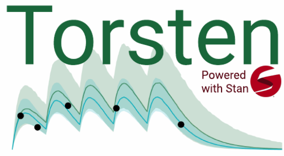
Making Bayesian pharmacometrics modeling simpler (but not too simple) with Torsten



Charles Margossian
Flatiron Institute

Fellow Torsten developers:

Yi Zhang

Bill Gillespie

https://github.com/metrumresearchgroup/Torsten

Features to support pmx modeling

mc-stan.org

Torsten

Powered with Stan

flexible base language

ODE solvers

matrix exponential

algebraic solver

DAE solver

within-chain parallelization

Features to support pmx modeling

mc-stan.org

flexible base language

ODE solvers

matrix exponential

algebraic solver

DAE solver

within-chain parallelization

Torsten

Powered with Stan

event schedule

PK analytical solutions

PK analytical + numerical

within-pop parallelization

`bbr.bayes`: R workflow

Features to support pmx modeling

mc-stan.org

flexible base language

ODE solvers

matrix exponential

algebraic solver

DAE solver

within-chain parallelization

Torsten

Powered with Stan

event schedule

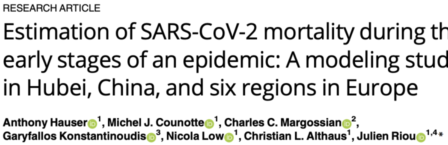
PK analytical solutions

PK analytical + numerical

within-pop parallelization

`bbr.bayes`: R workflow

- Goal: build a workflow that's intuitive but hackable

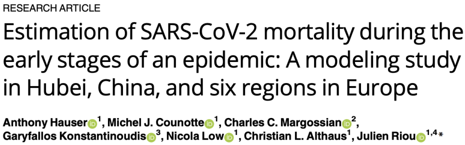


RESEARCH ARTICLE

Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe

Anthony Hauser[1], Michel J. Counotte[1], Charles C. Margossian[2], Garyfallos Konstantinoudis[3], Nicola Low[1], Christian L. Althaus[1], Julien Riou[1,4]*

- ODE-based model initially took 3 days to fit.
- Rewriting Stan code led to model fitting in 2 hours!



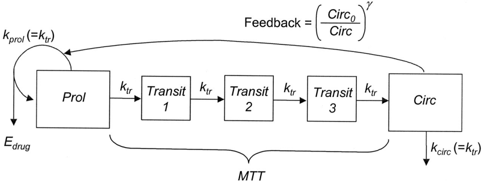
RESEARCH ARTICLE

Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe

Anthony Hauser[1], **Michel J. Counotte**[1], **Charles C. Margossian**[2], **Garyfallos Konstantinoudis**[3], **Nicola Low**[1], **Christian L. Althaus**[1], **Julien Riou**[1,4]*

- ODE-based model initially took 3 days to fit.
- Rewriting Stan code led to model fitting in 2 hours!

Drug-induced myelosuppression example



- Bound $\mathrm{E}_{\mathrm{drug}}$ to deal with “unreasonable” parameter values the Markov chain may encounter during warmup.
- Write the ODE as a baseline difference from y_0 to reduce the computational cost of automatic differentiation, etc.

Tutorial papers:

- Grinsztajn et al. Bayesian workflow for disease transmission in Stan. *Statistics in medicine*, 2021.
- M, Zhang and Gillespie. Bayesian modeling using Stan and Torsten, Part I.*CPT: PSP*, 2022 — part II in preperation.
- Elmokadem et al. Bayesian PBPK using R/Stan/Torsten and Julia/SciML/Turing.JL. *CPT: PSP*, 2022.

Tutorial papers:

- Grinsztajn et al. Bayesian workflow for disease transmission in Stan. *Statistics in medicine*, 2021.
- M, Zhang and Gillespie. Bayesian modeling using Stan and Torsten, Part I.*CPT: PSP*, 2022 — part II in preperation.
- Elmokadem et al. Bayesian PBPK using R/Stan/Torsten and Julia/SciML/Turing.JL. *CPT: PSP*, 2022.

- Conferences, workshops, meetups, etc.

Tutorial papers:

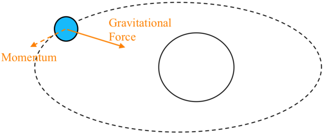
- Grinsztajn et al. Bayesian workflow for disease transmission in Stan. *Statistics in medicine*, 2021.
- M, Zhang and Gillespie. Bayesian modeling using Stan and Torsten, Part I.*CPT: PSP*, 2022 — part II in preperation.
- Elmokadem et al. Bayesian PBPK using R/Stan/Torsten and Julia/SciML/Turing.JL. *CPT: PSP*, 2022.

- Conferences, workshops, meetups, etc.

What are some elementary gaps in our workflow?

1. Tuning the ODE solver.
2. Initializing the Markov chain.
3. Picking the length of warmup and sampling phases.

Example: planetary motion[†]

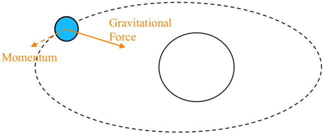


Let q be the planet's position
and p its momentum.

$$\begin{aligned} \dot{q}(t) &= p(t)/m \\ \dot{p}(t) &= -\frac{k}{r^3}(q - q_\odot) \end{aligned}$$

[†] M and Gelman, *Stan Case studies* (2020), Gelman et al, *arXiv:2011.01808* (2020)

Example: planetary motion[†]



Let q be the planet's position and p its momentum.

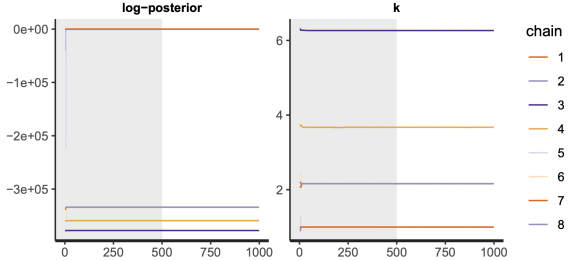
$$\begin{aligned} \dot{q}(t) &= p(t)/m \\ \dot{p}(t) &= -\frac{k}{r^3}(q - q_\odot) \end{aligned}$$

500 warmup + 500 sampling

Chain	**Time (s)**
1	10.56
2	3.40
3	4433.93
4	181.98

[†] M and Gelman, *Stan Case studies* (2020), Gelman et al, *arXiv:2011.01808* (2020)

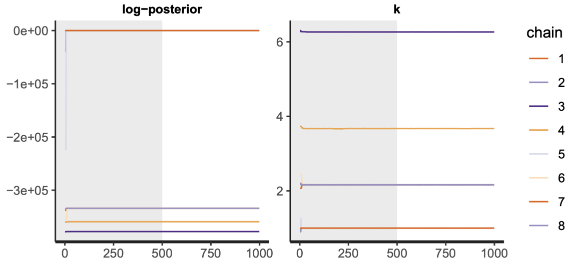
Example: planetary motion[†]



Problem is highly multimodal.

[†] M and Gelman, *Stan Case studies* (2020), Gelman et al, *arXiv:2011.01808* (2020)

Example: planetary motion[†]

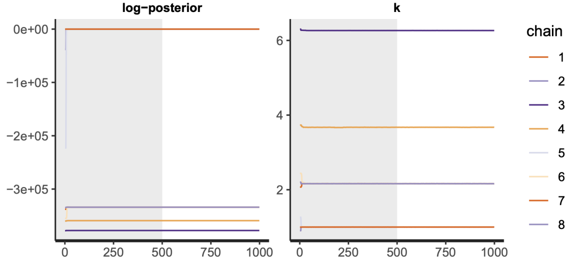


Problem is highly multimodal.

The ODE solver's error scales as $\mathcal{O}(\epsilon^t k)$, with $t \geq 2$.

[†] M and Gelman, *Stan Case studies* (2020), Gelman et al, *arXiv:2011.01808* (2020)

Example: planetary motion[†]



Problem is highly multimodal.

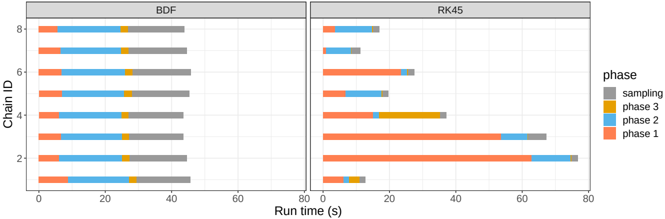
The ODE solver's error scales as $\mathcal{O}(\epsilon^t k)$, with $t \geq 2$.

Stan's default init: $\log k \sim \text{uniform}[-2, 2]$

Recommendation: use more careful initializations, e.g. sample from prior.

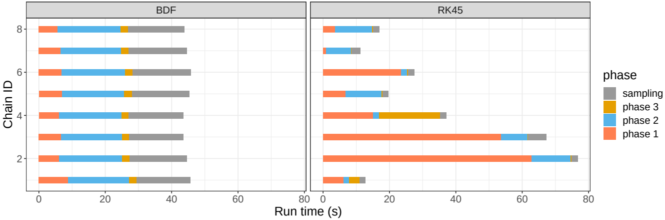
[†] M and Gelman, *Stan Case studies* (2020), Gelman et al, *arXiv:2011.01808* (2020)

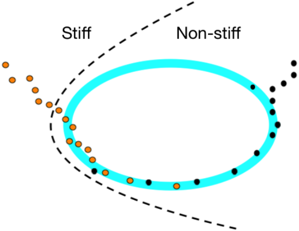
Example: non-linear pharmacokinetic model[†]



[†] M et al, *PAGE* (2021)

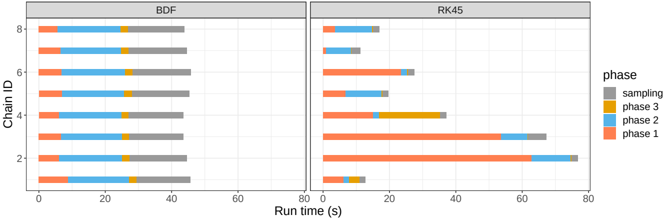
Example: non-linear pharmacokinetic model[†]





[†] M et al, *PAGE* (2021)

Example: non-linear pharmacokinetic model[†]



	Phase I	**Phase II**	**Phase III**	**Sampling**
RK45	RK45	RK45	RK45	RK45
BDF	BDF	BDF	BDF	BDF
Early switch	BDF	RK45	RK45	RK45
Late switch	BDF	BDF	RK45	RK45

[†] M et al, *PAGE* (2021)

What makes a good initialization for MCMC?

[†] Gelman and Rubin. *Statistical Science* (1992), M et al. *arXiv:2110.13017* (2022)

What makes a good initialization for MCMC?

With a good initialization, the bias of our Monte Carlo estimator decays faster.

Safeguards us against regions where ODE solver struggles.

[†] Gelman and Rubin. *Statistical Science* (1992), M et al. *arXiv:2110.13017* (2022)

What makes a good initialization for MCMC?

With a good initialization, the bias of our Monte Carlo estimator decays faster.

Safeguards us against regions where ODE solver struggles.

But for $\widehat{R}$ to be reliable, we need *overdispersed initialization* relative to the initial bias[†]

$$\text{Initial Variance} \geq \alpha + \beta(\text{Initial Bias})^2$$

[†] Gelman and Rubin. *Statistical Science* (1992), M et al. *arXiv:2110.13017* (2022)

What makes a good initialization for MCMC?

With a good initialization, the bias of our Monte Carlo estimator decays faster.

Safeguards us against regions where ODE solver struggles.

But for $\widehat{R}$ to be reliable, we need *overdispersed initialization* relative to the initial bias[†]

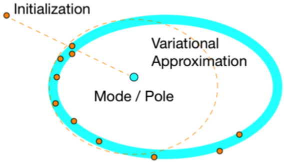
$$\text{Initial Variance} \geq \alpha + \beta(\text{Initial Bias})^2$$

We can expect the prior to verify this property.

But also any approximation that reduces the squared bias of the prior more than its variance.

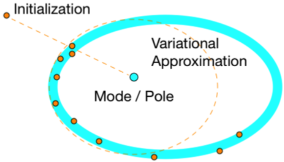
[†] Gelman and Rubin. *Statistical Science* (1992), M et al. *arXiv:2110.13017* (2022)

What makes a good initialization?

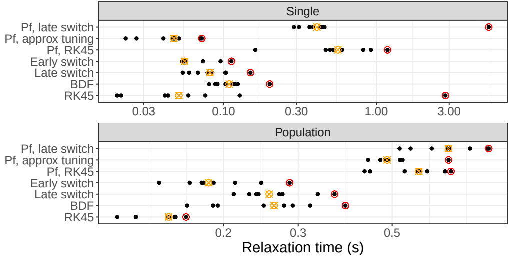


[†] Zhang et al. *JMLR* (2022), M et al. *PAGE* (2022)

What makes a good initialization?

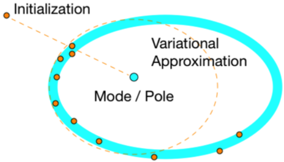


Experiments with pathfinder in R.[†]



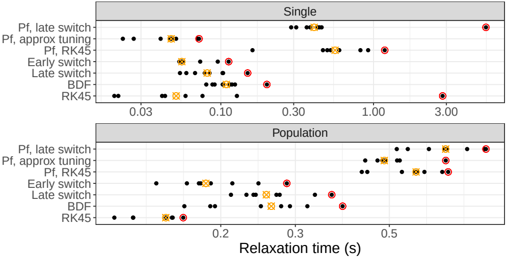
[†] Zhang et al. *JMLR* (2022), M et al. *PAGE* (2022)

What makes a good initialization?



Experiments with pathfinder in R.[†]

Stan released a C++ implementation.



[†] Zhang et al. *JMLR* (2022), M et al. *PAGE* (2022)

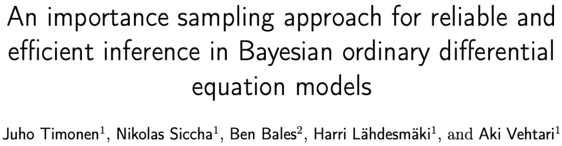
Picking a tolerance for the ODE solvers[†]

An importance sampling approach for reliable and efficient inference in Bayesian ordinary differential equation models

Juho Timonen[1], Nikolas Siccha[1], Ben Bales[2], Harri Lähdesmäki[1], and Aki Vehtari[1]

[†] Timonen et al. *arXiv:2205.09059* (2022)

Picking a tolerance for the ODE solvers[†]



An importance sampling approach for reliable and efficient inference in Bayesian ordinary differential equation models

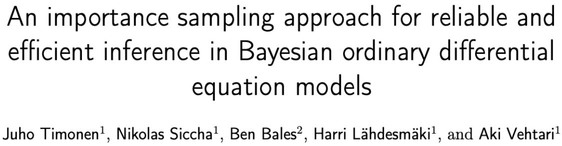
Juho Timonen[1], Nikolas Siccha[1], Ben Bales[2], Harri Lähdesmäki[1], and Aki Vehtari[1]

Due to solver's precision ϵ_1, only compute approximate likelihood

$$\tilde{\pi}(y \mid \theta; \epsilon_1) \neq \pi(y \mid \theta)$$

[†] Timonen et al. *arXiv:2205.09059* (2022)

Picking a tolerance for the ODE solvers[†]


An importance sampling approach for reliable and efficient inference in Bayesian ordinary differential equation models

Juho Timonen[1], Nikolas Siccha[1], Ben Bales[2], Harri Lähdesmäki[1], and Aki Vehtari[1]

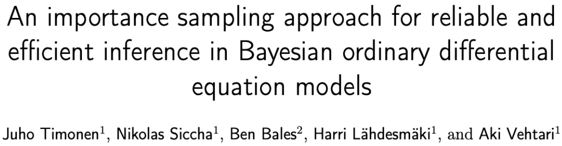
Due to solver's precision ϵ_1, only compute approximate likelihood

$$\tilde{\pi}(y \mid \theta; \epsilon_1) \neq \pi(y \mid \theta)$$

IS compares $\tilde{\pi}(y \mid \theta; \epsilon_1)$ to $\tilde{\pi}(y \mid \theta; \epsilon_2) \approx \pi(y \mid \theta)$, $\epsilon_2 \ll \epsilon_1$.

[†] Timonen et al. *arXiv:2205.09059* (2022)

Picking a tolerance for the ODE solvers[†]



An importance sampling approach for reliable and efficient inference in Bayesian ordinary differential equation models

Juho Timonen[1], Nikolas Siccha[1], Ben Bales[2], Harri Lähdesmäki[1], and Aki Vehtari[1]

Due to solver's precision ϵ_1, only compute approximate likelihood

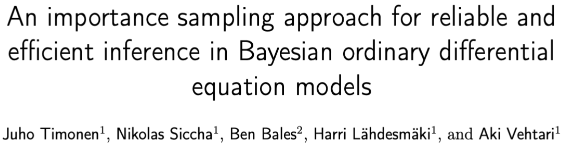
$$\tilde{\pi}(y \mid \theta; \epsilon_1) \neq \pi(y \mid \theta)$$

IS compares $\tilde{\pi}(y \mid \theta; \epsilon_1)$ to $\tilde{\pi}(y \mid \theta; \epsilon_2) \approx \pi(y \mid \theta)$, $\epsilon_2 \ll \epsilon_1$.

Provides a diagnostic for solver's tuning parameters.

[†] Timonen et al. *arXiv:2205.09059* (2022)

Picking a tolerance for the ODE solvers[†]



An importance sampling approach for reliable and efficient inference in Bayesian ordinary differential equation models

Juho Timonen[1], Nikolas Siccha[1], Ben Bales[2], Harri Lähdesmäki[1], and Aki Vehtari[1]

Due to solver's precision ϵ_1, only compute approximate likelihood

$$\tilde{\pi}(y \mid \theta; \epsilon_1) \neq \pi(y \mid \theta)$$

IS compares $\tilde{\pi}(y \mid \theta; \epsilon_1)$ to $\tilde{\pi}(y \mid \theta; \epsilon_2) \approx \pi(y \mid \theta)$, $\epsilon_2 \ll \epsilon_1$.

Provides a diagnostic for solver's tuning parameters.

Plan: experiment on range of PK models and implement in Torsten workflow.

[†] Timonen et al. *arXiv:2205.09059* (2022)

What are some elementary gaps in our workflow?

1. Tuning the ODE solver. ✓
2. Initializing the Markov chain. ✓
3. Picking the length of warmup and sampling phases.

(✓ means there is a course of action)

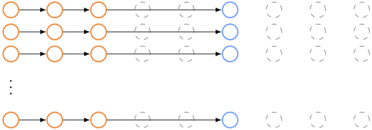
[†] Lao et al. *arXiv:2002.01184* (2020); M et al. *arXiv:2110.13017* (2022)

What are some elementary gaps in our workflow?

1. Tuning the ODE solver. ✓
2. Initializing the Markov chain. ✓
3. Picking the length of warmup and sampling phases.

(✓ means there is a course of action)

Many-short-chains regime[†]



[†] Lao et al. *arXiv:2002.01184* (2020); M et al. *arXiv:2110.13017* (2022)

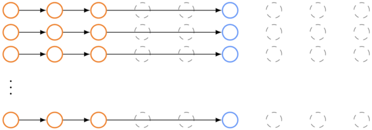
What are some elementary gaps in our workflow?

1. Tuning the ODE solver. ✓
2. Initializing the Markov chain. ✓
3. Picking the length of warmup and sampling phases.

(✓ means there is a course of action)

Many-short-chains regime[†]

Dynamically monitor convergence via nested $\widehat{R}$



[†] Lao et al. *arXiv:2002.01184* (2020); M et al. *arXiv:2110.13017* (2022)

What are some elementary gaps in our workflow?

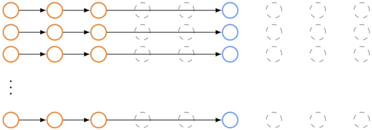
1. Tuning the ODE solver. ✓
2. Initializing the Markov chain. ✓
3. Picking the length of warmup and sampling phases.

(✓ means there is a course of action)

Many-short-chains regime[†]

Dynamically monitor convergence via nested $\widehat{R}$

Requires running many chains (on GPU or clusters of CPUs), which can be challenging with Stan and ODE-based models.



[†] Lao et al. *arXiv:2002.01184* (2020); M et al. *arXiv:2110.13017* (2022)