# New Insights into History Matching via Sequential Monte Carlo

Associate Professor Chris Drovandi

School of Mathematical Sciences
ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)
Queensland University of Technology

Collaborators: David Nott (National University of Singapore) and Dan Pagendam (CSIRO Data 61)

# Parameter Estimation

Denote $\theta$ as the parameter of a deterministic model $M(\theta)$ or stochastic model $M(\theta, u)$ where $u$ are the random numbers needed in the simulation.

Have some observed data $y$ believed to be generated by $M$.

Wish to obtain a collection of plausible $\theta$ based on $y$. Eg a Bayesian posterior, $\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$.

However, the model $M$ is too expensive to use say Markov chain Monte Carlo or sequential Monte Carlo.

# History Matching

Aim: Determine ‘plausible’ region of the parameter space relatively quickly.

History matching (eg Craig et al 1997) determines a non-implausible region of the parameter space relatively quickly by using an *emulator* of the model outputs or ‘distance’ between model outputs and $y$.

The emulator not only provides predictions at untrained points but also quantifies the uncertainty in the predictions.

# History Matching Notation/Details

Denote $\pi(\theta)$ as a 'prior' distribution.

For simplicity we will assume there is only $n = 1$ model output, eg:

- The model really does have only 1 output
- The output is a distance between outputs and data.
- The output is a (approximate) likelihood function.

Parameter $\theta$ is deemed as non-implausible if $\mathcal{I}(\theta) < c$ for cut-off $c$ where $\mathcal{I}(\theta)$ includes emulation uncertainty, eg Andrianakis et al 2015

$$\mathcal{I}(\theta) = \frac{|y_\theta - y_{\mathrm{obs}}|}{\sqrt{s^2_{m,\theta} + s^2_{e,\theta} + s^2_d}}. \quad (1)$$

Uncertainty: $s^2_{m,\theta}$ simulator, $s_{e,\theta}$ emulator, $s^2_d$ model.

# History Matching Procedure

Steps involved in the history matching algorithm:

1. Generate $N_w$ training samples $\{\theta_j\}_{j=1}^{N_w} \sim \pi(\theta)$ using a space filling design and simulate the model at each $\theta_j$ to generate the collection of outputs $\{y_{\theta_j}\}_{j=1}^{N_w}$.
2. Fit an emulator $E_w$ to the training data $\{\theta_j, y_{\theta_j}\}_{j=1}^{N_w}$.
3. Use the emulator $E_w$ to define an implausibility function $\mathcal{I}_w(\theta)$. If $\mathcal{I}_w(\theta) > c_w$ for some chosen $c_w$ then $\theta$ is deemed as implausible by emulator $E_w$.
4. Use all emulators $\{E_r\}_{r=1}^{w}$ to define the non-implausible region $\Theta_w = \{\theta \in \Theta | \cap_{r=1}^{w} \mathcal{I}_w(\theta) < c_w\}$.
5. Increase wave counter $w = w + 1$.
   Generate $N_w$ training samples $\{\theta_j\}_{j=1}^{N_w}$ from $\Theta_w$ and simulate the model at each $\theta_j$ to generate $\{y_{\theta_j}\}_{j=1}^{N_w}$.
6. If the stopping rule is satisfied then finish otherwise return to Line 2.

# Issues with History Matching

History matching does have at least two issues:

1. The cut-off values $c_w$ may not be easy to select in practice and there is no existing automated method for doing so.
2. Sampling uniformly from $\Theta_w$ as $w$ increases becomes increasingly difficult. History matching papers tend to “ignore" this.

Solution: We use sequential Monte Carlo (SMC, eg Chopin 2002) to help address the above 2 issues.

# Sequential Monte Carlo

SMC samples from a sequence of distributions (connecting easy and target distributions) by iteratively applying re-weighting, re-sampling and move steps.

Advantages of SMC approach over MCMC:

- Naturally adaptive
- Easily parallelisable
- More capable of dealing with multimodal or complex posterior distributions

# SMC History Matching

For history matching, we define the sequence of distributions as

$$p_w(\theta) \propto \pi(\theta) \prod_{k=1}^{w} \mathbb{I}(\mathcal{I}_k(\theta) \leq c_k).$$

Assume that we have a collection of 'particles', $\{W_w^i, \theta_w^i\}_{i=1}^M$ from $p_w(\theta)$. To push the particles to target $w+1$ we apply re-weighting step:

$$W_{w+1}^i \propto W_w^i \mathbb{I}(\mathcal{I}_{w+1}(\theta_w^i) \leq c_{w+1}).$$

Here $W_w$ are all equal, so the weights $W_{w+1}$ will either be constant or equal to zero. Thus we can select $c_{w+1}$ so that a certain proportion, $\alpha$, have non-zero weight. Same as ensuring that the effective sample size (ESS) $= 1/\sum_{i=1}^M (W_{w+1}^i)^2$, is $\alpha M$.

# SMC History Matching

After re-weighting the ESS drops to roughly $\alpha M$.

Resampling $M$ times from the surviving particles allows the ESS to return to $M$. However, the drawback is that some particles will be duplicated.

Diversify particles with an MCMC kernel $R$ times to each of the resampled particles. Can adaptively choose $R$. Can use population of particles to inform efficient MCMC proposal. This move step only uses the emulator (no expensive model simulations).

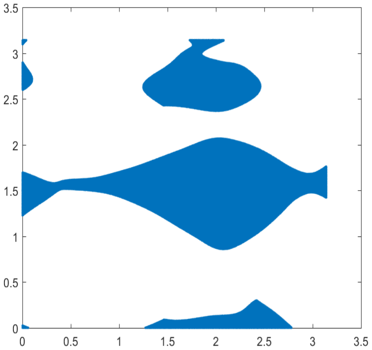
# SMC History Matching – Move Step

History matching produces irregular and multi-modal distribution, need efficient MCMC kernel.

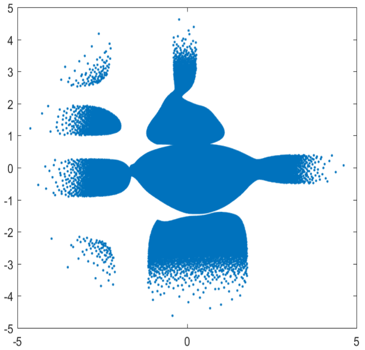
We use a kernel density estimate (kde) of each component. Using the cdf of the kde we transform each margin to roughly $\mathcal{N}(0, 1)$. Then we perform a random walk on the transformed space.

Modelling of the marginals can help bring modes closer together and produces a more regular space to perform a random walk.

# Modelling of Marginals




(a) before




(b) after

# Emulation

For emulation we use a Gaussian Process (GP). A 'prior' over functions when evaluated at a finite number of points has a multivariate normal distribution.

Characterised by a mean function $m_\beta(\boldsymbol{\theta})$ and covariance function $C_\gamma(\boldsymbol{\theta}, \boldsymbol{\theta}') = \text{cov}\{f(\boldsymbol{\theta}), f(\boldsymbol{\theta}')\}$, with hyperparameters $\beta$ and $\gamma$.

Here we take $m_\beta(\boldsymbol{\theta}) = 0$ and

$$C_\gamma(\boldsymbol{\theta}, \boldsymbol{\theta}') = \delta_c \exp\left\{-\frac{1}{2}\sum_{k=1}^{p}\frac{(\theta_k - \theta_k')^2}{r_k^2}\right\},$$

with hyperparameters $\gamma = (\delta_c, r_1, \ldots, r_p)^\top$.

# Emulation (Cont...)

Also include ‘nugget’ term:

$$\text{cov}\{\hat{f}(\boldsymbol{\theta}), \hat{f}(\boldsymbol{\theta}')\} = C_\gamma(\boldsymbol{\theta}, \boldsymbol{\theta}') + \delta 1(\boldsymbol{\theta} = \boldsymbol{\theta}')$$

Based on “training" data, $\Theta$, hyperparameters estimated via (marginal) maximum likelihood estimation. Obtain “posterior" over functions.

Can use to predict at untrained location and quantify prediction uncertainty:

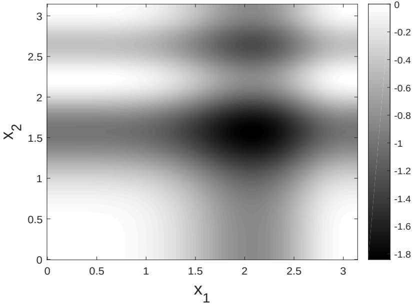
$$m^*(\boldsymbol{\theta}) = C_{\hat{\gamma}}(\Theta, \boldsymbol{\theta})^\top \{C_{\hat{\gamma}}(\Theta, \Theta) + \hat{\delta} I\}^{-1} \boldsymbol{f}(\Theta),$$
$$s^*(\boldsymbol{\theta})^2 = C_{\hat{\gamma}}(\boldsymbol{\theta}, \boldsymbol{\theta}) - C_{\hat{\gamma}}(\Theta, \boldsymbol{\theta})^\top \{C_{\hat{\gamma}}(\Theta, \Theta) + \hat{\delta} I\}^{-1} C_{\hat{\gamma}}(\Theta, \boldsymbol{\theta}).$$

# Simple Example

Consider the function

$$y = -\sin(x_1)\sin(x_1^2/\pi)^2 - \sin(x_2)\sin(2x_2^2/\pi)^2,$$

where $\theta = (x_1, x_2) \in (0,\pi) \times (0,\pi)$.
We wish to find the regions where $y$ is small.



# History Matching Parameters

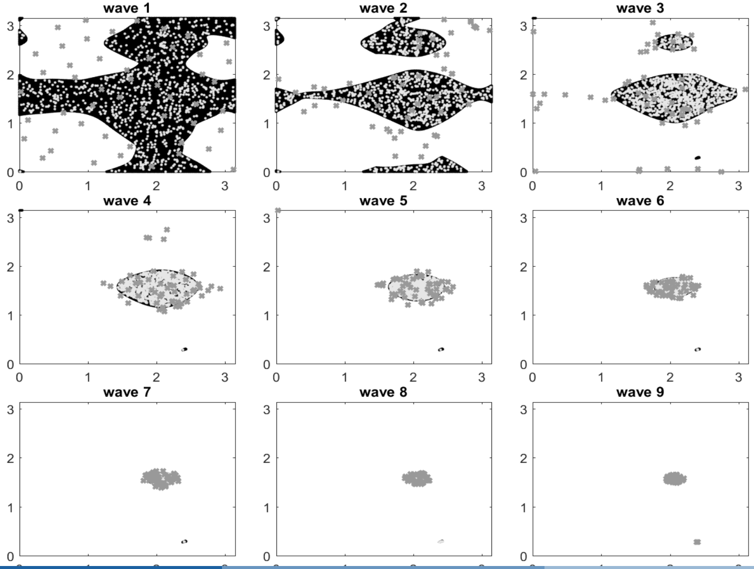
We use $N = 50$ training samples at each wave.

Implausability measure $\mathcal{I}(\theta) = y_p(\theta) - r \times s_p(\theta)$ where $y_p(\theta)$ and $s_p(\theta)$ is the prediction and the standard deviation from the currently fitted GP, respectively. Set $r = 3$.

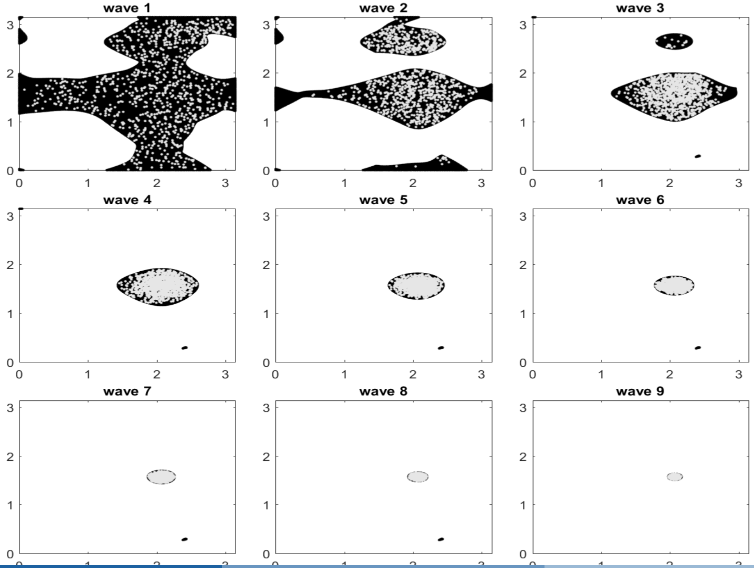
We solve the history matching problem 'perfectly' by taking $2^{20}$ draws from parameter space and use rejection sampling. The cut-offs $c_w$ are chosen so that half the remaining particles survive at each wave.

We investigate how well SMC can uniformly sample from non-implausible space implied by 'perfect' history matching.

# Results from SMC History Matching



# Results from Typical History Matching Approach



# More Challenging Example

Rainfall Run-off Model in Hydrology. 7 parameters.

Underlying model is deterministic. Output is distance between observed and “simulated" time series

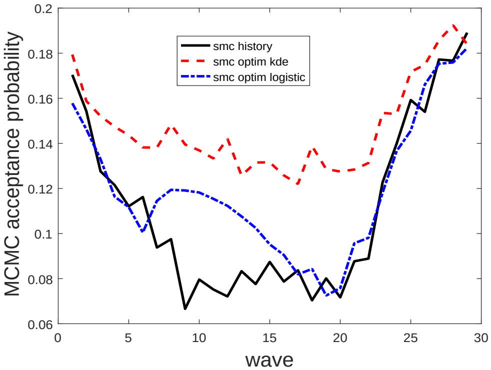
$$\rho_p(\theta) = \sum_{t=1}^{T} \frac{(y^t_{\mathrm{obs}} - y^t_\theta)^2}{y^t_{\mathrm{obs}}},$$

We compare SMC History Matching with standard SMC optimisation (no emulator):

$$p_w(\theta) \propto \pi(\theta)\mathbb{I}(\rho_p \le d_w), \text{ for } d_1 > d_2 > \cdots > d_w > \cdots,$$

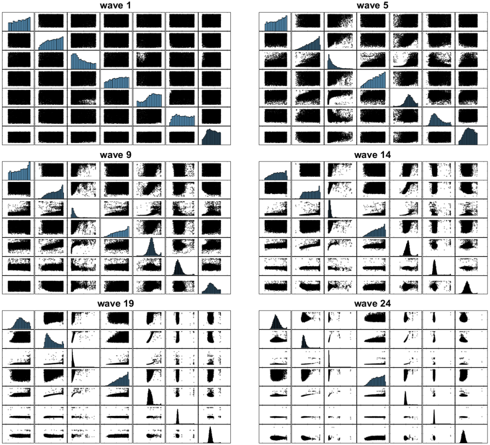
with $d_w$ determined adaptively based on quantiles of $\rho_p$.

# Results



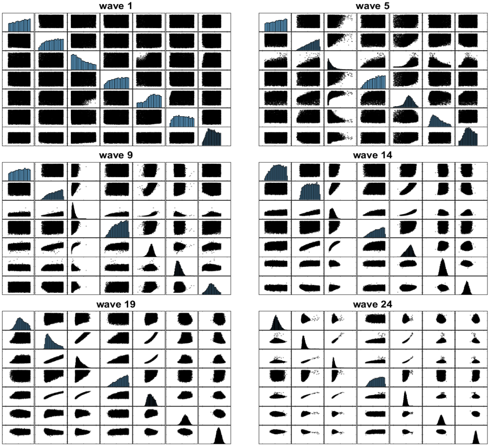

Acceptance probability of the MCMC step in SMC history matching method (solid) and SMC optimisation method (dash).

# Results




Bivariate scatterplots of the parameters with SMC history matching.

# Results




Bivariate scatterplots of the parameters for SMC optimisation.

# Results

SMC optimisation has less challenging sampling problem.

But... SMC History Matching uses 2 orders of magnitude less evaluations of expensive simulator.

# Take Home Message

History Matching is useful for finding non-implausible regions for complex simulators quickly. But...

It replaces the hard parameter calibration problem with a hard sampling problem.

But... SMC can help here, and create a more automatic history matching algorithm.

# References

Andrianakis et al (2015). Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on HIV in Uganda. PLOS Computational Biology, 11(1):e1003968.

Chopin (2002). A sequential particle filter method for static models. Biometrika, 89(3):539-552.

Craig et al (1997). Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments. In Case Studies in Bayesian Statistics, pages 37-93. Springer.

Drovandi et al (2017). New insights into history matching via sequential Monte Carlo. `https://eprints.qut.edu.au/112175/`.

Email: c.drovandi@qut.edu.au Web: chrisdrovandi@weebly.com Twitter: @chris_drovandi