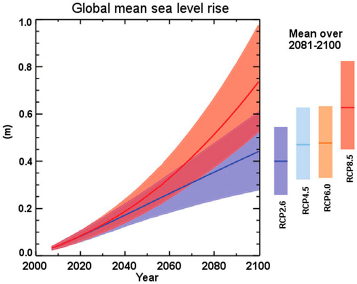
Some Bayesian methods for inverse problems

Ieva Kazlauskaitė

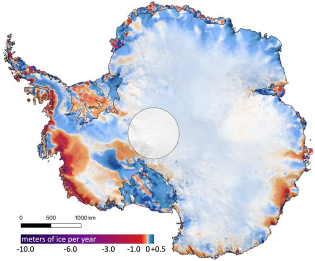
Department of Statistical Science, University College London

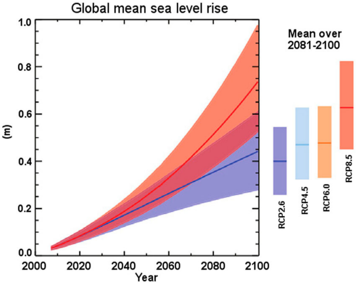
Motivation

Inverse problems & UQ in complex physical models

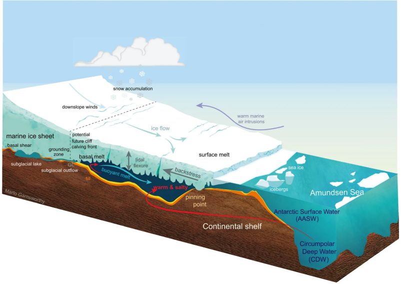


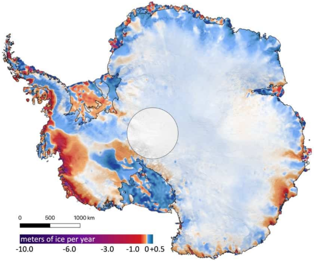
Inverse problems & UQ in complex physical models

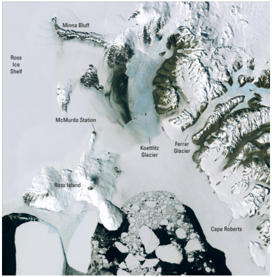


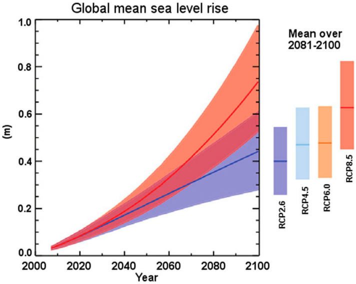


Inverse problems & UQ in complex physical models



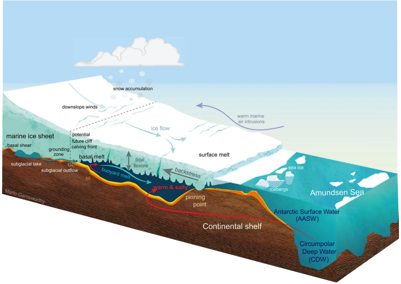


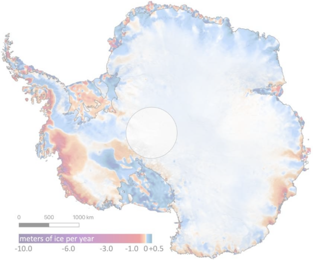


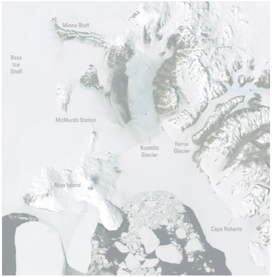


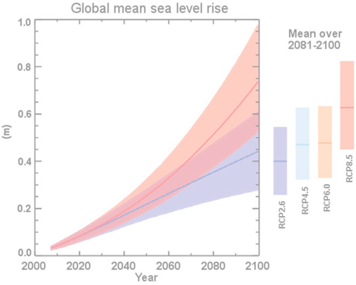
Images: BAS, NASA, B Smith et al/Science, IPCC

Inverse problems & UQ in complex physical models



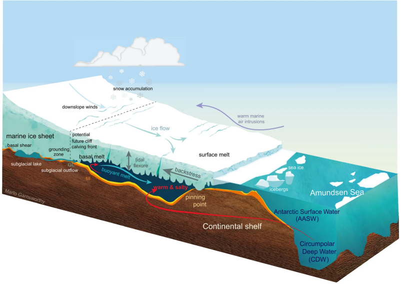


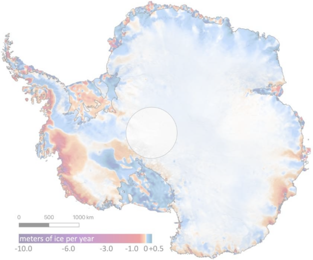


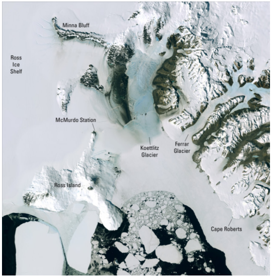


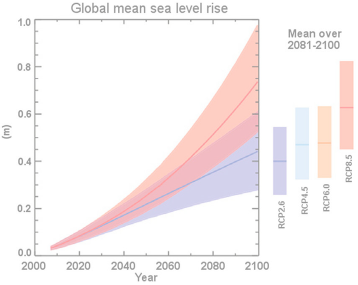
Images: BAS, NASA, B Smith et al/Science, IPCC

Inverse problems & UQ in complex physical models



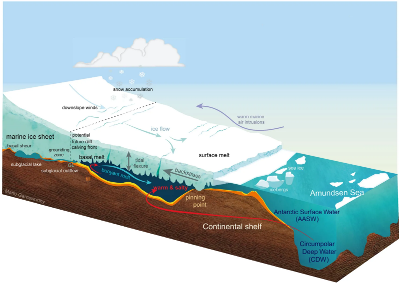


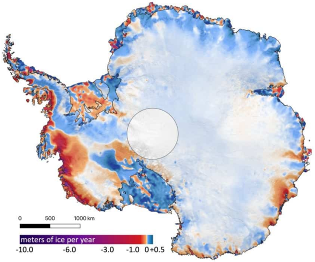


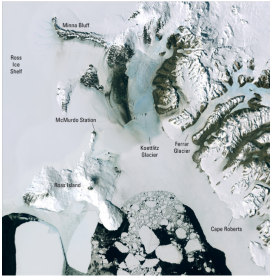


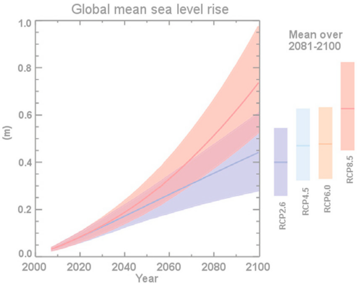
Images: BAS, NASA, B Smith et al/Science, IPCC

Inverse problems & UQ in complex physical models



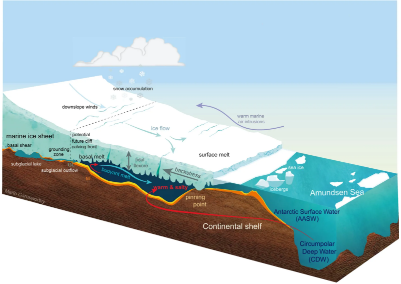


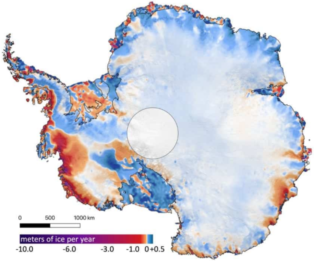


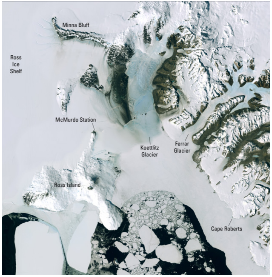


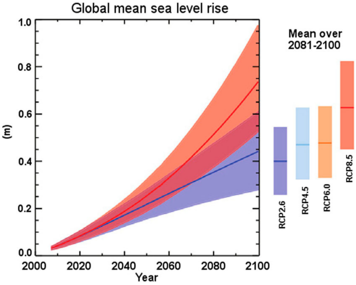
Images: BAS, NASA, B Smith et al/Science, IPCC

Inverse problems & UQ in complex physical models



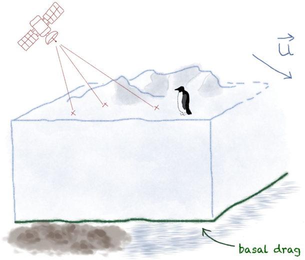




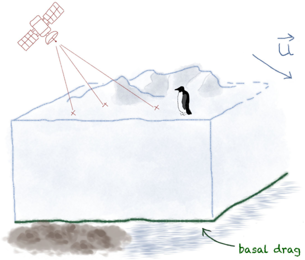


Images: BAS, NASA, B Smith et al/Science, IPCC

Inverse problems in ice sheet models



Inverse problems in ice sheet models

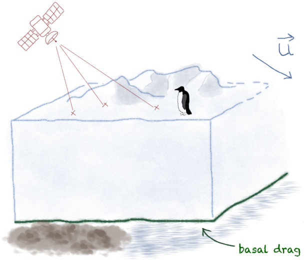


Stokes flow with Robin BC:

$$
\begin{aligned}
-\beta \Delta u + \nabla p &= f && \text{in } \mathcal{O} \\
\nabla \cdot u &= 0 && \text{in } \mathcal{O} \\
\partial_\nu u - p\nu + \theta u &= 0 && \text{on } \partial\mathcal{O}
\end{aligned}
$$

where f is forcing, β is viscocity, and θ is the unknown boundary parameter.

Inverse problems in ice sheet models

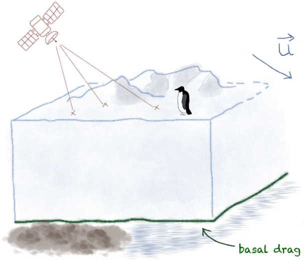


Stokes flow with Robin BC:

$$
\begin{aligned}
-\beta \Delta u + \nabla p = f & \qquad \text{in } \mathcal{O} \\
\nabla \cdot u = 0 & \qquad \text{in } \mathcal{O} \\
\partial_\nu u - p\nu + \theta u = 0 & \qquad \text{on } \partial\mathcal{O}
\end{aligned}
$$

where f is forcing, β is viscocity, and θ is the unknown boundary parameter. The forward map $\mathcal{G}_\theta : \theta \mapsto u$.

Inverse problems in ice sheet models



Stokes flow with Robin BC:

$$\begin{aligned} -\beta\Delta u + \nabla p &= f && \text{in } \mathcal{O} \\ \nabla \cdot u &= 0 && \text{in } \mathcal{O} \\ \partial_\nu u - p\nu + \theta u &= 0 && \text{on } \partial\mathcal{O} \end{aligned}$$

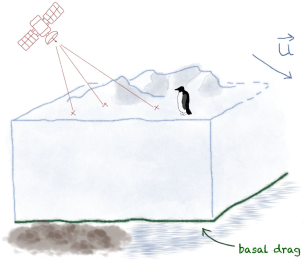
where f is forcing, β is viscocity, and θ is the unknown boundary parameter.
The forward map $\mathcal{G}_\theta : \theta \mapsto u$.

Regression problem:

$$Y_i = \mathcal{G}_\theta(X_i) + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$$

with data $D_N = (Y_i, X_i)_{i=1}^N$.

Inverse problems in ice sheet models



Stokes flow with Robin BC:

$$
\begin{aligned}
-\beta\Delta u + \nabla p &= f && \text{in } \mathcal{O}\\
\nabla \cdot u &= 0 && \text{in } \mathcal{O}\\
\partial_\nu u - p\nu + \theta u &= 0 && \text{on } \partial\mathcal{O}
\end{aligned}
$$

where f is forcing, β is viscocity, and θ is the unknown boundary parameter. The forward map $\mathcal{G}_\theta : \theta \mapsto u$.

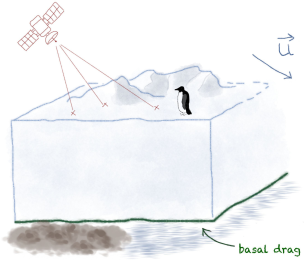
Regression problem:

$$Y_i = \mathcal{G}_\theta(X_i) + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$$

with data $D_N = (Y_i, X_i)_{i=1}^N$.

Challenges:

Inverse problems in ice sheet models



Stokes flow with Robin BC:

$$\begin{aligned} -\beta \Delta u + \nabla p &= f && \text{in } \mathcal{O} \\ \nabla \cdot u &= 0 && \text{in } \mathcal{O} \\ \partial_\nu u - p\nu + \theta u &= 0 && \text{on } \partial\mathcal{O} \end{aligned}$$

where f is forcing, β is viscocity, and θ is the unknown boundary parameter. The forward map $\mathcal{G}_\theta : \theta \mapsto u$.

Regression problem:

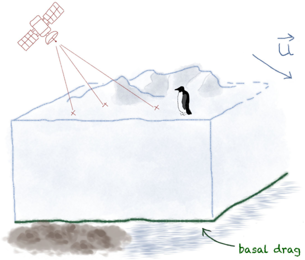
$$Y_i = \mathcal{G}_\theta(X_i) + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$$

with data $D_N = (Y_i, X_i)_{i=1}^N$.

Challenges:

- No closed form inversion formula $\mathcal{G}^{-1}$ for inferring θ from u.

Inverse problems in ice sheet models



Stokes flow with Robin BC:

$$\begin{aligned} -\beta\Delta u + \nabla p &= f && \text{in } \mathcal{O} \\ \nabla \cdot u &= 0 && \text{in } \mathcal{O} \\ \partial_\nu u - p\nu + \theta u &= 0 && \text{on } \partial\mathcal{O} \end{aligned}$$

where f is forcing, β is viscocity, and θ is the unknown boundary parameter. The forward map $\mathcal{G}_\theta : \theta \mapsto u$.

Regression problem:

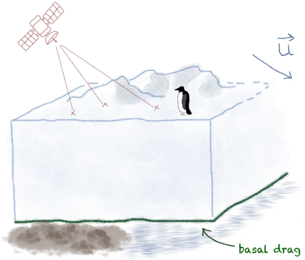
$$Y_i = \mathcal{G}_\theta(X_i) + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$$

with data $D_N = (Y_i, X_i)_{i=1}^N$.

Challenges:

- No closed form inversion formula $\mathcal{G}^{-1}$ for inferring θ from u.
- Solutions $\mathcal{G}_\theta$ for different θ available only through simulation.

Inverse problems in ice sheet models



Stokes flow with Robin BC:

$$
\begin{aligned}
-\beta\Delta u + \nabla p &= f && \text{in } \mathcal{O}\\
\nabla \cdot u &= 0 && \text{in } \mathcal{O}\\
\partial_\nu u - p\nu + \theta u &= 0 && \text{on } \partial\mathcal{O}
\end{aligned}
$$

where f is forcing, β is viscocity, and θ is the unknown boundary parameter. The forward map $\mathcal{G}_\theta : \theta \mapsto u$.

Regression problem:

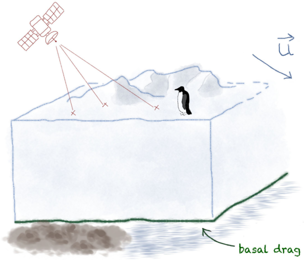
$$Y_i = \mathcal{G}_\theta(X_i) + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$$

with data $D_N = (Y_i, X_i)_{i=1}^N$.

Challenges:

- No closed form inversion formula $\mathcal{G}^{-1}$ for inferring θ from u.
- Solutions $\mathcal{G}_\theta$ for different θ available only through simulation.
- In general, D_N does not uniquely identify θ.

Inverse problems in ice sheet models



Stokes flow with Robin BC:

$$\begin{aligned} -\beta\Delta u + \nabla p = f \quad & \text{in } \mathcal{O} \\ \nabla \cdot u = 0 \quad & \text{in } \mathcal{O} \\ \partial_\nu u - p\nu + \theta u = 0 \quad & \text{on } \partial\mathcal{O} \end{aligned}$$

where f is forcing, β is viscocity, and θ is the unknown boundary parameter. The forward map $\mathcal{G}_\theta : \theta \mapsto u$.

Regression problem:

$$Y_i = \mathcal{G}_\theta(X_i) + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$$

with data $D_N = (Y_i, X_i)_{i=1}^N$.

Challenges:

- No closed form inversion formula $\mathcal{G}^{-1}$ for inferring θ from u.
- Solutions $\mathcal{G}_\theta$ for different θ available only through simulation.
- In general, D_N does not uniquely identify θ.
- $\mathcal{G}$ is 'smoothing'.

Ice sheet simulator / model / solver $G_\theta(x_i)$

$$G_\theta(x) = u$$

viscosity velocity

Observation model :

$$y_i = u_i + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma)$$

$$y_i = G_\theta(x_i) + \varepsilon_i$$

Regression problem!

Three Bayesian approaches

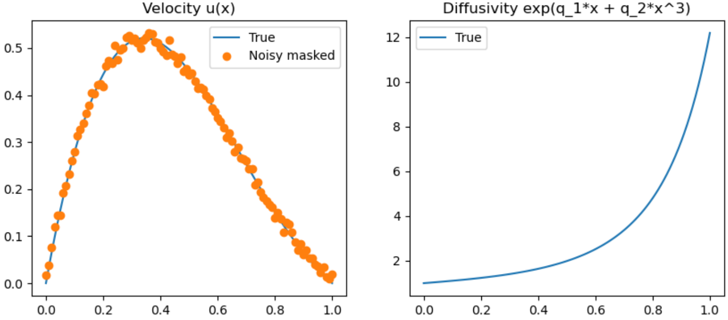
1. Bayesian approximate computation (ABC); likelihood-free inference
2. (Full) Bayesian inference
3. Bayesian optimisation (BO)

1D Poisson problem

Consider a simpler problem in 1D:

$$
\begin{aligned}
-\nabla \cdot \beta(x)\nabla u(x) &= f(x) && \text{in } \mathcal{O} \\
u(x) &= 0 && \text{on } \partial\mathcal{O}.
\end{aligned}
$$

Parameterise diffusivity as: $\beta(x) = exp(q_1 x + q_2 x^3)$.



I: Approximate Bayesian computation

for $i = 1$ **to** M **do**

Pick θ

Run $\mathcal{G}_\theta(X)$

Compare $u = \mathcal{G}_\theta(X_i)$ to observations y_i

Reject if error is large

end for

I: Approximate Bayesian computation

for $i = 1$ **to** M **do**

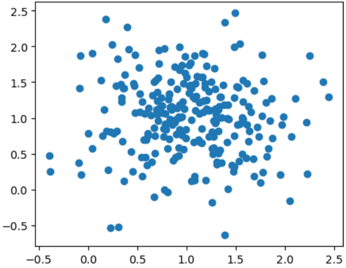
Pick $\theta \propto \text{Prior}(\theta)$

Run $\mathcal{G}_\theta(X)$

Compare $u = \mathcal{G}_\theta(X_i)$ to observations y_i

Reject if error is large

end for



I: Approximate Bayesian computation

for $i = 1$ **to** M **do**

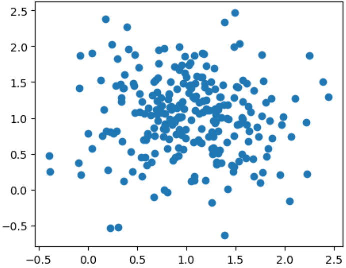
  Pick $\theta \propto \text{Prior}(\theta)$

  Run $\mathcal{G}_\theta(X)$

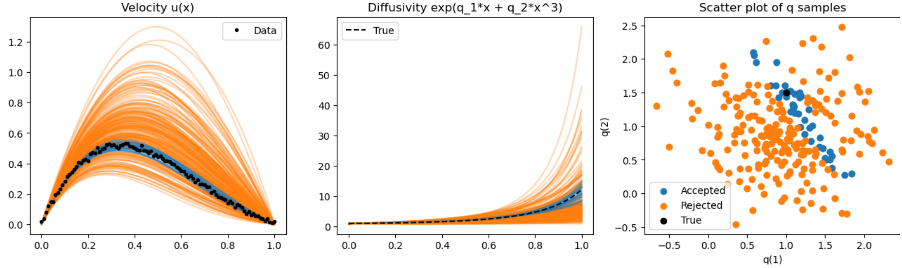
  Compare $R = |\mathcal{G}_\theta(X_i) - y_i|^2$

  Accept if $R < e_{\text{threshold}}$

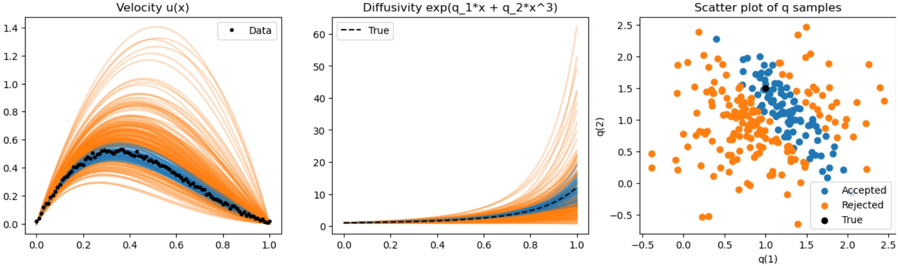
end for



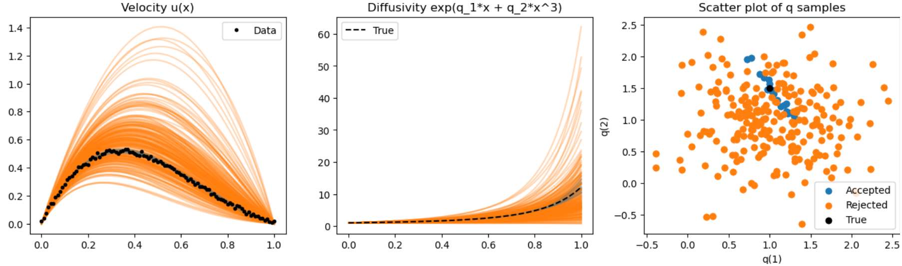
Approximate Bayesian computation



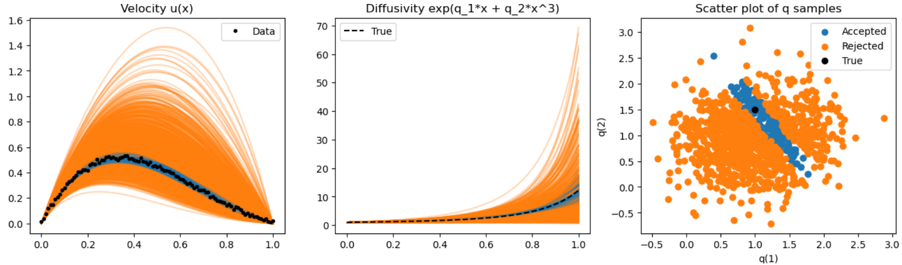
Approximate Bayesian computation: higher error threshold



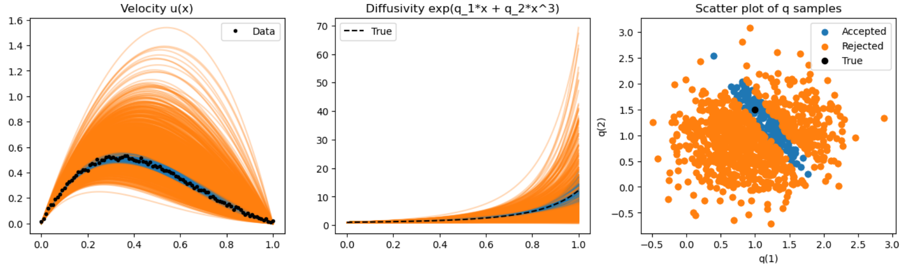
Approximate Bayesian computation: lower error threshold



Approximate Bayesian computation: More samples

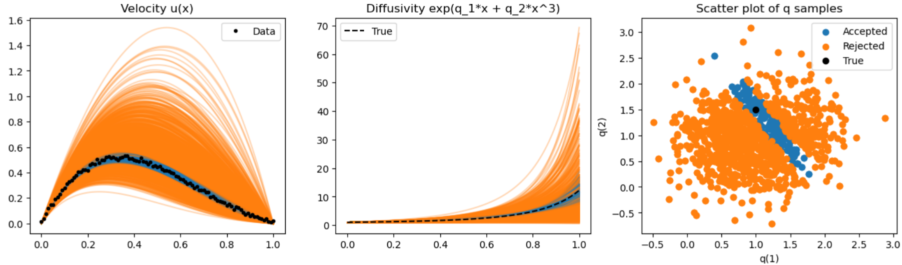


Approximate Bayesian computation: More samples



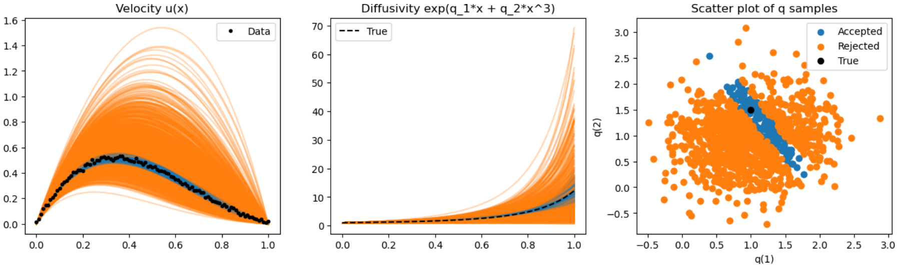
- Does not need explicit likelihood

Approximate Bayesian computation: More samples



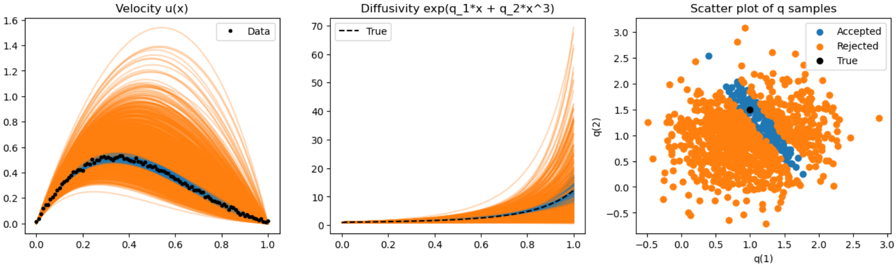
- Does not need explicit likelihood
- Easy to implement

Approximate Bayesian computation: More samples



- Does not need explicit likelihood
- Easy to implement
- Computationally expensive/wasteful

Approximate Bayesian computation: More samples



- Does not need explicit likelihood
- Easy to implement
- Computationally expensive/wasteful
- Approximately Bayesian

II: Bayesian inference

Recall Bayes' rule:

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

II: Bayesian inference

Recall Bayes' rule:

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

- Posterior probability: $p(\theta \mid D)$

II: Bayesian inference

Recall Bayes' rule:

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

- Posterior probability: $p(\theta \mid D)$
- Likelihood: $p(D \mid \theta)$

II: Bayesian inference

Recall Bayes' rule:

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

- Posterior probability: $p(\theta \mid D)$
- Likelihood: $p(D \mid \theta)$
- Prior probability: $p(\theta)$

II: Bayesian inference

Recall Bayes' rule:

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

- Posterior probability: $p(\theta \mid D)$
- Likelihood: $p(D \mid \theta)$
- Prior probability: $p(\theta)$
- Marginal likelihood (or model evidence): $p(D)$

Inference in general:

II: Bayesian inference

Recall Bayes' rule:

$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

- Posterior probability: $p(\theta \mid D)$
- Likelihood: $p(D \mid \theta)$
- Prior probability: $p(\theta)$
- Marginal likelihood (or model evidence): $p(D)$

Inference in general:

- Sampling (Markov chain Monte Carlo)

II: Bayesian inference

Recall Bayes' rule:

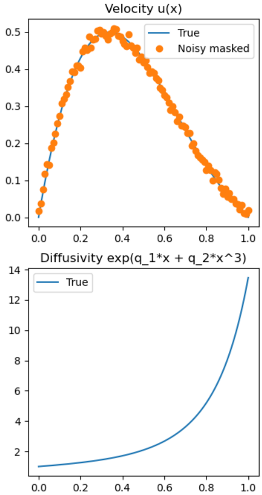
$$p(\theta \mid D) = \frac{p(D \mid \theta)p(\theta)}{p(D)}$$

- Posterior probability: $p(\theta \mid D)$
- Likelihood: $p(D \mid \theta)$
- Prior probability: $p(\theta)$
- Marginal likelihood (or model evidence): $p(D)$

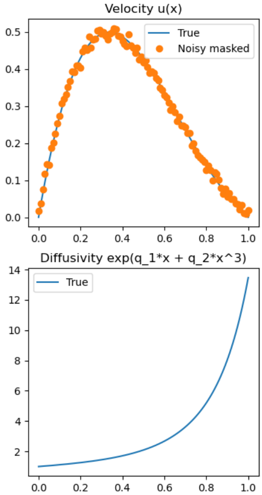
Inference in general:

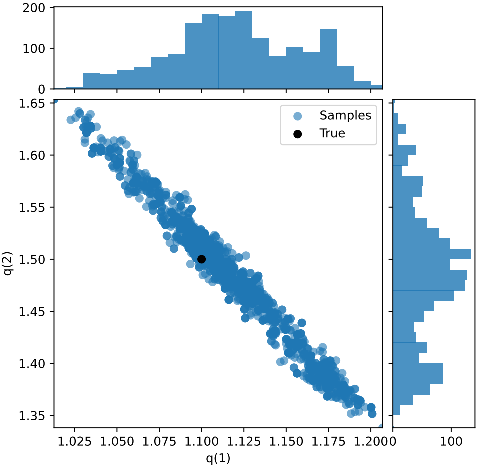
- Sampling (Markov chain Monte Carlo)
- Variational inference

Metropolis-Hastings sampling for inference

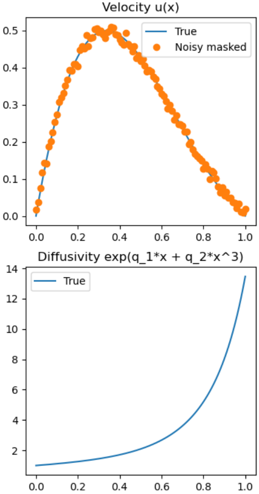


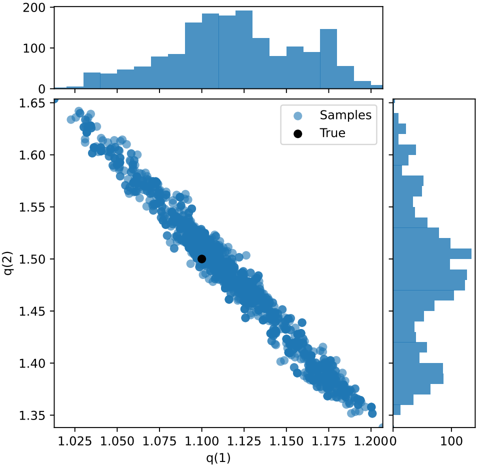
Metropolis-Hastings sampling for inference





Metropolis-Hastings sampling for inference





Expensive!

III: Bayesian optimisation

- Global optimisation algorithm

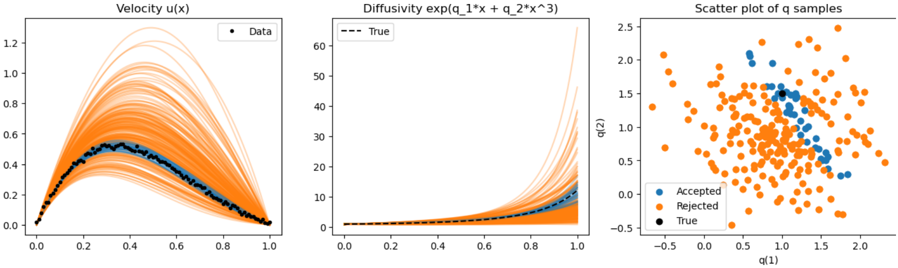
III: Bayesian optimisation

- Global optimisation algorithm
- Balances exploration and exploitation

III: Bayesian optimisation

- Global optimisation algorithm
- Balances exploration and exploitation

Recall, for approximate Bayesian computation:



III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

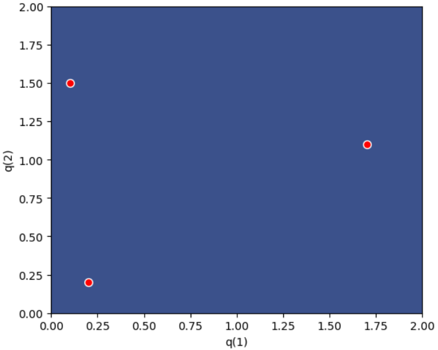
Want an algorithm that:

III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

Want an algorithm that:

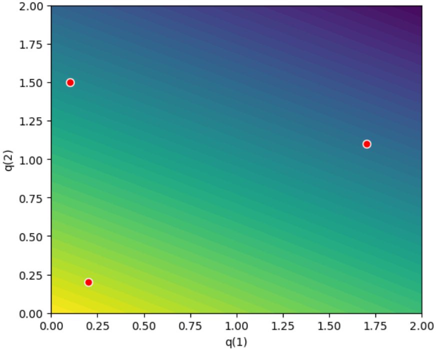


III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

Want an algorithm that:



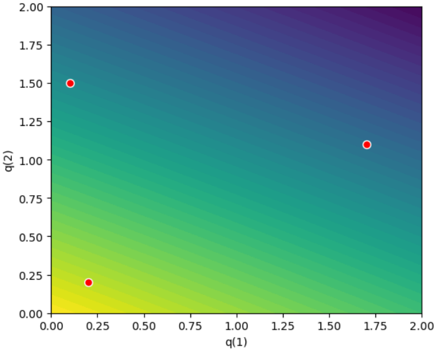
III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2} \sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

Want an algorithm that:

- **Exploits** promising parameter values, and



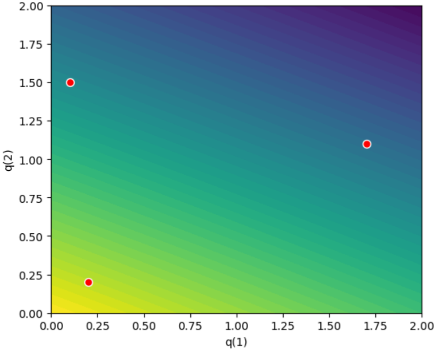
III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

Want an algorithm that:

- **Exploits** promising parameter values, and
- **Explores** unknown parameter space.



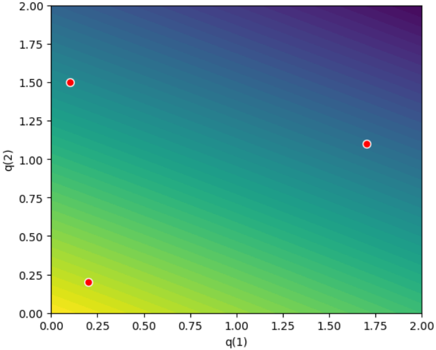
III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

Want an algorithm that:

- **Exploits** promising parameter values, and
- **Explores** unknown parameter space.



- Interpolate existing data

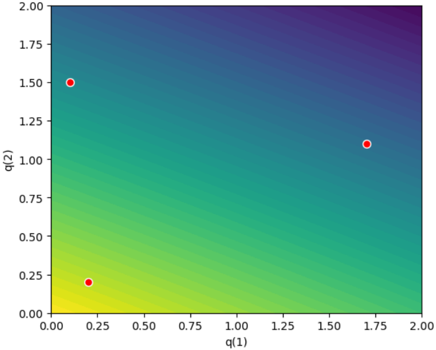
III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

Want an algorithm that:

- **Exploits** promising parameter values, and
- **Explores** unknown parameter space.



- Interpolate existing data
- Quantify distance amongst data

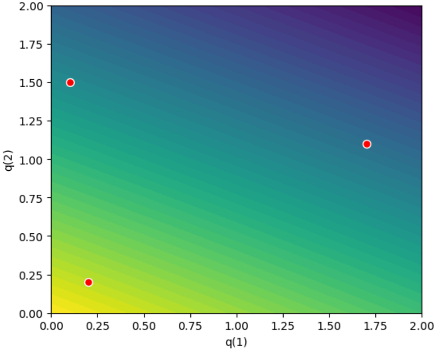
III: Bayesian optimisation

Recall, we want θ that minimises the negative log-likelihood:

$$\ell_N(\theta) = \frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathcal{G}_\theta(X_i)|_V^2.$$

Want an algorithm that:

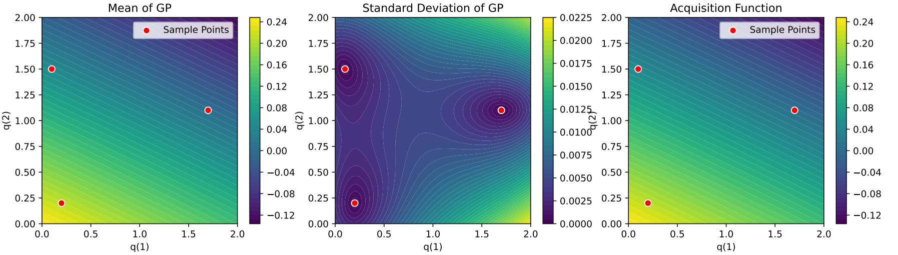
- **Exploits** promising parameter values, and
- **Explores** unknown parameter space.



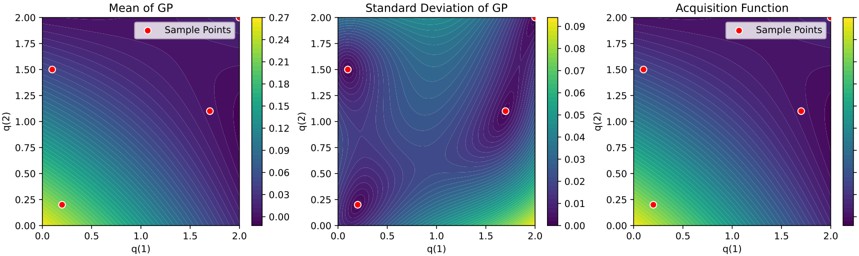
- Interpolate existing data
- Quantify distance amongst data

Gaussian process!

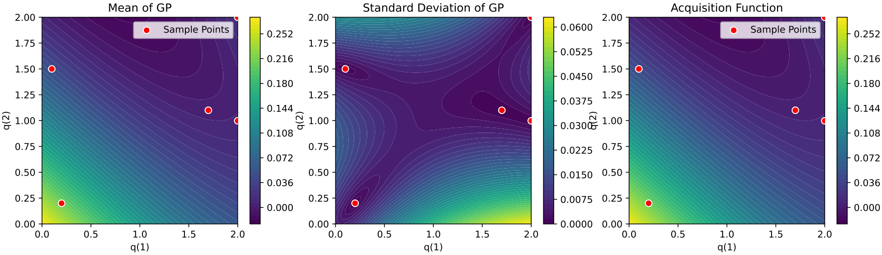
III: Bayesian optimisation: Exploitation



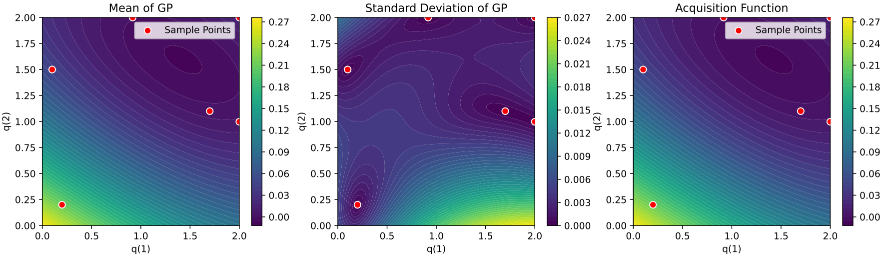
III: Bayesian optimisation: Exploitation



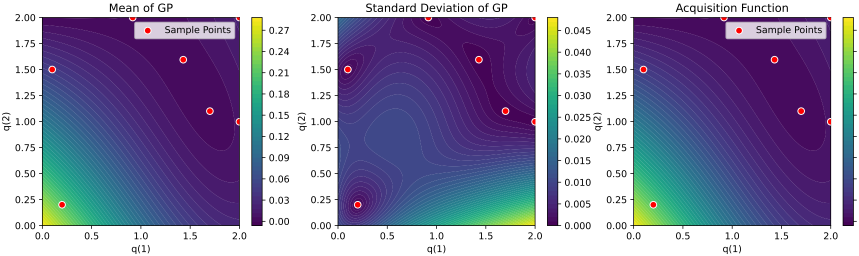
III: Bayesian optimisation: Exploitation



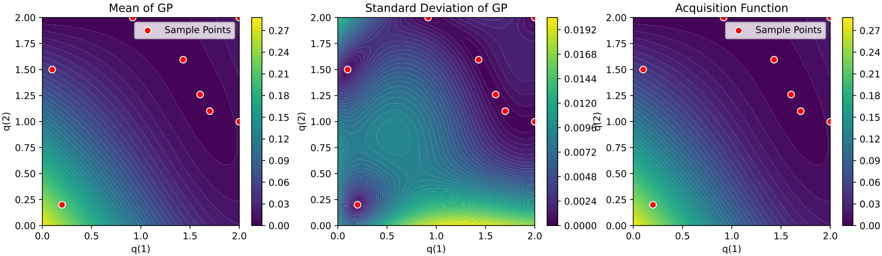
III: Bayesian optimisation: Exploitation



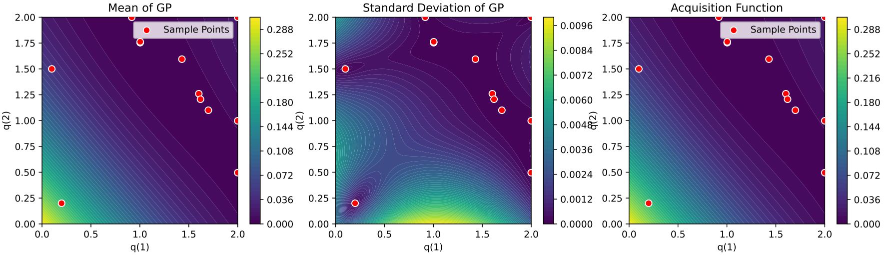
III: Bayesian optimisation: Exploitation



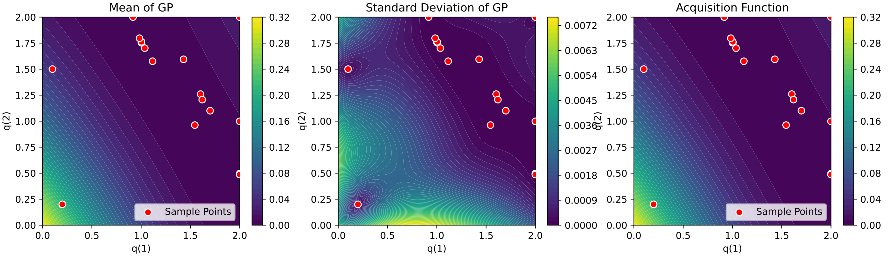
III: Bayesian optimisation: Exploitation



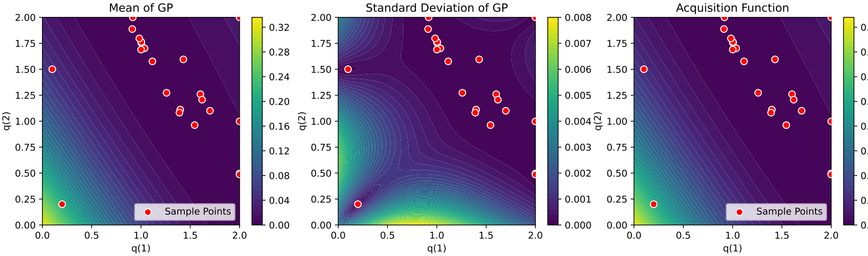
III: Bayesian optimisation: Exploitation



III: Bayesian optimisation: Exploitation

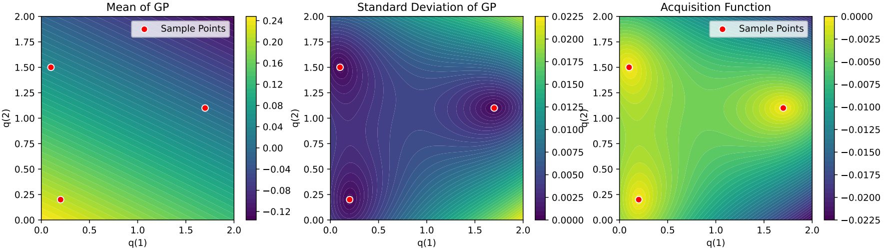


III: Bayesian optimisation: Exploitation

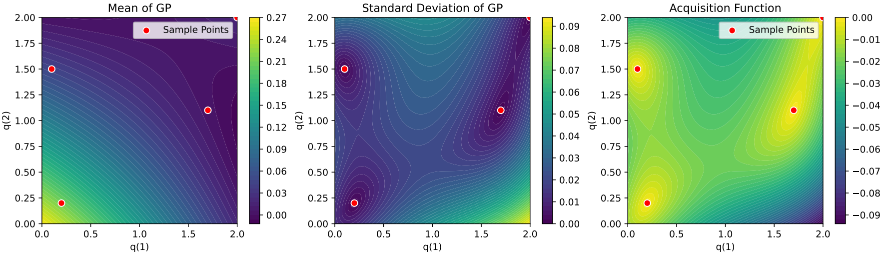


Exploitation

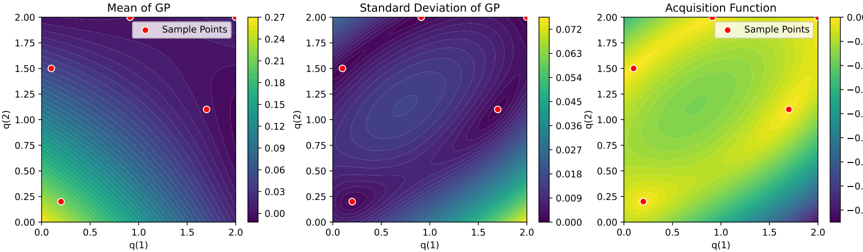
III: Bayesian optimisation: Exploration



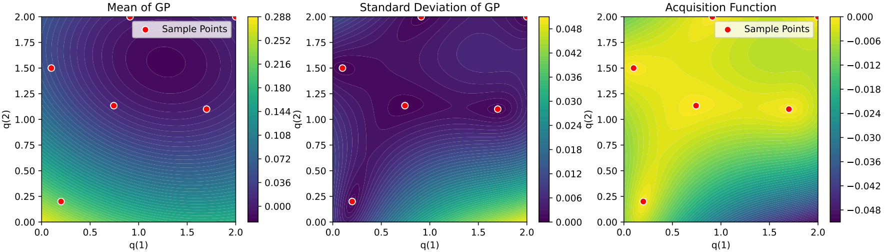
III: Bayesian optimisation: Exploration



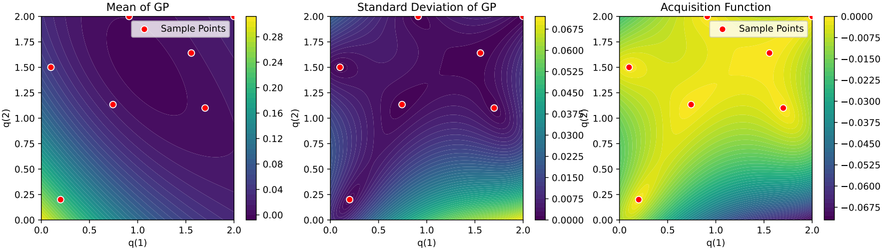
III: Bayesian optimisation: Exploration



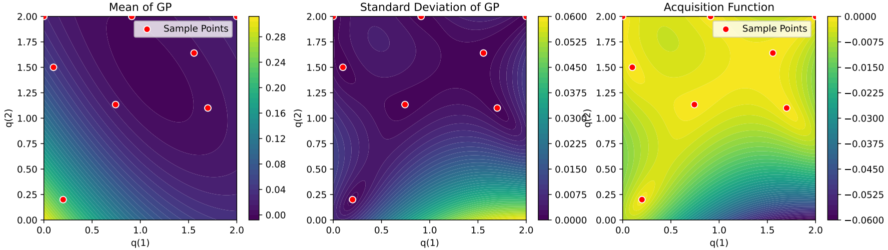
III: Bayesian optimisation: Exploration



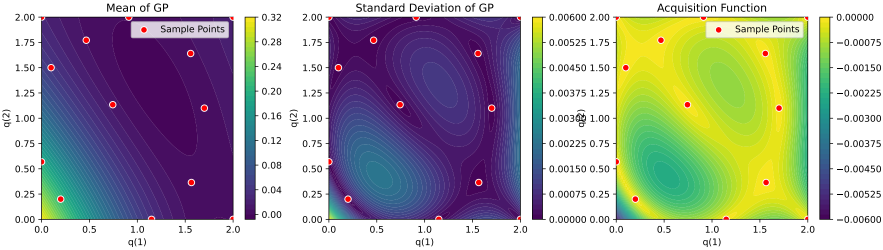
III: Bayesian optimisation: Exploration



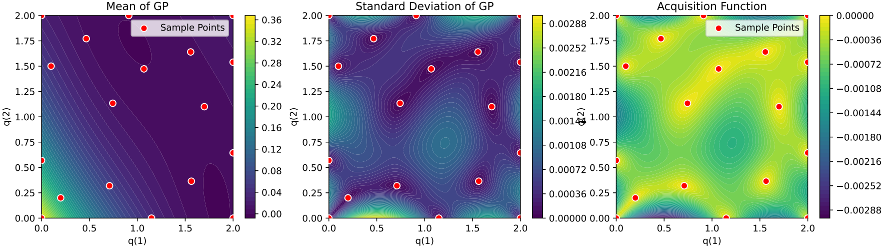
III: Bayesian optimisation: Exploration



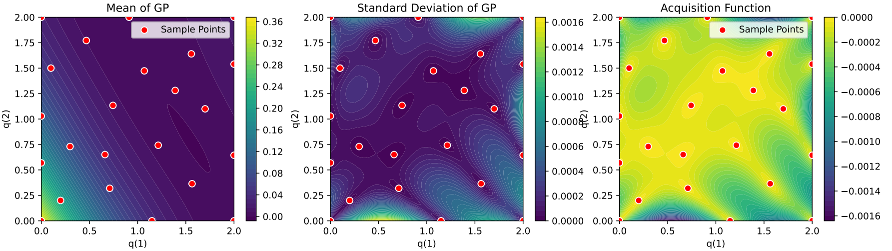
III: Bayesian optimisation: Exploration



III: Bayesian optimisation: Exploration

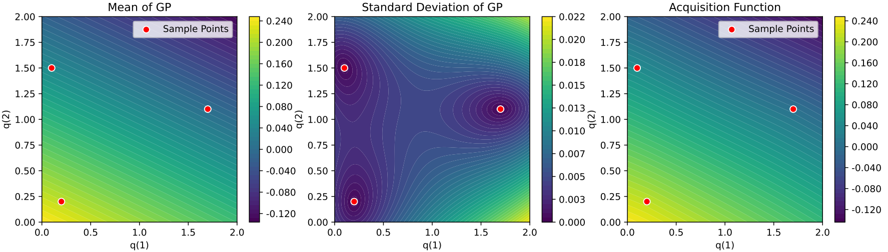


III: Bayesian optimisation: Exploration

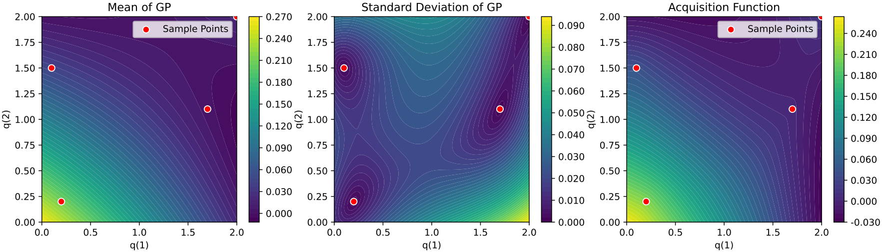


Exploration

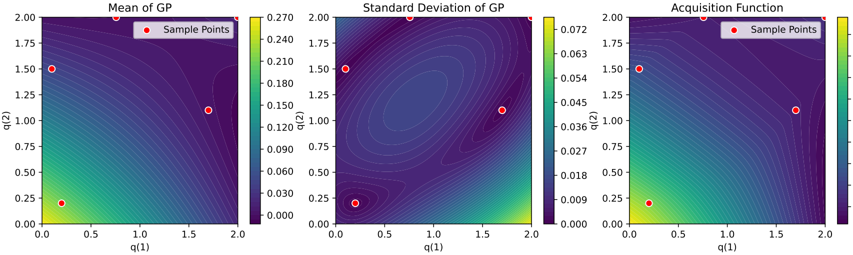
III: Bayesian optimisation: Exploration & exploitation



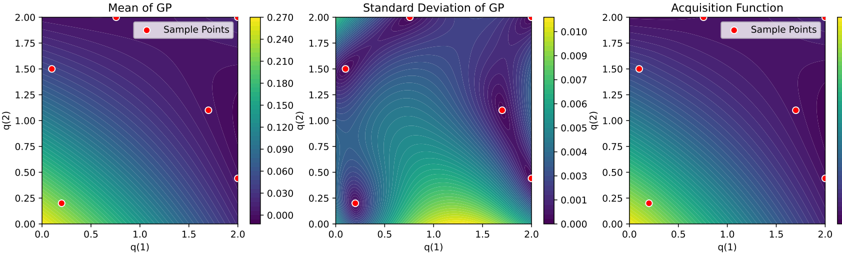
III: Bayesian optimisation: Exploration & exploitation



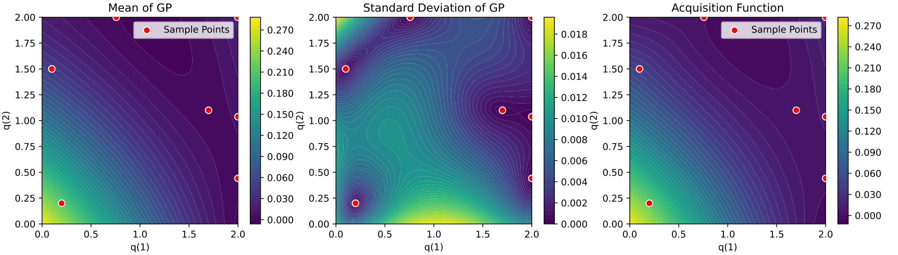
III: Bayesian optimisation: Exploration & exploitation



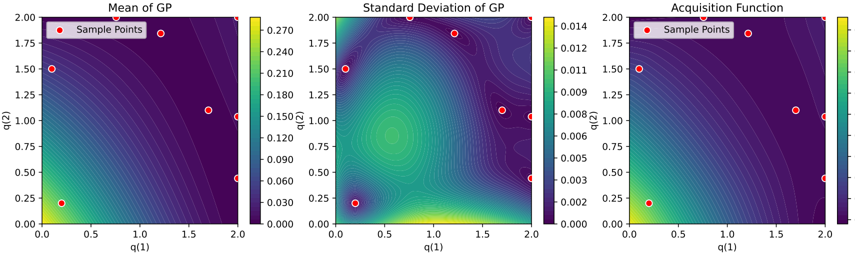
III: Bayesian optimisation: Exploration & exploitation



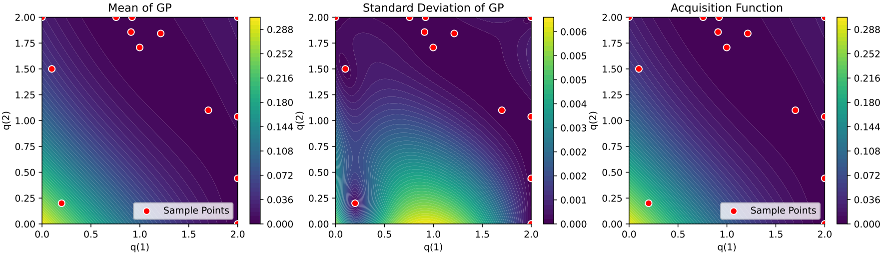
III: Bayesian optimisation: Exploration & exploitation



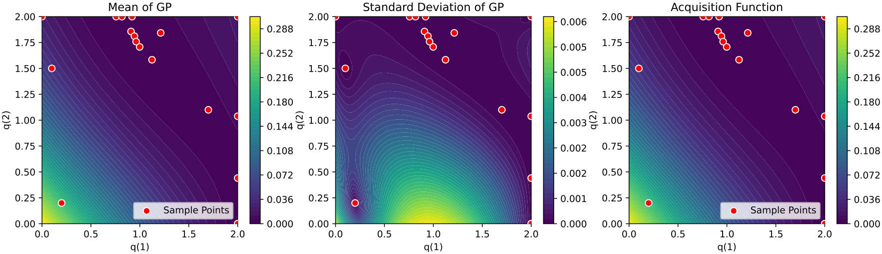
III: Bayesian optimisation: Exploration & exploitation



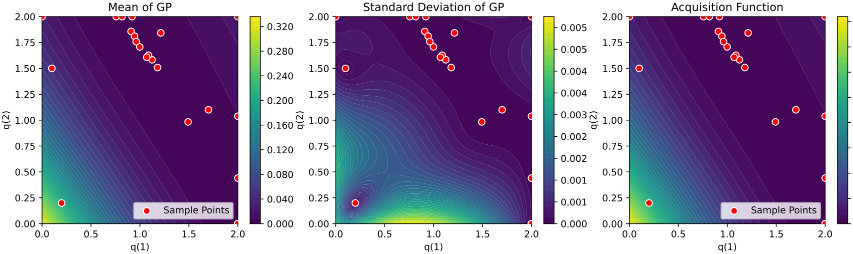
III: Bayesian optimisation: Exploration & exploitation



III: Bayesian optimisation: Exploration & exploitation



III: Bayesian optimisation: Exploration & exploitation



Exploration & exploitation

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
    - Easy to implement

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
    - Easy to implement
    - Computationally expensive

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
    - Easy to implement
    - Computationally expensive
    - Approximately Bayesian

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
    - Easy to implement
    - Computationally expensive
    - Approximately Bayesian

2. (Full) Bayesian inference

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
    - Easy to implement
    - Computationally expensive
    - Approximately Bayesian
2. (Full) Bayesian inference
    - Computationally expensive

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
   - Easy to implement
   - Computationally expensive
   - Approximately Bayesian

2. (Full) Bayesian inference
   - Computationally expensive
   - Often unnecessary

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
    - Easy to implement
    - Computationally expensive
    - Approximately Bayesian
2. (Full) Bayesian inference
    - Computationally expensive
    - Often unnecessary
3. Bayesian optimisation (BO)

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
    - Easy to implement
    - Computationally expensive
    - Approximately Bayesian
2. (Full) Bayesian inference
    - Computationally expensive
    - Often unnecessary
3. Bayesian optimisation (BO)
    - Balanced exploration & exploitation of parameters

Three Bayesian approaches to inverse problems

1. Bayesian approximate computation (ABC)
   - Easy to implement
   - Computationally expensive
   - Approximately Bayesian

2. (Full) Bayesian inference
   - Computationally expensive
   - Often unnecessary

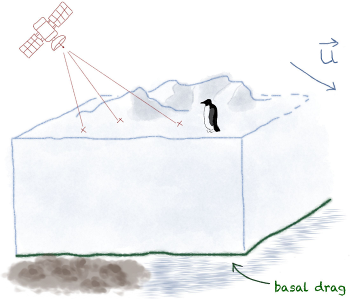
3. Bayesian optimisation (BO)
   - Balanced exploration & exploitation of parameters
   - Point estimate

Open questions

1. Are all inverse problems solvable?

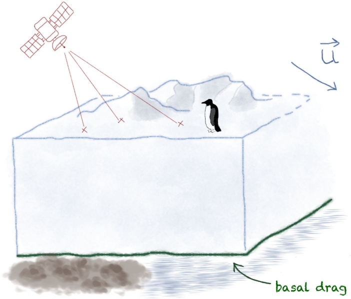
Open questions

1. Are all inverse problems solvable?



Open questions

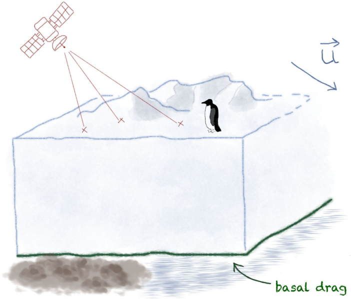
1. Are all inverse problems solvable?



2. What are emulators and when to use them?

Open questions

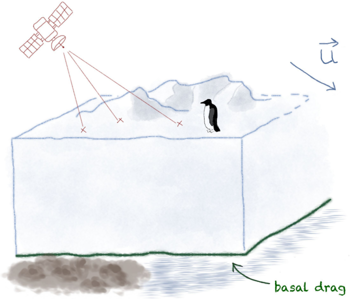
1. Are all inverse problems solvable?



2. What are emulators and when to use them?
   - Unknown physics (parameterisations)

Open questions

1. Are all inverse problems solvable?



2. What are emulators and when to use them?
   - Unknown physics (parameterisations)
   - Plentiful data allows to replace the physical model

Thank you