

Bayesian workflow

Elizaveta Semenova,
University of Oxford

Nordic Probabilistic AI School, Helsinki (ProbAI)
13 June 2022

Preface

- ▶ Objectives:
 - ▶ understand the basic concepts of Bayesian workflow,
 - ▶ consider specific modelling examples.
- ▶ Prerequisites to follow hands-on examples:
 - ▶ basic programming with R,
 - ▶ general understanding of Bayesian inference,
 - ▶ Stan (a PPL).
- ▶ Links to materials are available on GitHub:
github.com/elizavetasemenova/ProbAI-2022

Outline

Bayesian inference

Principles of Bayesian workflow

Example: Disease transmission modelling

Example: Concentration-response curve fitting

Conclusions

Bayesian inference

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Bayesian inference

$$\underbrace{p(\theta|y)}_{\text{posterior}} \propto \underbrace{p(y|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

Bayesian inference

$$\underbrace{p(\theta|y)}_{\text{posterior}} \propto \underbrace{p(y|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

What can possibly go wrong?

General principle

General principle of Bayesian inference:

- ▶ specify a complete Bayesian model
 - consider data $y = \{y_1, \dots, y_n\}$ and parameter θ

General principle

General principle of Bayesian inference:

- ▶ specify a complete Bayesian model
 - consider data $y = \{y_1, \dots, y_n\}$ and parameter θ
 - specify an **observation model**, e.g.

$$p(y|\theta) = \prod_i \mathcal{N}(y_i|\theta, 1)$$

General principle

General principle of Bayesian inference:

- ▶ specify a complete Bayesian model
 - consider data $y = \{y_1, \dots, y_n\}$ and parameter θ
 - specify an **observation model**, e.g.

$$p(y|\theta) = \prod_i \mathcal{N}(y_i|\theta, 1)$$

- complete the model with a **prior distribution**, e.g.

$$p(\theta) = \mathcal{N}(0, 1)$$

- ▶ sample the **posterior distribution** of the parameter θ .

General principle

General principle of Bayesian inference:

- ▶ specify a complete Bayesian model
 - consider data $y = \{y_1, \dots, y_n\}$ and parameter θ
 - specify an **observation model**, e.g.

$$p(y|\theta) = \prod_i \mathcal{N}(y_i|\theta, 1)$$

- complete the model with a **prior distribution**, e.g.

$$p(\theta) = \mathcal{N}(0, 1)$$

- ▶ sample the **posterior distribution** of the parameter θ .

Sometimes posterior is available in a closed form. But rarely.

Probabilistic programming languages (PPLs)

PPLs from a user's perspective:

- ▶ PPLs are designed to let the user focus on modelling while inference happens automatically.
- ▶ Users need to specify
 1. prior,
 2. likelihood.
- ▶ Inference is performed via powerful algorithms such as MCMC variations.
- ▶ Availability of diagnostic tools.

Diagnosing MCMC outputs

- ▶ Convergence diagnostics

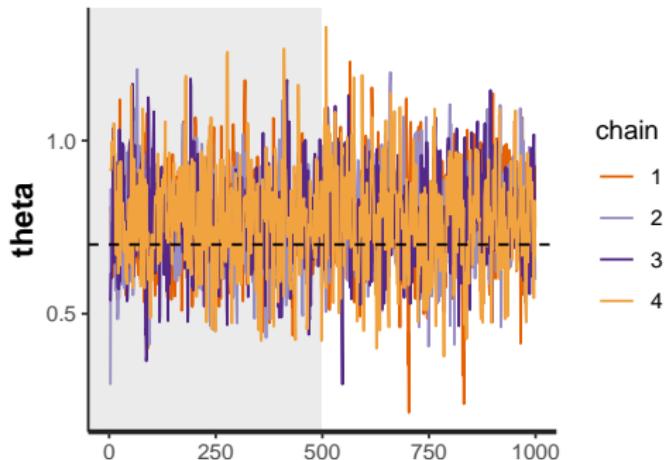
- \hat{R} ,
- traceplots

- ▶ Effective sample size (ESS):

- samples will be typically **autocorrelated** within a chain, which increases the uncertainty of the estimation of posterior quantities
- ESS – number of **independent** samples required to obtain the same level of uncertainty as from the available dependent samples

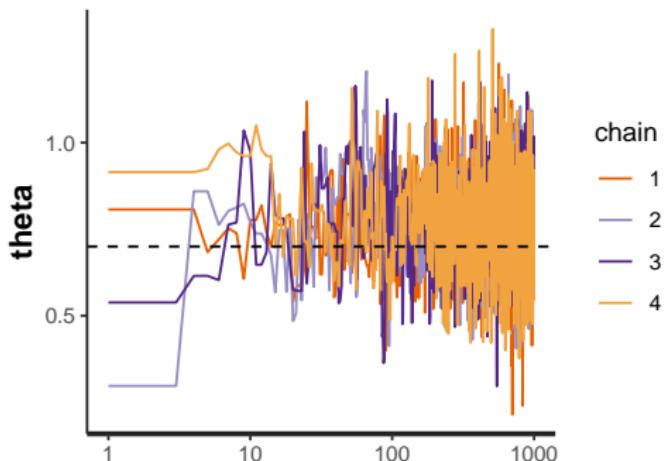
Diagnosing MCMC outputs

We use **multiple chains** and inspect convergence after warm-up



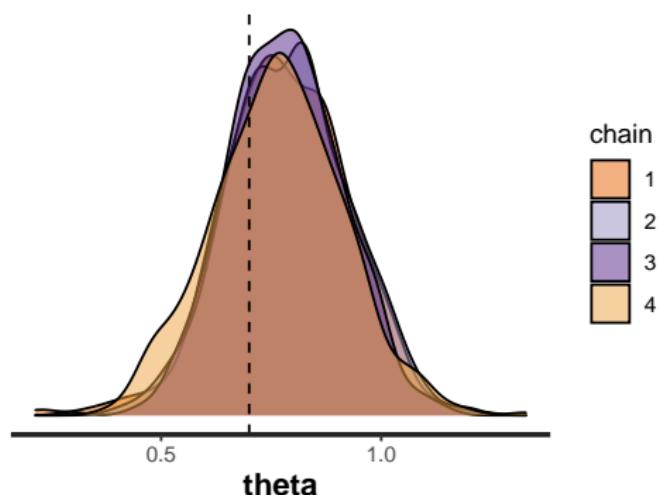
Diagnosing MCMC outputs

We use **multiple chains** and inspect convergence after warm-up



Diagnosing MCMC outputs

The post-warm-up samples of θ approximate its posterior distribution



Outline

Bayesian inference

Principles of Bayesian workflow

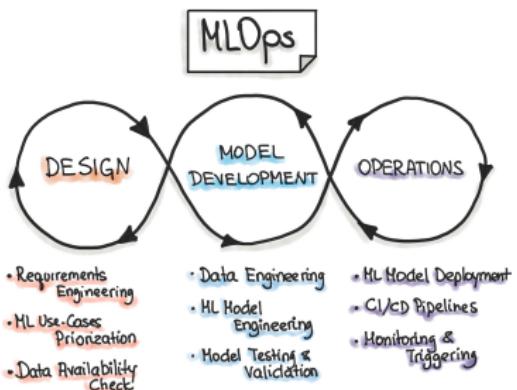
Example: Disease transmission modelling

Example: Concentration-response curve fitting

Conclusions

Workflows as 'good practice' standards

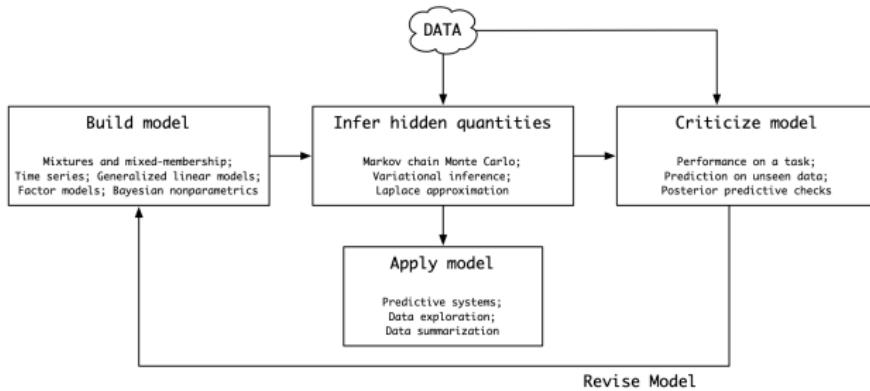
Workflows exist in a variety of disciplines. For example, in machine learning workflow standards are being formalised under the name of MLOps:



Credit: <https://ml-ops.org/>

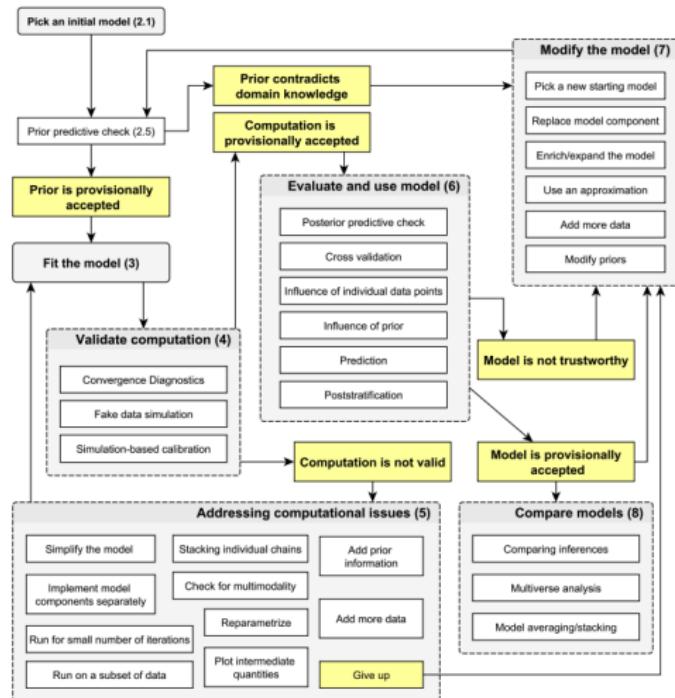
Box's loop

In the 1960's, the statistician Box formulated the notion of a loop to understand the nature of the scientific method. This loop is called Box's loop by Blei et. al. (2014):



Modern Bayesian workflow

A systematic review of the steps within the modern Bayesian workflow, described in Gelman et al. (2020):



Prior predictive checks

Prior predictive checking consists in simulating data from the priors:

- ▶ visualize priors (especially after transformation)
- ▶ this shows the range of data compatible with the model
- ▶ it helps understand the adequacy of the chosen priors, as it is often easier to elicit expert knowledge on measureable quantities of interest rather than abstract parameter values

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the **domain** and problem,

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the **domain** and problem,
- ▶ Formulate the model **mathematically**,

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the **domain** and problem,
- ▶ Formulate the model **mathematically**,
- ▶ Implement model, test, **debug**,

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the **domain** and problem,
- ▶ Formulate the model **mathematically**,
- ▶ Implement model, test, **debug**,
- ▶ Perform **prior predictive** check,

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the domain and problem,
- ▶ Formulate the model mathematically,
- ▶ Implement model, test, debug,
- ▶ Perform prior predictive check,
- ▶ Fit the model,

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the domain and problem,
- ▶ Formulate the model mathematically,
- ▶ Implement model, test, debug,
- ▶ Perform prior predictive check,
- ▶ Fit the model,
- ▶ Assess convergence diagnostics,

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the domain and problem,
- ▶ Formulate the model mathematically,
- ▶ Implement model, test, debug,
- ▶ Perform prior predictive check,
- ▶ Fit the model,
- ▶ Assess convergence diagnostics,
- ▶ Perform posterior predictive check,

Iterative model building

A possible realisation of the Bayesian workflow loop:

- ▶ Understand the domain and problem,
- ▶ Formulate the model mathematically,
- ▶ Implement model, test, debug,
- ▶ Perform prior predictive check,
- ▶ Fit the model,
- ▶ Assess convergence diagnostics,
- ▶ Perform posterior predictive check,
- ▶ Improve the model iteratively: from baseline to complex and computationally efficient models.

Outline

Bayesian inference

Principles of Bayesian workflow

Example: Disease transmission modelling

Example: Concentration-response curve fitting

Conclusions

Introduction

There exist many classes of disease transmission models:

- ▶ Interpretability: **mechanistic**
- ▶ Scale: **population-based**, agent-based
- ▶ Framework: **deterministic**, stochastic
- ▶ Data-generating mechanisms: incubation, contagion, immunity, vaccination, mobility...

Introduction

There exist many classes of disease transmission models:

- ▶ Interpretability: mechanistic
- ▶ Scale: population-based, agent-based
- ▶ Framework: deterministic, stochastic
- ▶ Data-generating mechanisms: incubation, contagion, immunity, vaccination, mobility...

Mechanistic + population-based + deterministic

→ ordinary differential equations (ODE)-based compartmental model

Introduction

ODE-based compartmental model:

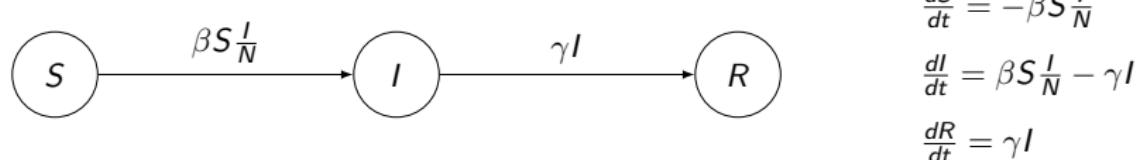
- ▶ Divide the population into homogeneous groups (**compartments**)
- ▶ Define the **flows** between compartments with ODEs
- ▶ Define initial conditions (at t_0)
- ▶ Solve for the time-dependent volume in each compartment

Introduction

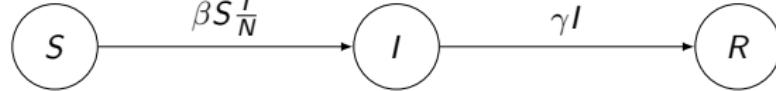
ODE-based compartmental model:

- ▶ Divide the population into homogeneous groups (**compartments**)
- ▶ Define the **flows** between compartments with ODEs
- ▶ Define initial conditions (at t_0)
- ▶ Solve for the time-dependent volume in each compartment

The **susceptible-infectious-recovered** (SIR) model:



The SIR model



$$\frac{dS}{dt} = -\beta S \frac{I}{N}$$

$$\frac{dI}{dt} = \beta S \frac{I}{N} - \gamma I$$

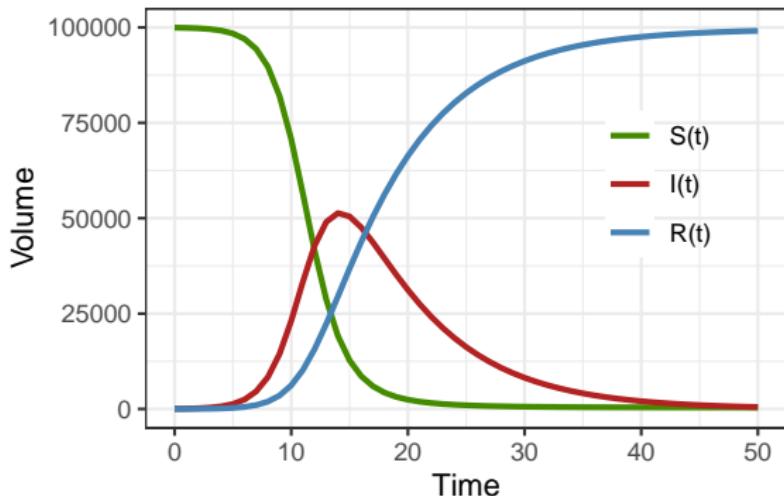
$$\frac{dR}{dt} = \gamma I$$

Where:

- ▶ $S(t)$ is the number of people **susceptible** to infection
- ▶ $I(t)$ is the number of people **infected** (i.e. the prevalence)
- ▶ $R(t)$ is the number of people **recovered** (lifelong immunity)
- ▶ N is the population size ($S(t) + I(t) + R(t) = N$ for any t)
- ▶ β is the **infectious contact rate** (per day per person)
- ▶ γ is the **recovery rate** (1/infectious period)

Simulate a SIR

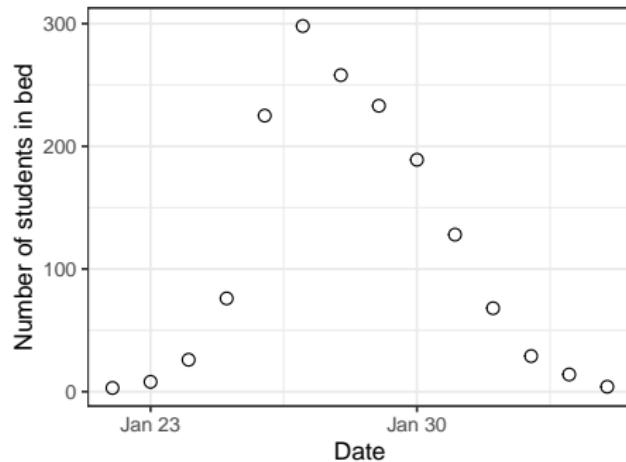
with $\beta = 0.8$; $\gamma = 1/7$; $S_0 = 100000 - 50$; $I_0 = 50$ and $R_0 = 0$



Example dataset

Outbreak of influenza A (H1N1) at a British boarding school in 1978 (available in R package outbreaks)

- ▶ 763 students, 512 had symptoms
- ▶ daily number of students in bed over 14 days (prevalence)



Specify the model

Points to consider:

- ▶ prevalence data: \mathbb{I}_t with $t \in \{1, \dots, 14\}$
- ▶ inputs that will remain fixed: $\{S_0 = 762, I_0 = 1, R_0 = 0\}$
- ▶ map data \mathbb{I}_t to SIR model output $I(t)$ using an observation model with an appropriate likelihood:

$$p(\mathbb{I}|\theta) = \prod_{t=1}^{14} \text{NegBin}(\mathbb{I}_t | I(t), \phi)$$

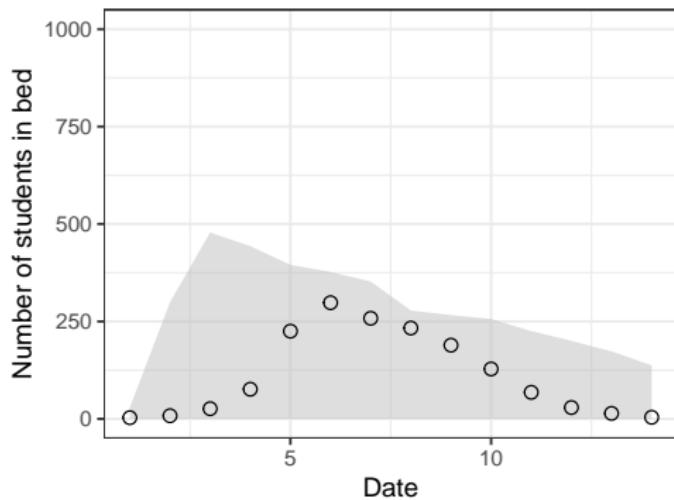
- ▶ parameters to estimate: $\theta = \{\beta, \gamma, \phi\}$
- ▶ prior distributions $p(\beta) = \text{Exp}(1)$

$$p(1/\gamma) = \mathcal{N}^+(2, 0.5)$$

$$p(1/\phi) = \text{Exp}(5)$$

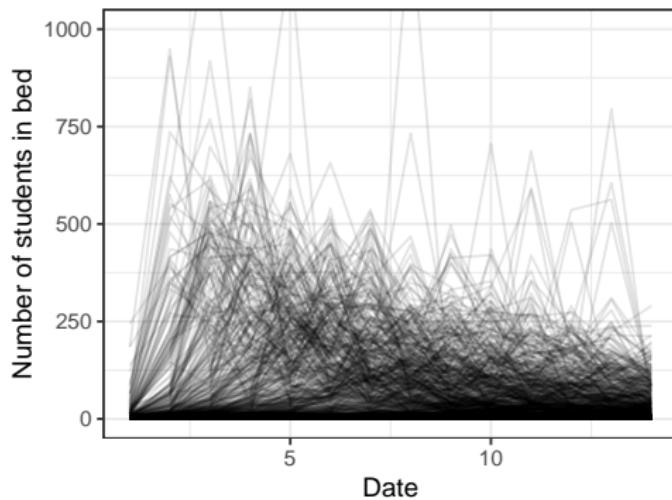
Prior predictive checks

Prior predictive check: simulating potential epidemic trajectories



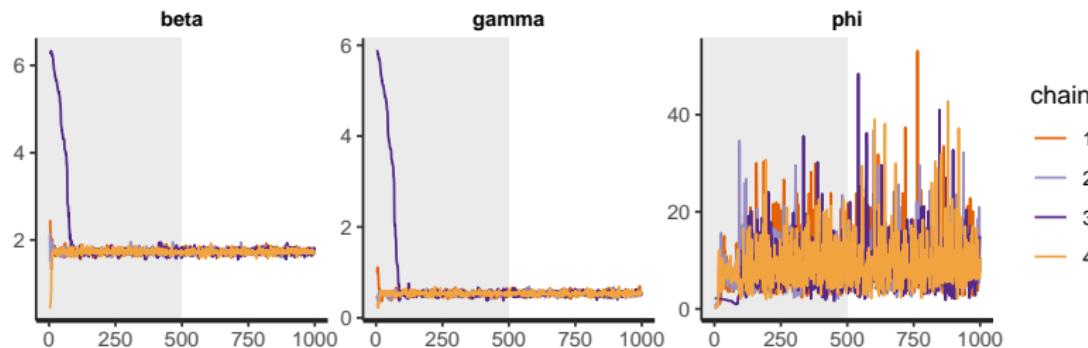
Prior predictive checks

Prior predictive check: simulating potential epidemic trajectories



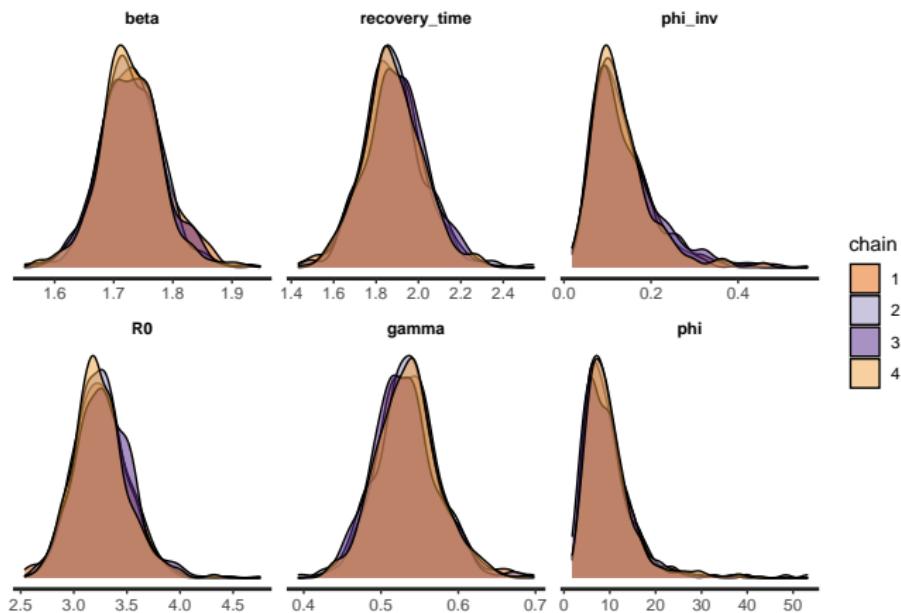
Diagnostics

Examine trace plots:



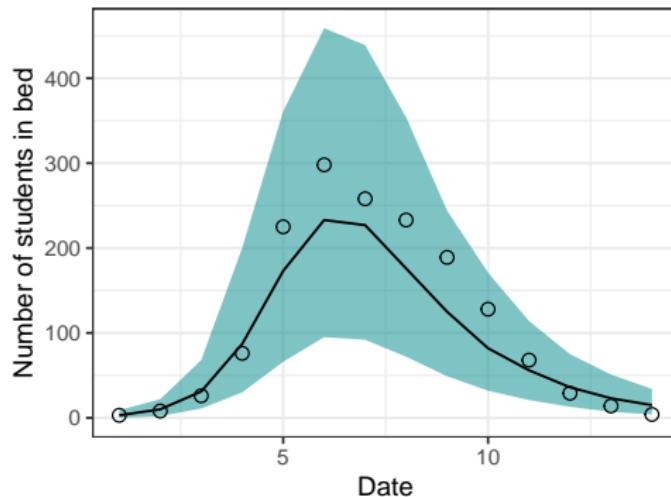
Diagnostics

Examine chain mixing:

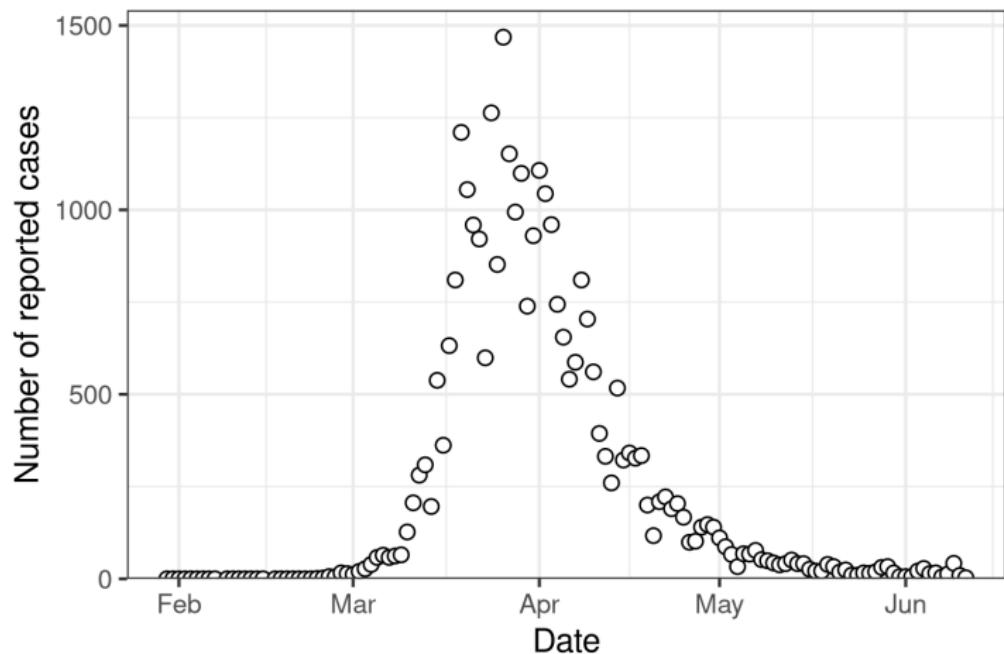


Diagnostics

Posterior predictive checking:

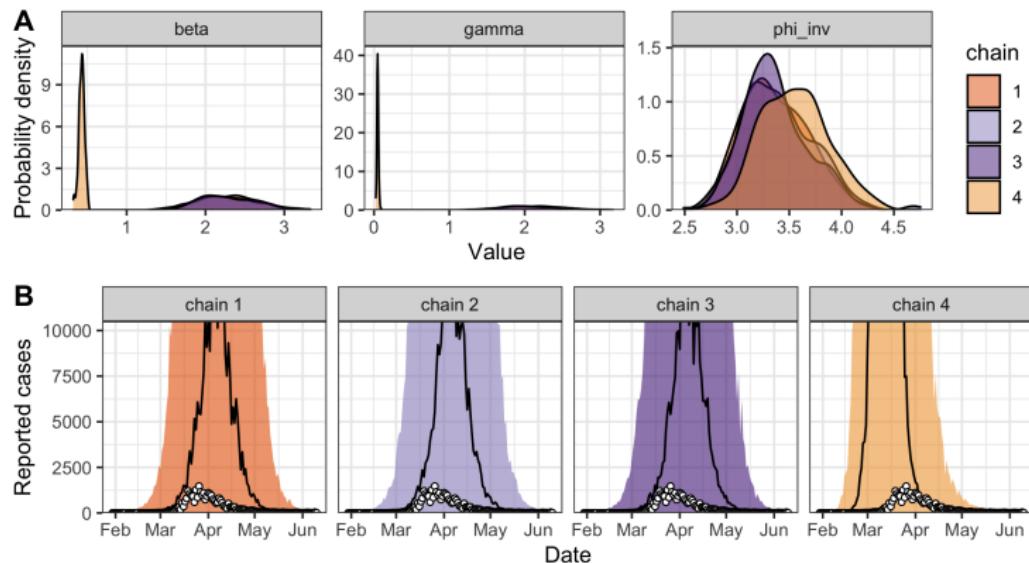


New Covid-19 cases in Switzerland in 2020

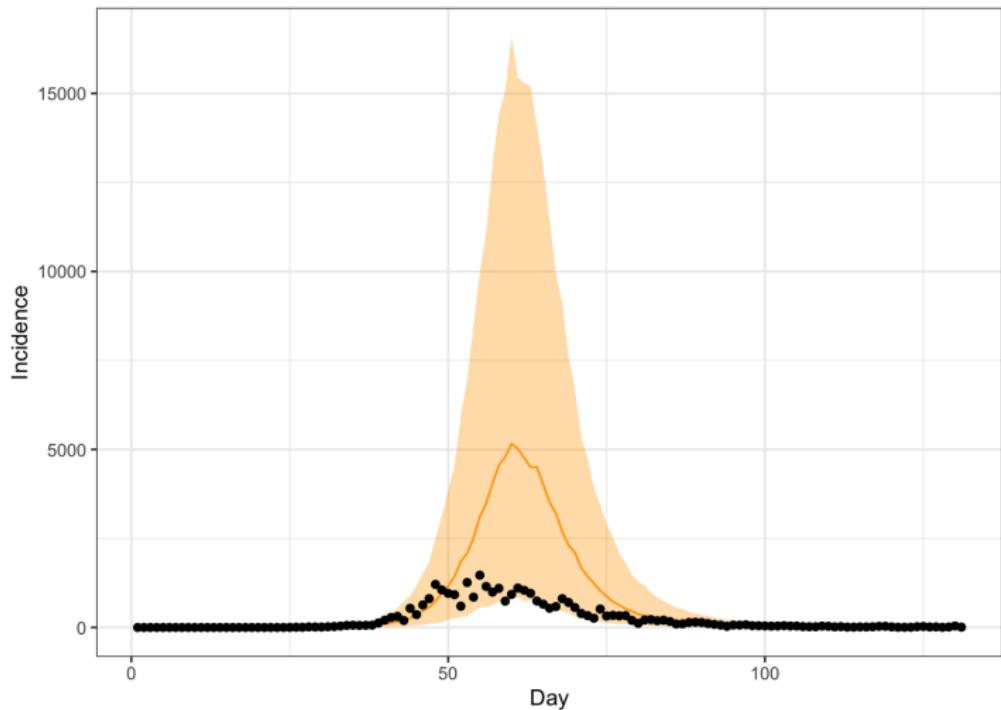


Basic SIR

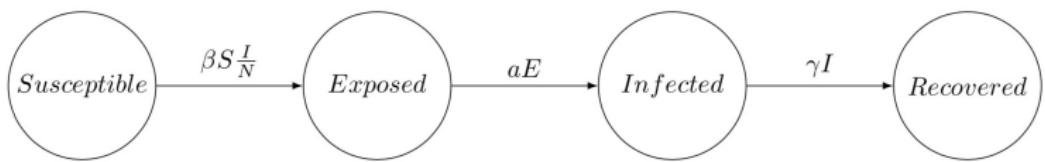
```
## Warning: The largest R-hat is 1.54, indicating chains have not mixed.  
## Running the chains for more iterations may help. See  
## http://mc-stan.org/misc/warnings.html#r-hat
```



Adding underreporting: posterior predictive check

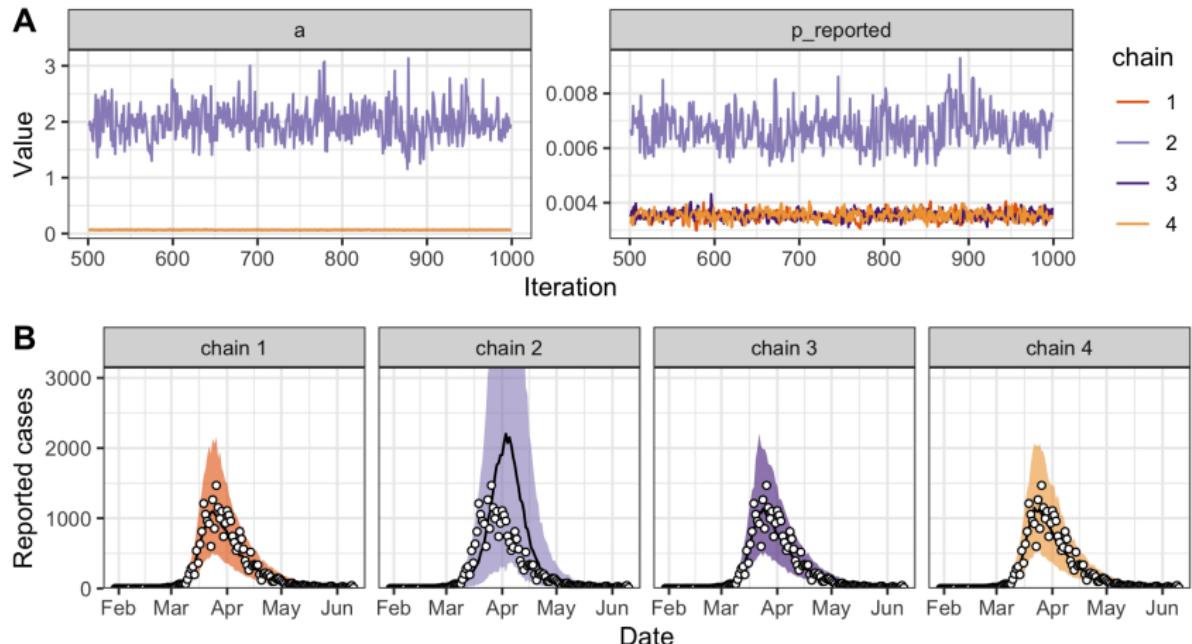


Adding incubation time

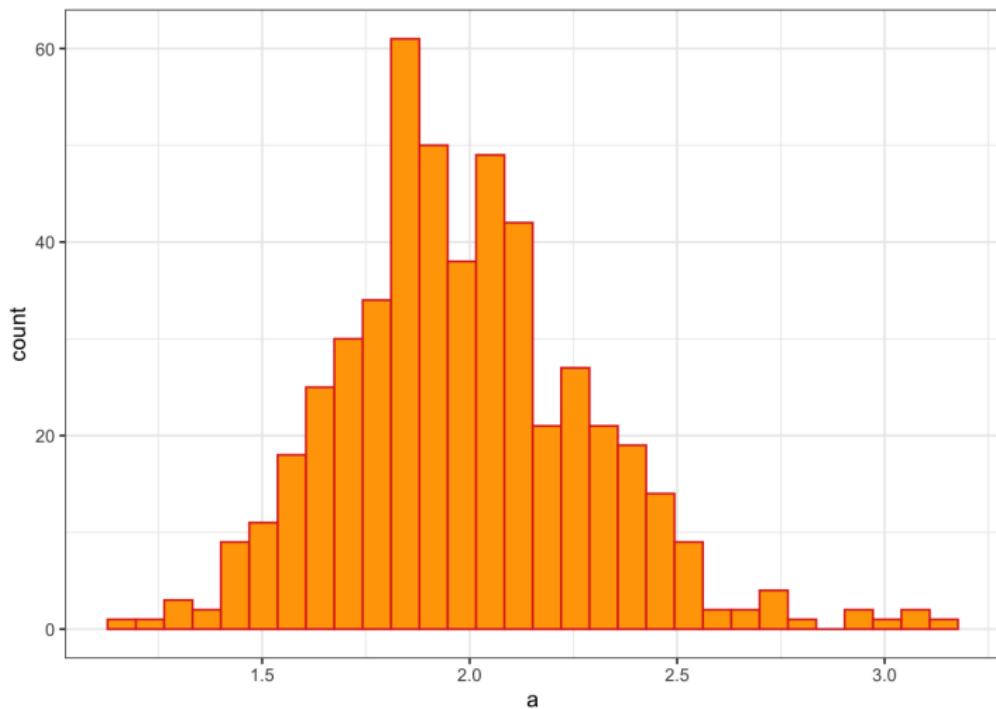


$a \sim \mathcal{N}^+(0.4, 0.5),$
1/a corresponds to incubation time

Adding incubation time



Adding incubation time

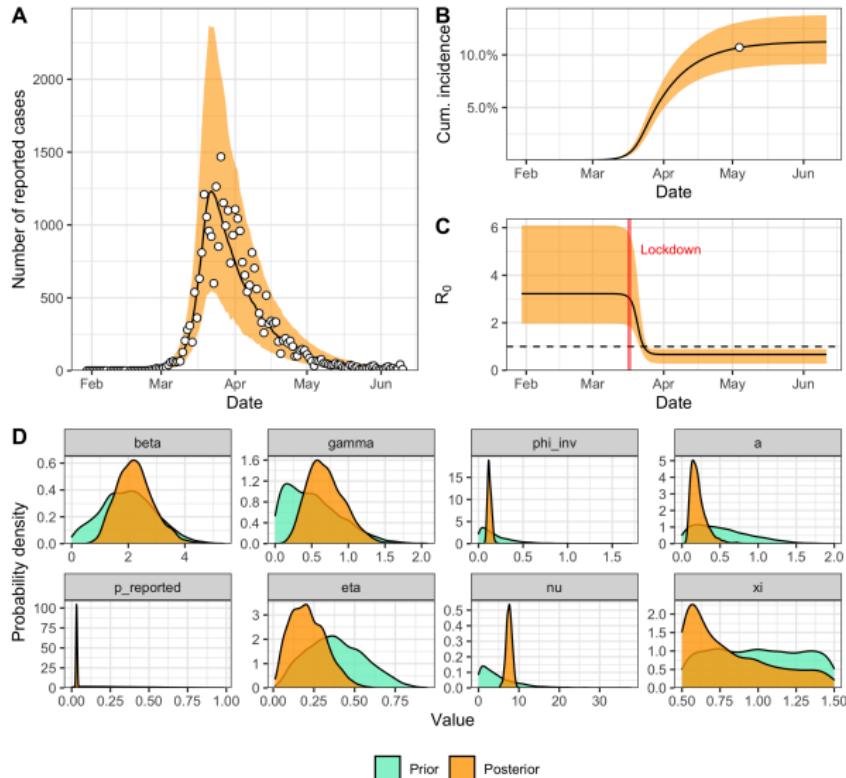


Domain knowledge $\rightarrow \frac{1}{a} \sim \mathcal{N}^+(6, 1)$

Additional considerations

- ▶ Modeling control measures: $\beta(t) = f(t) * \beta$,
- ▶ Using serological survey data: survey data from May 4 to May 7 in Geneva (83 / 775 have antibodies).

Final model



Outline

Bayesian inference

Principles of Bayesian workflow

Example: Disease transmission modelling

Example: Concentration-response curve fitting

Conclusions

Introduction

- ▶ Concentration-response (CR) experiments are used to rank drug candidates.

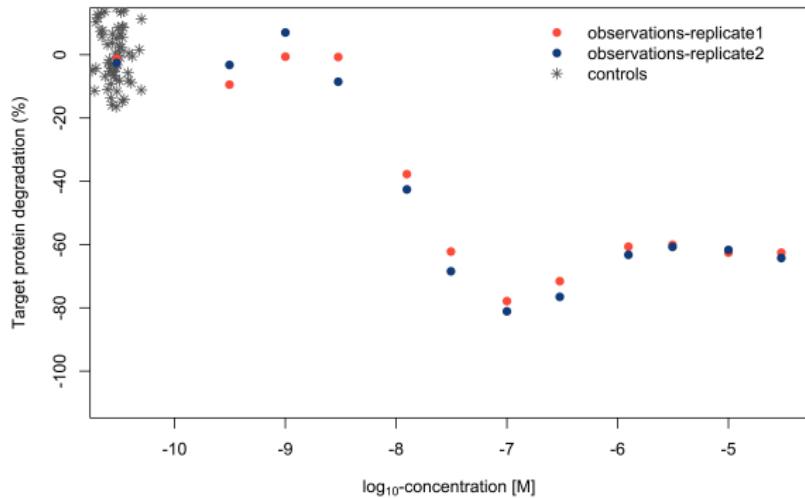
Introduction

- ▶ Concentration-response (CR) experiments are used to rank drug candidates.
- ▶ Traditional small molecules typically yield sigmoidal curves, characterized by a plateau at high drug concentrations.

Introduction

- ▶ Concentration-response (CR) experiments are used to rank drug candidates.
- ▶ Traditional small molecules typically yield sigmoidal curves, characterized by a plateau at high drug concentrations.
- ▶ CR curves of a new drug modality show a loss of efficacy at higher doses, known as the 'hook effect'.

Understanding data



Domain understanding

We are looking to fit a curve which

- ▶ is flat at low concentrations (no compound activity),
- ▶ is able to capture curve characteristics at higher concentrations (the 'hook effect').

Likelihood

To build a likelihood, we account for two sources of uncertainty:

- ▶ curve uncertainty,
- ▶ replicate-to-replicate variation.

We denote our mean predicted curve by \underline{y} .

Traditional Hill's (4PL) model

$$\underline{y}(x) = d + \frac{a - d}{1 + \exp(-b(x-c))}$$

d : degradation at zero concentration,

a : D_{\max} - maximal degradation,

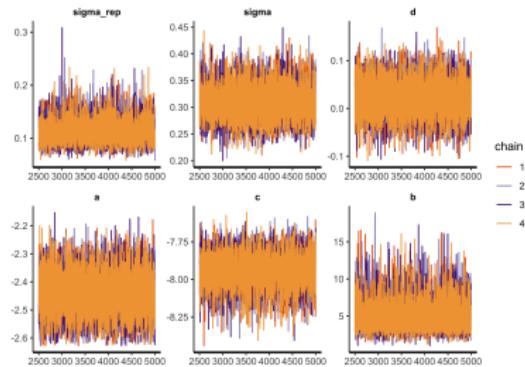
c : $\log_{10}(DC_{50})$ - concentration of half-degradation,

x : dose on the \log_{10} -scale,

b : Hill's slope (slope at the half-degradation point).

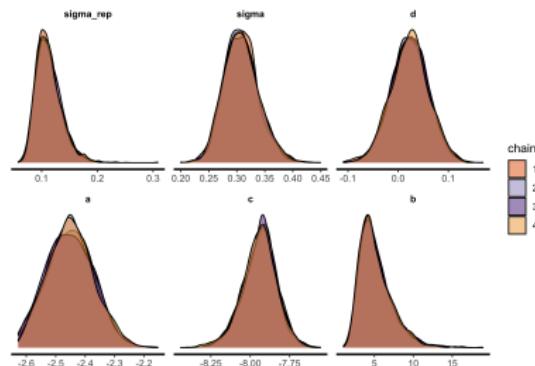
Diagnostics

Examine trace plots:



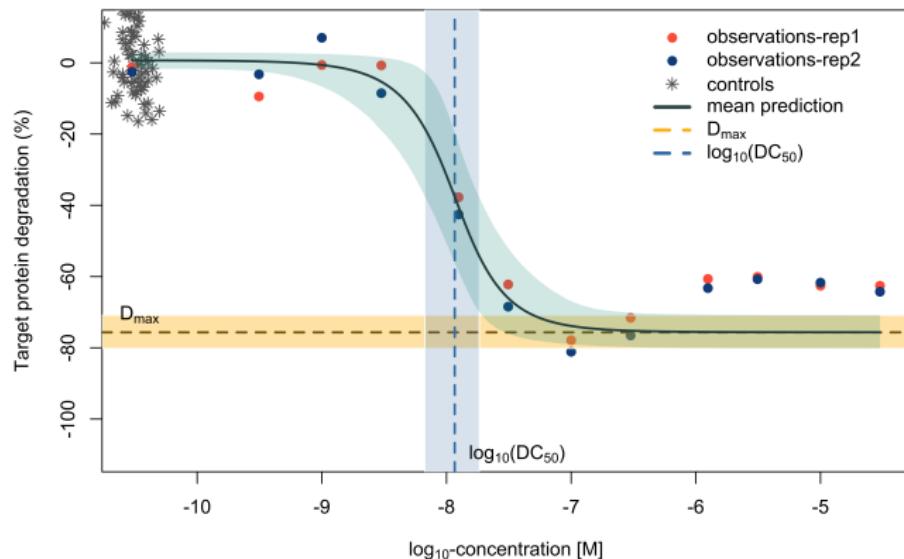
Diagnostics

Examine chain mixing:



Diagnostics

Posterior predictive:



Gaussian Process (GP) model

Gaussian Process model allows to fit **flexible curve shapes**. It is defined as

$$f \sim \text{GP}(0, k).$$

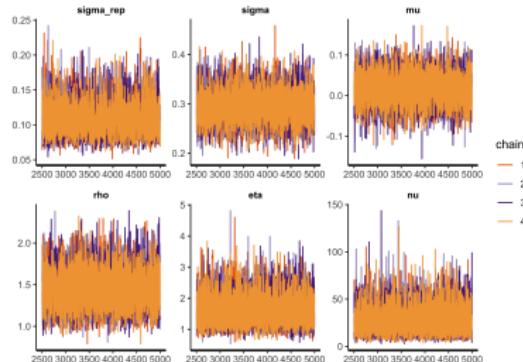
On a finite set of points is evaluated as a **multivariate normal** with covariance matrix K . For example

$$K[i, j] = \eta^2 \exp\left(-\frac{(x_i - x_j)^2}{\rho^2}\right).$$

Parameters η and ρ define the amplitude and lengthscale of the curve, correspondingly.

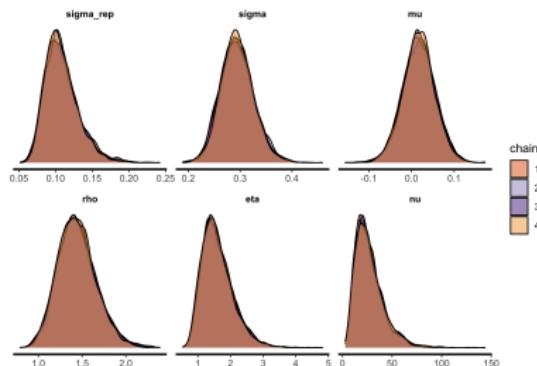
Diagnostics

Examine trace plots:



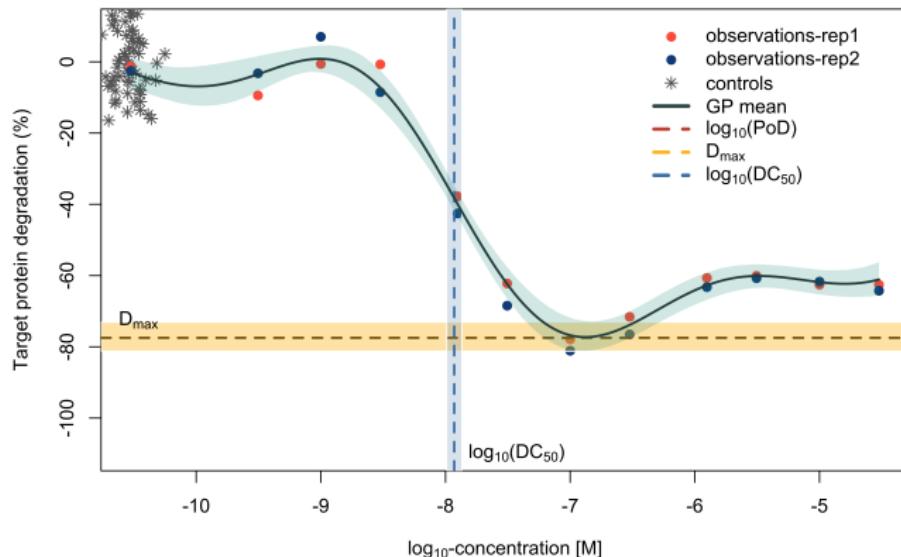
Diagnostics

Examine chain mixing:



Diagnostics

Posterior predictive:



Changepoint Gaussian process

Kernel design allows to specify a wider range of GP priors. Given two GPs

$$\begin{aligned}f_1(x) &\sim \text{GP}(0, k_1), \\f_2(x) &\sim \text{GP}(0, k_2),\end{aligned}$$

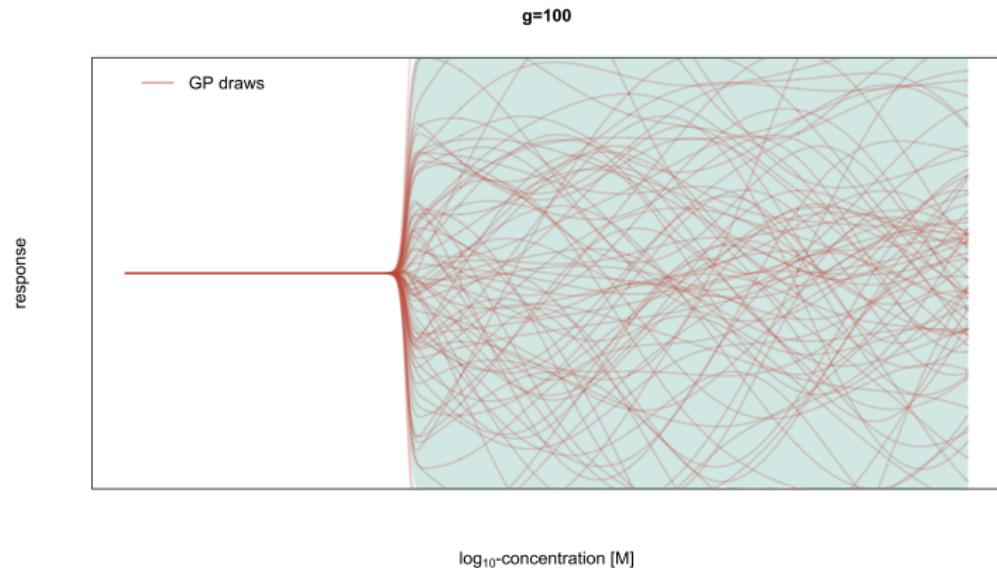
we can construct a new one

$$\begin{aligned}f_\theta(x) &= (1 - w_\theta(x))f_1(x) + w_\theta(x)f_2(x), \\w_\theta(x) &= \sigma(g(x - \theta)), g > 1.\end{aligned}$$

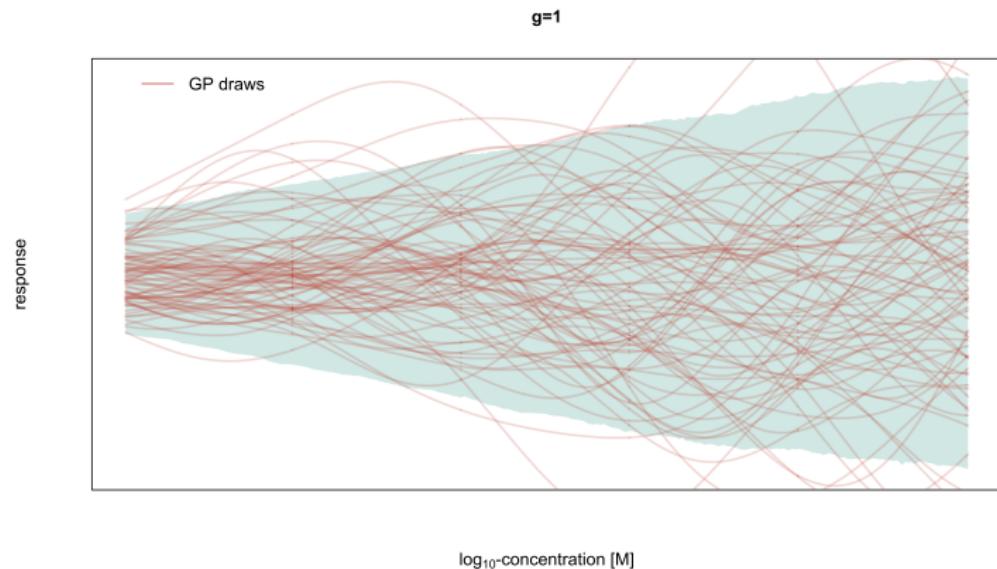
Then

$$\begin{aligned}f_\theta(x) &\sim \text{GP}(0, k_\theta), \\k_\theta(x, x') &= (1 - w_\theta(x))k_1(x, x')(1 - w_\theta(x')) + w_\theta(x)k_2(x, x')w_\theta(x').\end{aligned}$$

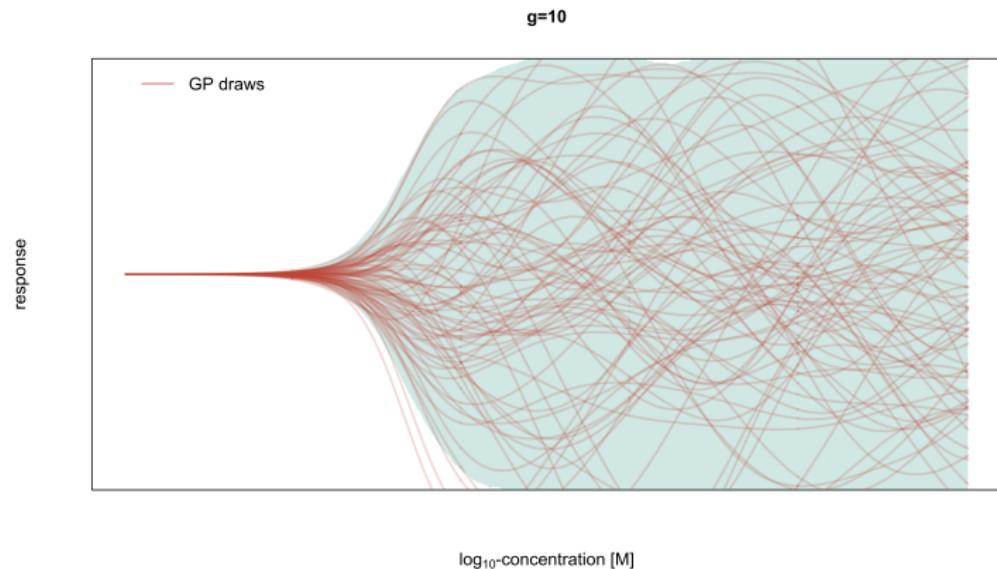
Changepoint GP priors



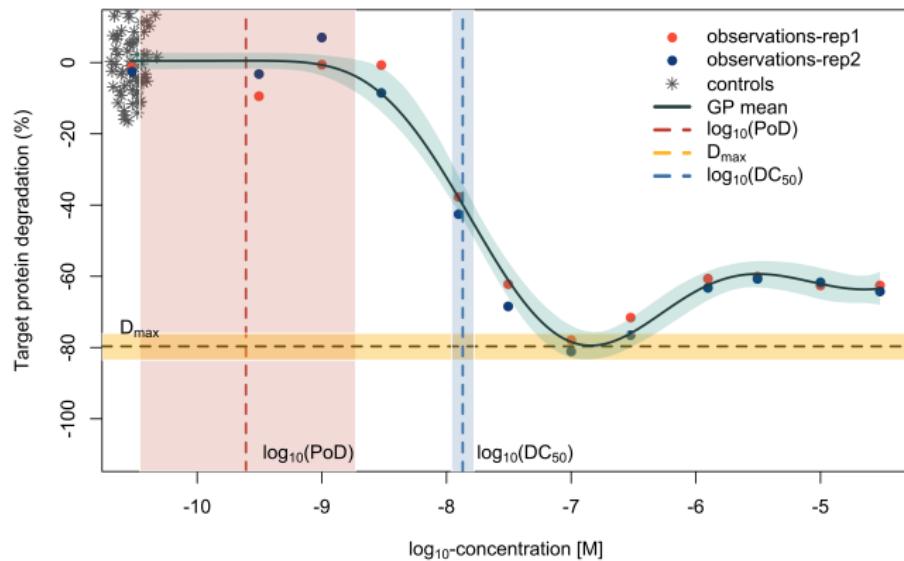
Changepoint GP priors



Changepoint GP priors



Changepoint GP prediction



Outline

Bayesian inference

Principles of Bayesian workflow

Example: Disease transmission modelling

Example: Concentration-response curve fitting

Conclusions

Conclusions

We have considered two examples from different application domains. The emerging **issues** were solved via a variety of different steps and included

- ▶ using of **domain understanding**,
- ▶ priors informed by **domain knowledge**,
- ▶ using **external data**,
- ▶ using a **more complex model**.

There is no "One size fits all" solution. That is why we need the Bayesian workflow.

References

- ▶ David M Blei, *Build, compute, critique, repeat: Data analysis with latent variable models*
- ▶ Andrew Gelman et al., *Bayesian workflow*
- ▶ Jonah Gabry et al., *Visualization in Bayesian workflow*
- ▶ Leo Grinsztajn et al., *Bayesian workflow for disease transmission modeling in Stan*
- ▶ Semenova et al., *Flexible Fitting of PROTAC Concentration–Response Curves with Changepoint Gaussian Processes*

Thank You!

Please send any feedback to
elizaveta.p.semenova@gmail.com

Announcement



We are hiring!

- ▶ Postdocs at Oxford,
- ▶ All levels from PhD to assistant professor at University of Copenhagen,
- ▶ A disease transmission modeller in London.

Contacts:

- ▶ Samir Bhatt: samir.bhatt@sund.ku.dk
- ▶ Seth Flaxman: seth.flaxman@cs.ox.ac.uk
- ▶ MLGlobalHealth@gmail.com