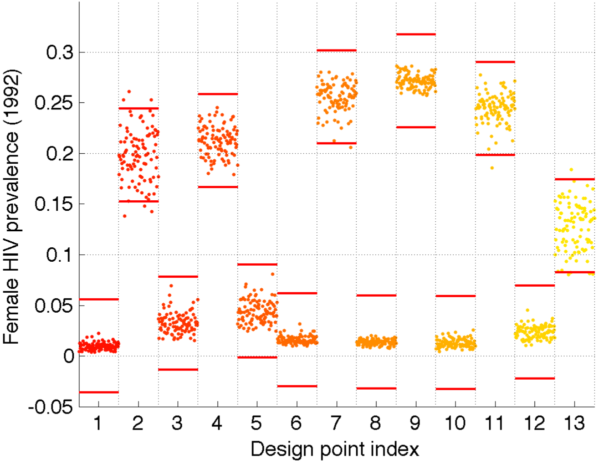
The variance $s(\mathbf{x})$

The variance $s(\mathbf{x})$ is a function of $\mathbf{x}$ and this has to be taken into account in history matching.



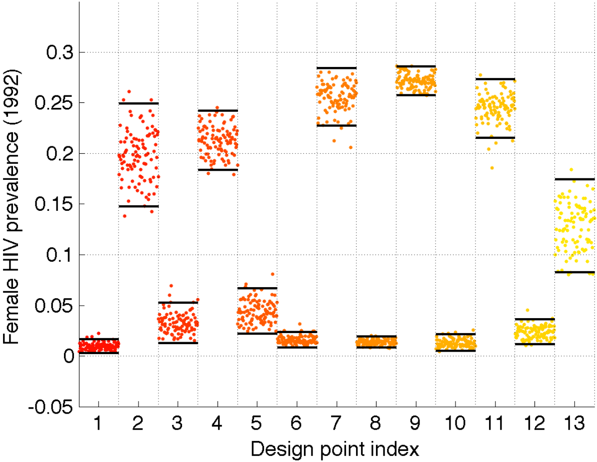
The variance $s(\mathbf{x})$

One approach is to assume a fixed variance across $\mathbf{x}$.

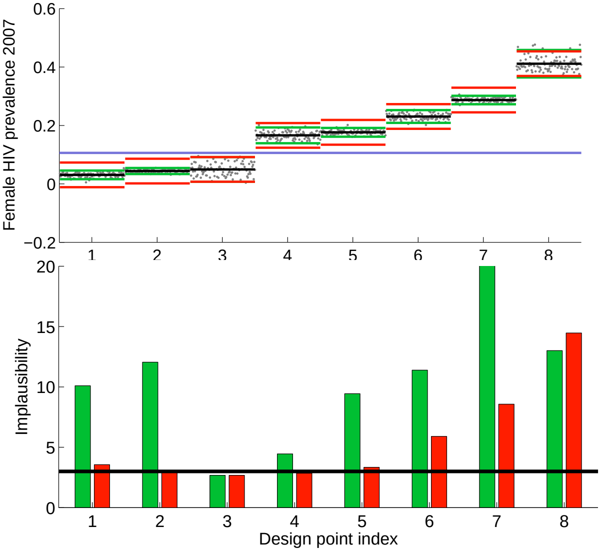


The variance $s(\mathbf{x})$

Estimating the variance can improve the results.



An example



0.6
0.4
0.2
0
−0.2
Female HIV prevalence 2007
1
2
3
4
5
6
7
8
20
15
10
5
0
Implausibility
Design point index

Emulation

- Evaluate $f(\mathbf{x})$ K times at each of N different inputs $\mathbf{x}_n$.
- Calculate means and log-variances

$$\hat{g}(\mathbf{x}_n) = \frac{1}{k}\sum_{k=1}^{K} f_k(\mathbf{x}_n), \quad \hat{\xi}(\mathbf{x}_n) = \ln\left(\frac{1}{K-1}\sum_{k=1}^{K}(f_k(\mathbf{x}_n) - \hat{g}(\mathbf{x}_n))^2\right).$$

- Gather training data $D = \{\mathbf{x}_n, \hat{g}(\mathbf{x}_n)\}$ and $D' = \{\mathbf{x}_n, \hat{\xi}(\mathbf{x}_n)\}$.
- Use a GP prior on g and ξ, i.e. $g(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$, $\xi(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$.
- Calculate posteriors: $\mathrm{E}^*[g(\mathbf{x})]$, $\mathrm{Var}^*[g(\mathbf{x})]$, $\mathrm{E}^*[\xi(\mathbf{x})]$ and $\mathrm{Var}^*[\xi(\mathbf{x})]$.

Emulation

- Evaluate $f(\mathbf{x})$ K times at each of N different inputs $\mathbf{x}_n$.
- Calculate means and log-variances

$$\hat{g}(\mathbf{x}_n) = \frac{1}{k}\sum_{k=1}^{K} f_k(\mathbf{x}_n), \quad \hat{\xi}(\mathbf{x}_n) = \ln\left(\frac{1}{K-1}\sum_{k=1}^{K}(f_k(\mathbf{x}_n) - \hat{g}(\mathbf{x}_n))^2\right).$$

- Gather training data $D = \{\mathbf{x}_n, \hat{g}(\mathbf{x}_n)\}$ and $D' = \{\mathbf{x}_n, \hat{\xi}(\mathbf{x}_n)\}$.
- Use a GP prior on g and ξ, i.e. $g(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$, $\xi(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$.
- Calculate posteriors: $\mathrm{E}^*[g(\mathbf{x})]$, $\mathrm{Var}^*[g(\mathbf{x})]$, $\mathrm{E}^*[\xi(\mathbf{x})]$ and $\mathrm{Var}^*[\xi(\mathbf{x})]$.

Emulation

- Evaluate $f(\mathbf{x})$ K times at each of N different inputs $\mathbf{x}_n$.
- Calculate means and log-variances

$$\hat{g}(\mathbf{x}_n) = \frac{1}{k}\sum_{k=1}^{K} f_k(\mathbf{x}_n), \quad \hat{\xi}(\mathbf{x}_n) = \ln\left(\frac{1}{K-1}\sum_{k=1}^{K}(f_k(\mathbf{x}_n) - \hat{g}(\mathbf{x}_n))^2\right).$$

- Gather training data $D = \{\mathbf{x}_n, \hat{g}(\mathbf{x}_n)\}$ and $D' = \{\mathbf{x}_n, \hat{\xi}(\mathbf{x}_n)\}$.
- Use a GP prior on g and ξ, i.e. $g(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$, $\xi(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$.
- Calculate posteriors: $\mathrm{E}^*[g(\mathbf{x})]$, $\mathrm{Var}^*[g(\mathbf{x})]$, $\mathrm{E}^*[\xi(\mathbf{x})]$ and $\mathrm{Var}^*[\xi(\mathbf{x})]$.

Emulation

- Evaluate $f(\mathbf{x})$ K times at each of N different inputs $\mathbf{x}_n$.
- Calculate means and log-variances

$$\hat{g}(\mathbf{x}_n) = \frac{1}{k}\sum_{k=1}^{K} f_k(\mathbf{x}_n), \quad \hat{\xi}(\mathbf{x}_n) = \ln\left(\frac{1}{K-1}\sum_{k=1}^{K}(f_k(\mathbf{x}_n) - \hat{g}(\mathbf{x}_n))^2\right).$$

- Gather training data $D = \{\mathbf{x}_n, \hat{g}(\mathbf{x}_n)\}$ and $D' = \{\mathbf{x}_n, \hat{\xi}(\mathbf{x}_n)\}$.
- Use a GP prior on g and ξ, i.e. $g(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$, $\xi(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$.
- Calculate posteriors: $\mathrm{E}^*[g(\mathbf{x})]$, $\mathrm{Var}^*[g(\mathbf{x})]$, $\mathrm{E}^*[\xi(\mathbf{x})]$ and $\mathrm{Var}^*[\xi(\mathbf{x})]$.

Emulation

- Evaluate $f(\mathbf{x})$ K times at each of N different inputs $\mathbf{x}_n$.
- Calculate means and log-variances

$$\hat{g}(\mathbf{x}_n) = \frac{1}{k}\sum_{k=1}^{K} f_k(\mathbf{x}_n), \quad \hat{\xi}(\mathbf{x}_n) = \ln\left(\frac{1}{K-1}\sum_{k=1}^{K}(f_k(\mathbf{x}_n) - \hat{g}(\mathbf{x}_n))^2\right).$$

- Gather training data $D = \{\mathbf{x}_n, \hat{g}(\mathbf{x}_n)\}$ and $D' = \{\mathbf{x}_n, \hat{\xi}(\mathbf{x}_n)\}$.
- Use a GP prior on g and ξ, i.e. $g(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$, $\xi(\mathbf{x}) \sim \mathcal{GP}(\cdot,\cdot)$.
- Calculate posteriors: $\mathrm{E}^*[g(\mathbf{x})]$, $\mathrm{Var}^*[g(\mathbf{x})]$, $\mathrm{E}^*[\xi(\mathbf{x})]$ and $\mathrm{Var}^*[\xi(\mathbf{x})]$.

Some emulator details

- $g(\mathbf{x}), \xi(\mathbf{x}) \sim \mathcal{GP}(h(\mathbf{x})\beta, \sigma^2(c(\mathbf{x}, \mathbf{x}') + \nu))$
- $h(\mathbf{x}) = 1 + \mathbf{x} + \mathbf{x}^2 + \mathbf{x}^3$.
- $c(\mathbf{x}, \mathbf{x}')$ is the Matérn 3/2 correlation function.
- β, σ^2 are marginalised with $p(\beta, \sigma^2) \propto \sigma^{-2}$.
- Correlation lengths δ and nugget ν are estimated with maximum likelihood.
- All outputs are emulated independently.

Some emulator details

- $g(\mathbf{x}), \xi(\mathbf{x}) \sim \mathcal{GP}(h(\mathbf{x})\beta, \sigma^2(c(\mathbf{x}, \mathbf{x}') + \nu))$
- $h(\mathbf{x}) = 1 + \mathbf{x} + \mathbf{x}^2 + \mathbf{x}^3$.
- $c(\mathbf{x}, \mathbf{x}')$ is the Matérn 3/2 correlation function.
- β, σ^2 are marginalised with $p(\beta, \sigma^2) \propto \sigma^{-2}$.
- Correlation lengths δ and nugget ν are estimated with maximum likelihood.
- All outputs are emulated independently.

Some emulator details

- $g(\mathbf{x}), \xi(\mathbf{x}) \sim \mathcal{GP}(h(\mathbf{x})\beta, \sigma^2(c(\mathbf{x}, \mathbf{x}') + \nu))$
- $h(\mathbf{x}) = 1 + \mathbf{x} + \mathbf{x}^2 + \mathbf{x}^3$.
- $c(\mathbf{x}, \mathbf{x}')$ is the Matérn 3/2 correlation function.
- β, σ^2 are marginalised with $p(\beta, \sigma^2) \propto \sigma^{-2}$.
- Correlation lengths δ and nugget ν are estimated with maximum likelihood.
- All outputs are emulated independently.

Some emulator details

- $g(\mathbf{x}), \xi(\mathbf{x}) \sim \mathcal{GP}(h(\mathbf{x})\beta, \sigma^2(c(\mathbf{x}, \mathbf{x}') + \nu))$
- $h(\mathbf{x}) = 1 + \mathbf{x} + \mathbf{x}^2 + \mathbf{x}^3$.
- $c(\mathbf{x}, \mathbf{x}')$ is the Matérn 3/2 correlation function.
- β, σ^2 are marginalised with $p(\beta, \sigma^2) \propto \sigma^{-2}$.
- Correlation lengths δ and nugget ν are estimated with maximum likelihood.
- All outputs are emulated independently.

Some emulator details

- $g(\mathbf{x}), \xi(\mathbf{x}) \sim \mathcal{GP}(h(\mathbf{x})\beta, \sigma^2(c(\mathbf{x}, \mathbf{x}') + \nu))$
- $h(\mathbf{x}) = 1 + \mathbf{x} + \mathbf{x}^2 + \mathbf{x}^3$.
- $c(\mathbf{x}, \mathbf{x}')$ is the Matérn 3/2 correlation function.
- β, σ^2 are marginalised with $p(\beta, \sigma^2) \propto \sigma^{-2}$.
- Correlation lengths δ and nugget ν are estimated with maximum likelihood.
- All outputs are emulated independently.

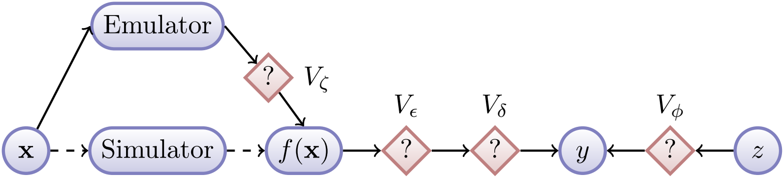
Some emulator details

- $g(\mathbf{x}), \xi(\mathbf{x}) \sim \mathcal{GP}(h(\mathbf{x})\beta, \sigma^2(c(\mathbf{x}, \mathbf{x}') + \nu))$
- $h(\mathbf{x}) = 1 + \mathbf{x} + \mathbf{x}^2 + \mathbf{x}^3$.
- $c(\mathbf{x}, \mathbf{x}')$ is the Matérn 3/2 correlation function.
- β, σ^2 are marginalised with $p(\beta, \sigma^2) \propto \sigma^{-2}$.
- Correlation lengths δ and nugget ν are estimated with maximum likelihood.
- All outputs are emulated independently.

Outline

- Motivation.
- Emulation.
- History matching.
- Results.

Uncertainty structure



z : Observations

y : Physical process

$f(\mathbf{x})$: Simulator's output

$\mathbf{x}$: Simulator's input

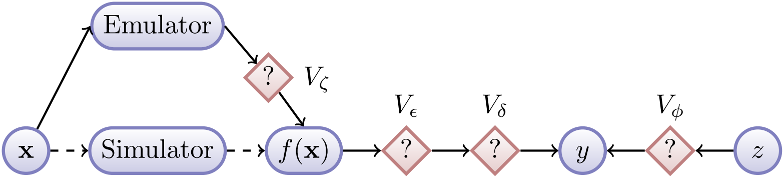
V_ζ: Code Uncertainty

V_δ: Model Discrepancy

V_ϵ: Ensemble Variability

V_ϕ: Observation Uncertainty

Uncertainty structure



- We link z with the posterior expectation of $g(\mathbf{x})$'s emulator via

$$z = \mathrm{E}^*[g(\mathbf{x})] + \zeta + \epsilon + \delta + \phi$$

where $\zeta, \epsilon, \delta, \phi$ are zero mean unimodal random variables.

- The variances V_δ, V_ϕ are provided by the model experts/data.
- $V_\zeta(\mathbf{x}) = \mathrm{Var}^*[g(\mathbf{x})]$ and $V_\epsilon(\mathbf{x}) = \exp(\mathrm{E}^*[\xi(\mathbf{x})])$.

The implausibility measure

- The link function $z = \mathrm{E}^*[g(\mathbf{x})] + \zeta + \epsilon + \delta + \phi$ allows us to write the implausibility measure as:

$$I(\mathbf{x}) = \frac{|z - \mathrm{E}^*[g(\mathbf{x})])|}{(V_\zeta(\mathbf{x}) + V_\epsilon(\mathbf{x}) + V_\delta + V_\phi)^{1/2}}$$

- A large value of $I(\mathbf{x})$, indicates that $\mathbf{x}$ is unlikely to result in a good match between the model and the data.
- A small value of $\mathbf{x}$ does not imply that $\mathbf{x}$ is good! We do not know yet.
- The magnitude of $I(\mathbf{x})$ is often judged based on Pukelsheim's 3σ rule.
- The use of emulators allows to evaluate the implausibility almost instantaneously.

The implausibility measure

- The link function $z = \mathrm{E}^*[g(\mathbf{x})] + \zeta + \epsilon + \delta + \phi$ allows us to write the implausibility measure as:

$$I(\mathbf{x}) = \frac{|z - \mathrm{E}^*[g(\mathbf{x})])|}{(V_\zeta(\mathbf{x}) + V_\epsilon(\mathbf{x}) + V_\delta + V_\phi)^{1/2}}$$

- A large value of $I(\mathbf{x})$, indicates that $\mathbf{x}$ is unlikely to result in a good match between the model and the data.

- A small value of $\mathbf{x}$ does not imply that $\mathbf{x}$ is good! We do not know yet.

- The magnitude of $I(\mathbf{x})$ is often judged based on Pukelsheim's 3σ rule.

- The use of emulators allows to evaluate the implausibility almost instantaneously.

The implausibility measure

- The link function $z = \mathrm{E}^*[g(\mathbf{x})] + \zeta + \epsilon + \delta + \phi$ allows us to write the implausibility measure as:

$$I(\mathbf{x}) = \frac{|z - \mathrm{E}^*[g(\mathbf{x})])|}{(V_\zeta(\mathbf{x}) + V_\epsilon(\mathbf{x}) + V_\delta + V_\phi)^{1/2}}$$

- A large value of $I(\mathbf{x})$, indicates that $\mathbf{x}$ is unlikely to result in a good match between the model and the data.
- A small value of $\mathbf{x}$ does not imply that $\mathbf{x}$ is good! We do not know yet.
- The magnitude of $I(\mathbf{x})$ is often judged based on Pukelsheim's 3σ rule.
- The use of emulators allows to evaluate the implausibility almost instantaneously.

The implausibility measure

- The link function $z = \mathrm{E}^*[g(\mathbf{x})] + \zeta + \epsilon + \delta + \phi$ allows us to write the implausibility measure as:

$$I(\mathbf{x}) = \frac{|z - \mathrm{E}^*[g(\mathbf{x})])|}{(V_\zeta(\mathbf{x}) + V_\epsilon(\mathbf{x}) + V_\delta + V_\phi)^{1/2}}$$

- A large value of $I(\mathbf{x})$, indicates that $\mathbf{x}$ is unlikely to result in a good match between the model and the data.

- A small value of $\mathbf{x}$ does not imply that $\mathbf{x}$ is good! We do not know yet.

- The magnitude of $I(\mathbf{x})$ is often judged based on Pukelsheim's 3σ rule.

- The use of emulators allows to evaluate the implausibility almost instantaneously.

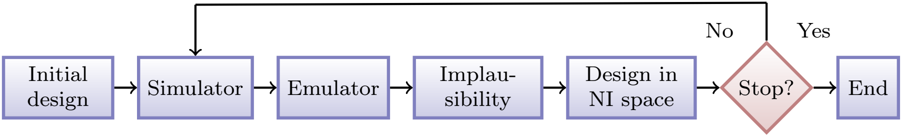
The implausibility measure

- The link function $z = \mathrm{E}^*[g(\mathbf{x})] + \zeta + \epsilon + \delta + \phi$ allows us to write the implausibility measure as:

$$I(\mathbf{x}) = \frac{|z - \mathrm{E}^*[g(\mathbf{x})])|}{(V_\zeta(\mathbf{x}) + V_\epsilon(\mathbf{x}) + V_\delta + V_\phi)^{1/2}}$$

- A large value of $I(\mathbf{x})$, indicates that $\mathbf{x}$ is unlikely to result in a good match between the model and the data.
- A small value of $\mathbf{x}$ does not imply that $\mathbf{x}$ is good! We do not know yet.
- The magnitude of $I(\mathbf{x})$ is often judged based on Pukelsheim's 3σ rule.
- The use of emulators allows to evaluate the implausibility almost instantaneously.

History matching procedure



Outline

- Motivation.
- Emulation.
- History matching.
- Results.

Results

- We compare results of the fixed variance and emulated variance approaches.
- At the end of both history matches, the non-implausible samples had a 70% probability to match all outputs.
- Emulating the variance required 3 fewer waves and 43% fewer simulator evaluations.

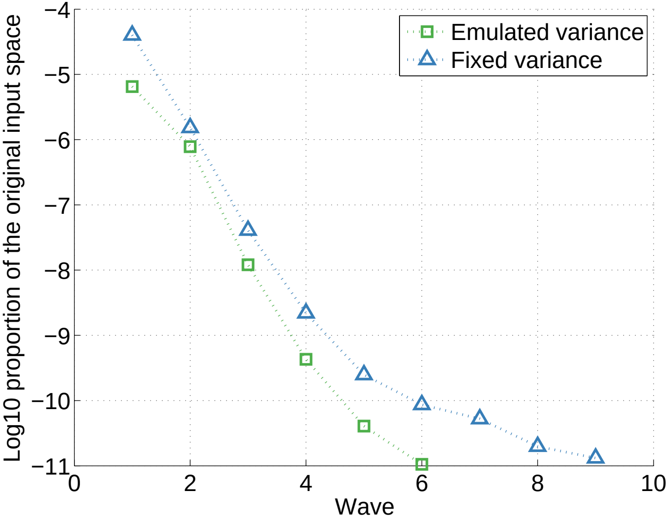
Results

- We compare results of the fixed variance and emulated variance approaches.
- At the end of both history matches, the non-implausible samples had a 70% probability to match all outputs.
- Emulating the variance required 3 fewer waves and 43% fewer simulator evaluations.

Results

- We compare results of the fixed variance and emulated variance approaches.
- At the end of both history matches, the non-implausible samples had a 70% probability to match all outputs.
- Emulating the variance required 3 fewer waves and 43% fewer simulator evaluations.

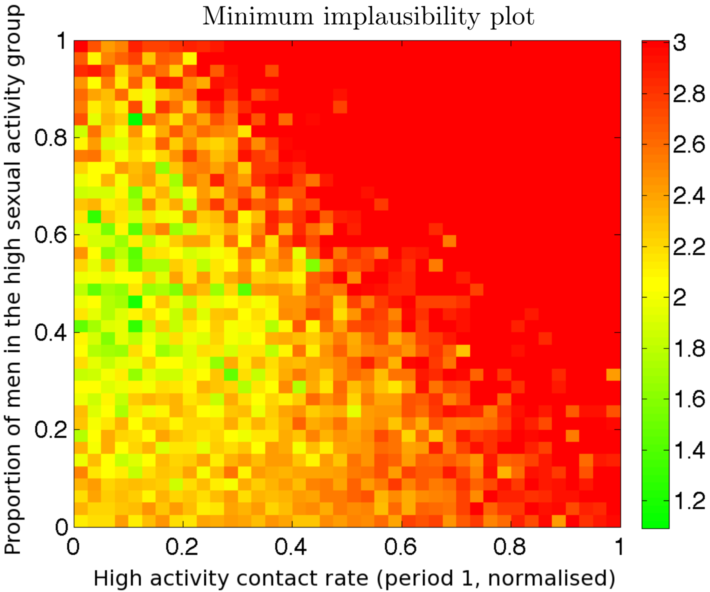
Rejection rates



Visualising the implausible space

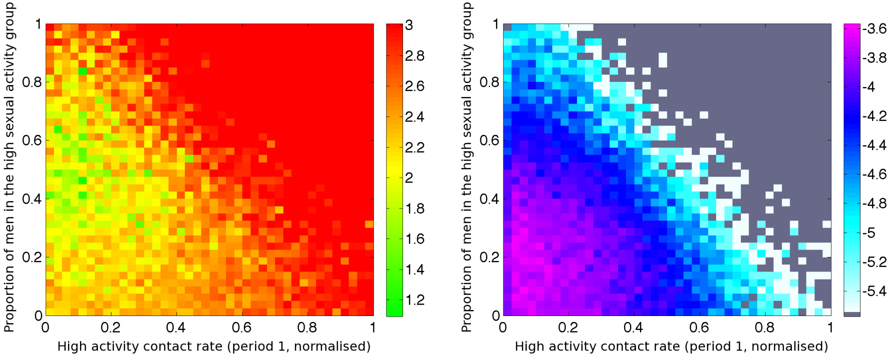
The implausible space can be visualised with minimum implausibility and optical depth plots.

Visualising the implausible space

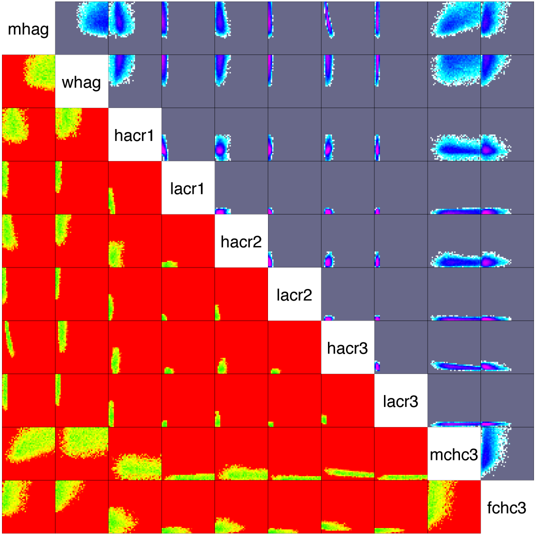


Visualising the implausible space

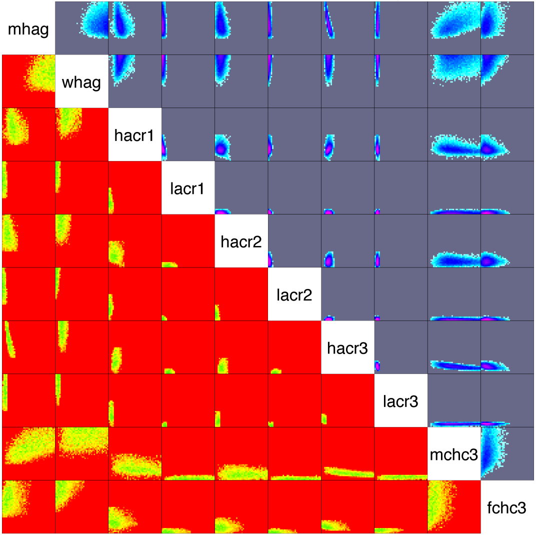
Optical depth plot



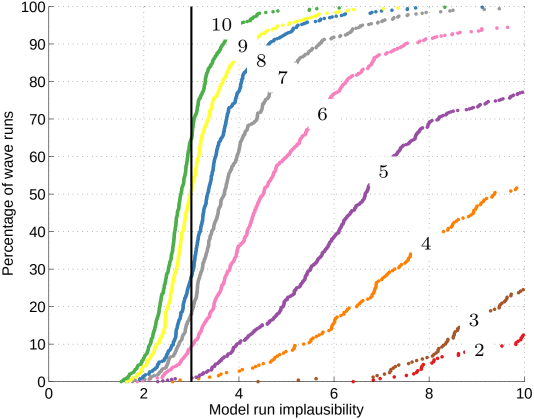
Implausibility plots wave 9 (fixed variance)



Implausibility plots wave 6



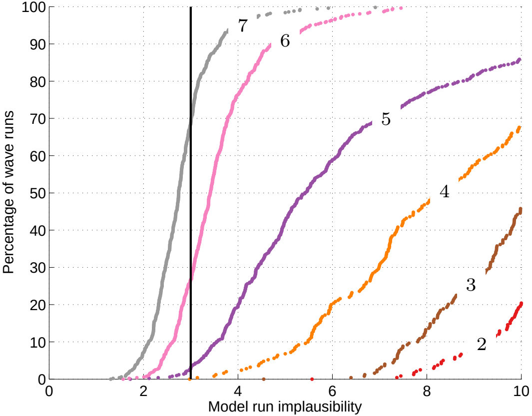
Implausibility of model runs (fixed variance)



Implausibility of the *actual* simulator runs (no emulation involved)

$$I(\mathbf{x}) = \frac{|z - \hat{g}(\mathbf{x})|}{(\hat{s}(\mathbf{x}) + V_\delta + V_\phi)^{1/2}}$$

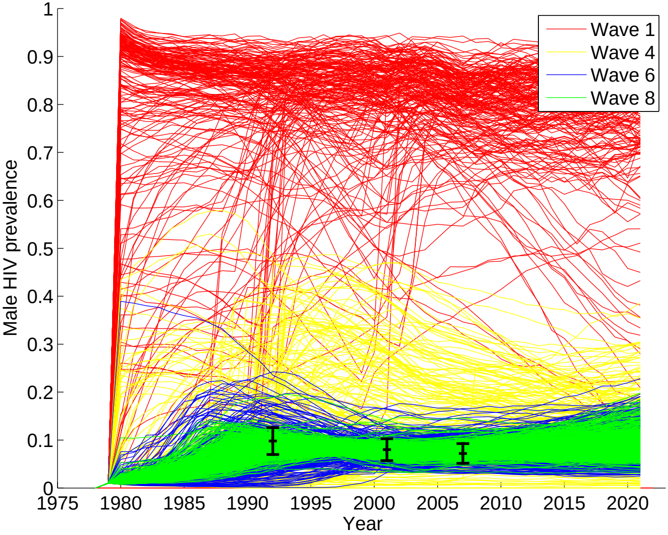
Implausibility of model runs



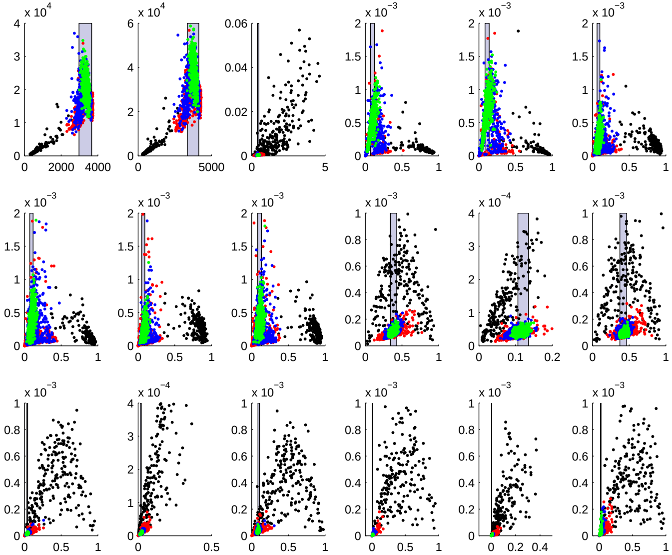
Implausibility of the *actual* simulator runs (no emulation involved)

$$I(\mathbf{x}) = \frac{|z - \hat{g}(\mathbf{x})|}{(\hat{s}(\mathbf{x}) + V_\delta + V_\phi)^{1/2}}$$

Output matching



Output matching



Conclusion

- We extended history matching so that it can efficiently handle stochastic models.
- The mean and variance of the simulator's output were both emulated using a Gaussian process.
- The simulator was calibrated in 6 waves instead of 9, requiring ~ 2000 simulator evaluations instead of ~ 3500.
- Linear regression models can be used for the variance instead of full GP ones (Boukouvalas 2014).
- Variance emulation can inform the number of replications needed at each design point.

Conclusion

- We extended history matching so that it can efficiently handle stochastic models.
- The mean and variance of the simulator's output were both emulated using a Gaussian process.
- The simulator was calibrated in 6 waves instead of 9, requiring ~ 2000 simulator evaluations instead of ~ 3500.
- Linear regression models can be used for the variance instead of full GP ones (Boukouvalas 2014).
- Variance emulation can inform the number of replications needed at each design point.

Conclusion

- We extended history matching so that it can efficiently handle stochastic models.
- The mean and variance of the simulator's output were both emulated using a Gaussian process.
- The simulator was calibrated in 6 waves instead of 9, requiring ~ 2000 simulator evaluations instead of ~ 3500.
- Linear regression models can be used for the variance instead of full GP ones (Boukouvalas 2014).
- Variance emulation can inform the number of replications needed at each design point.

Conclusion

- We extended history matching so that it can efficiently handle stochastic models.
- The mean and variance of the simulator's output were both emulated using a Gaussian process.
- The simulator was calibrated in 6 waves instead of 9, requiring ~ 2000 simulator evaluations instead of ~ 3500.
- Linear regression models can be used for the variance instead of full GP ones (Boukouvalas 2014).
- Variance emulation can inform the number of replications needed at each design point.

Conclusion

- We extended history matching so that it can efficiently handle stochastic models.
- The mean and variance of the simulator's output were both emulated using a Gaussian process.
- The simulator was calibrated in 6 waves instead of 9, requiring ~ 2000 simulator evaluations instead of ~ 3500.
- Linear regression models can be used for the variance instead of full GP ones (Boukouvalas 2014).
- Variance emulation can inform the number of replications needed at each design point.

Mukwano 2.0

- Our epidemiologists came up with a larger version of Mukwano.(96 inputs, 50 outputs)
- Simulates HIV transmission under various ART treatment strategies.
- Predicts the effects of ART on mortality and transmission over the next 15-20 years.
- A simplified version of history matching was applied.
- History matching was carried out using linear regression instead of emulators.
- This work was submitted in the Journal of Uncertainty Quantification (Sep. 2016).

Mukwano 2.0

- Our epidemiologists came up with a larger version of Mukwano.(96 inputs, 50 outputs)
- Simulates HIV transmission under various ART treatment strategies.
- Predicts the effects of ART on mortality and transmission over the next 15-20 years.
- A simplified version of history matching was applied.
- History matching was carried out using linear regression instead of emulators.
- This work was submitted in the Journal of Uncertainty Quantification (Sep. 2016).

Mukwano 2.0

- Our epidemiologists came up with a larger version of Mukwano.(96 inputs, 50 outputs)
- Simulates HIV transmission under various ART treatment strategies.
- Predicts the effects of ART on mortality and transmission over the next 15-20 years.
- A simplified version of history matching was applied.
- History matching was carried out using linear regression instead of emulators.
- This work was submitted in the Journal of Uncertainty Quantification (Sep. 2016).

Mukwano 2.0

- Our epidemiologists came up with a larger version of Mukwano.(96 inputs, 50 outputs)
- Simulates HIV transmission under various ART treatment strategies.
- Predicts the effects of ART on mortality and transmission over the next 15-20 years.
- A simplified version of history matching was applied.
- History matching was carried out using linear regression instead of emulators.
- This work was submitted in the Journal of Uncertainty Quantification (Sep. 2016).

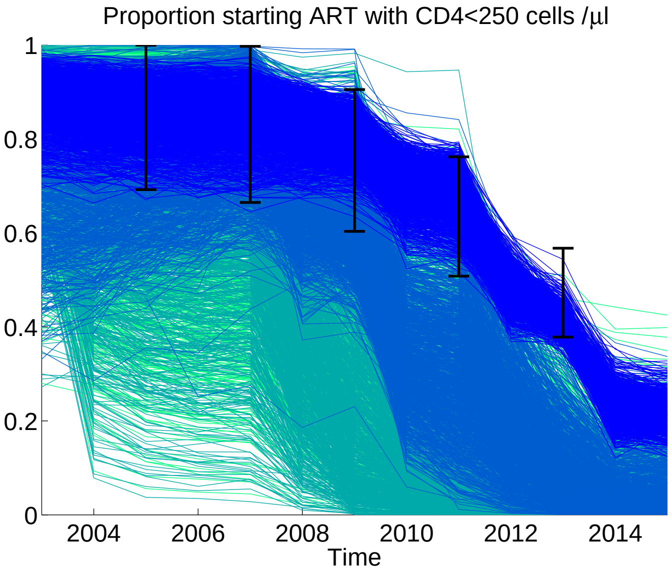
Mukwano 2.0

- Our epidemiologists came up with a larger version of Mukwano.(96 inputs, 50 outputs)
- Simulates HIV transmission under various ART treatment strategies.
- Predicts the effects of ART on mortality and transmission over the next 15-20 years.
- A simplified version of history matching was applied.
- History matching was carried out using linear regression instead of emulators.
- This work was submitted in the Journal of Uncertainty Quantification (Sep. 2016).

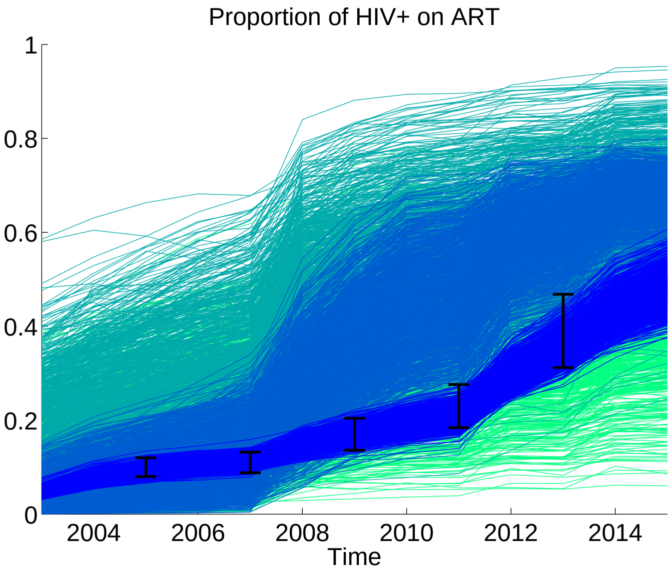
Mukwano 2.0

- Our epidemiologists came up with a larger version of Mukwano.(96 inputs, 50 outputs)
- Simulates HIV transmission under various ART treatment strategies.
- Predicts the effects of ART on mortality and transmission over the next 15-20 years.
- A simplified version of history matching was applied.
- History matching was carried out using linear regression instead of emulators.
- This work was submitted in the Journal of Uncertainty Quantification (Sep. 2016).

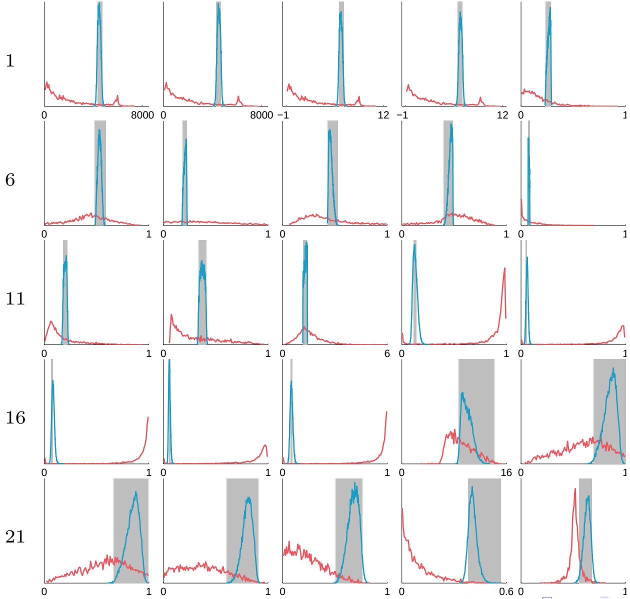
Output matching



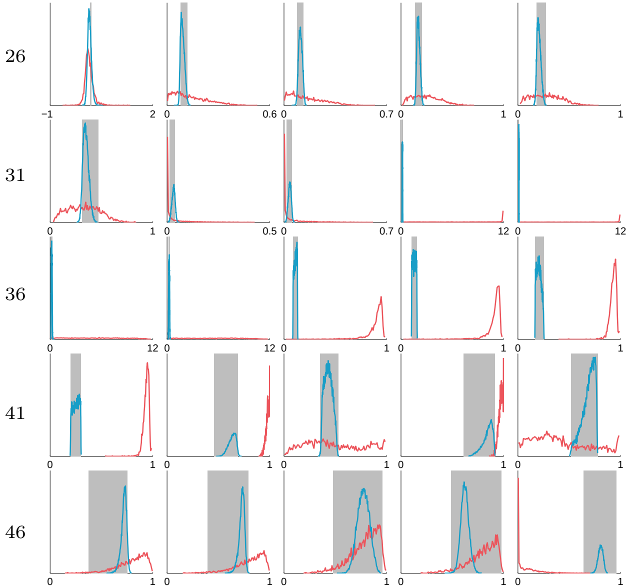
Output matching



Output matching



Output matching



Conclusion

- History matching provided hundreds of input points that match all the outputs simultaneously.
- These inputs are used to run the simulator into the future and predict the effect of different ART interventions to mortality, HIV prevalence etc.
- It allows incorporating in the predictions the uncertainty about the values of the input parameters.
- The results feed into a number of other research projects that quantify the effect of different ART deployment strategies, costs, etc.

Conclusion

- History matching provided hundreds of input points that match all the outputs simultaneously.
- These inputs are used to run the simulator into the future and predict the effect of different ART interventions to mortality, HIV prevalence etc.
- It allows incorporating in the predictions the uncertainty about the values of the input parameters.
- The results feed into a number of other research projects that quantify the effect of different ART deployment strategies, costs, etc.

Conclusion

- History matching provided hundreds of input points that match all the outputs simultaneously.
- These inputs are used to run the simulator into the future and predict the effect of different ART interventions to mortality, HIV prevalence etc.
- It allows incorporating in the predictions the uncertainty about the values of the input parameters.
- The results feed into a number of other research projects that quantify the effect of different ART deployment strategies, costs, etc.

Conclusion

- History matching provided hundreds of input points that match all the outputs simultaneously.
- These inputs are used to run the simulator into the future and predict the effect of different ART interventions to mortality, HIV prevalence etc.
- It allows incorporating in the predictions the uncertainty about the values of the input parameters.
- The results feed into a number of other research projects that quantify the effect of different ART deployment strategies, costs, etc.