Centre for Data Science

Bayesian approaches for Hawkes processes, and their applications

Raiha Browning

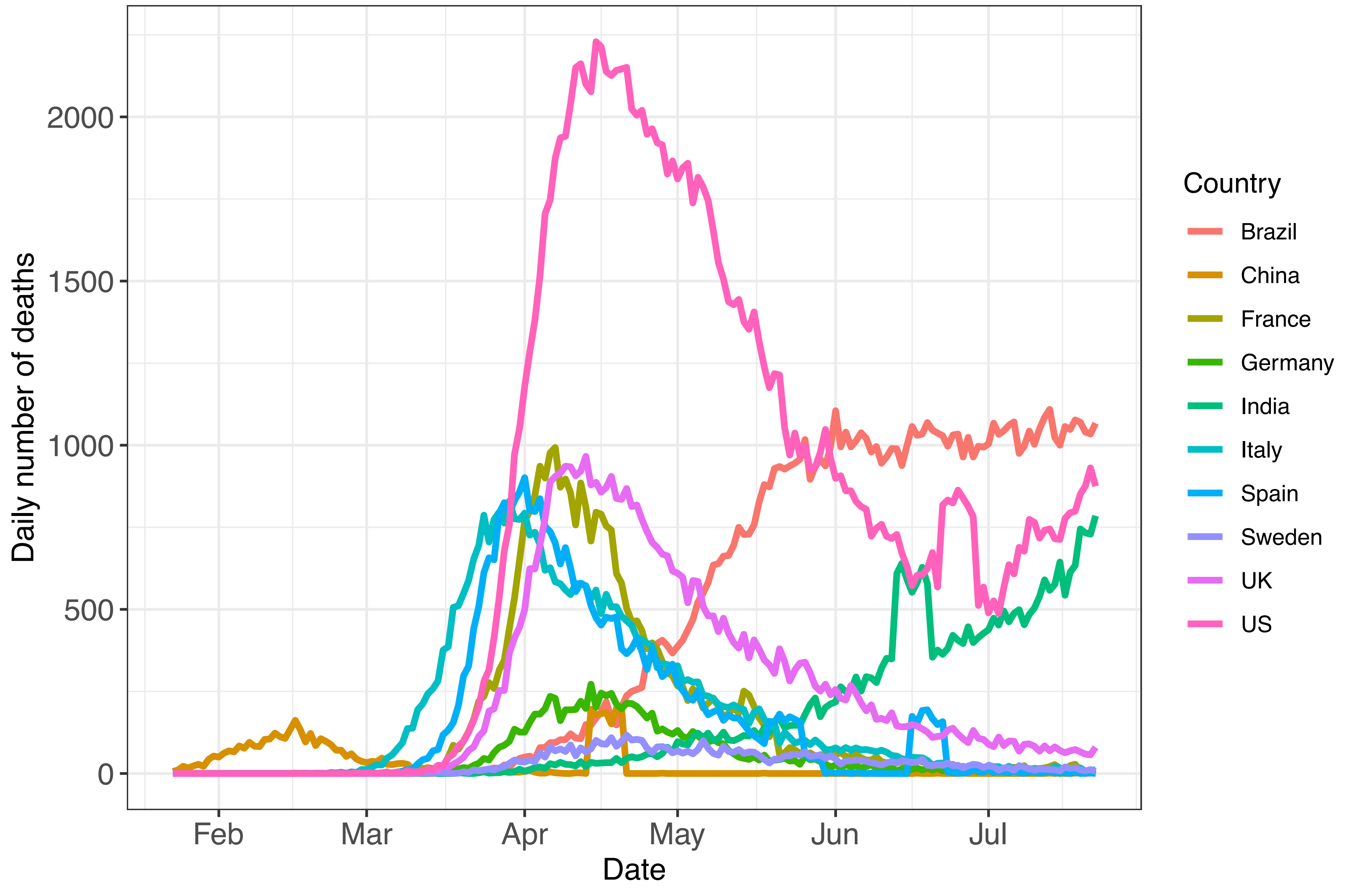
Final Seminar, 15 July 2022

research.qut.edu.au/qutcds

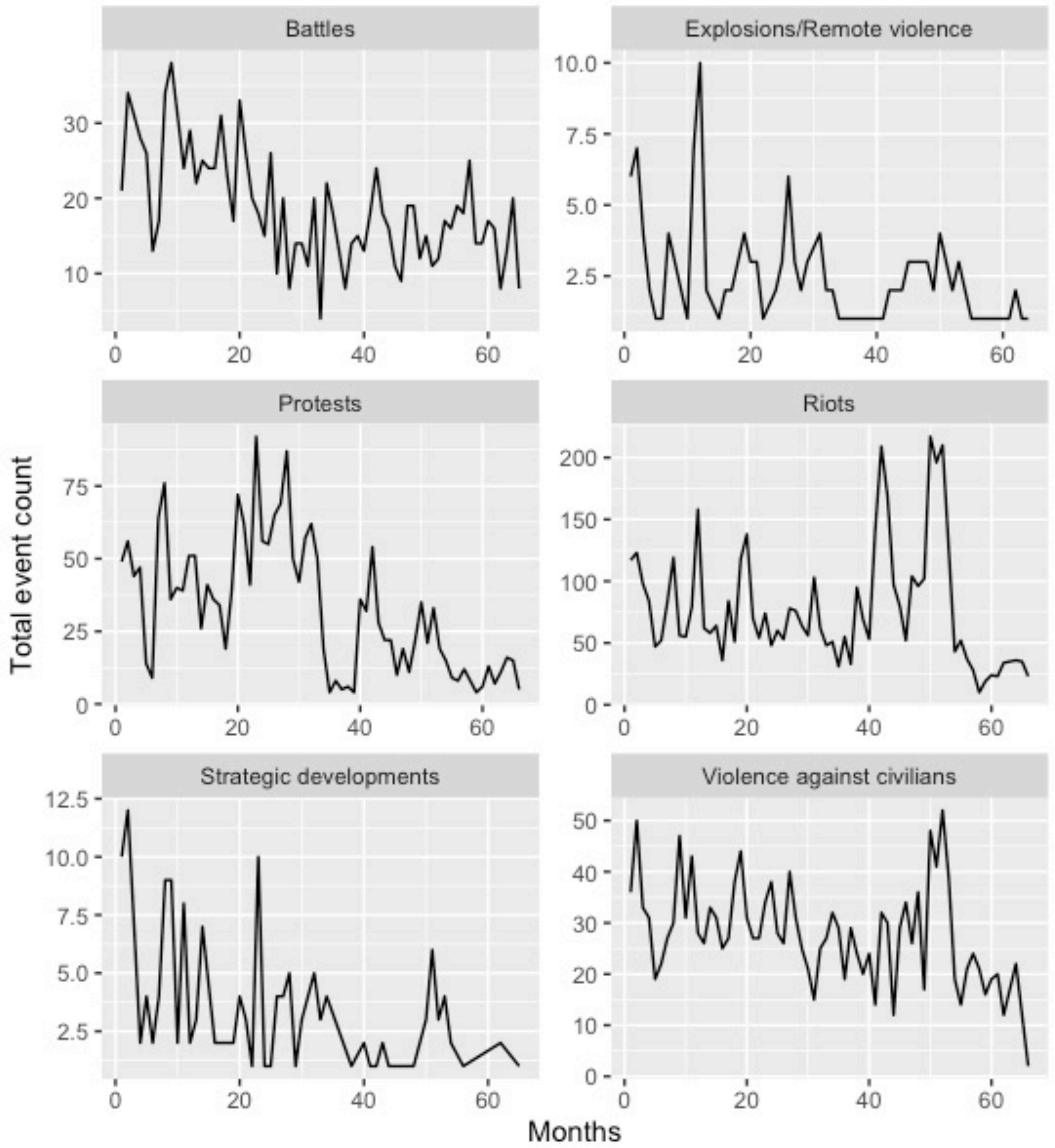
Motivation

Hawkes process

- A Hawkes process (HP) [Hawkes, 1971] is a self-exciting point process.
- Describes phenomena where current events increase the probability of future events.
- Useful in a number of applications including:
  - Seismic activity
  - Social media interactions
  - Neuron spikes
  - Financial markets
  - Conflict and terrorism
  - Infectious diseases



Daily COVID-19 mortality counts



Daily conflict event counts in Bangladesh

Hawkes process

- Consider a univariate Hawkes process $N(t)$
- The HP is characterised by its conditional intensity function,

$$\lambda(t \mid \mathscr{H}_t) = \lim_{h \to 0} \frac{\mathbb{E}[N(t+h) - N(t) \mid \mathscr{H}_t]}{h}$$

$$= \underbrace{\mu(t)}_{\text{baseline process}} + \underbrace{\alpha \sum_{i:t_i<t} g(t - t_i)}_{\text{self-exciting process}}$$

where $\mathscr{H}(t)$ is the history of the process prior to time t,
α is a magnitude parameter, and
$g(\cdot)$ is the triggering kernel

Hawkes process

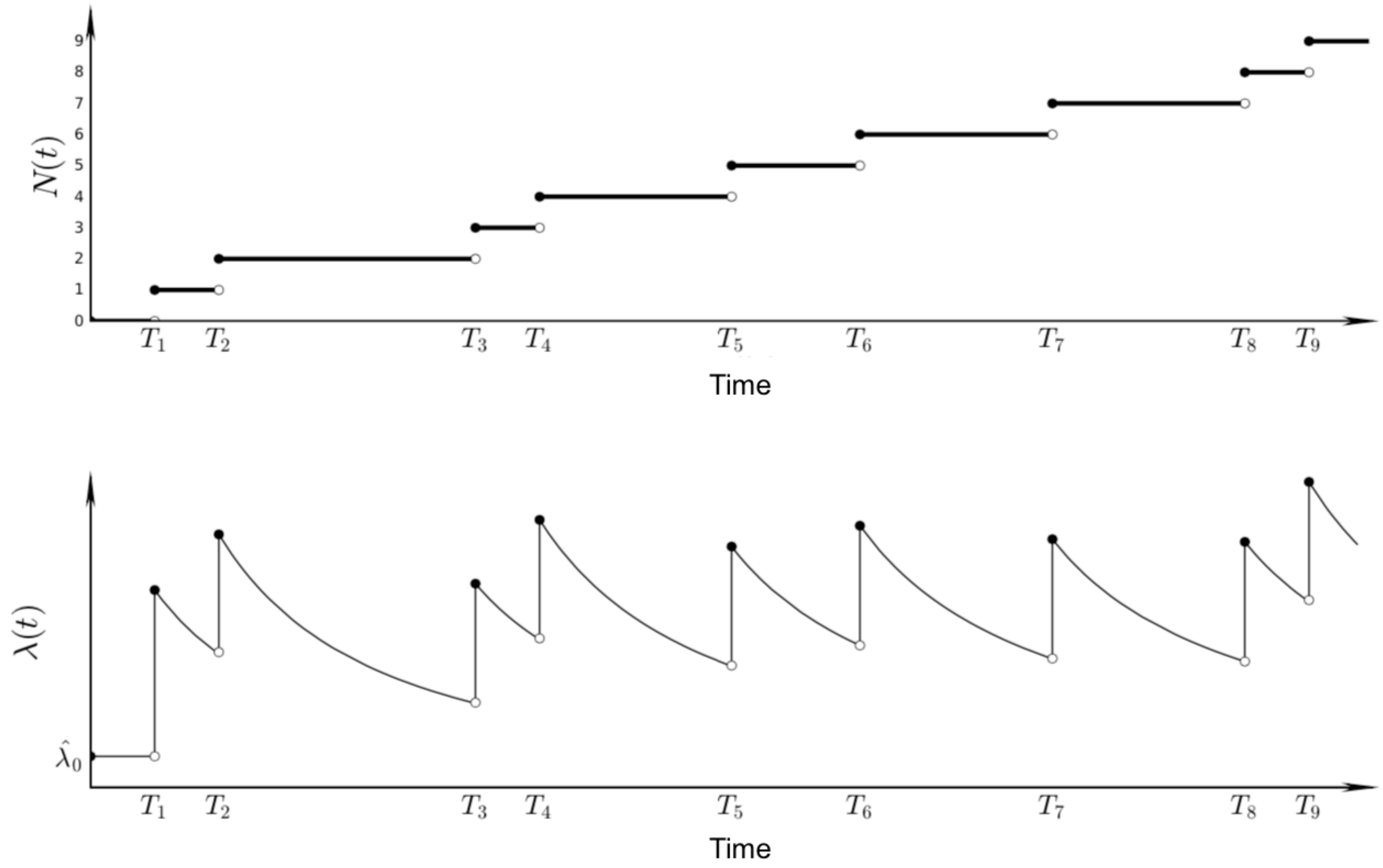


Image source: Rizoiu, M. et al., 2017. A Tutorial on Hawkes Processes for Events in Social Media. arXiv, stat.ML

Multivariate Hawkes process

- Extend to a K-dimensional Hawkes process $\boldsymbol{N(t)} = (N_1(t), N_2(t), \ldots, N_K(t))$
- The conditional intensity function is then,

$$\lambda^k(t \mid \mathscr{H}_t) = \underbrace{\mu^k(t)}_{\text{baseline process}} + \underbrace{\sum_{l=1}^{K} \alpha^{lk} \sum_{i:t_i^l<t} g^{lk}(t - t_i^l)}_{\text{self-exciting process}}$$

Properties of Hawkes processes

- Rich literature on theoretical properties of HPs
- Types of results:
  - Stationarity:

    - there exists a unique stationary distribution for $\boldsymbol{N(t)}$ with intensities $\lambda^k(t \mid \mathscr{H}_t)$ if the $K \times K$ matrix ρ where,

    $$\rho_{lk} = \alpha^{lk}, \quad l, k = 1, \ldots, K$$

    has spectral radius strictly smaller than 1 [Brémaud & Massoulié, 1996]

  - Asymptotic properties i.e. posterior concentration rates for:

    - linear multivariate HPs [Donnet et. al., 2020]

    - multivariate HPs with inhibition [Sulem et. al., 2021]

Discrete-time Hawkes process

- Often event data is only collected as counts over regular time intervals.
  - This violates the assumptions of a point process.
  - Discrete-time Hawkes processes (DTHPs) can be used in this instance
- DTHPs are relatively understudied in the literature compared to their continuous-time counterpart

Discrete-time Hawkes process

- Same construction as continuous-time HP.
- Number of events in interval t in dimension k, $N^k(t)$, is Poisson distributed with rate $\lambda^k(t \mid \mathscr{H}_t)$,

$$N^k(t) \sim \mathscr{P}(\lambda^k(t \mid \mathscr{H}_t)) \qquad t \in \mathbb{Z}_+$$

$$\begin{aligned}\lambda^k(t \mid \mathscr{H}_t) &= \mathbb{E}[N^k(t) - N^k(t-1) \mid \mathscr{H}_t] \\ &= \mu + \sum_{l=1}^{K} \alpha^{lk} \sum_{i:t_i^l<t} g^{lk}(t - t_i^l)\end{aligned}$$

Flexible Hawkes processes

- The triggering kernel usually has a simple parametric form
  - Common examples: exponential, power law, geometric
- These are often not sufficient to capture complex behaviours
- There are very few flexible models for HPs, and those that exist are for continuous-time processes

Aims and objectives

Overall aim

- Develop and apply new statistical methods for modelling multivariate, discrete-time Hawkes processes using flexible Bayesian approaches

- Three overall objectives:

  1. Methodological objective

  2. Applied objective

  3. Computational objective

1. Methodological objective

- Develop new methods for modelling flexible, multivariate discrete-time Hawkes processes:

  A. Develop Bayesian semi-parametric and nonparametric representations of the triggering kernel for discrete-time Hawkes processes

  B. Extend the framework for multivariate, Bayesian discrete-time Hawkes processes to incorporate spatial information

2. Applied objective

- Apply the framework to substantive case studies:

  A. Determine how Hawkes processes can be used to explain the novel coronavirus COVID-19

  B. Develop a new risk model for various types of conflict events (protests, riots etc.) using both spatial and temporal dependence

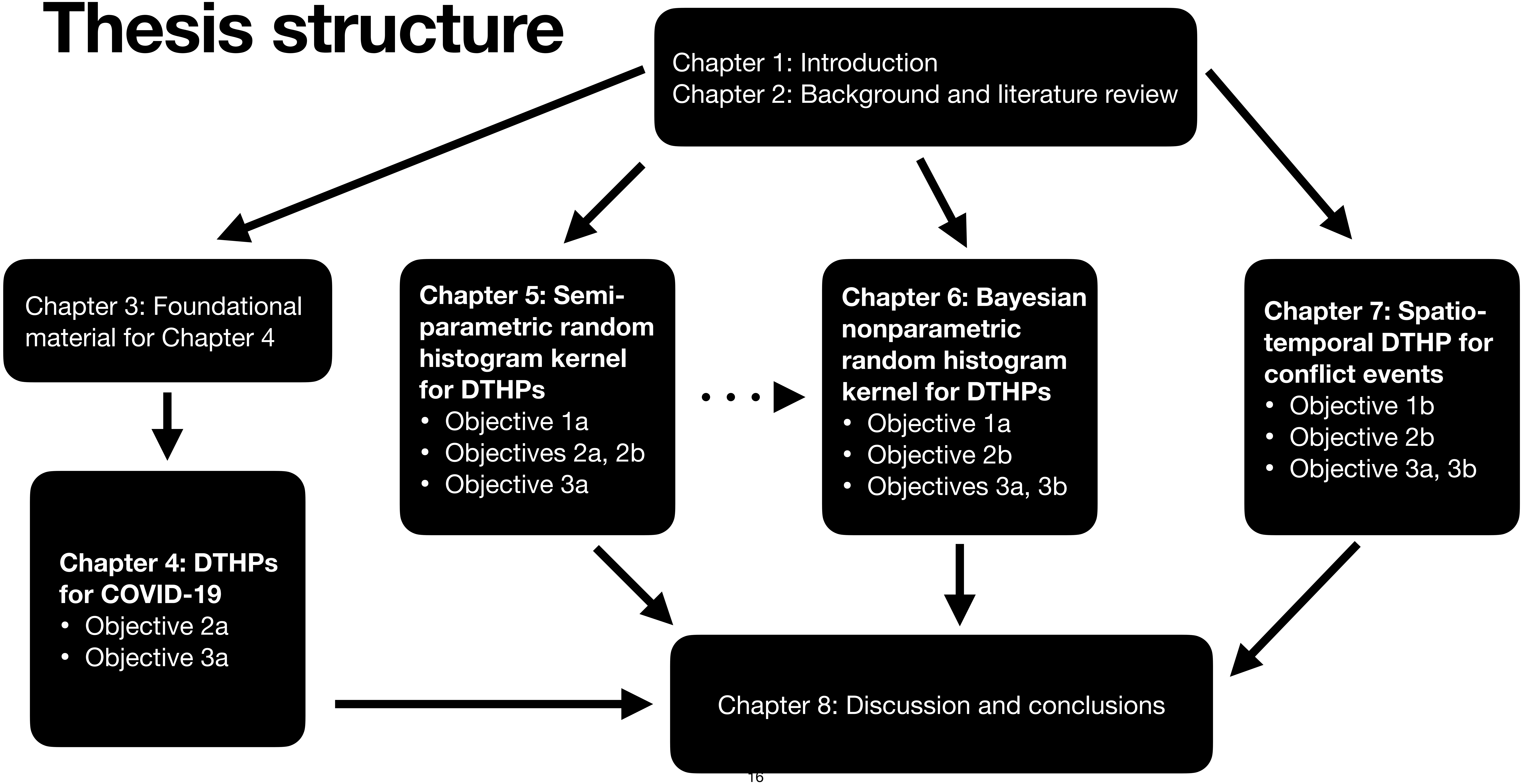
3. Computational objective

- Efficiently implement the proposed methods:

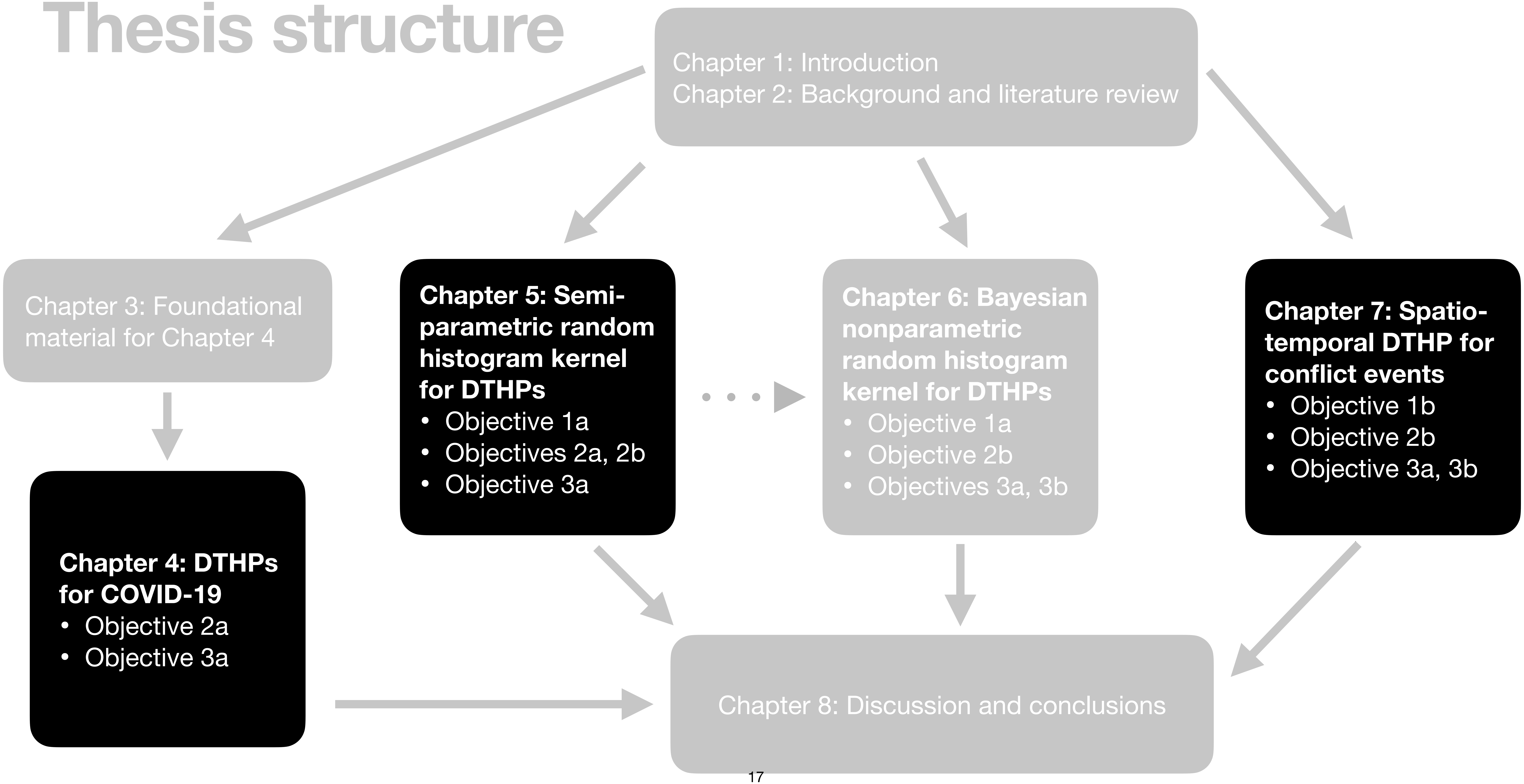
  A. Design samplers for the models stated in the Methodological Objective and implement them using statistical software packages

  B. Improve computational efficiency of sampling algorithms for discrete-time Hawkes processes

Thesis structure



Thesis structure



Simple discrete-time self-exciting models can describe complex dynamic processes: a case study of COVID-19

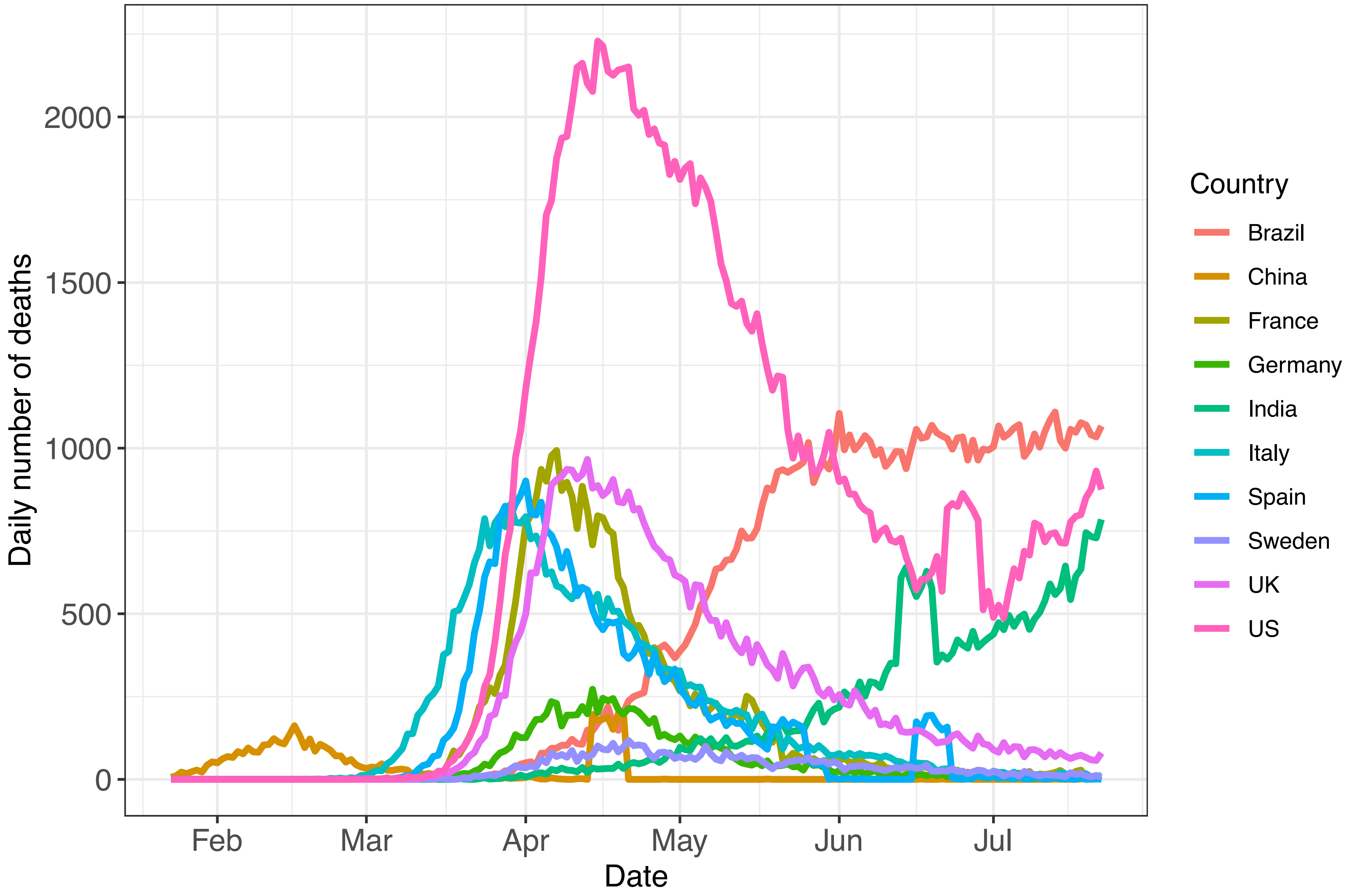
Joint work with Deborah Sulem (Oxford), Vincent Rivoirard (Paris Dauphine), Judith Rousseau (Oxford, Paris Dauphine) and Kerrie Mengersen (QUT)

Contributions

- Propose a multiple phase DTHP to a case study of COVID-19 for a range of countries with differing severity and non-pharmaceutical intervention
- Develop publicly available code for implementing DTHPs
- Gain alternative and complementary insights to standard epidemiological models, for example an indication of the level of imported cases and the temporal excitation pattern
- Proposed model adapts extremely well to volatility in the data - meaning it could also be useful for short term prediction

Case study: COVID-19

- Use Hawkes processes to model the novel coronavirus COVID-19
- Modelling mortality counts [CSSE JHU, 2020]
    - Countries considered: Brazil, China, France, Germany, India, Italy, Spain, Sweden, UK, US
- Initially consider a 2-phase model with data from Jan 2020 - July 2020:
    - Phase 1 - initial exponential growth
    - Phase 2 - political and social interventions
- Extend to subsequent phases up to Feb 2021 (not discussed in this talk)



Daily COVID-19 mortality counts

Multi-phase model

- Introduce change point T_1 to distinguish between Phase 1 and Phase 2
- Number of events on day t given by $N(t) \sim \mathcal{P}(\lambda(t \mid \mathcal{H}_t))$ where

$$\lambda(t \mid \mathcal{H}_t) = \begin{cases} \mu_1 + \alpha_1 \sum_{i:t_i<t} y_{t_i} g_1(t - t_i), & t \leq T_1 \\ \mu_2 + \alpha_2 \sum_{i:t_i<t} y_{t_i} g_2(t - t_i), & t > T_1 \end{cases}$$

and $g(t; \beta) = \beta(1 - \beta)^{t-1}$

A note on using mortality data

- There is a link between the dynamics of the death and infection processes (under some simplifications)
- Can interpret the parameters of our model with respect to infections, rather than deaths
  - μ - the external infection rate multiplied by the probability of death given infection
  - α - the average number of people contaminated by an infected person
  - β - inverse of the mean excitation time for the infections process

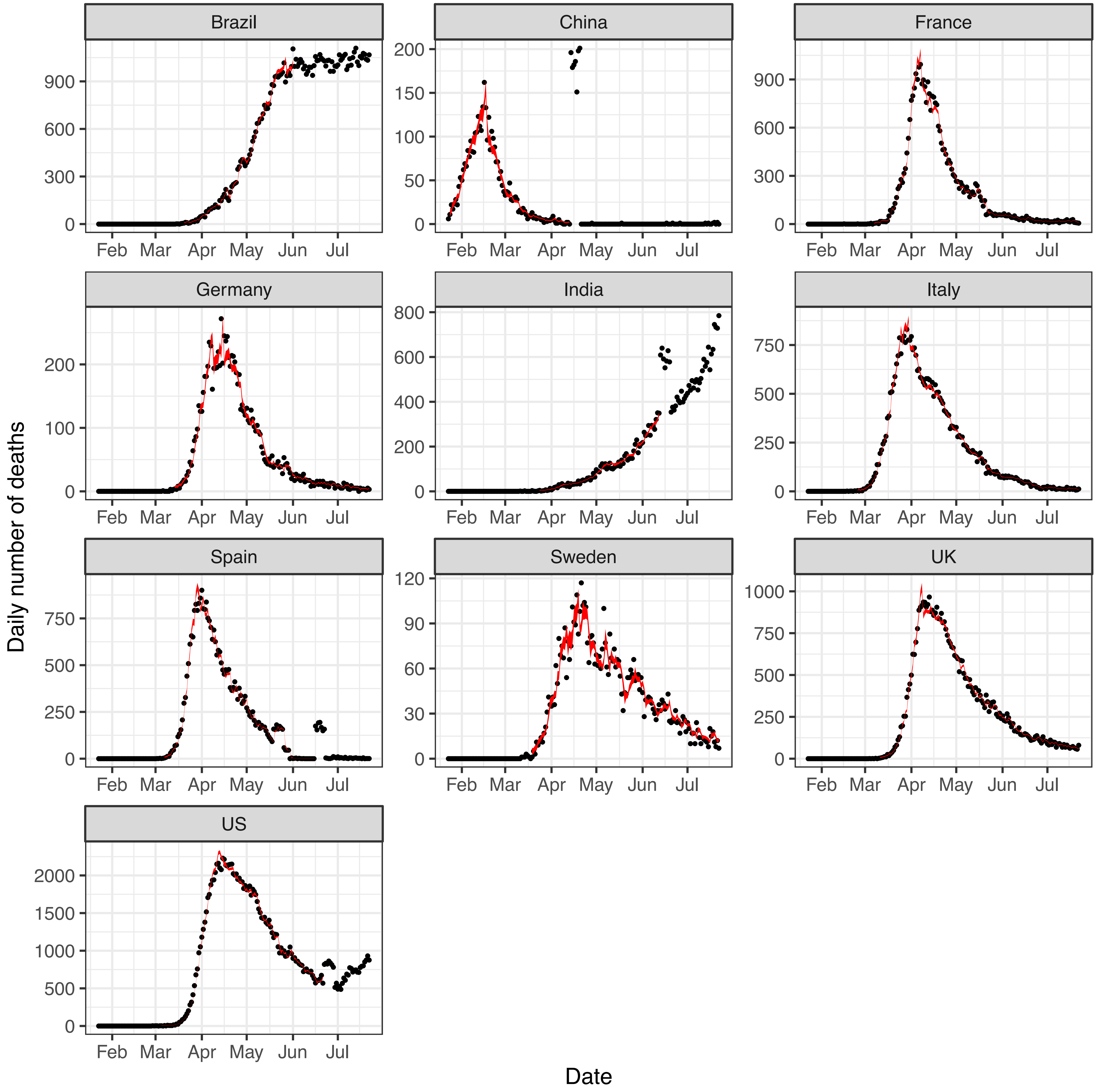
Inference

- Standard random walk Metropolis-Hastings update for μ
- Metropolis-adjusted Langevin algorithm (MALA) [Roberts & Tweedie, 1996] to jointly update α and β
    - MALA is a gradient-based MCMC algorithm that uses Langevin dynamics to guide the random walk to high density regions of the posterior

Results

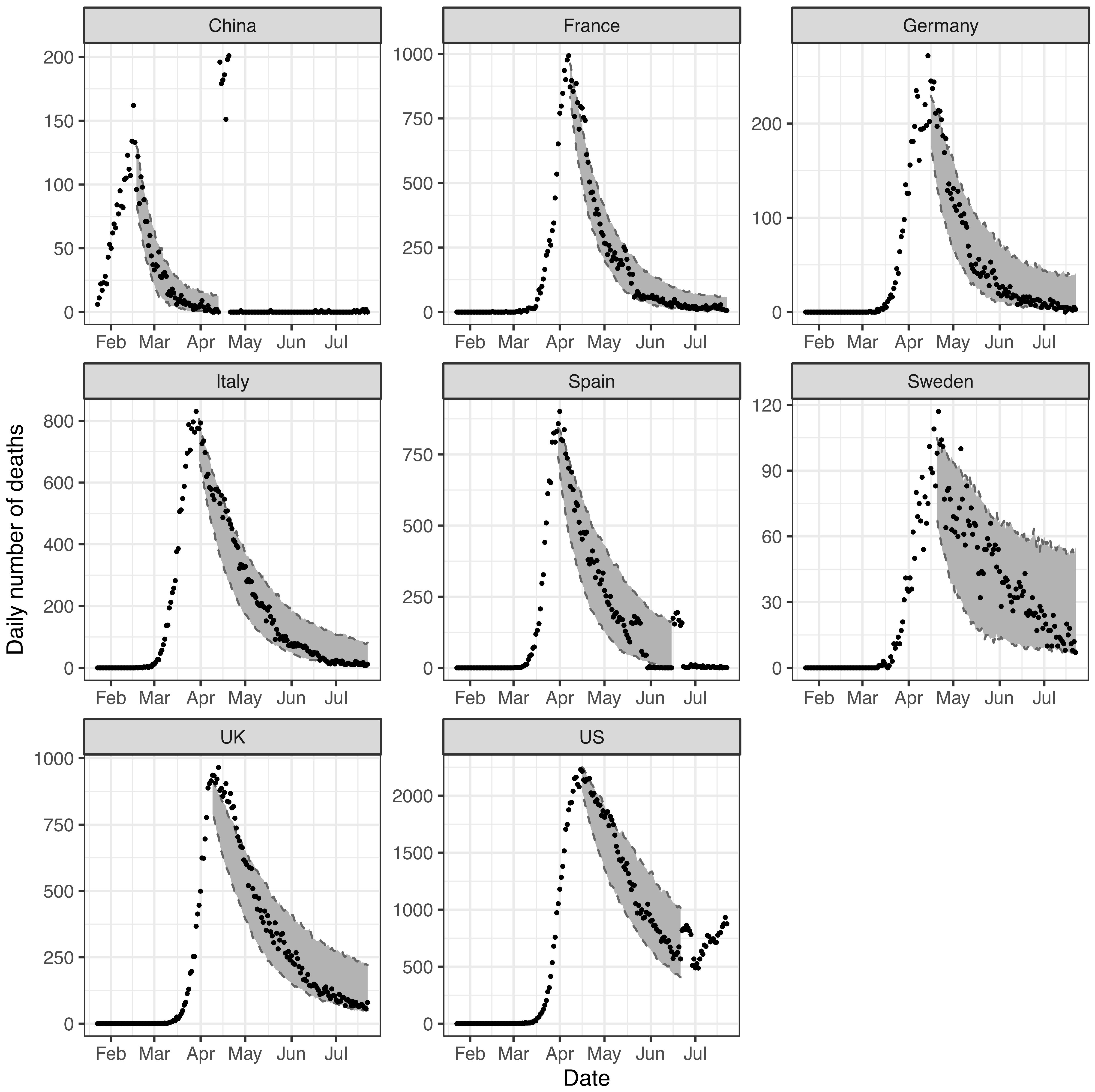
- The background rate μ generally reduces from the first to the second phase, increasing again in subsequent phases
  - Reflects the initial hard lockdowns, followed by increased travel/mobility
- The magnitude parameter α behaves as expected
  - $\alpha > 1$ indicates an "explosive" process i.e. $E(\lambda(\infty \,|\, \mathscr{H}_t)) = \infty$
- The triggering kernel parameter β is generally higher in the first phase than subsequent phases.
  - Large β could be an indication of instability
  - Difficult to predict and discern patterns in the data at the beginning of the pandemic

Results



Estimated 95% credible interval for $\lambda(t \mid \mathscr{H}_t)$ (red ribbon) versus observed data (black dots)

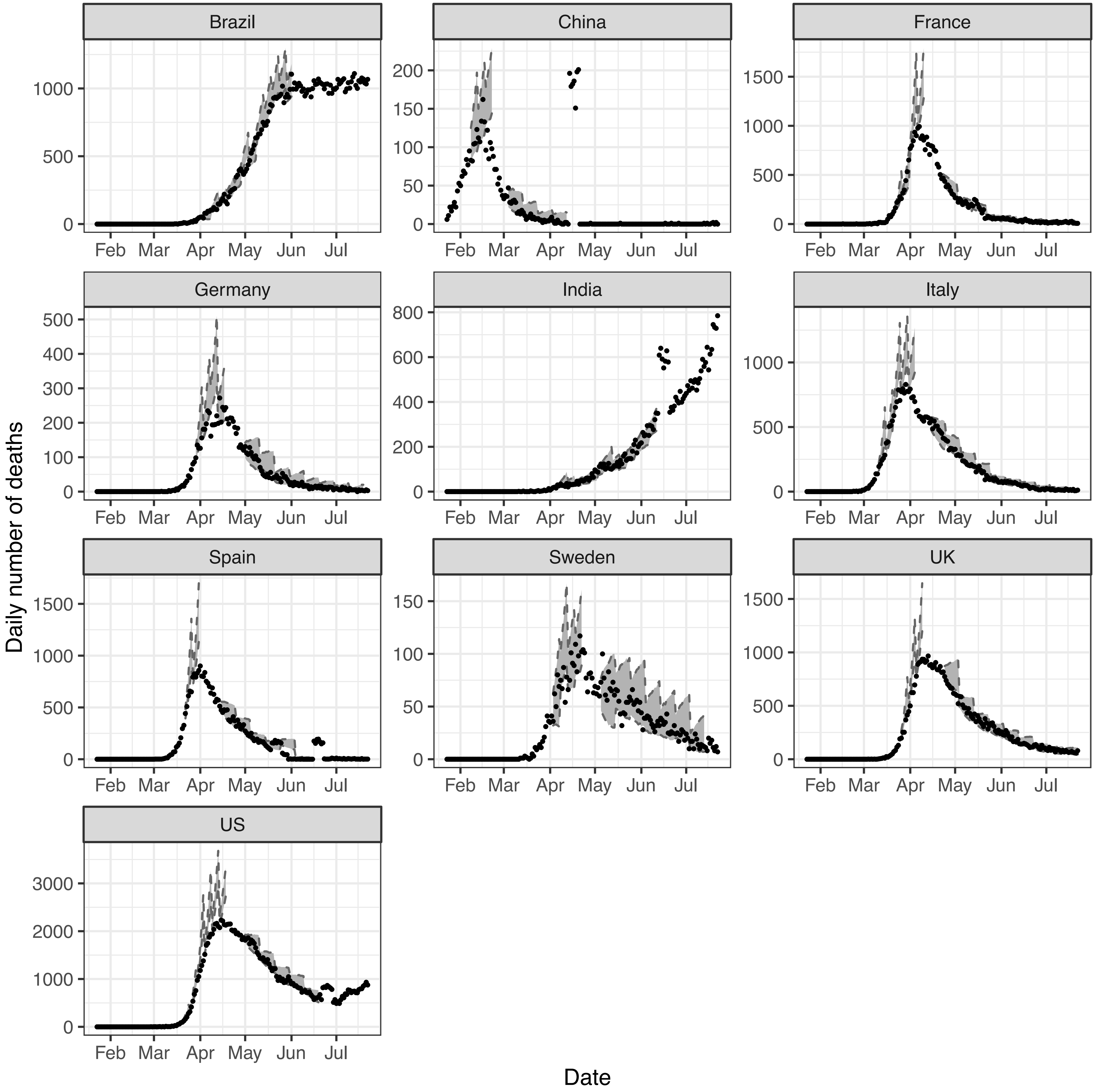
Results



In sample predictive checks.

Observed deaths (black dots) versus 95% posterior predictive interval for the estimated expected number of events $\lambda(t \mid \mathscr{H}_t)$ (grey ribbon).

Results



Out of sample predictive checks.

Observed deaths (black dots) versus 95% posterior predictive interval for the estimated expected number of events $\lambda(t \mid \mathscr{H}_t)$ (grey ribbon).

Conclusions

- Discrete-time Hawkes processes (DTHP) are flexible and can describe very complex phenomena surprisingly well
- Distinct phases in the progression of the pandemic can be captured adequately using a simple change point model
- Our model is complementary to standard epidemiological models (not a replacement)

PLOS ONE

RESEARCH ARTICLE

Simple discrete-time self-exciting models can describe complex dynamic processes: A case study of COVID-19

Raiha Browning[1,2]*, Deborah Sulem[3], Kerrie Mengersen[1,2]☯, Vincent Rivoirard[4]☯, Judith Rousseau[3,4]☯

1 School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, 2 Australian Research Council, Centre of Excellence for Mathematical and Statistical Frontiers, Brisbane, Australia, 3 Department of Statistics, University of Oxford, Oxford, United Kingdom, 4 Ceremade, Université Paris-Dauphine, Paris, France

☯ These authors contributed equally to this work.
* raihatuitaura.browning@qut.edu.au

Check for updates

OPEN ACCESS

Citation: Browning R, Sulem D, Mengersen K, Rivoirard V, Rousseau J (2021) Simple discrete-

Abstract

Hawkes processes are a form of self-exciting process that has been used in numerous applications, including neuroscience, seismology, and terrorism. While these self-exciting processes have a simple formulation, they can model incredibly complex phenomena. Tra-

A flexible, random histogram kernel for discrete-time Hawkes processes

Joint work with Judith Rousseau (Oxford, Paris Dauphine) and Kerrie Mengersen (QUT)

Contributions

- Extend the existing framework for DTHPs using standard parametric distributions by proposing a flexible, random histogram kernel for DTHPs
- Apply the proposed model both COVID-19 data [CSSE JHU, 2021] and data on terrorist activity in Southeast Asia [START, 2022]
- Develop a bespoke trans-dimensional sampler for these models
- Provide an approach to determine whether standard parametric triggering kernels are sufficient

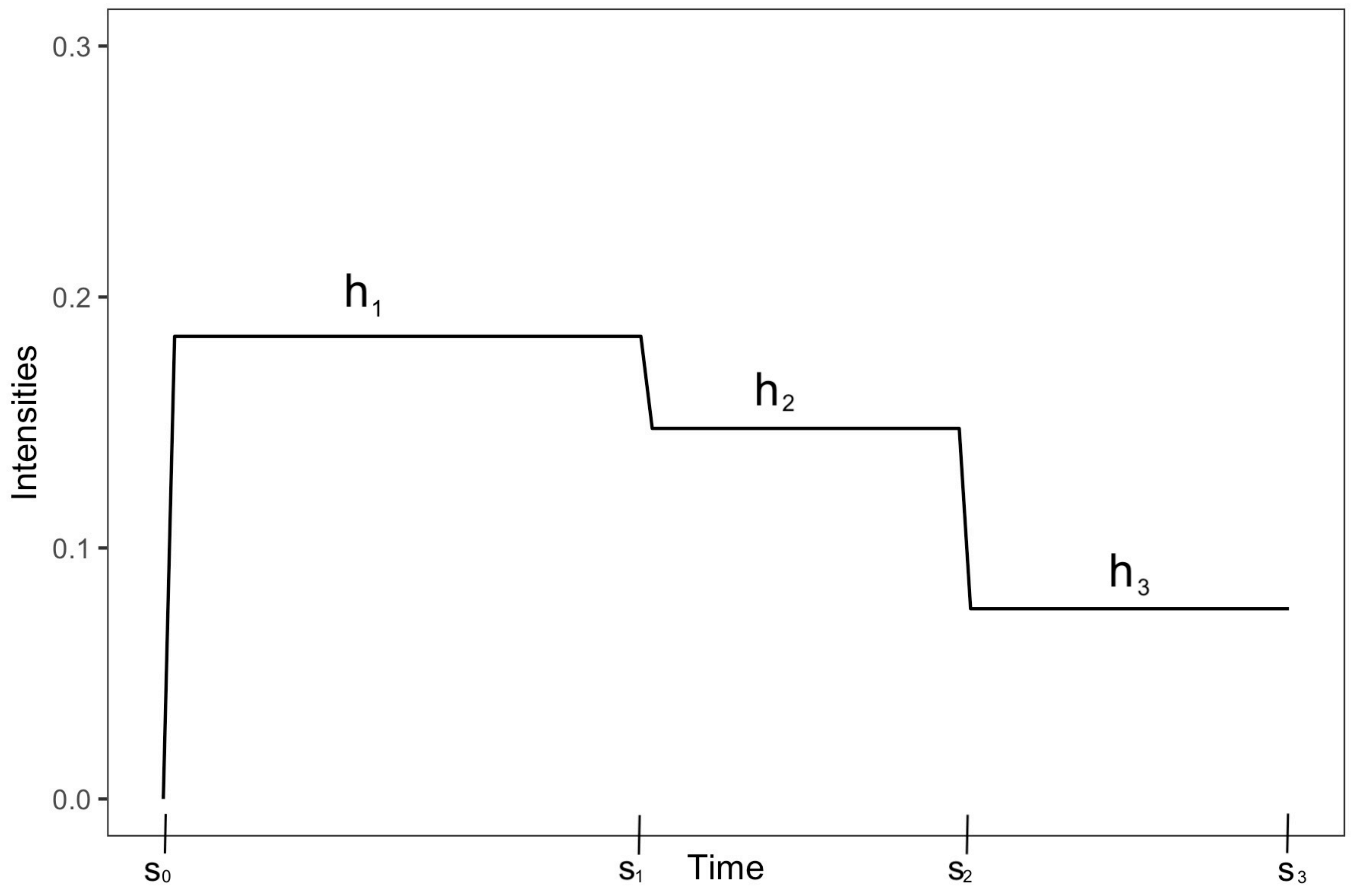
Random histogram prior

- Histogram kernel adapted from Donnet et. al. [2020]

- Interaction functions $g^{lk}(\cdot)$ are histogram densities,

$$g^{lk}(t) = \sum_{j=1}^{J} \frac{h_j^{lk}}{\sum_{i=1}^{J}(s_i^{lk} - s_{i-1}^{lk})h_i^{lk}} \mathbb{1}_{t \in I_j^{lk}}$$

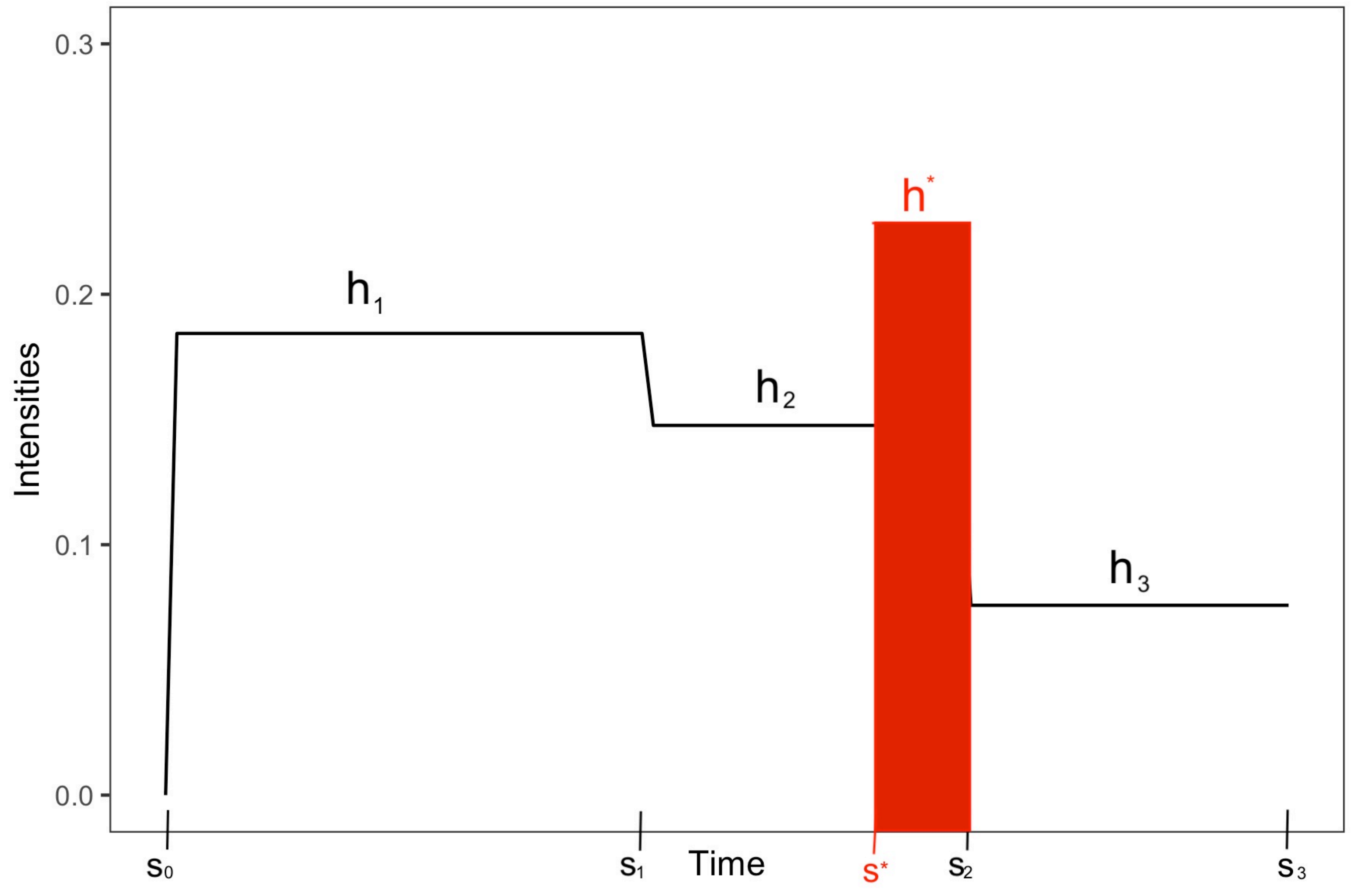
  with unknown number of components, J.



Example histogram (J=3)

Random histogram prior

- Reversible jump MCMC [Green, 1995] with birth-death step used to update the histogram
  - Birth move: select a vacant change point $s*$ and draw new height $h*$
  - Death move: select a current change point $s*$ and remove, along with the corresponding height $h*$



Example histogram with birth/death step

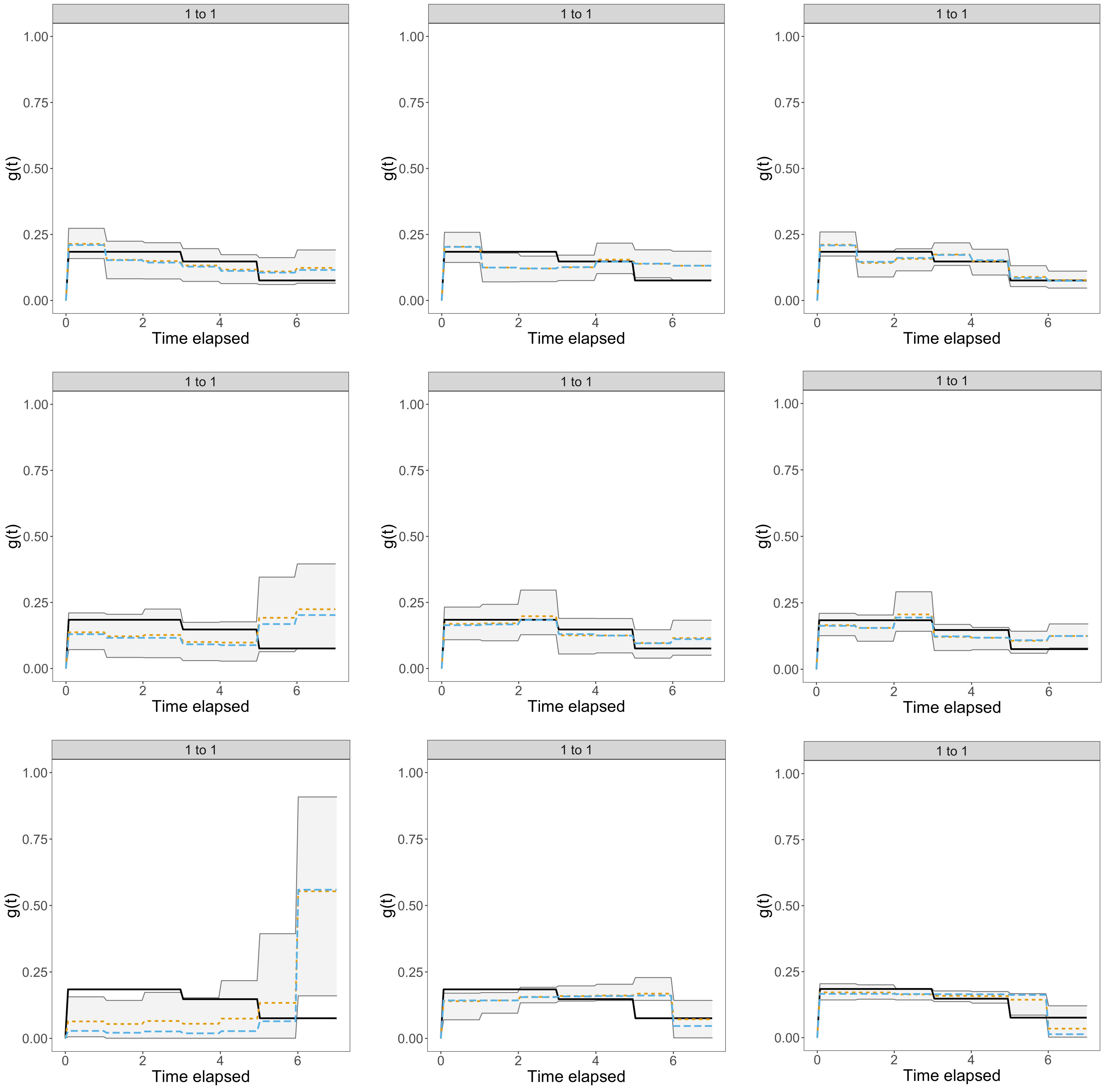
Inference

Algorithm 1: Reversible jump MCMC (RJMCMC) pseudocode

Initialise parameters;
for $i = 1{:}N$ **do**
  Update $\log(\mu)$, $\log(\alpha)$ and $\log(h_i)$'s using RWMH;
  Update each s_k with proposal density $q(s_k) = \frac{1}{n_{\text{vacant}}}$;
  Sample $m \in \{birth, death\}$ with probability p_{birth} and p_{death};
  if $m = birth$ **then**
    Propose new knot point, s^* from set of vacant knot points, with probability $\frac{1}{s_{\max} - J}$;
    Propose new height h^* from $\mathcal{N}(\frac{\sum_{i=2}^{J} h_i}{J-1}, \sigma)$;
    Accept move with probability A;
  else if $m = death$ **then**
    Propose knot point to remove, s^* with probability $\frac{1}{J-1}$;
    Remove corresponding h^* with proposal density the same as the birth move;
    Accept move with probability A^{-1};
end

Results

50 days | 200 days | 500 days



Informative prior setting

Relatively informative prior setting

Quite uninformative prior setting

Estimated triggering kernels.

Black line: true histogram
Blue dashed line: posterior median
Orange dashed line: posterior mean
Grey ribbon: 80% posterior interval

Case study: COVID-19

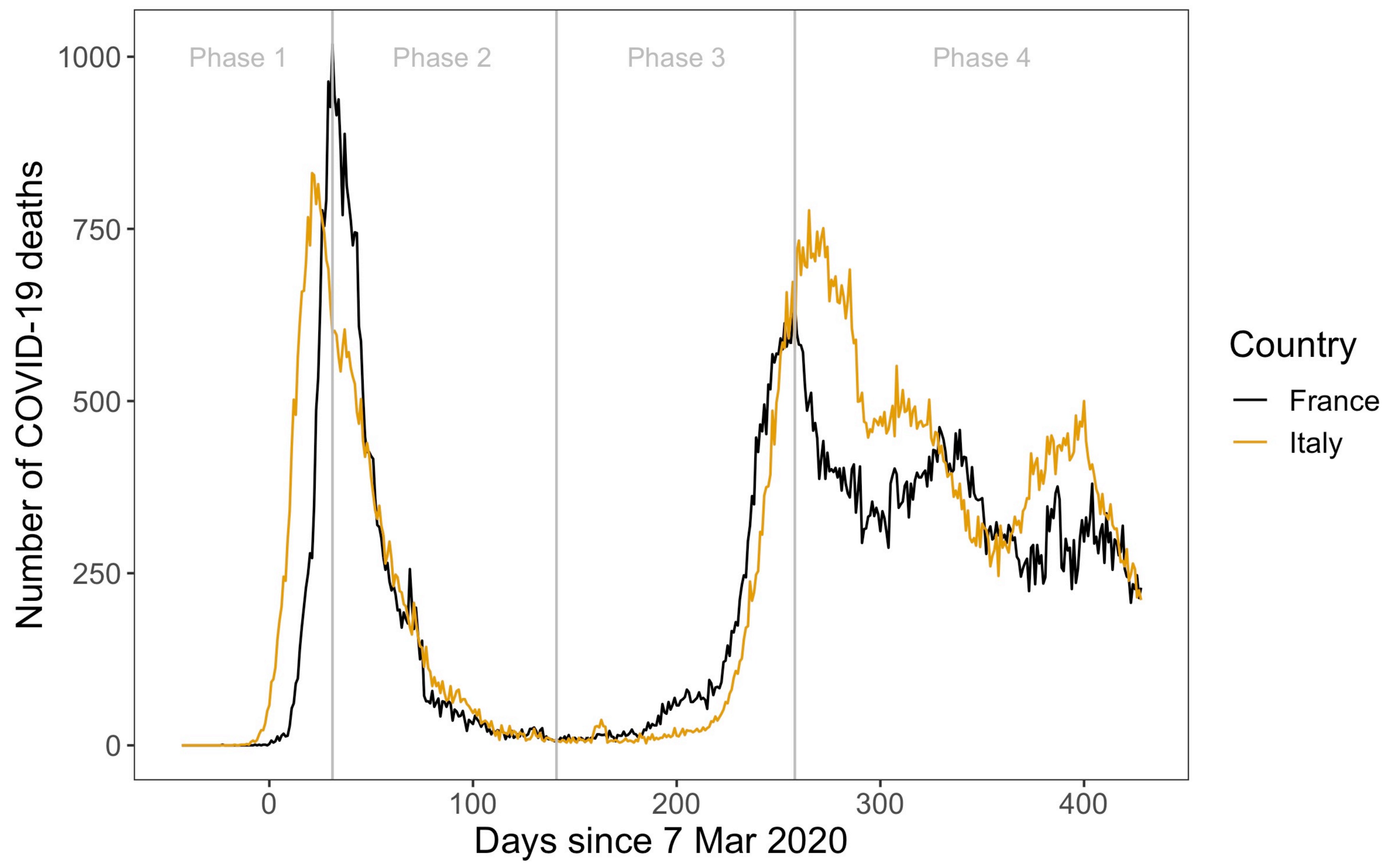
- Want to characterise movement and interactions between countries or regions

- For this case study:

  - Consider spread of COVID-19 between Italy and France

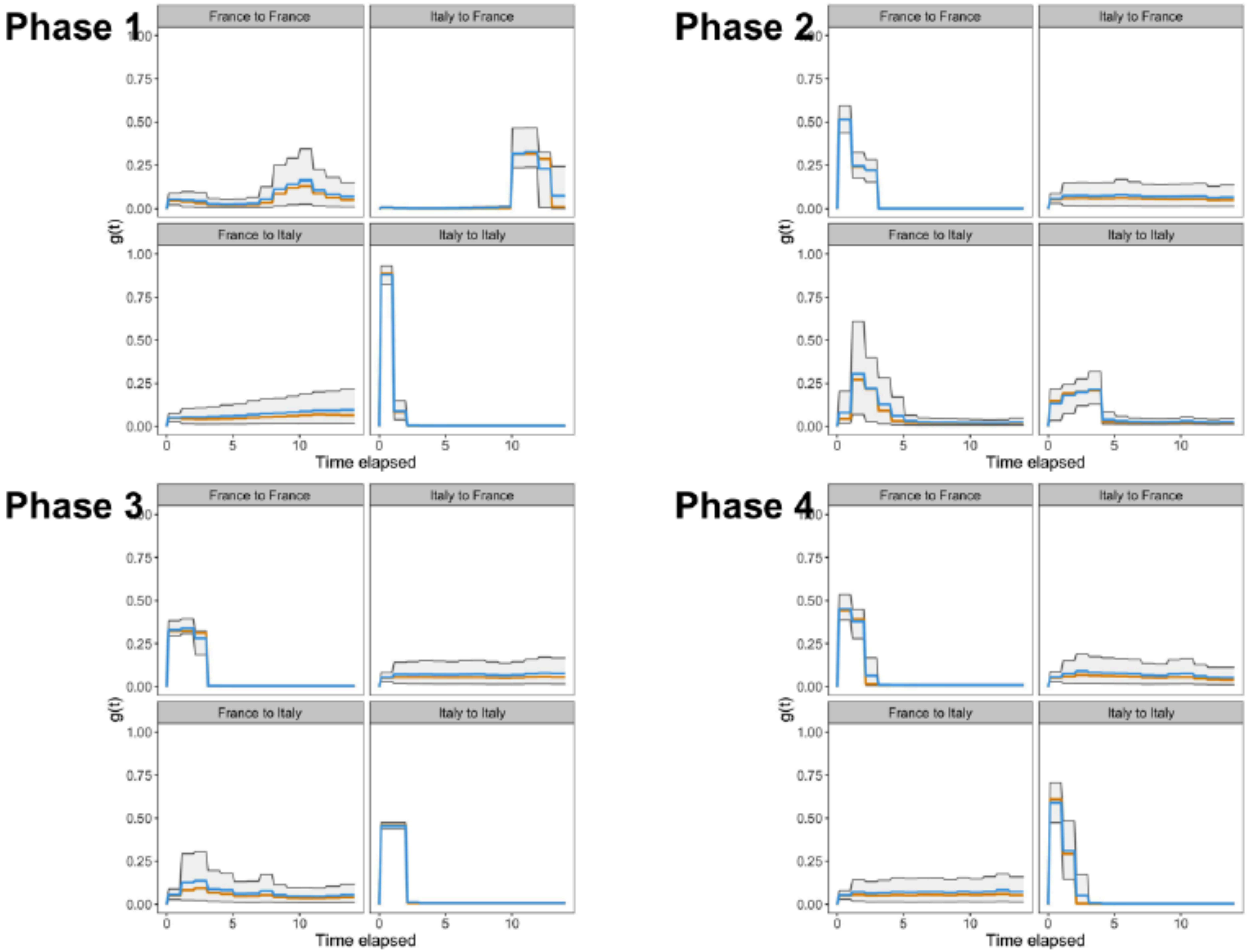
  - Observation window: March 2020 to May 2021

  - Time series partitioned into 4 phases



Daily COVID-19 mortality counts

Results



Estimated triggering kernels.

Blue line: posterior median
Orange line: posterior mean
Grey ribbon: 80% posterior interval

Conclusions

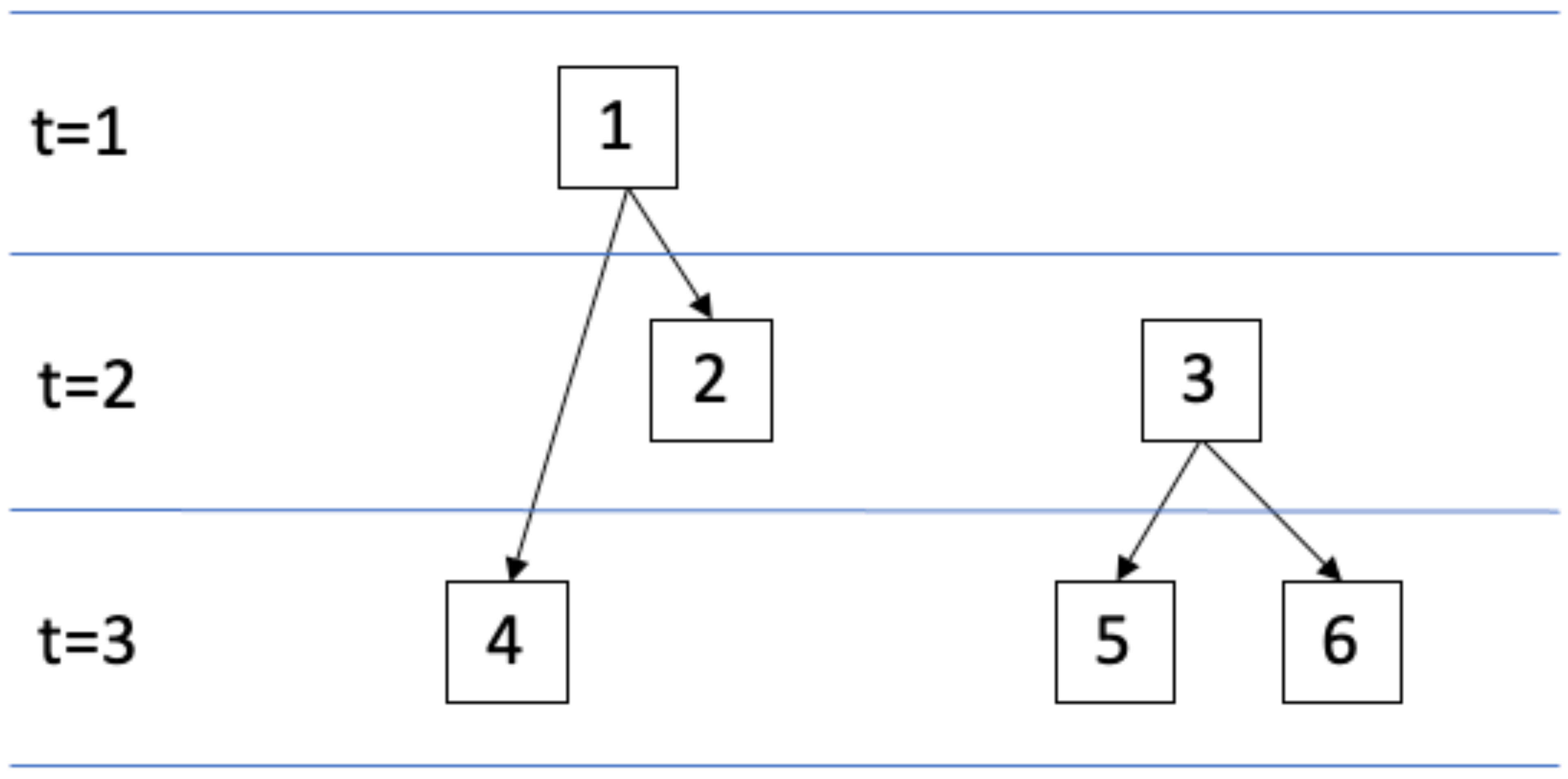
- We propose an approach for flexible estimation of the excitation kernel of the DTHP
- Avoid making assumptions of simple parametric distributions that may be restrictive
- Multivariate approach allows for inference of self-exciting and mutually-exciting processes

A fully-Gibbs sampler for flexible estimation of the temporal excitation pattern of discrete-time self-exciting processes

Joint work with Judith Rousseau (Oxford, Paris Dauphine) and Kerrie Mengersen (QUT)

Contributions

- Extend the random histogram kernel with a Bayesian nonparametric alternative by exploiting the branching structure of DTHPs
- Develop a novel Gibbs sampler [Gelfand & Smith, 1990] with conjugate priors for this flexible model with publicly available source code
    - A faster alternative to reversible-jump MCMC since decomposing the likelihood function makes the algorithm more parallelisable
- Removes the manual selection of the maximum excitation time



Branching process example

A spatiotemporal triggering kernel for discrete-time Hawkes processes for conflict events in South Asia

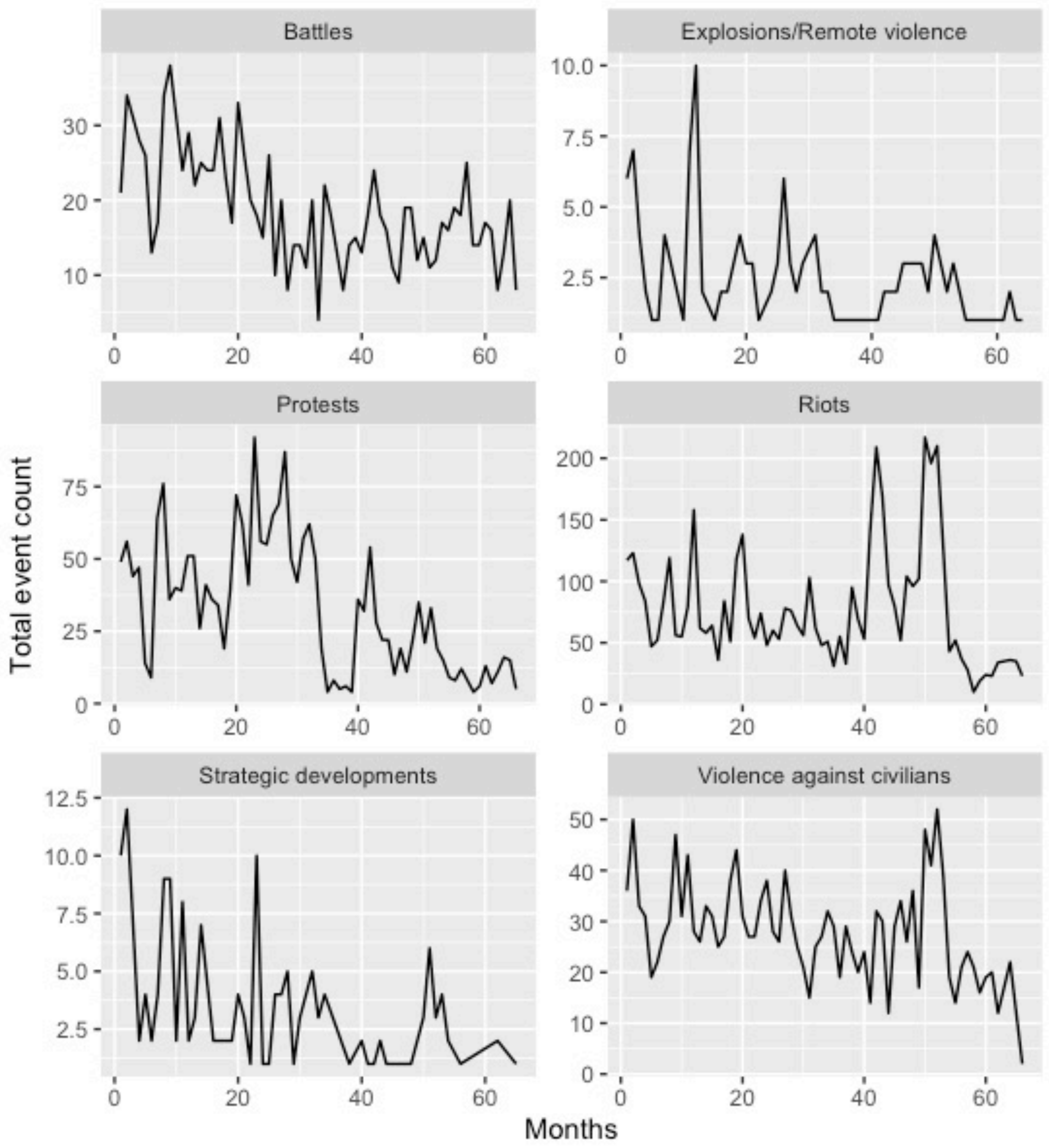
Joint work with Hamish Patten (United Nations Office for Disaster Risk Reduction), Judith Rousseau (Oxford, Paris Dauphine) and Kerrie Mengersen (QUT)

Contributions

- Present a spatiotemporal model for DTHPs
- Incorporate spatial triggering kernel to model the impact of conflict events in nearby regions in South Asia [ACLED, 2022]
- Provide code for estimating this model in efficient Bayesian inference software (Stan)
- Provide new risk metrics, such as the risk of events over the next 3 months for a given location

Case study: conflict events

- The Armed Conflict Location and Event Data (ACLED) project provides a comprehensive database containing details of political violence and protests worldwide
- ACLED team currently uses a simplistic average to classify the risk of conflict events in a given region
    - Propose a new, more sophisticated risk model
- Explore both temporal and spatial dynamics between events
- For this case study:
    - Consider monthly counts of conflict events in South Asia (Bangladesh, Nepal, Sri Lanka and Pakistan)
    - Observation window: Jan 2010 - Dec 2014



Daily conflict event counts in Bangladesh

Event types

- **Battles**: violent interaction between two organised, armed groups
- **Explosions/remote violence:** one-sided violence events in which the tool for engaging in conflict creates asymmetry by taking away the ability of the target to respond
- **Protests:** a public demonstration against a political entity, government institution, policy or group in which the participants are not violent
- **Riots:** violent events where demonstrators or mobs engage in disruptive acts or disorganised acts of violence against property or people
- **Violence against civilians:** violent events where an organised armed group deliberately inflicts violence upon unarmed non-combatants
- **Strategic development:** accounts for often non-violent activity by conflict and other agents within the context of the war/dispute. Recruitment, looting and arrests included.

Spatiotemporal triggering kernel

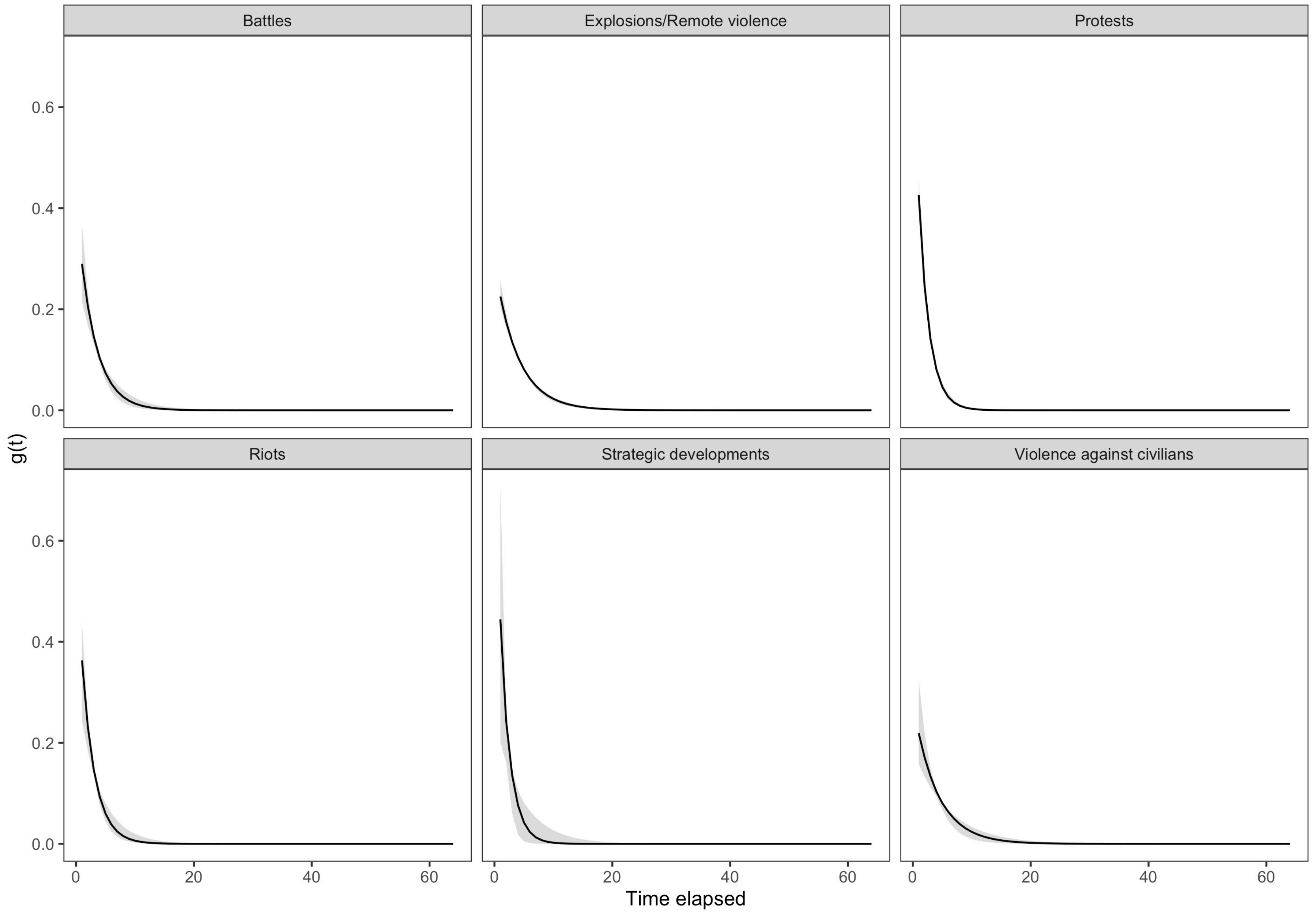
- The rate parameter $\lambda(t, x, y \mid \mathscr{H}_t)$ at time t and location (x, y) for the univariate spatiotemporal DTHP is,

$$\lambda(t, x, y \mid \mathscr{H}_t) = \mu + \alpha \sum_{i:t_i<i} g(t - t_i)\; h(d(x - x_i, y - y_i))$$

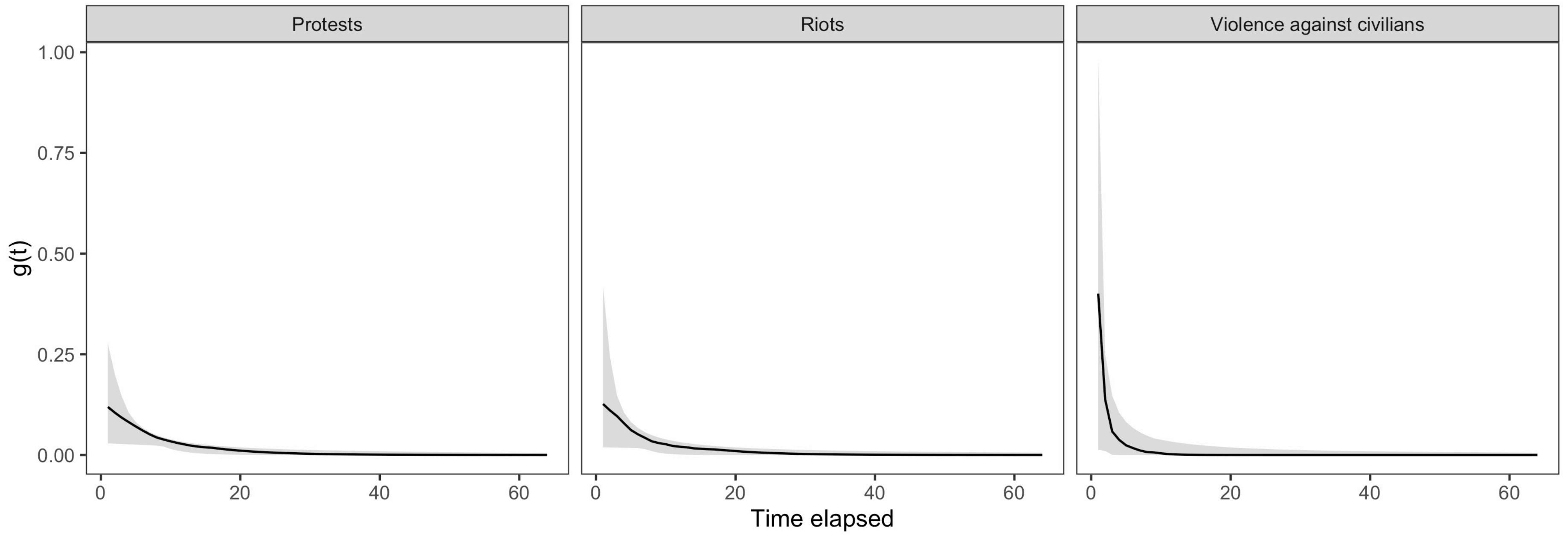
- $g(\cdot)$ is the geometric density
- $d(\cdot)$ is the Euclidean distance function (evaluated between all pairwise locations)
- $h(\cdot)$ is the radial basis function kernel,

$$h(d(x - x_i, y - y_i)) = \exp\left(-\frac{d(x - x_i, y - y_i)}{2\sigma^2}\right)$$

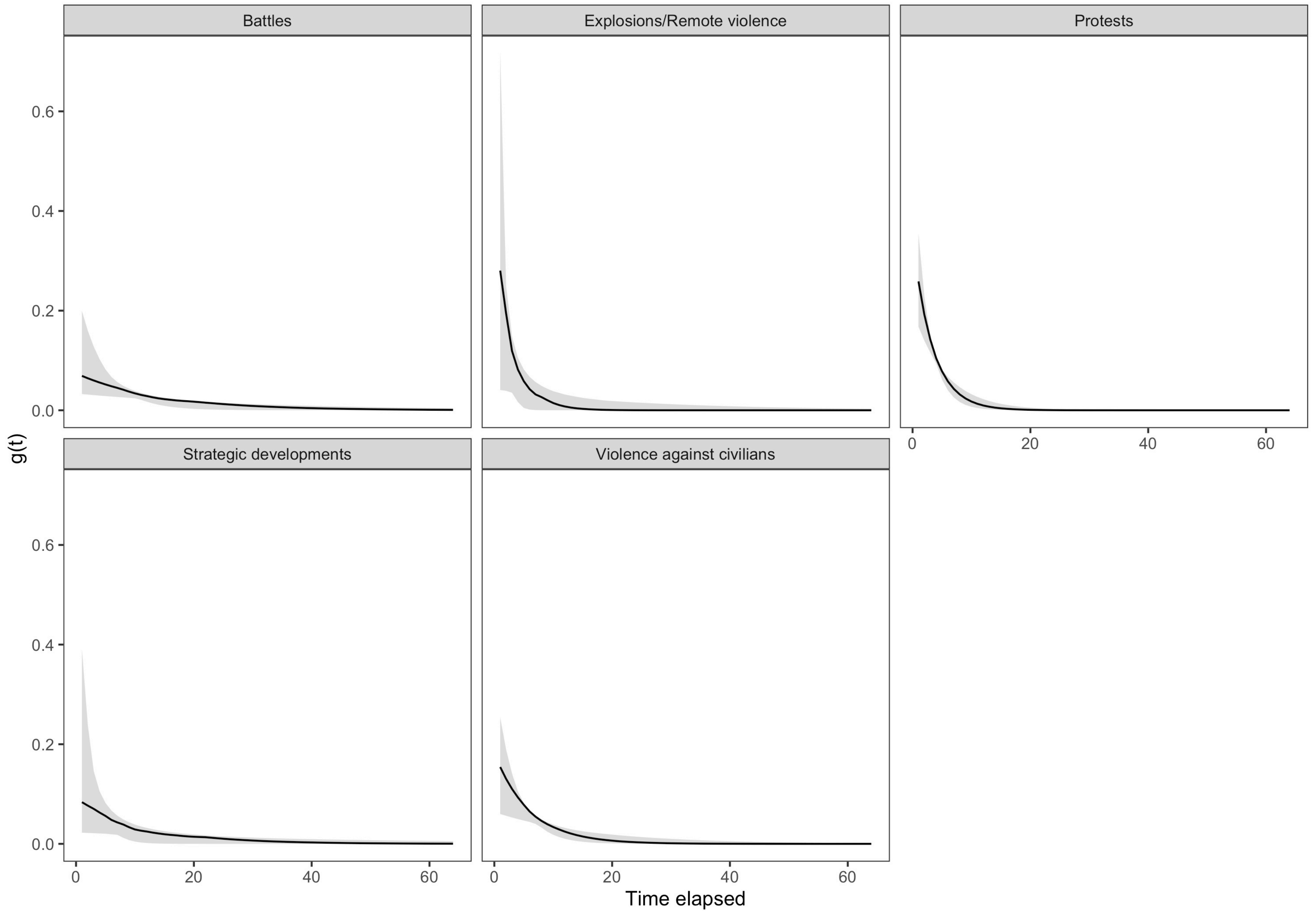
Temporal excitation kernels



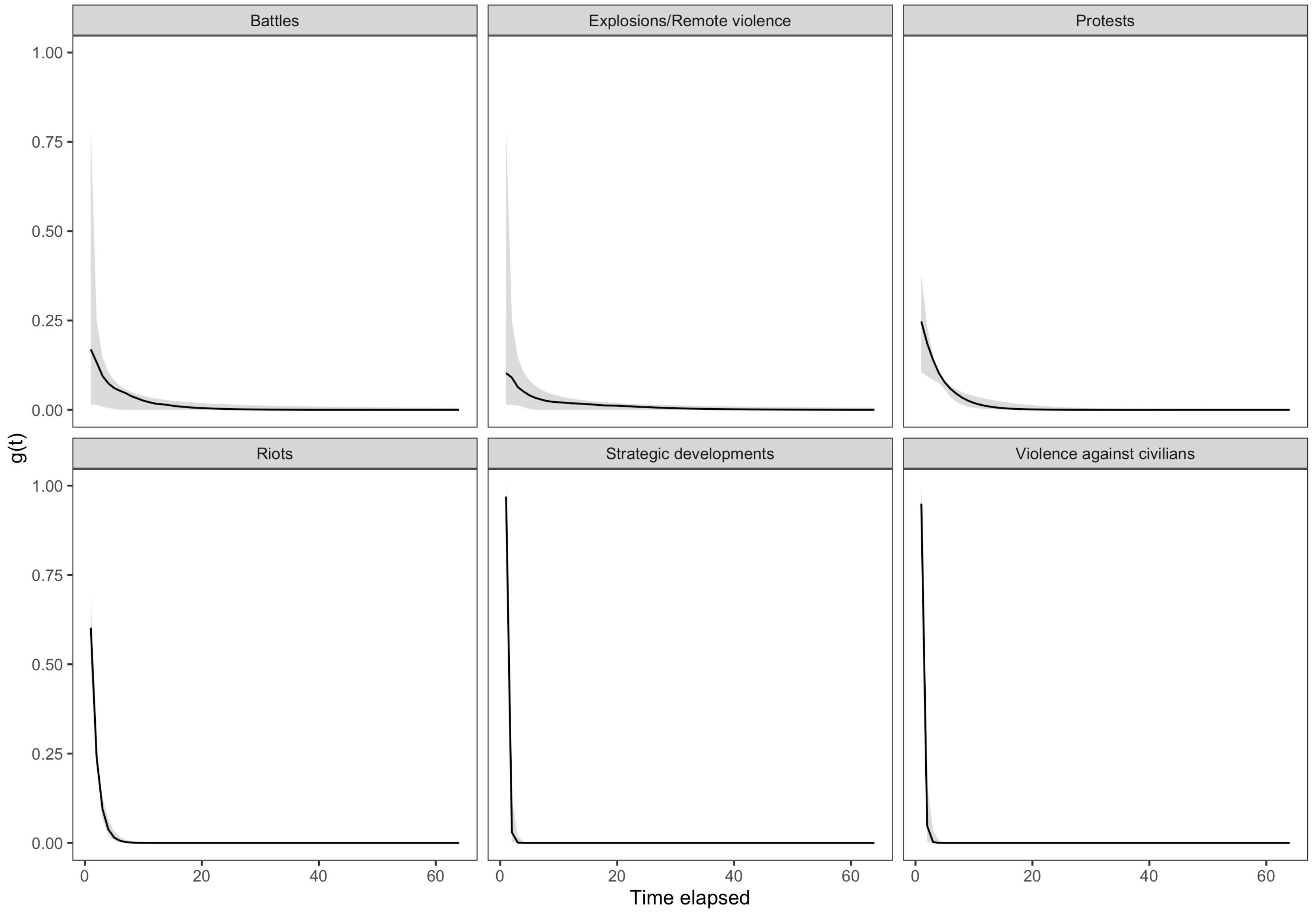
Pakistan



Sri Lanka



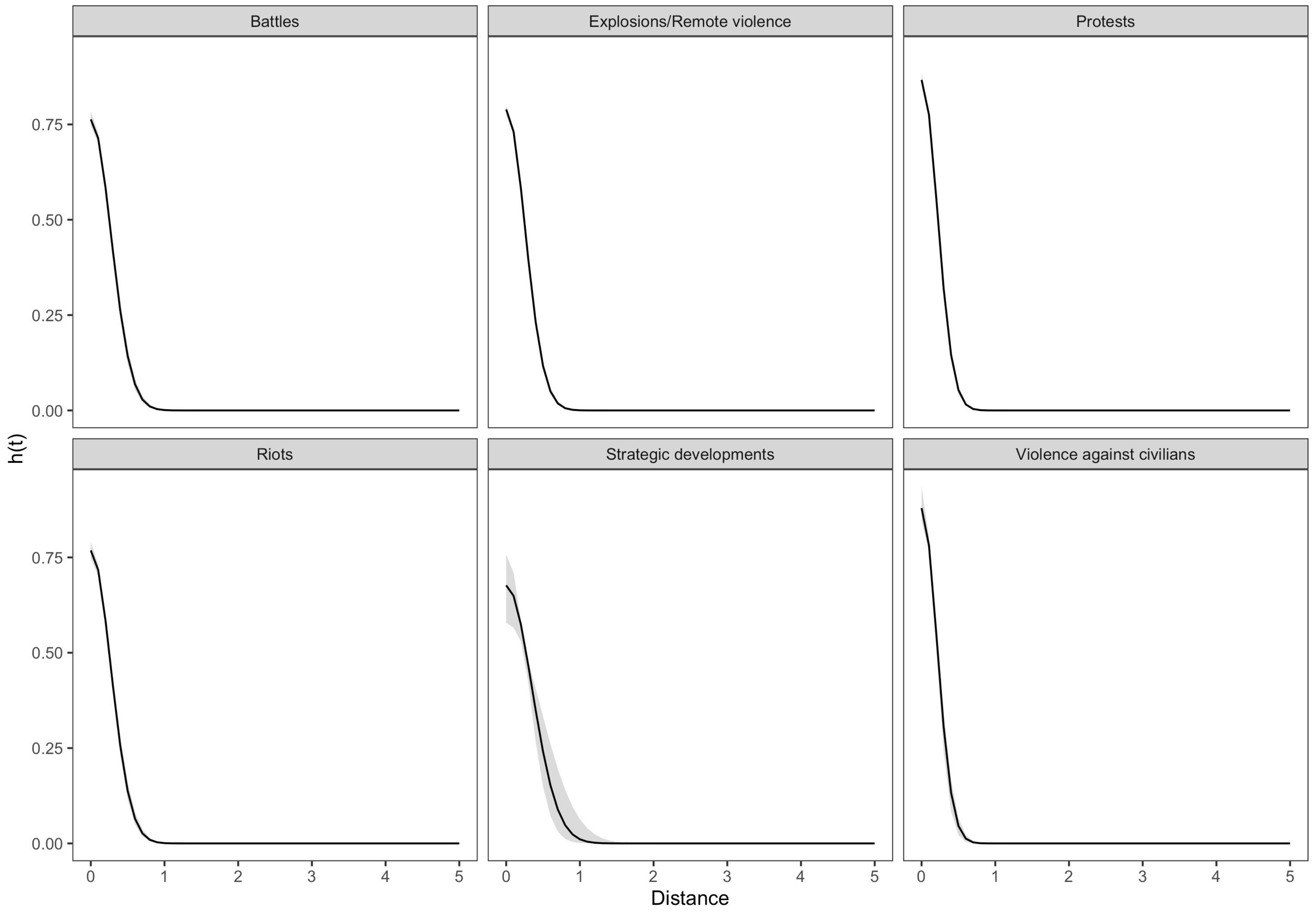
Bangladesh



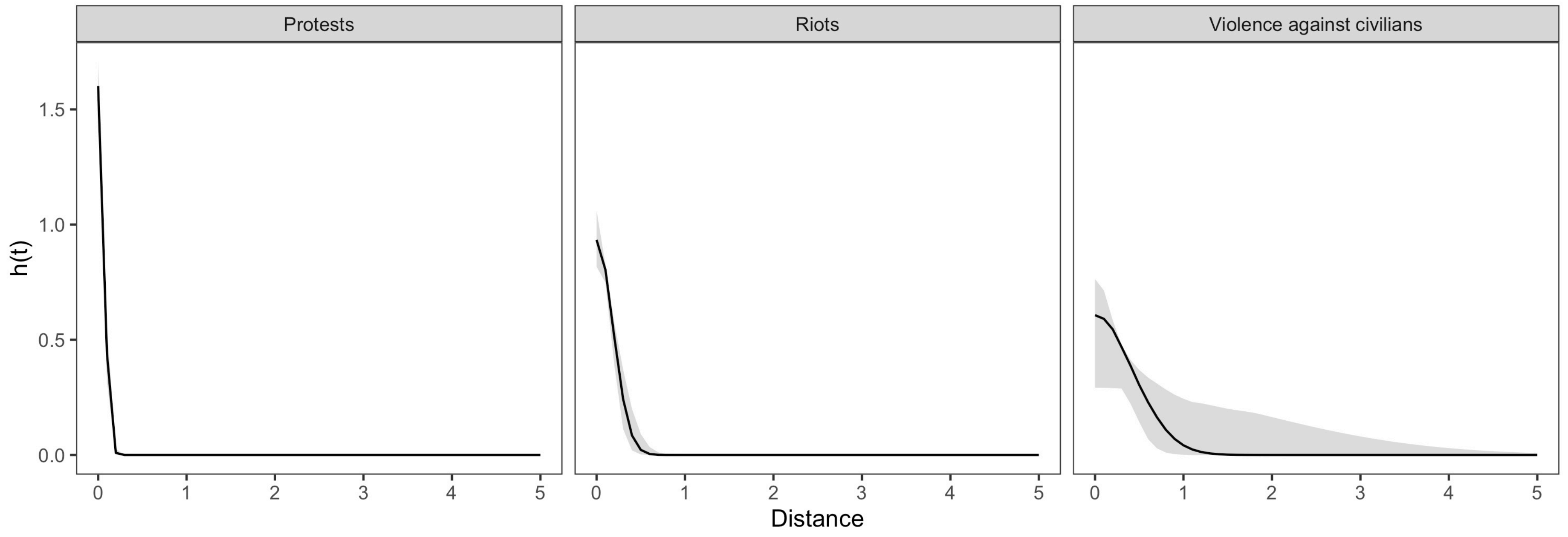
Nepal

Estimated temporal excitation kernels with 95% posterior intervals.

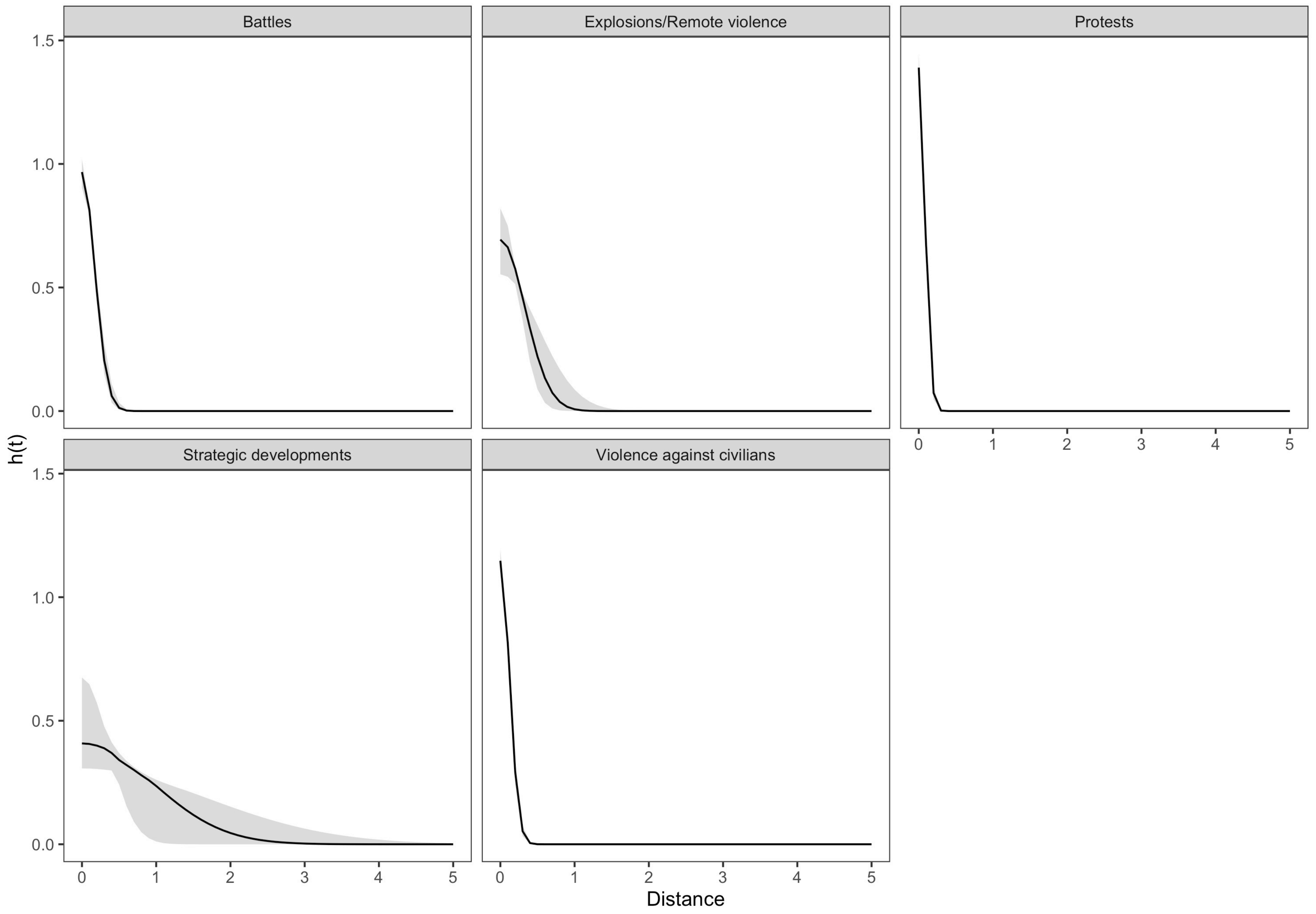
Spatial excitation kernels



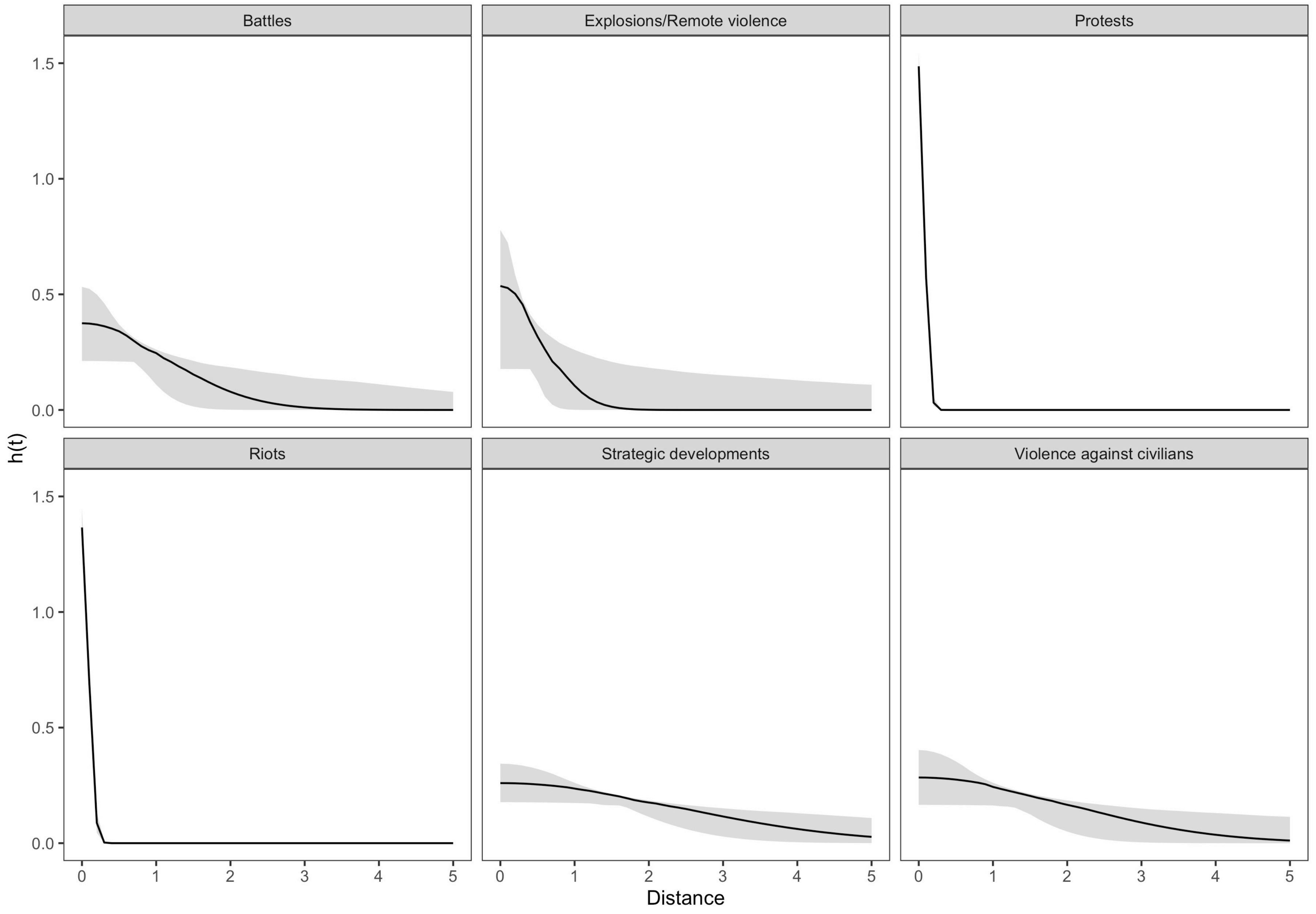
Pakistan



Sri Lanka



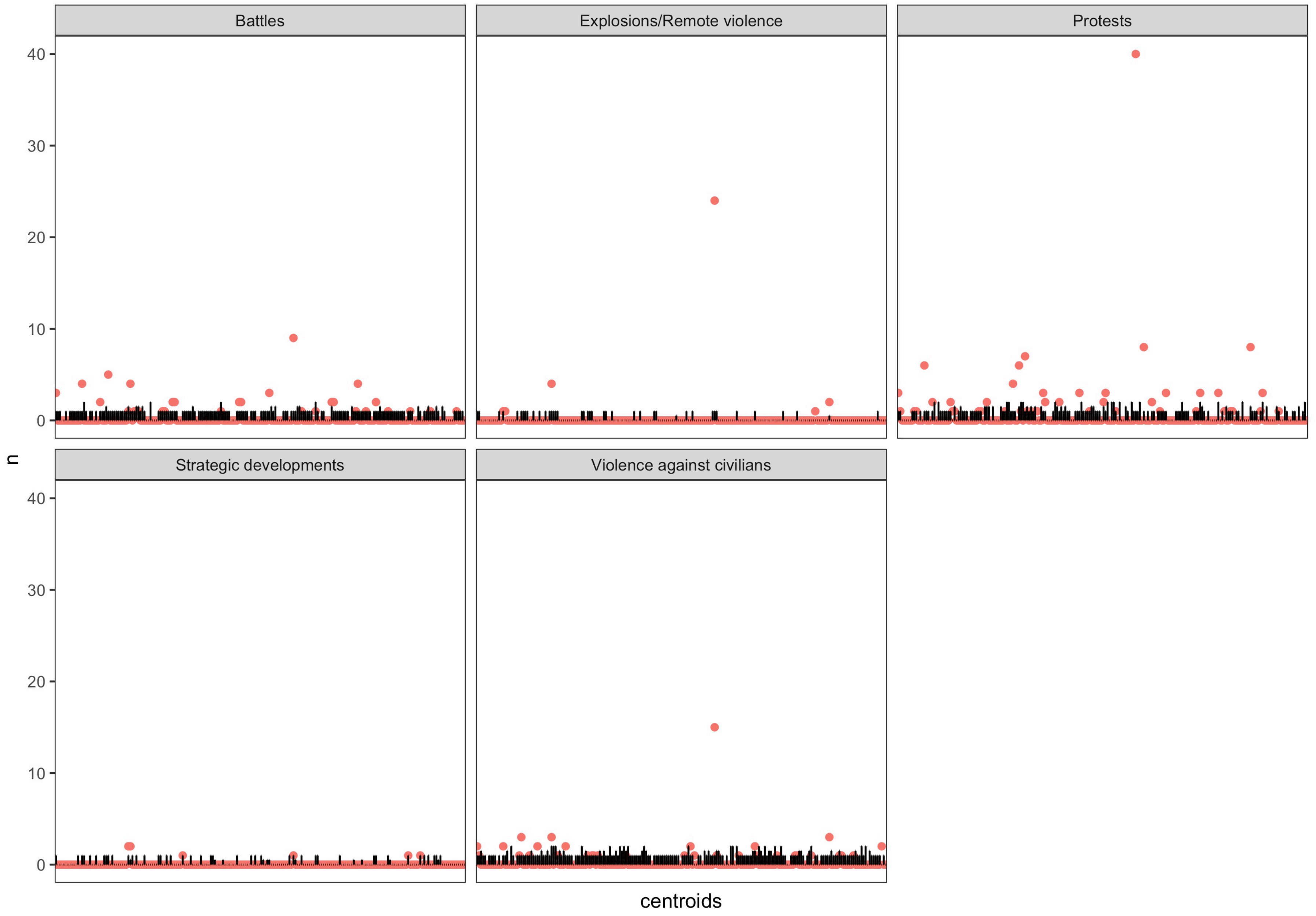
Bangladesh



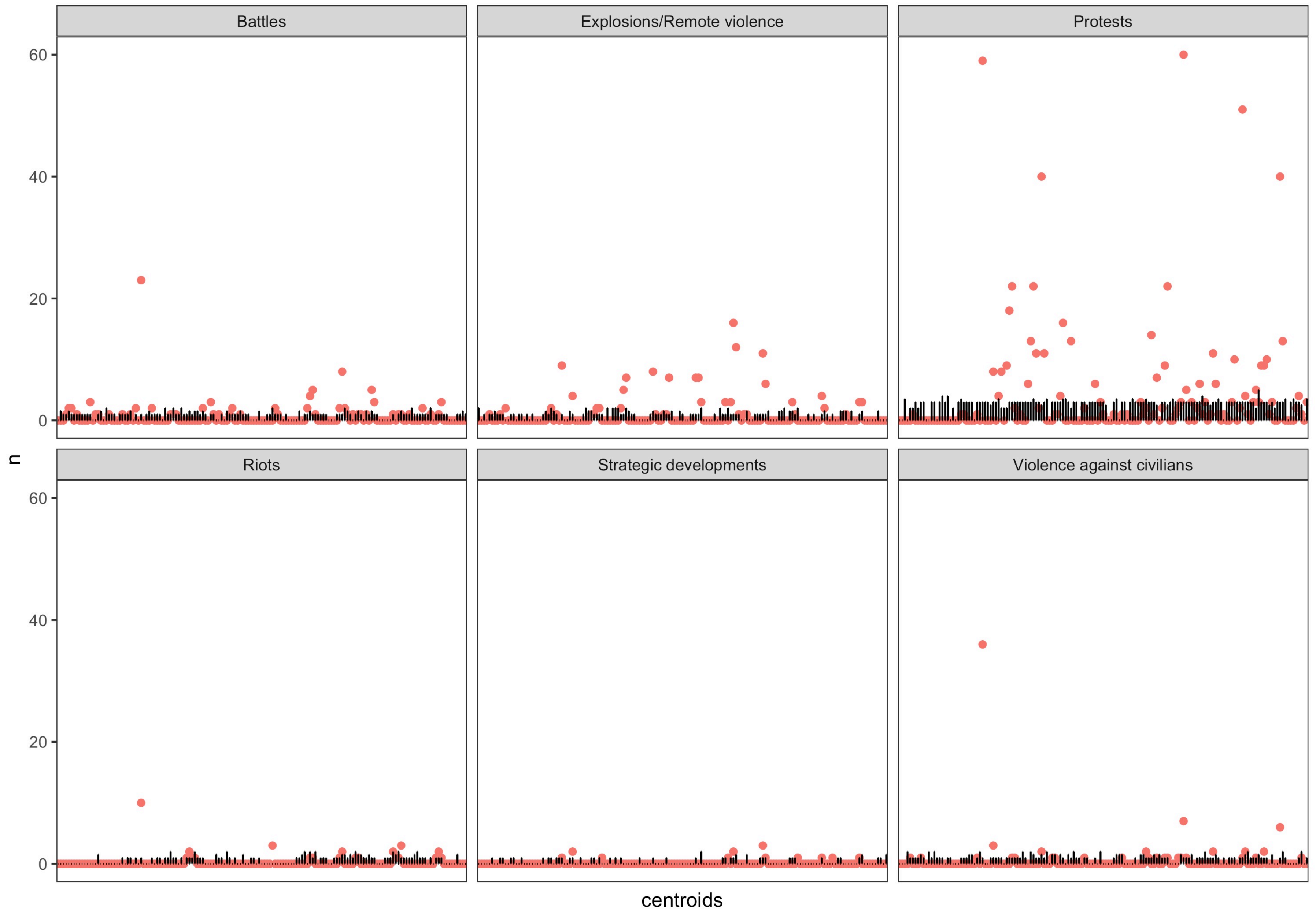
Nepal

Estimated spatial excitation kernels with 95% posterior intervals.

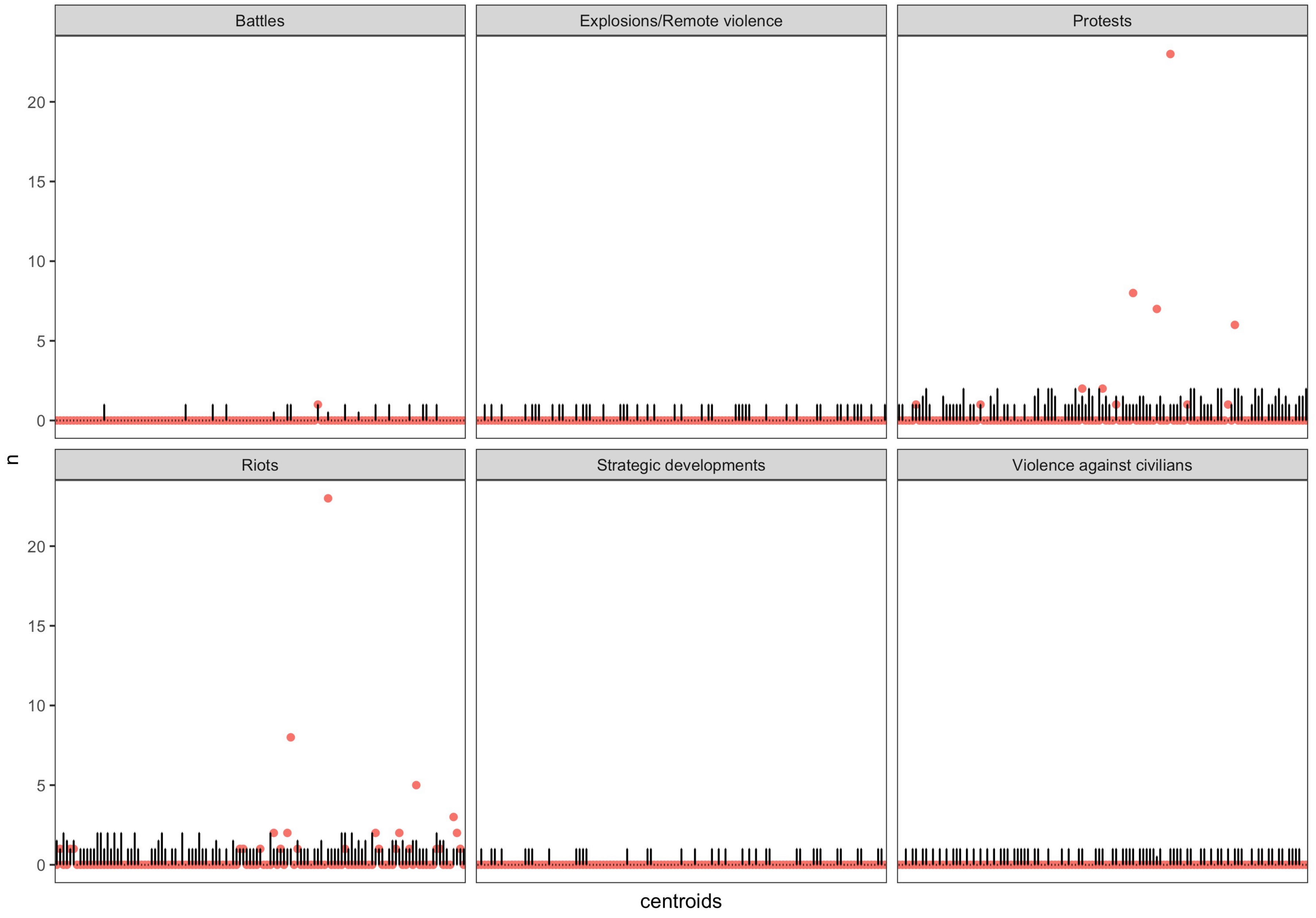
By location



Bangladesh



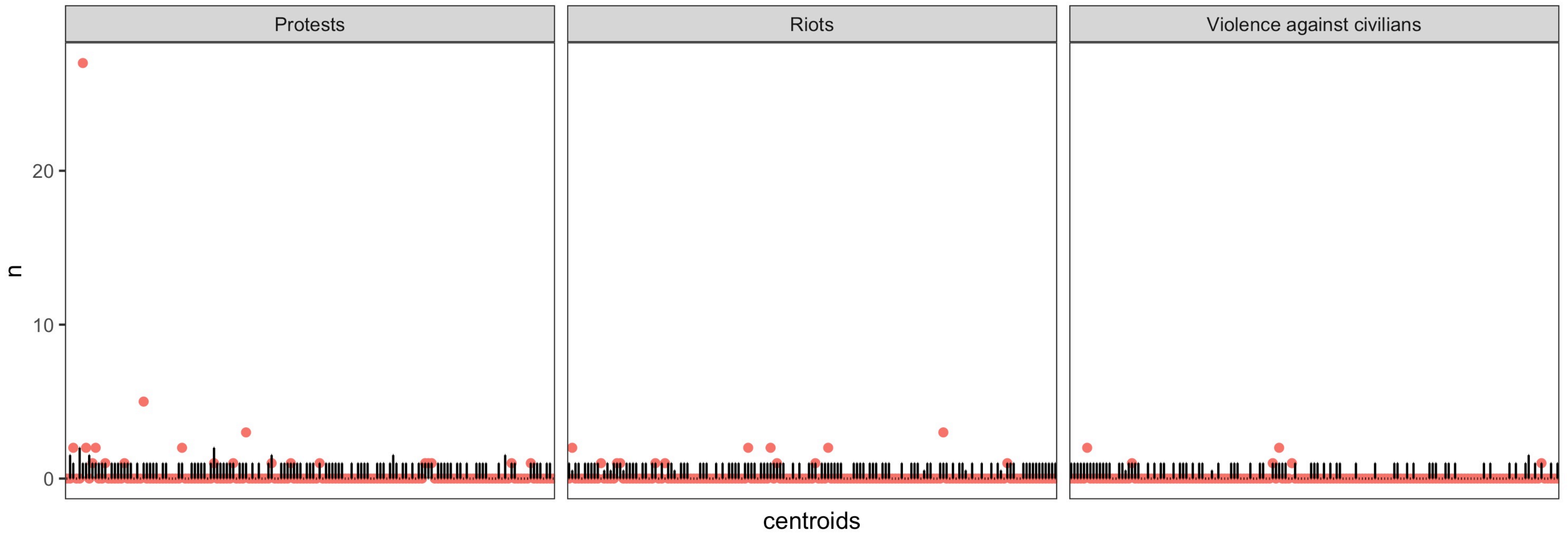
Pakistan



Nepal

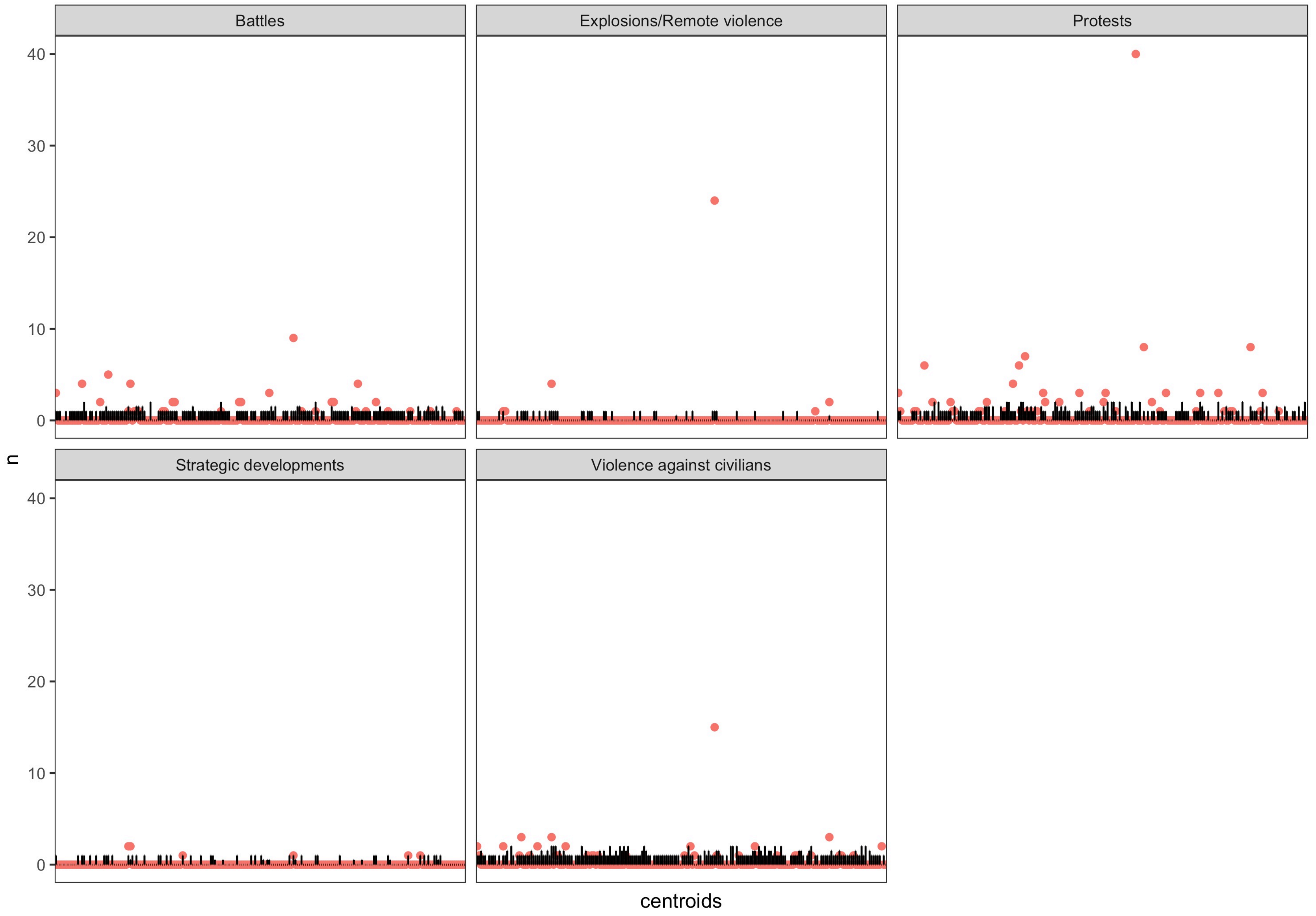
Out of sample predictive checks by location.

The red dots show the observed number of events. The bars show the 95% posterior predictive interval from 100 simulations.

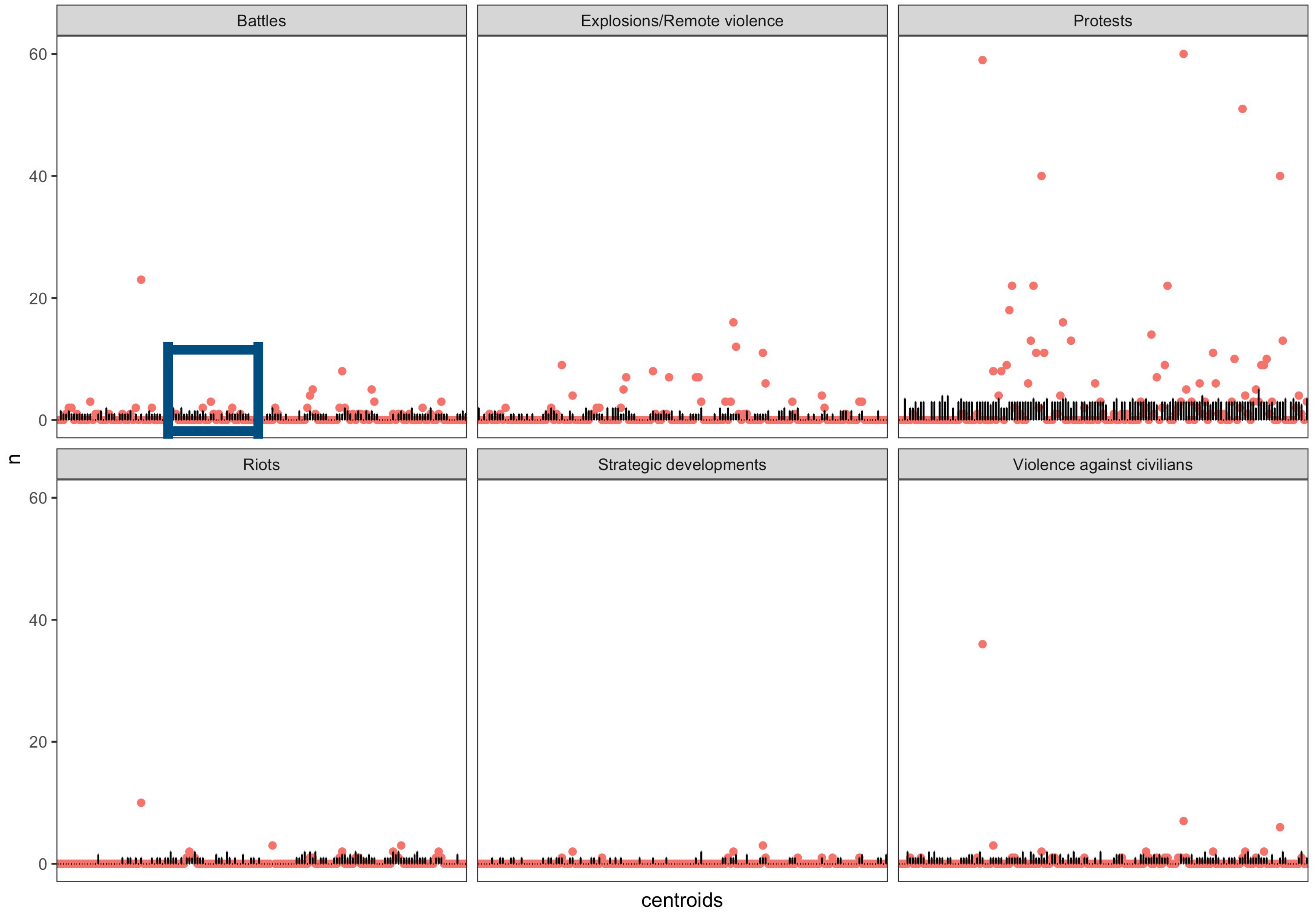


Sri Lanka

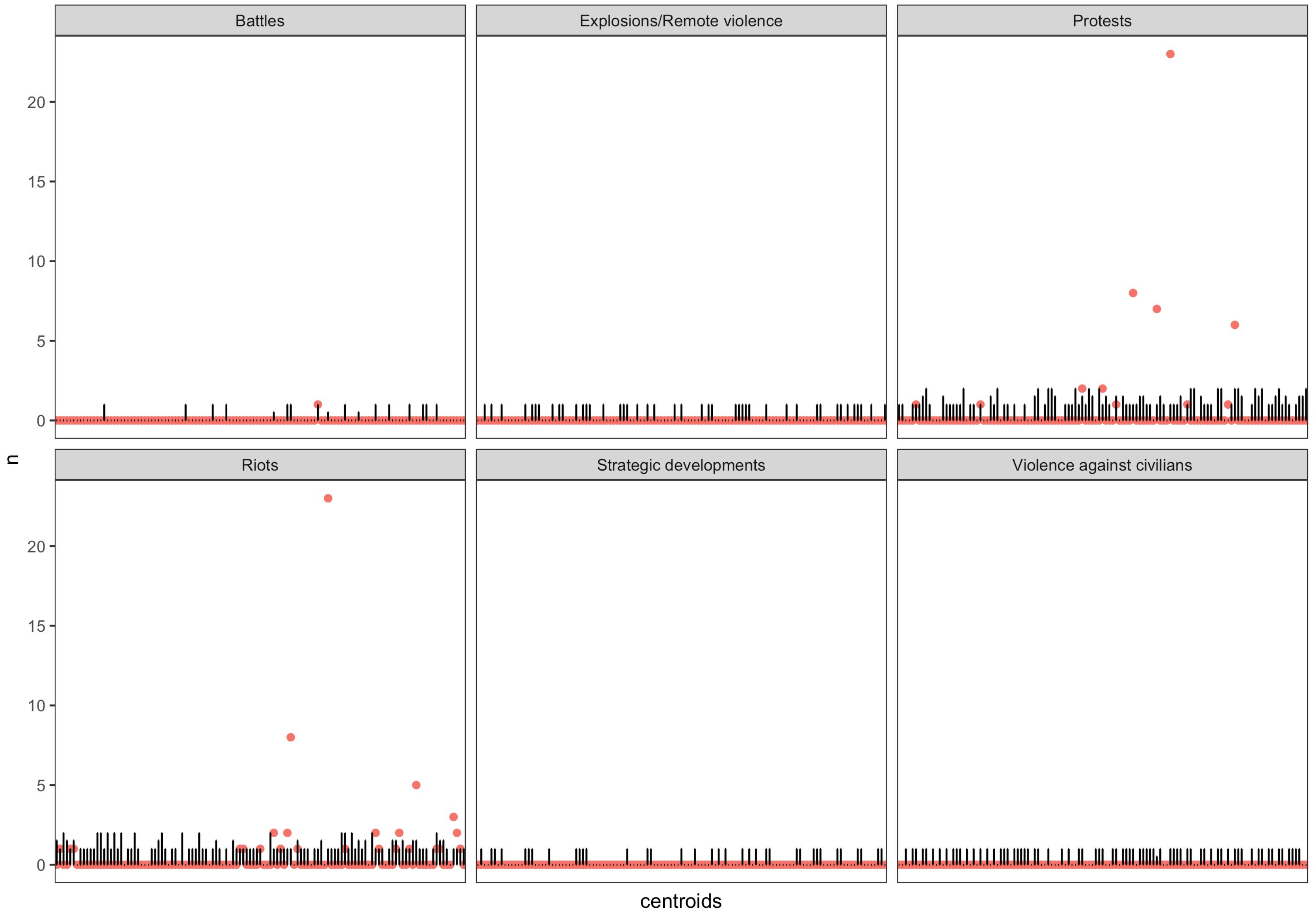
By location



Bangladesh



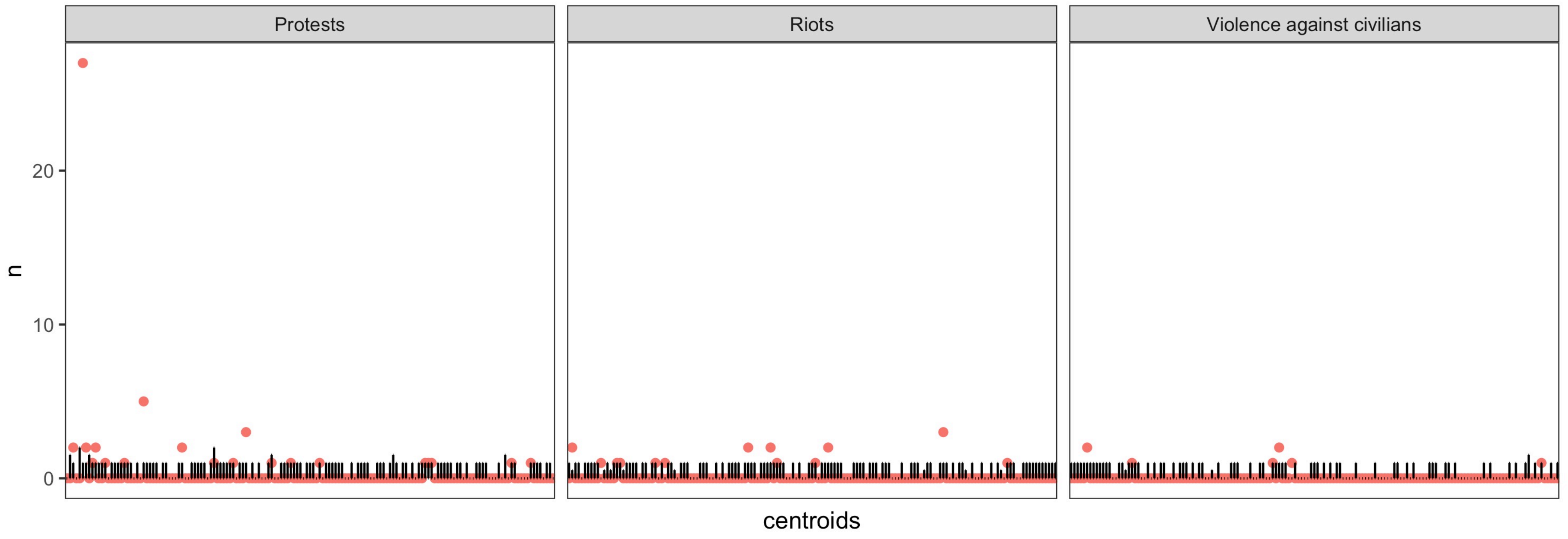
Pakistan



Nepal

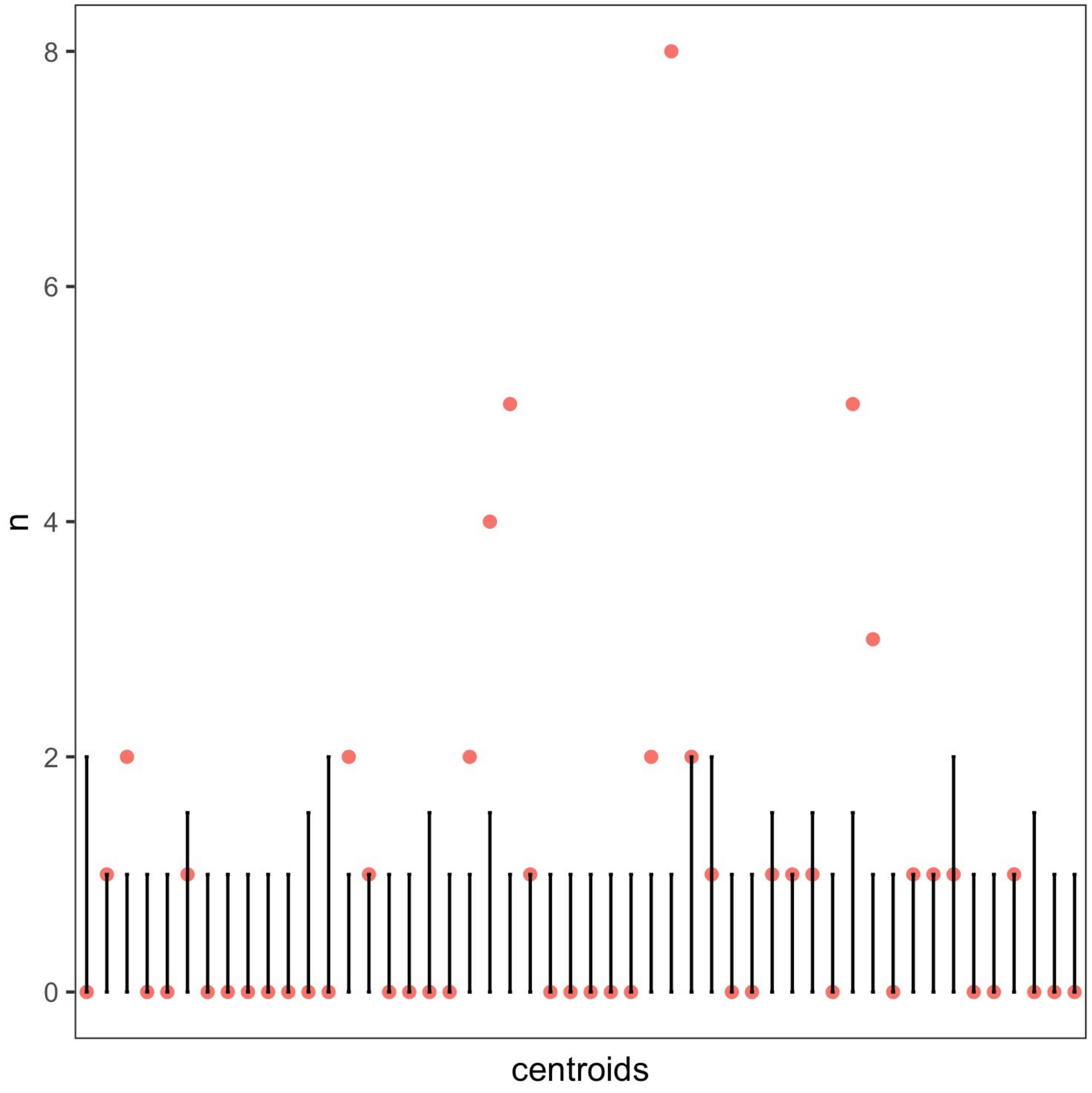
Out of sample predictive checks by location.

The red dots show the observed number of events. The bars show the 95% posterior predictive interval from 100 simulations.



Sri Lanka

By location



Predictions for Battles in Pakistan (for 50 locations)

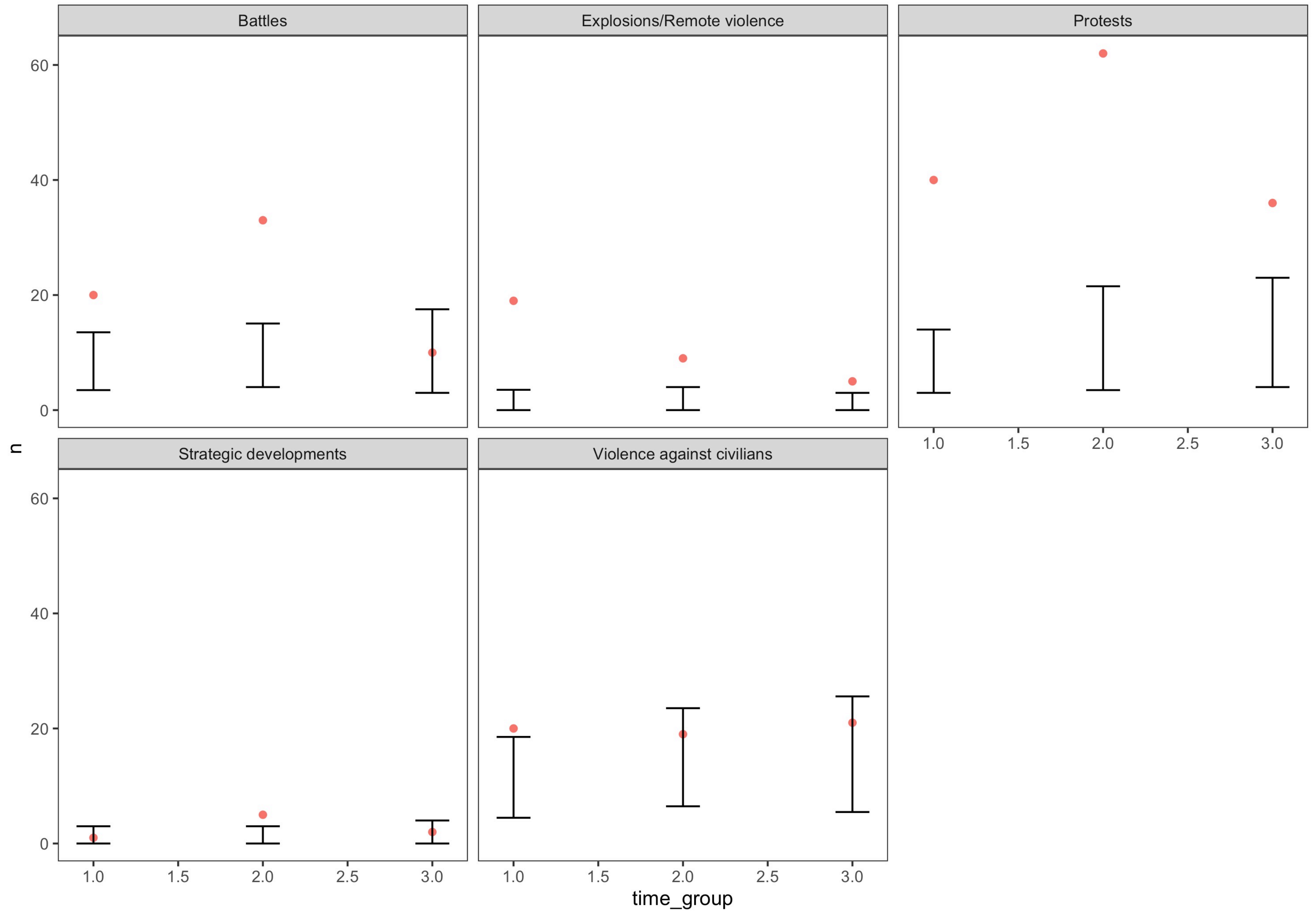
Out of sample predictive checks by location.

The red dots show the observed number of events. The bars show the 95% posterior predictive interval from 100 simulations.

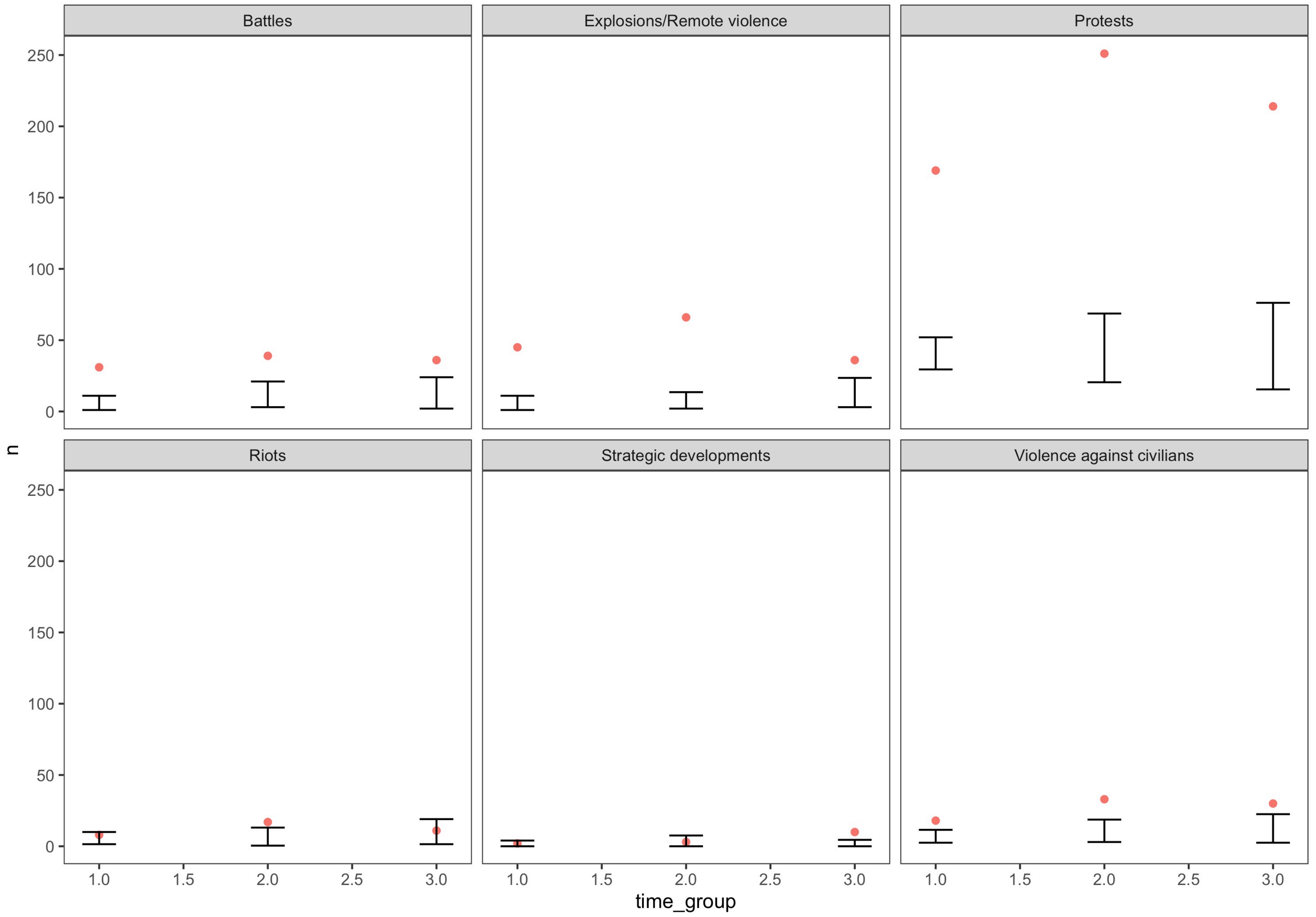
By time

Out of sample predictive checks by month.

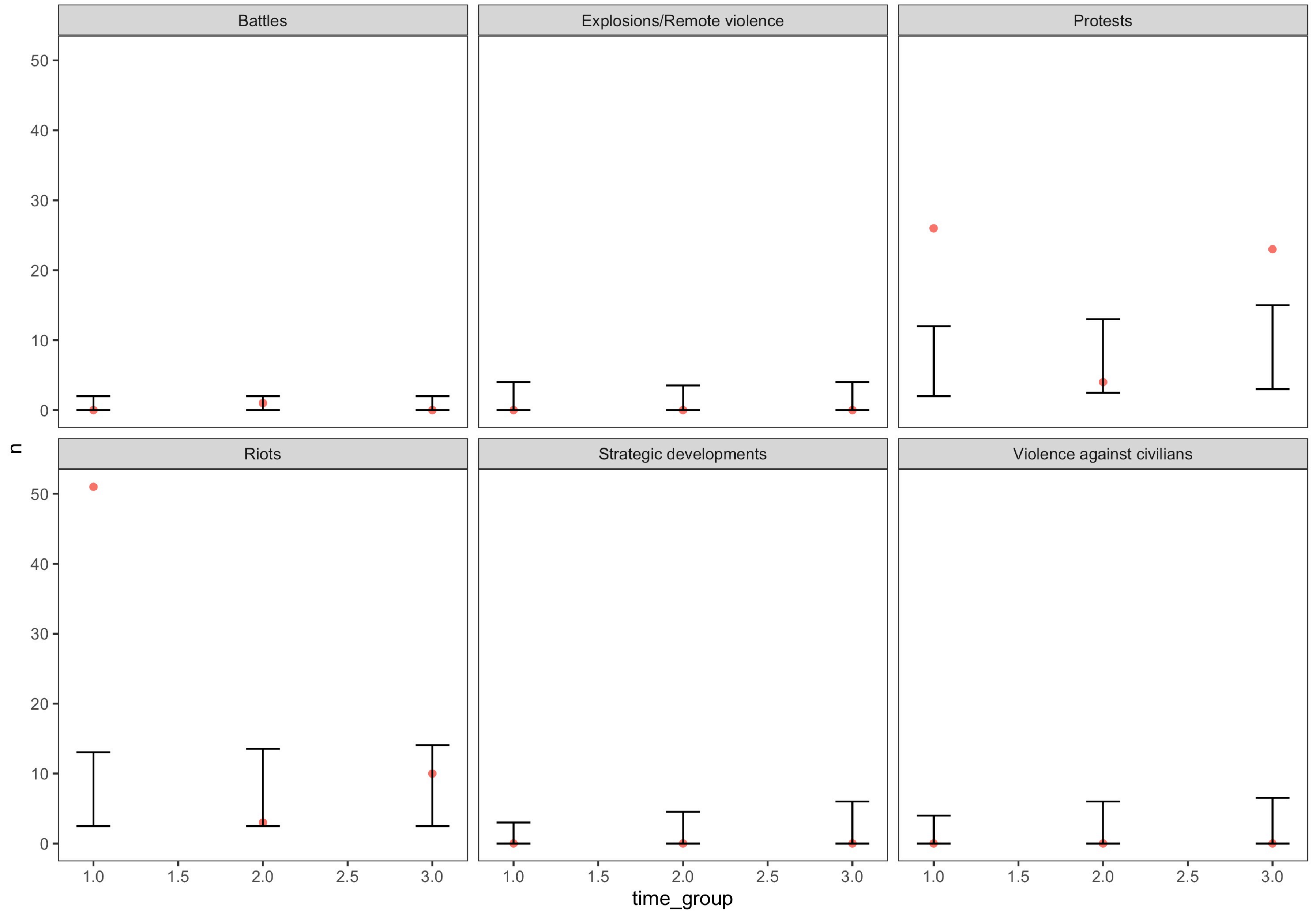
The red dots show the observed number of events. The bars show the 95% posterior predictive interval from 100 simulations.



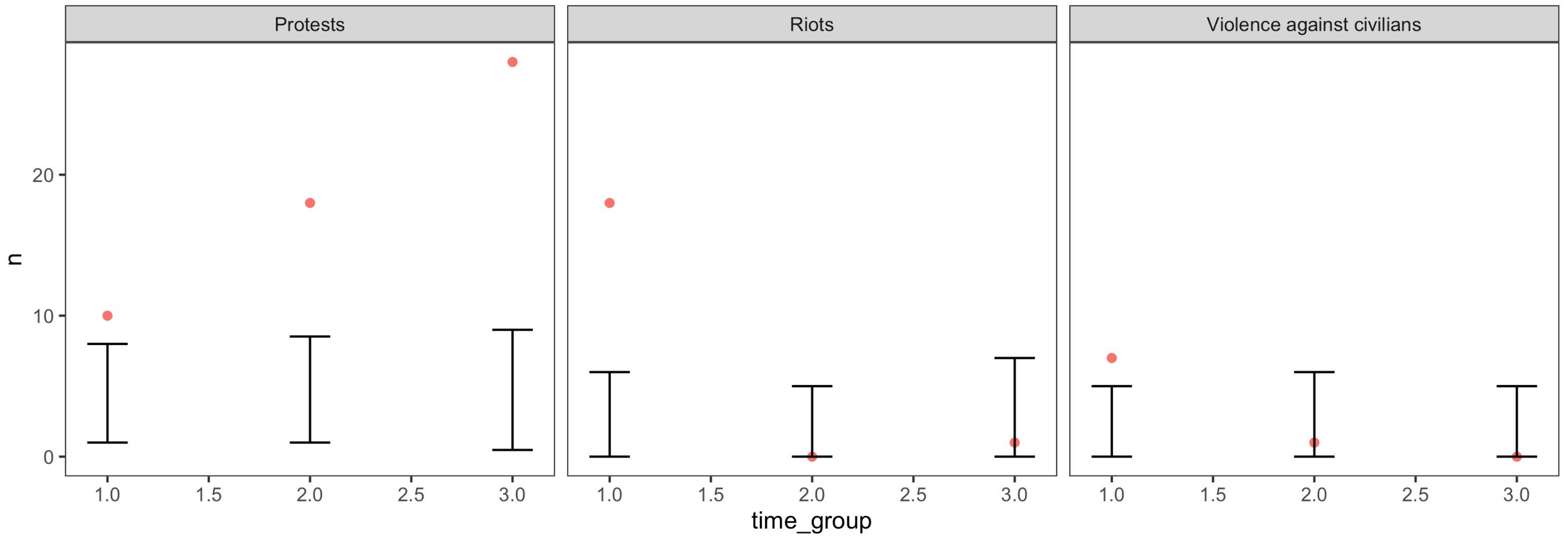
Bangladesh



Pakistan



Nepal



Sri Lanka

Conclusions

- Propose spatiotemporal DTHP to model conflict data from the ACLED project
- Potential to be adopted as the official ACLED risk model
- Strong dependence on external factors means these types of events are extremely difficult to predict - but monitoring the level of risk through time and space is still important
  - Still considering how to summarise goodness of fit for these models

Overall conclusions

Summary

- Propose several new models for DTHPs:
  - Variations of random histogram kernels for temporal triggering kernels
  - Spatiotemporal kernel for DTHPs
- Random histogram kernels avoid assumption of simple parametric distributions that may be restrictive
- Applied these models to two substantive case studies
- Explored and proposed new algorithms for implementing DTHPs

Future work

- Design scalable algorithms to enable high-dimensional Bayesian nonparametric inference of HPs
- Quantify the approximation error that results from the discretisation of the time domain
- Compare this approximation with the errors introduced by imputing data collected in discrete time onto a continuous timeline
- Time varying parameters

Acknowledgements

- **Supervisors:** Kerrie Mengersen, Judith Rousseau and Gentry White
- **Panel members:** Adam Clements, Leah South
- **PhD stipend and travel funding:** from QUT, CDS, ACEMS, QUT SMS and my supervisor
- **Communities:** ACEMS, BRAG and CDS
- **Collaborators:** Deborah Sulem, Vincent Rivoirard, Hamish Patten
- **Fellow PhD students, friends, family, and my husband**

References

Hawkes, Alan G. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika. 58* (1): 83–90. https://doi.org/10.2307/2334319

Brémaud, P., & Massoulié, L. (1996). Stability of Nonlinear Hawkes Processes. *The Annals of Statistics*, *24*, 1563–1588. https://doi.org/10.2307/2244985

Donnet, S., Rivoirard, V., & Rousseau, J. (2020). Nonparametric Bayesian estimation for multivariate Hawkes processes. *The Annals of Statistics*, *48*(5), 2698–2727. https://doi.org/10.1214/19-aos1903

Sulem, D., Rivoirard, V., & Rousseau, J. (2021). Bayesian estimation of nonlinear Hawkes process. *ArXiv*.

Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. COVID- 19 data repository; 2020. Accessed: 26 July 2020. https://github.com/CSSEGISandData/COVID-19

Roberts, G. O., & Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, *2*(4), 341–363. https://doi.org/https://doi.org/10.2307/3318418

START (National Consortium for the Study of Terrorism and Responses to Terrorism). (2022). *Global Terrorism Database 1970 - 2020*. https://www.start.umd.edu/gtd

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732. https://academic.oup.com/biomet/article-abstract/82/4/711/252058

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, *85*(410), 398. https://doi.org/10.2307/2289776

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. (2010). “Introducing ACLED-Armed Conflict Location and Event Data.” Journal of Peace Research 47(5) 651-660