Assessing model mismatch and model selection in a Bayesian uncertainty quantification analysis of a fluid-dynamics model of pulmonary blood circulation

Mihaela Paun

Glasgow University, Scotland

Jointly with …

Dirk Husmeier

Nick Hill

Glasgow University

Mette Olufsen

Mitchel Colebank

North Carolina State University

Talk based on work presented in our recent publication:

INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Check for updates

Assessing model mismatch and model selection in a Bayesian uncertainty quantification analysis of a fluid-dynamics model of pulmonary blood circulation

L. Mihaela Paun[1], Mitchel J. Colebank[2], Mette S. Olufsen[2], Nicholas A. Hill[1] and Dirk Husmeier[1]

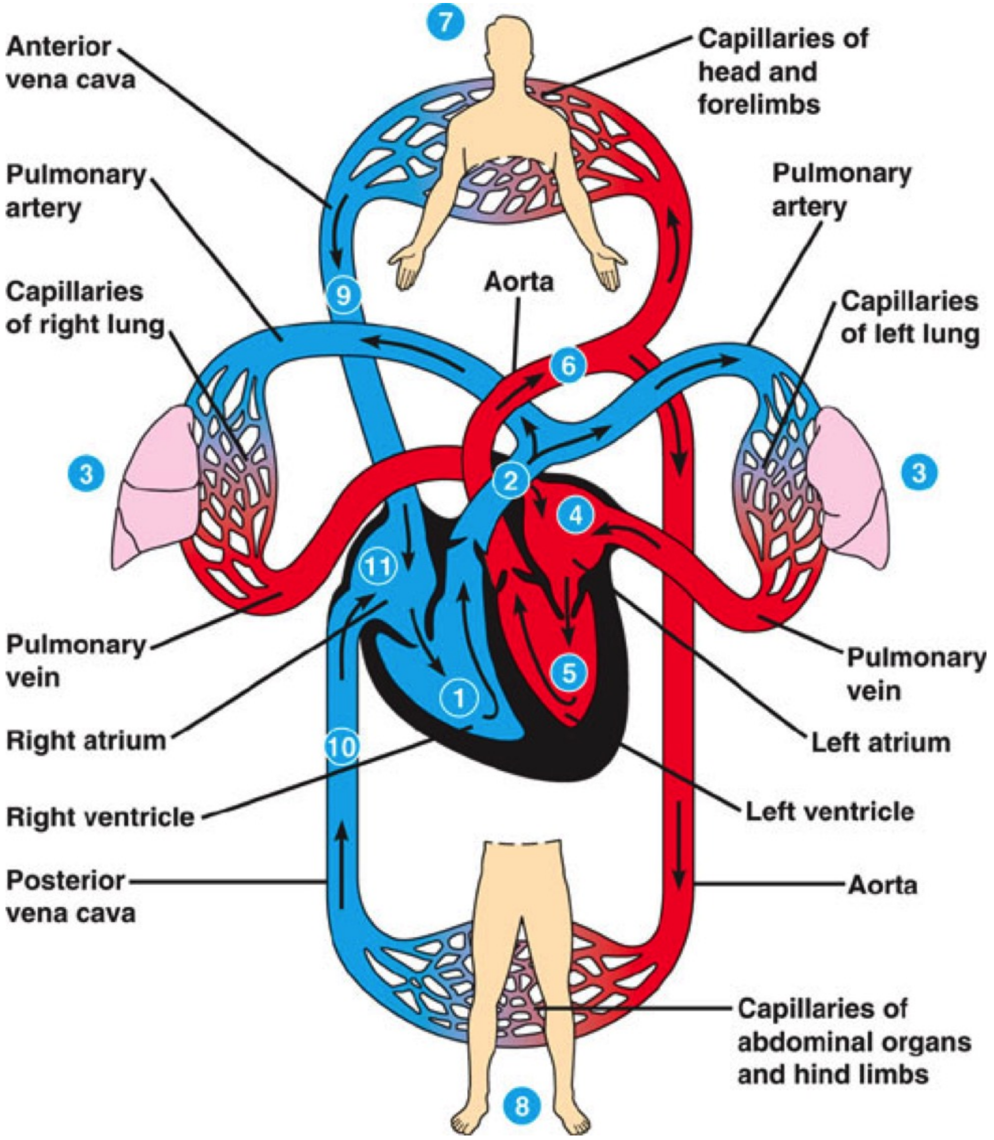
Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Blood Circulation

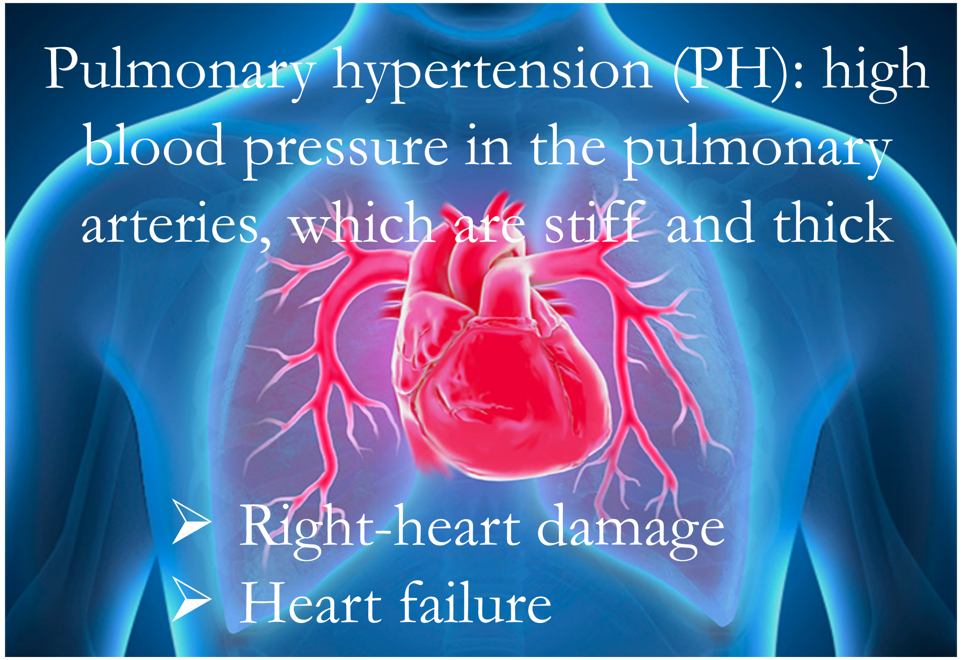


Pulmonary circulation

Systemic circulation

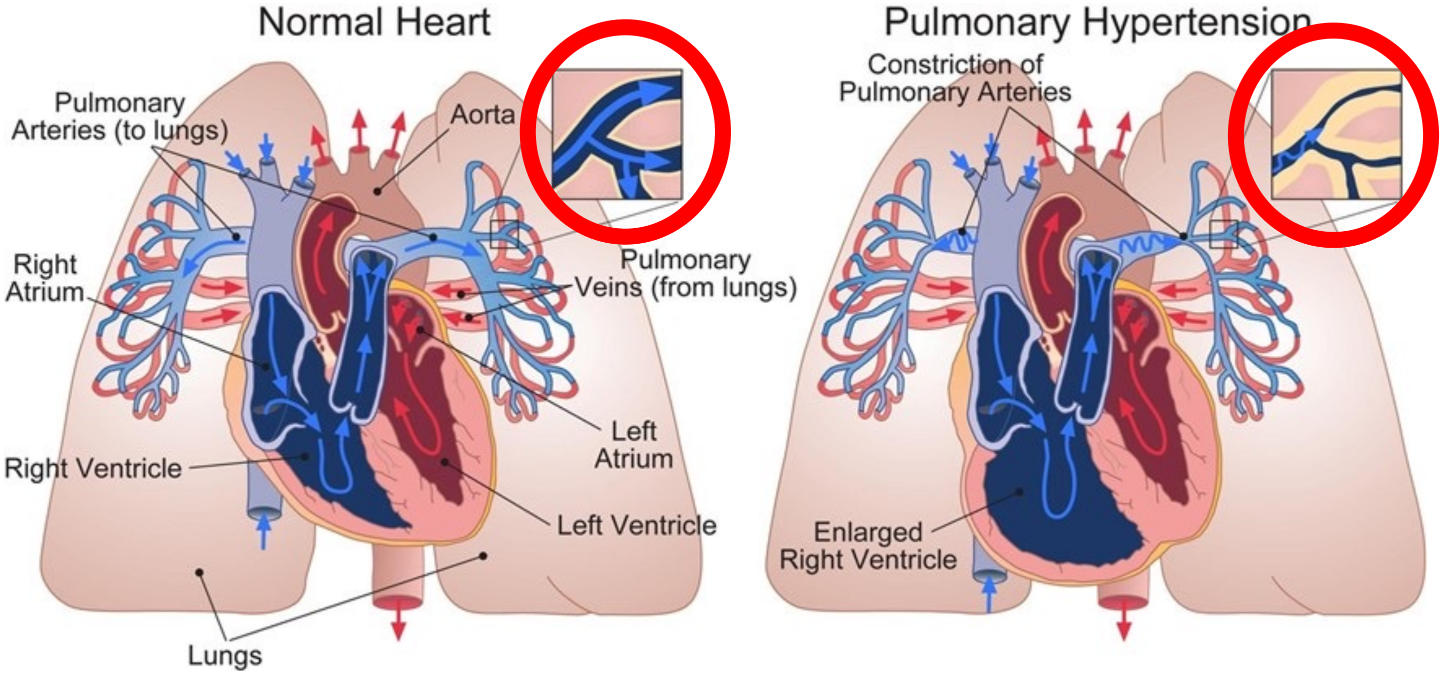
Outline

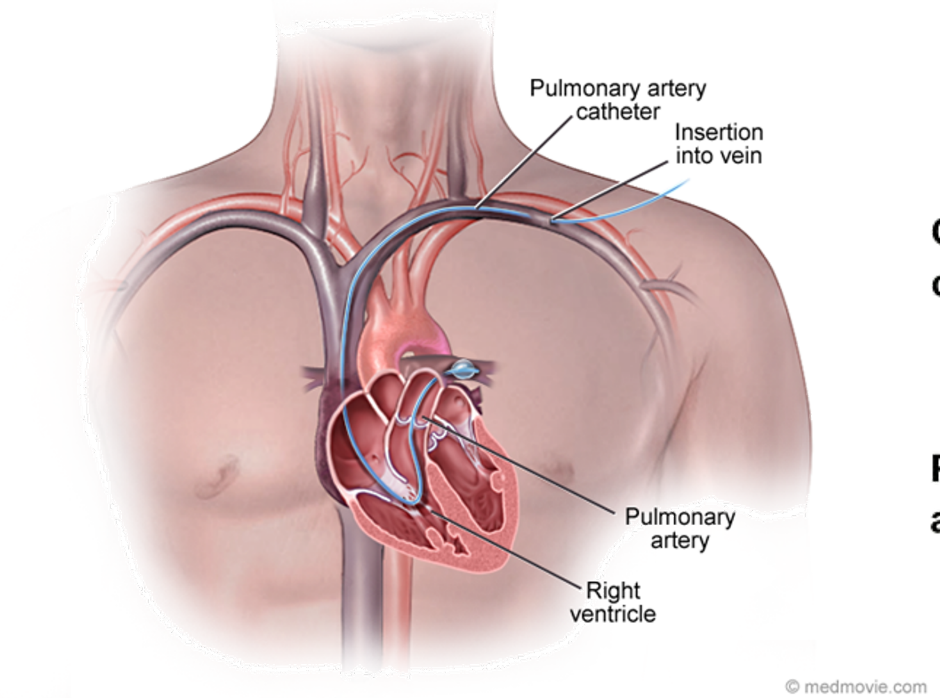
- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

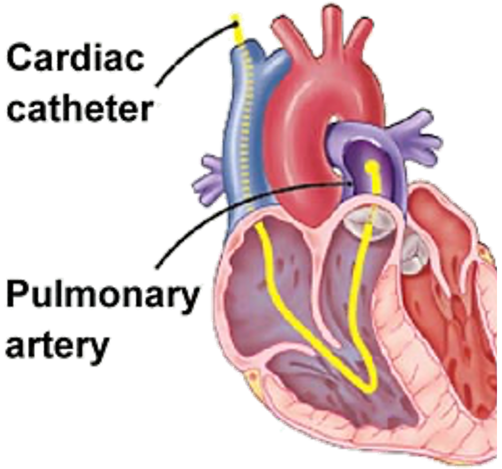


Pulmonary hypertension (PH): high blood pressure in the pulmonary arteries, which are stiff and thick

- Right-heart damage
- Heart failure







PH: diagnosed by **invasively measuring** pulmonary pressure with right-heart catheterisation

Risks: excessive bleeding; partial lung collapse

=> Develop a non-invasive alternative

Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Aim

Combine

- Haemodynamic data (pulmonary blood flow – measured non-invasively with MRI)
- Imaging data (computed tomography scan giving pulmonary vessel network)
- Mathematical modelling
- Statistical modelling

to:

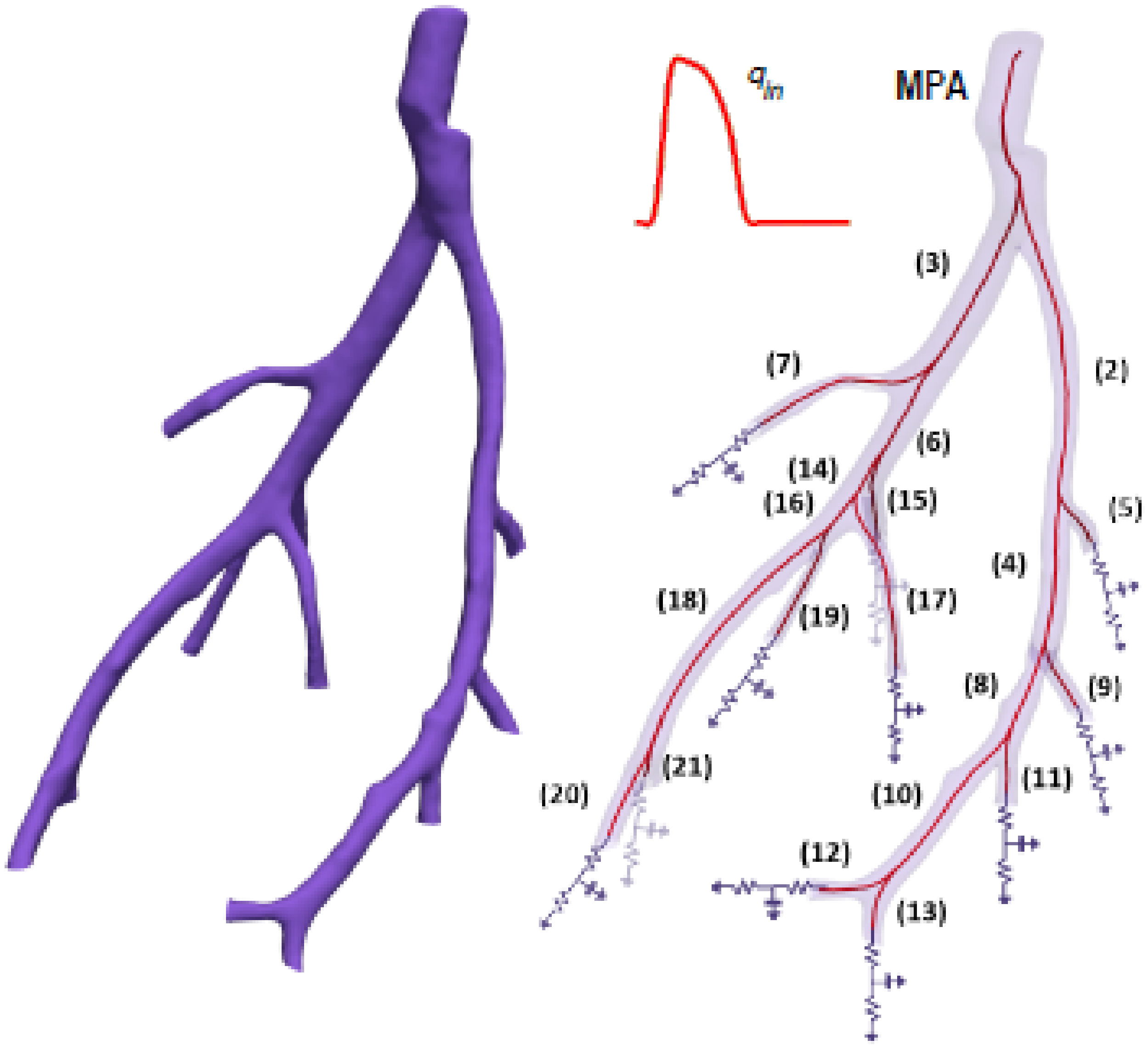
❖Predict pulmonary pressure (rather than measure it)

❖Build a patient-specific model allowing personalised medicine

Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Data



Data from a healthy mouse

Imaging data:

- Gives information on pulmonary vessel network and geometry

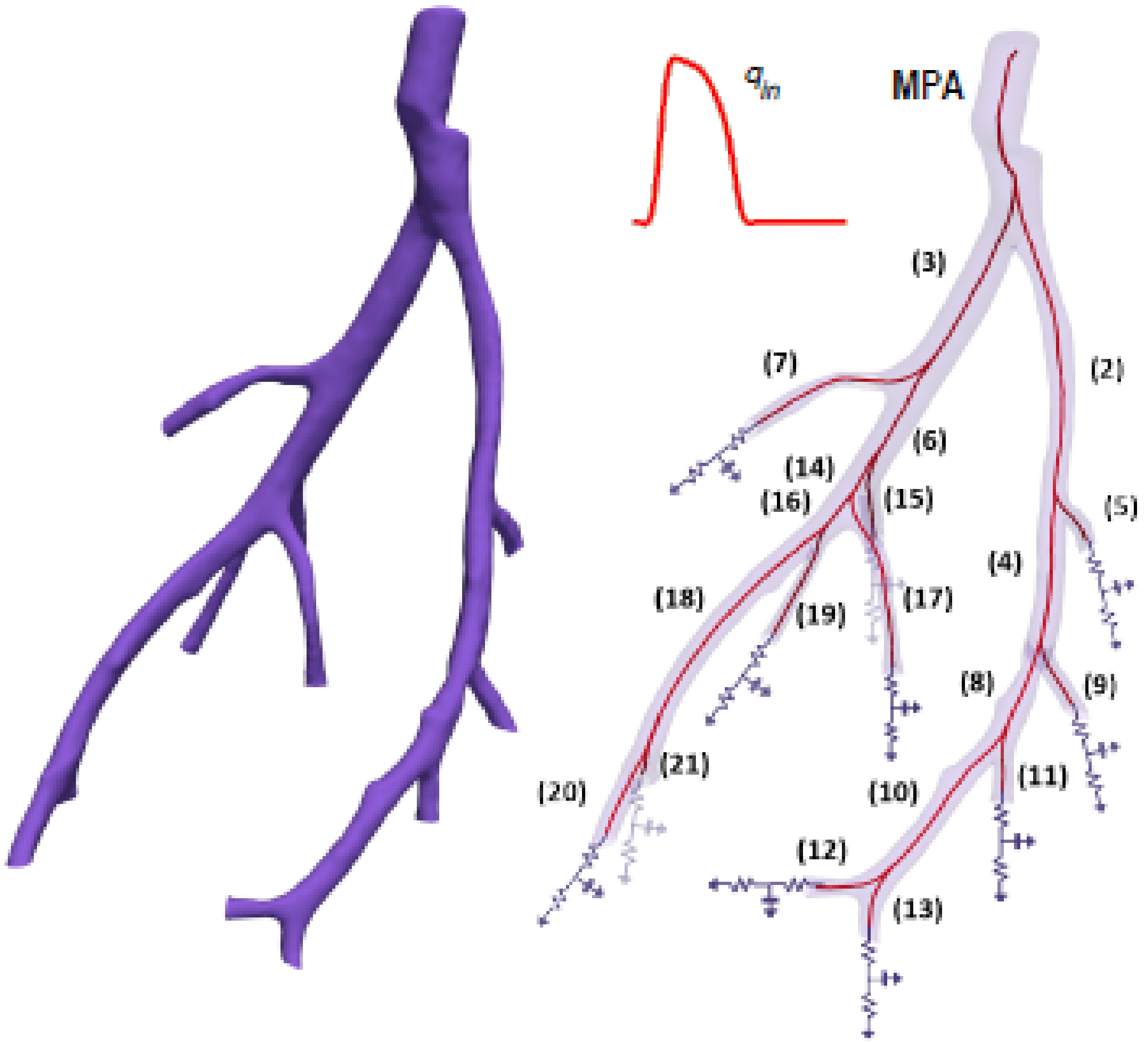
Haemodynamic data:

- Pulmonary blood pressure in the main pulmonary artery (MPA).
- Pulmonary blood flow in MPA.

Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Mathematical model



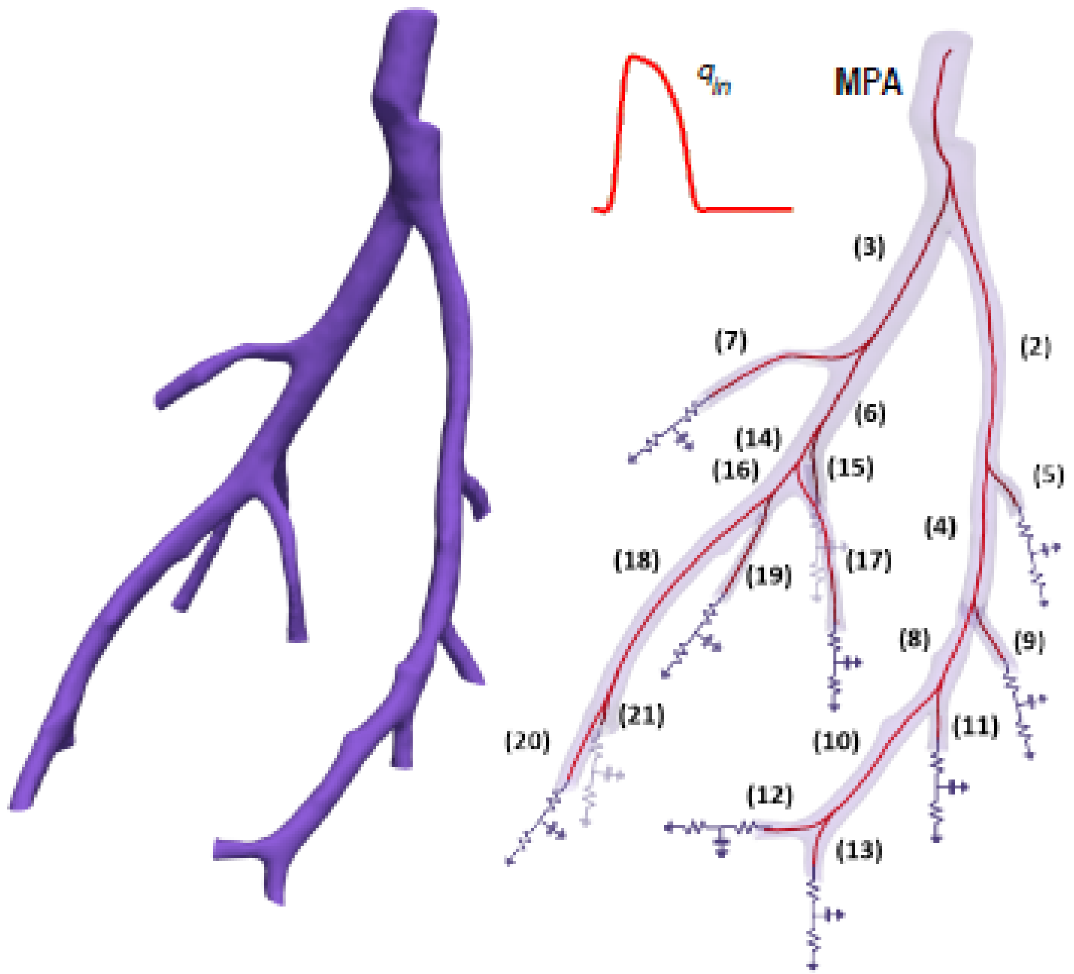
PDE model of the blood pressure and flow wave propagation in vessels

coupled with:

3-element Windkessel models at the outlet

Mathematical model

Inlet boundary conditions: Flow in MPA



PDE model of blood pressure and flow

Flow

$$\frac{\partial A}{\partial t} + \frac{\partial q}{\partial x} = 0,$$

$$\frac{\partial q}{\partial t} + \frac{\partial}{\partial x}\left(\frac{q^2}{A}\right) + \frac{A}{\rho}\frac{\partial p}{\partial x} = -\frac{2\pi\mu r}{\rho\delta}\frac{q}{A},$$

Pressure

Cross-sectional area

$$p = \frac{4}{3}\frac{Eh}{r_0}\left(\sqrt{\frac{A}{A_0}} - 1\right)$$

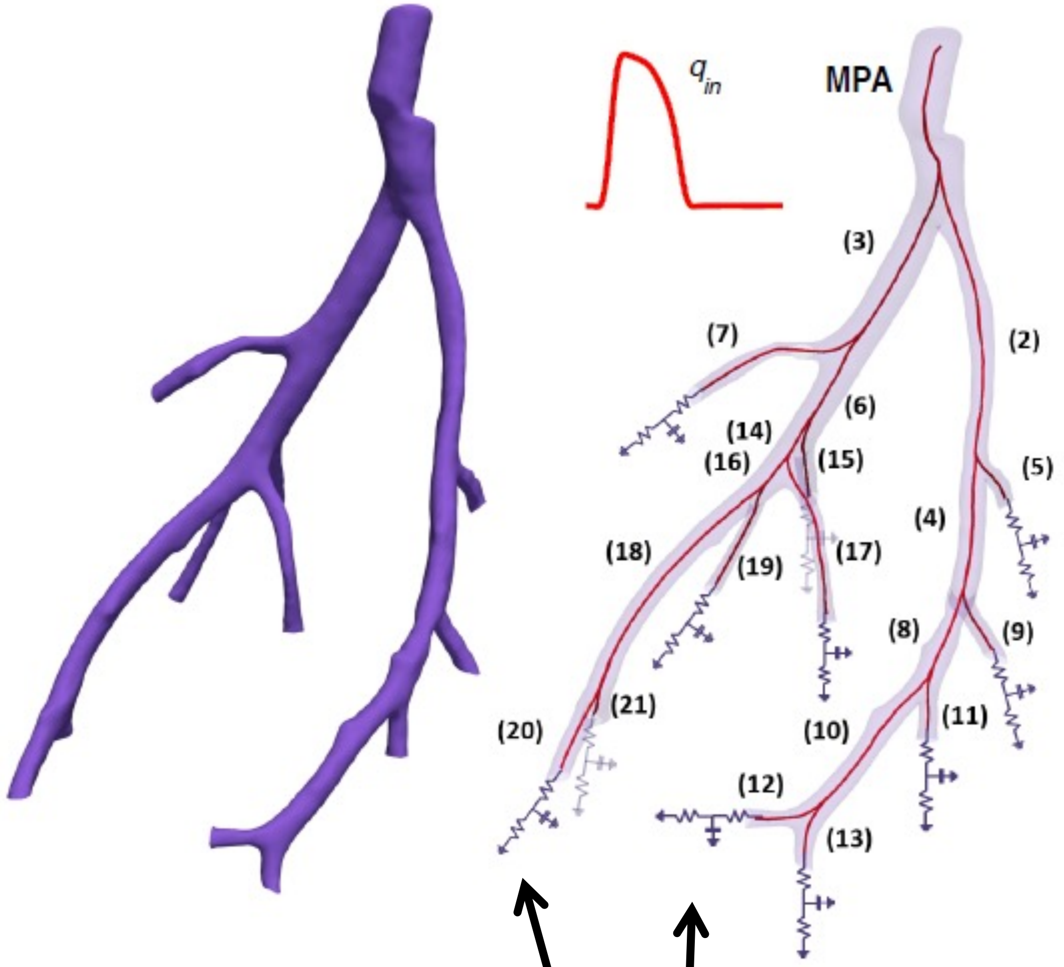
Boundary conditions at vessel junctions:

$$p_p = p_{d_i}$$

where p – parent, d_1 and d_2 - daughters

$$q_p = \sum_i q_i$$

Mathematical model



Windkessel models

3-element Windkessel model
(outlet boundary conditions)

Pressure

Flow

Resistances

$$\frac{dp(L,t)}{dt} - R_1 \frac{dq(L,t)}{dt} = q(L,t)\left(\frac{R_1 + R_2}{R_2 C}\right) - \frac{p(L,t)}{R_2 C}$$

Capacitance

Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Parameter estimation

$$p = \frac{4}{3}\frac{Eh}{r_0}\left(\sqrt{\frac{A}{A_0}} - 1\right)$$

Unknown parameters

$$\frac{dp(L,t)}{dt} - R_1\frac{dq(L,t)}{dt} = q(L,t)\left(\frac{R_1+R_2}{R_2C}\right) - \frac{p(L,t)}{R_2C}$$

$R_1^j = r_1 R_{01}^j, \quad R_2^j = r_2 R_{02}^j, \quad C^j = cC_0^j.$ for j^{th} terminal vessel.

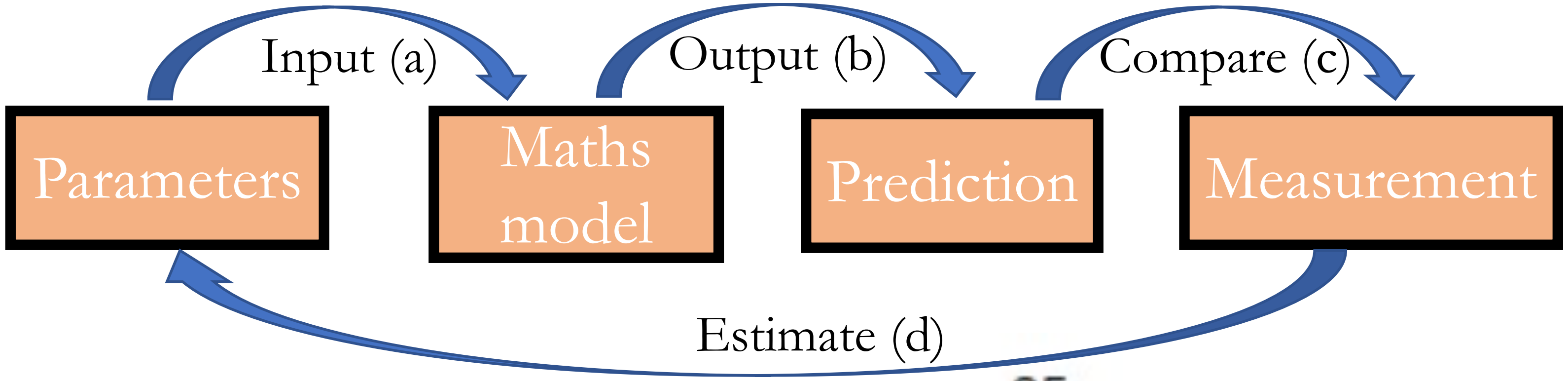
global scaling factors (common to all terminal vessels)

Estimate parameters from measured data ('inverse problem')

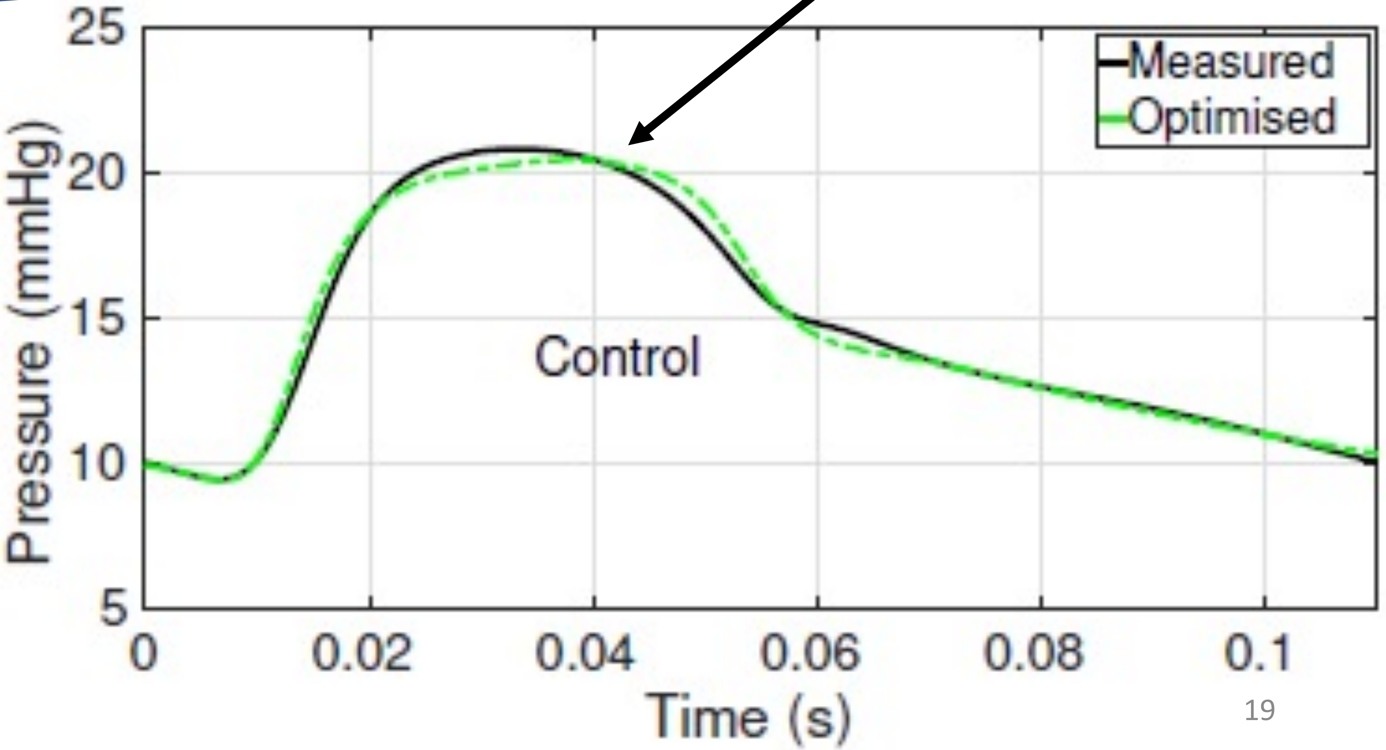
4D parameter vector:

$\frac{Eh}{r_0}$, r_1, r_2, c

Conventional parameter estimation method



MODEL MISMATCH



Conventional method: maximise agreement between measurement and prediction ('least squares fit')

WRONG for mis-specified maths models!

Conventional parameter estimation method

Measurement

PDE solution (prediction)

Variance of measurement errors

Iid errors: $\mathbf{y} \sim \mathcal{MVN}(\mathbf{m}(\underline{\theta}), \sigma^2 \mathbf{I})$, i.e.

Parameters

Iid: independent and identically distributed
MVN: multivariate normal distribution

Data likelihood

$$p(\mathbf{y}|\theta, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (y_i - m_i(\theta))^2}{2\sigma^2}\right),$$

where

Sum of squared errors (SSE)

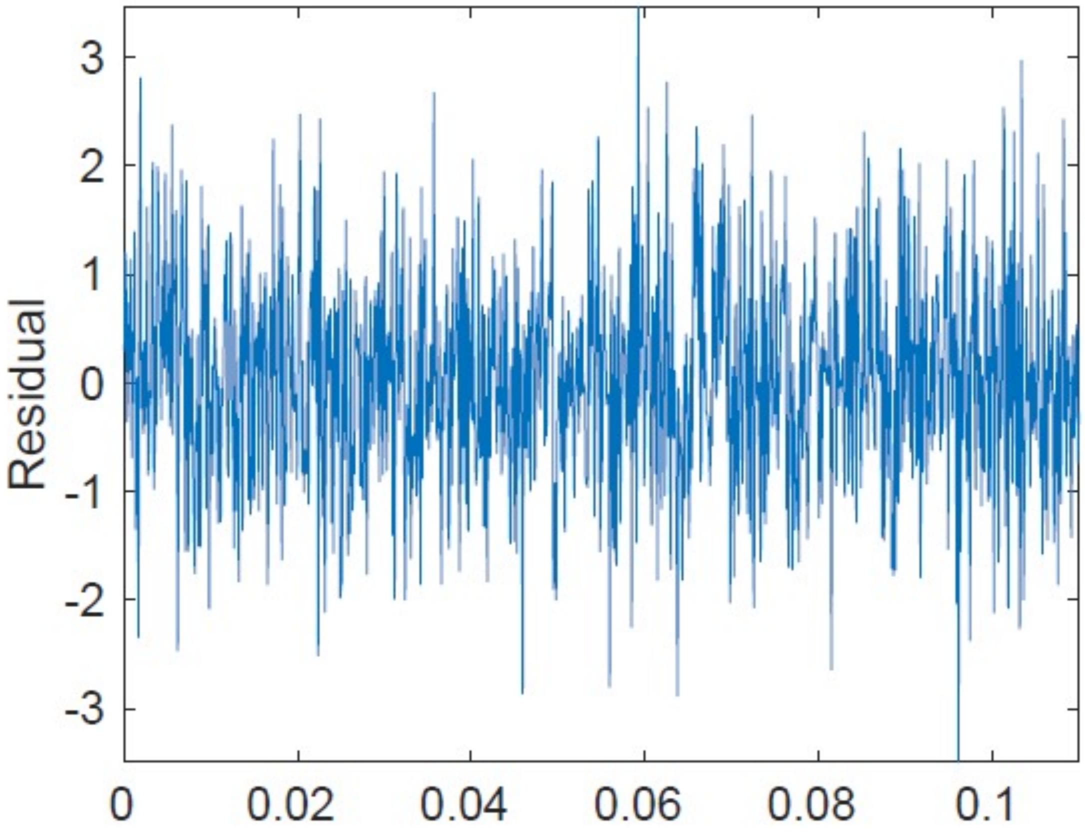
$$\sum_{i=1}^n (y_i - m_i(\theta))^2 = (\mathbf{y} - \mathbf{m}(\theta))^{\mathrm{T}} (\mathbf{y} - \mathbf{m}(\theta))$$

The conventional method minimises SSE
(maximises data likelihood for iid errors)

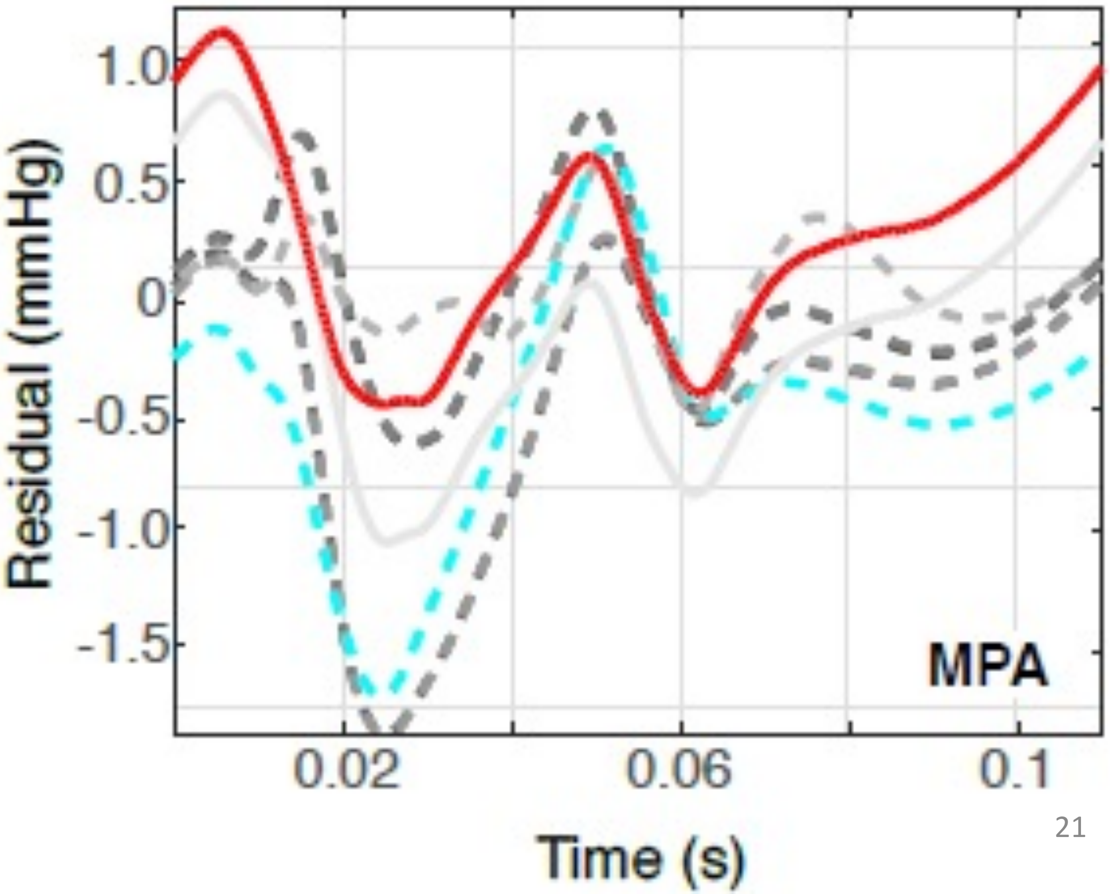
Visualising errors (residuals)

Errors: $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{m}(\boldsymbol{\theta})$

NO MODEL MISMATCH
IID ERRORS



MODEL MISMATCH
CORRELATED ERRORS



Parameter estimation method when model mismatch is present

Correlated errors: $\mathbf{y} \sim \mathscr{MVN}(\mathbf{m}(\theta), \mathbf{C})$, i.e.

Data likelihood

$$p(\mathbf{y}|\theta, \mathbf{C}) = \det(2\pi\mathbf{C})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{m}(\theta))^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{y} - \mathbf{m}(\theta))\right)$$

Maximise data likelihood

Now estimate θ and **C**
(**C** found using Gaussian Processes)

Parameters that maximise data likelihood do **not** minimise SSE !

Modelling model mismatch

- Ignoring model mismatch (iid errors):

$$\mathbf{y}(\mathbf{t}) = \mathbf{m}(\boldsymbol{\theta}, \mathbf{t}) + \mathbf{u}(\mathbf{t}), \quad \mathbf{u}(\mathbf{t}) \sim \mathcal{MVN}(\mathbf{0}, \sigma^2 \mathbf{I})$$

- Modelling model mismatch (correlated errors):

model mismatch function

$$\mathbf{y}(\mathbf{t}) = \mathbf{m}(\boldsymbol{\theta}, \mathbf{t}) + \underline{\boldsymbol{\Gamma}(\mathbf{t})} = \mathbf{m}(\boldsymbol{\theta}, \mathbf{t}) + \mathbf{f}(\mathbf{t}) + \mathbf{u}(\mathbf{t}),$$

$$\mathbf{f}(\mathbf{t}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}|\boldsymbol{\eta}), \quad \mathbf{u}(\mathbf{t}) \sim \mathcal{MVN}(\mathbf{0}, \sigma_n^2 \mathbf{I}),$$

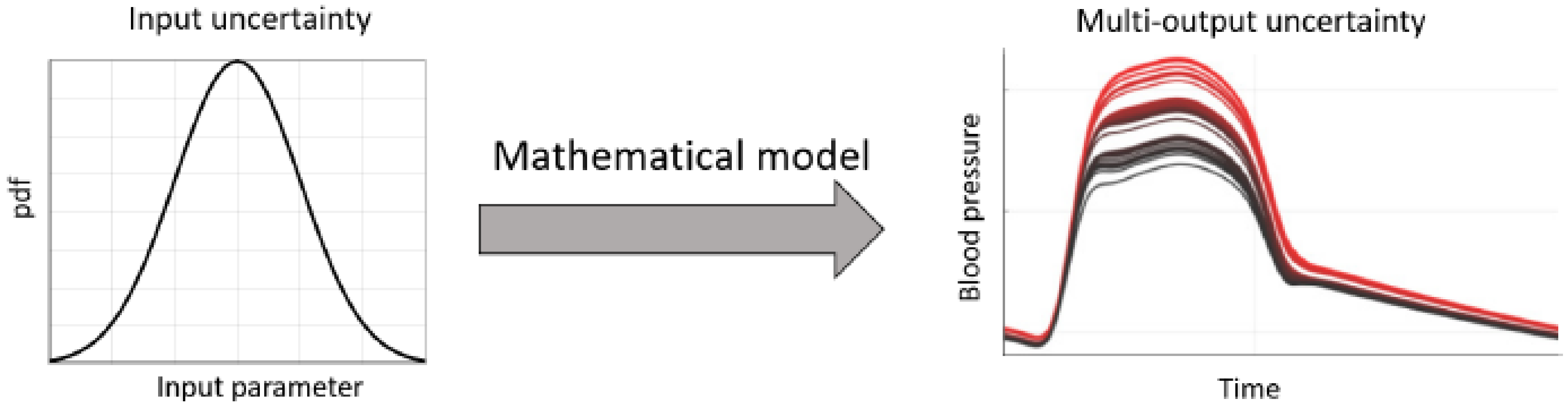
- Covariance matrix of the residuals:

$$\mathbf{C} = \mathbf{K} + \sigma_n^2 \mathbf{I}$$

Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Uncertainty quantification (UQ)



Uncertainty in parameters due to: natural variability, limited and noisy data

Uncertainty in parameters propagates to uncertainty in predictions

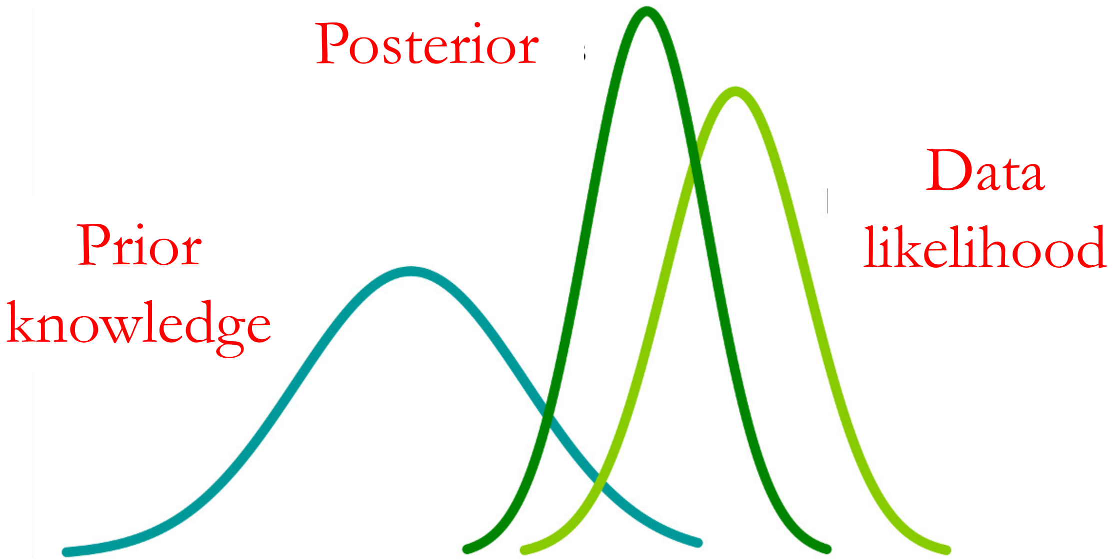
Quantify the uncertainty in parameters and predictions

Bayesian analysis

Posterior distribution — Data likelihood — Prior distribution

$$p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\eta})p(\boldsymbol{\theta}, \boldsymbol{\eta}),$$

where $\boldsymbol{\theta}$ are the biophysical parameters and $\boldsymbol{\eta}$ are the error parameters.



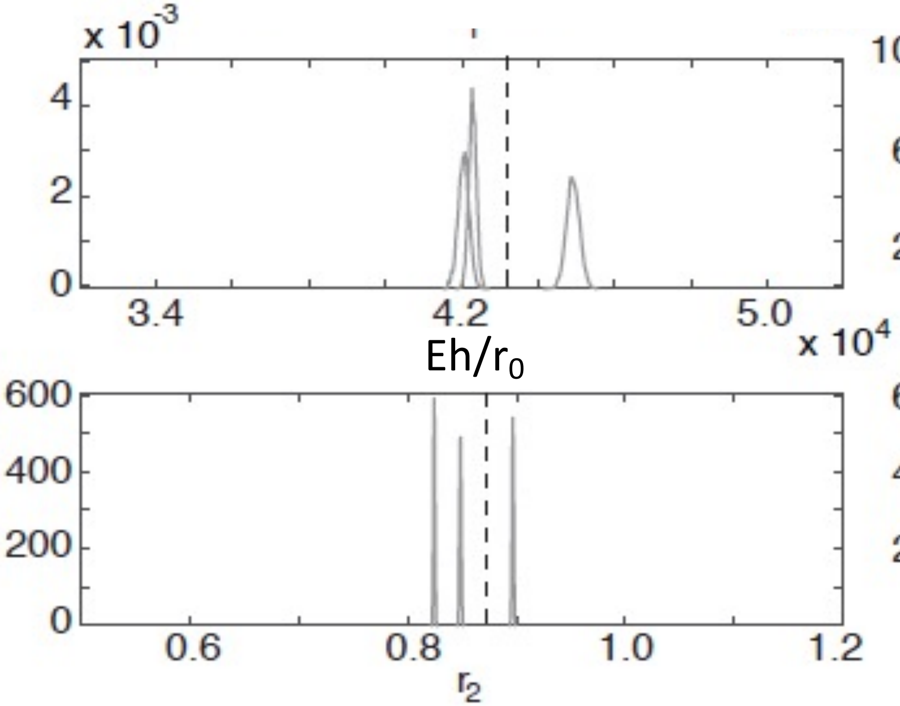
Define C, covariance matrix of residuals (modelled with Gaussian Processes)

Idea: Sample parameter values (approximately) from the posterior distribution (using Markov Chain Monte Carlo, MCMC).

Outline

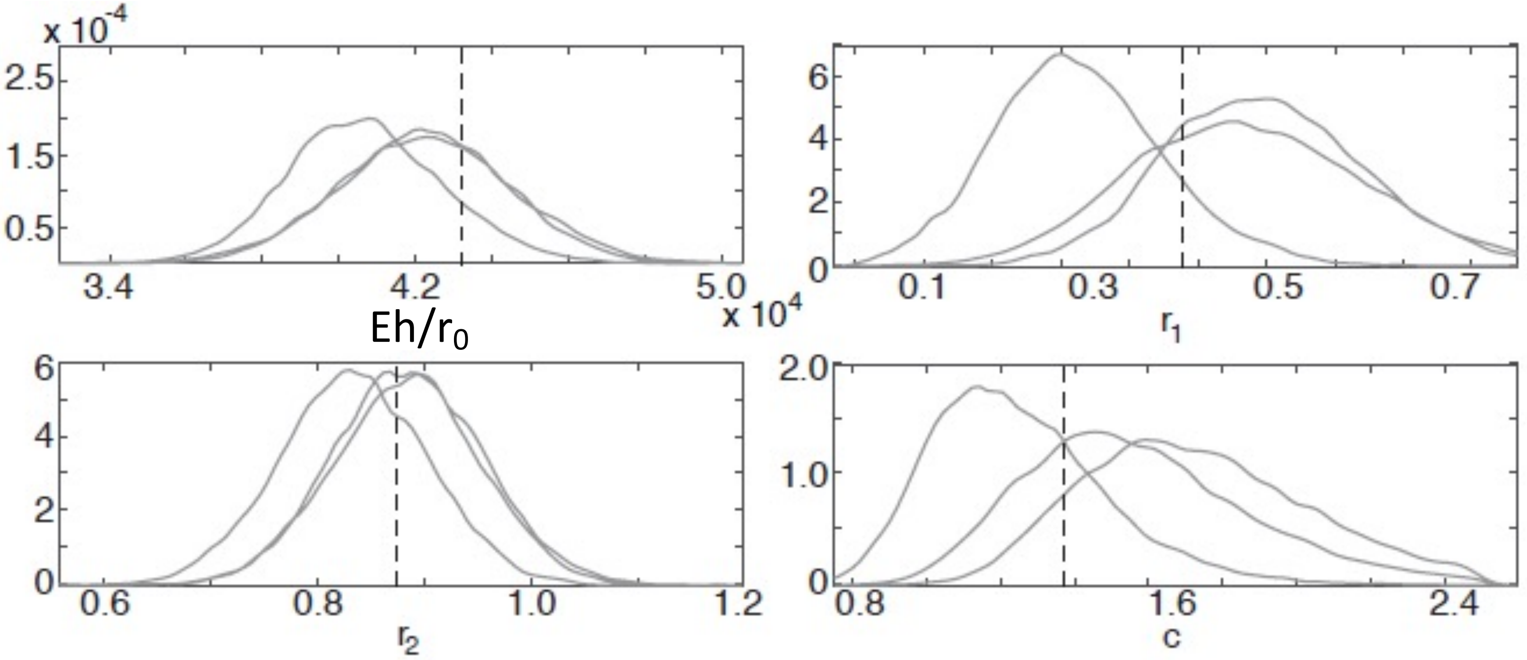
- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Synthetic data
(3 data sets generated with the same ‘true’ parameter values)



No model mismatch

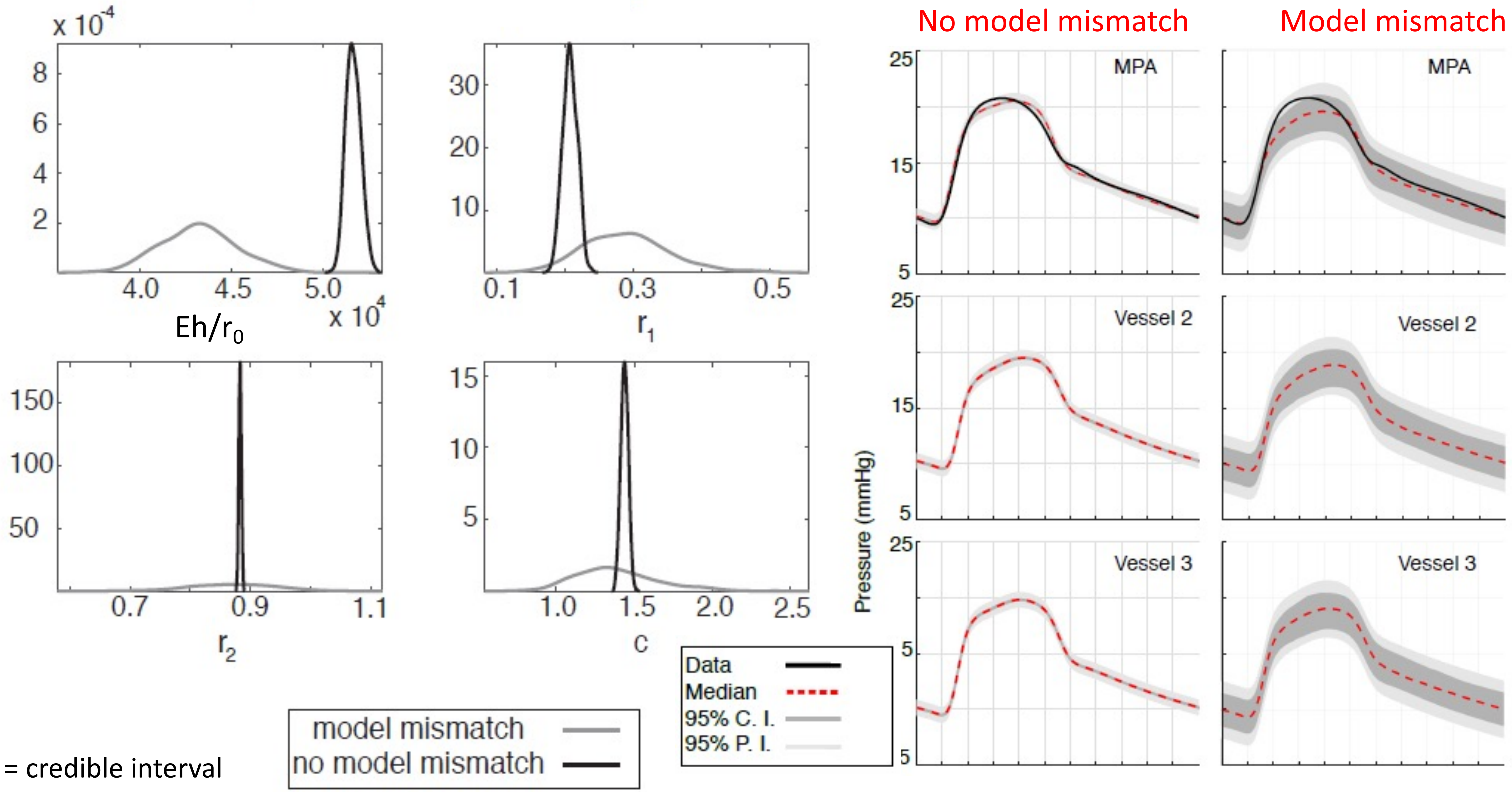
Model mismatch



Vertical dashed line: true parameter values

Measured (real) data

OUTPUT SPACE



C.I. = credible interval

P.I. = prediction interval

Euclidean distance: 66 VS 571

Statistical model selection

- Watanabe–Akaike information criterion (WAIC) computed based on MCMC posterior samples.

- WAIC:
  - Ignoring model mismatch: 408
  - Accounting for model mismatch: -4515

Model with lower WAIC score is preferred.

Outline

- **Background**
- **Motivation**
- **Aim**
- **Data**
- **Mathematical Model**
- **Parameter estimation**
- **Uncertainty quantification**
- **Numerical results**
- **Conclusions**

Conclusions

For mis-specific mathematical models, optimising the sum of squared errors ignores the model mismatch, and leads to:

- wrong parameter values and model predictions,
- uncertainty underestimation in input and output space.

- These issues can be overlooked if a point-estimate analysis is used.

Thank you! ☺

Any questions?

Gaussian Processes

$$
\begin{aligned}
\mathbf{y}|\mathbf{f} &\sim \mathcal{MVN}(\mathbf{f}, \sigma^2\mathbf{I}), \\
\mathbf{f}(\mathbf{X})|\gamma &\sim \mathcal{GP}(\mathbf{m}(\mathbf{X}), \mathbf{K}|\gamma), \\
\gamma, \sigma^2 &\sim p(\gamma)p(\sigma^2),
\end{aligned}
$$

$$
\mathbf{C} = \begin{pmatrix}
\mathrm{Cov}(r_1, r_1) & \mathrm{Cov}(r_1, r_2) & \mathrm{Cov}(r_1, r_3) & \ldots & \mathrm{Cov}(r_1, r_n) \\
\mathrm{Cov}(r_2, r_1) & \mathrm{Cov}(r_2, r_2) & \mathrm{Cov}(r_2, r_3) & \ldots & \mathrm{Cov}(r_2, r_n) \\
. & . & . & \ldots & . \\
. & . & . & \ldots & . \\
. & . & . & \ldots & . \\
\mathrm{Cov}(r_n, r_1) & \mathrm{Cov}(r_n, r_2) & \mathrm{Cov}(r_n, r_3) & \ldots & \mathrm{Cov}(r_n, r_n)
\end{pmatrix}
$$

$$
k(\mathbf{x}, \mathbf{x}'|\gamma) = \sigma_m^2 \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2l^2}\right)
$$