Stan

Software Ecosystem for Modern Bayesian Inference

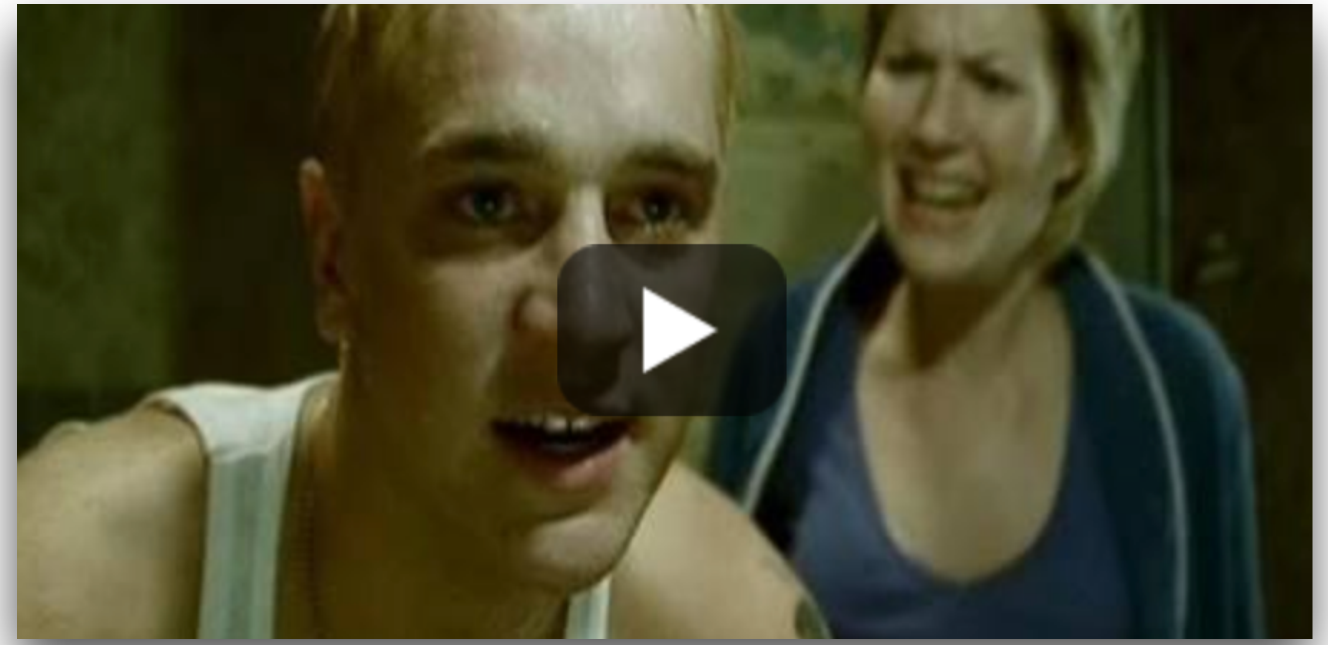
StanCon 2018 Helsinki

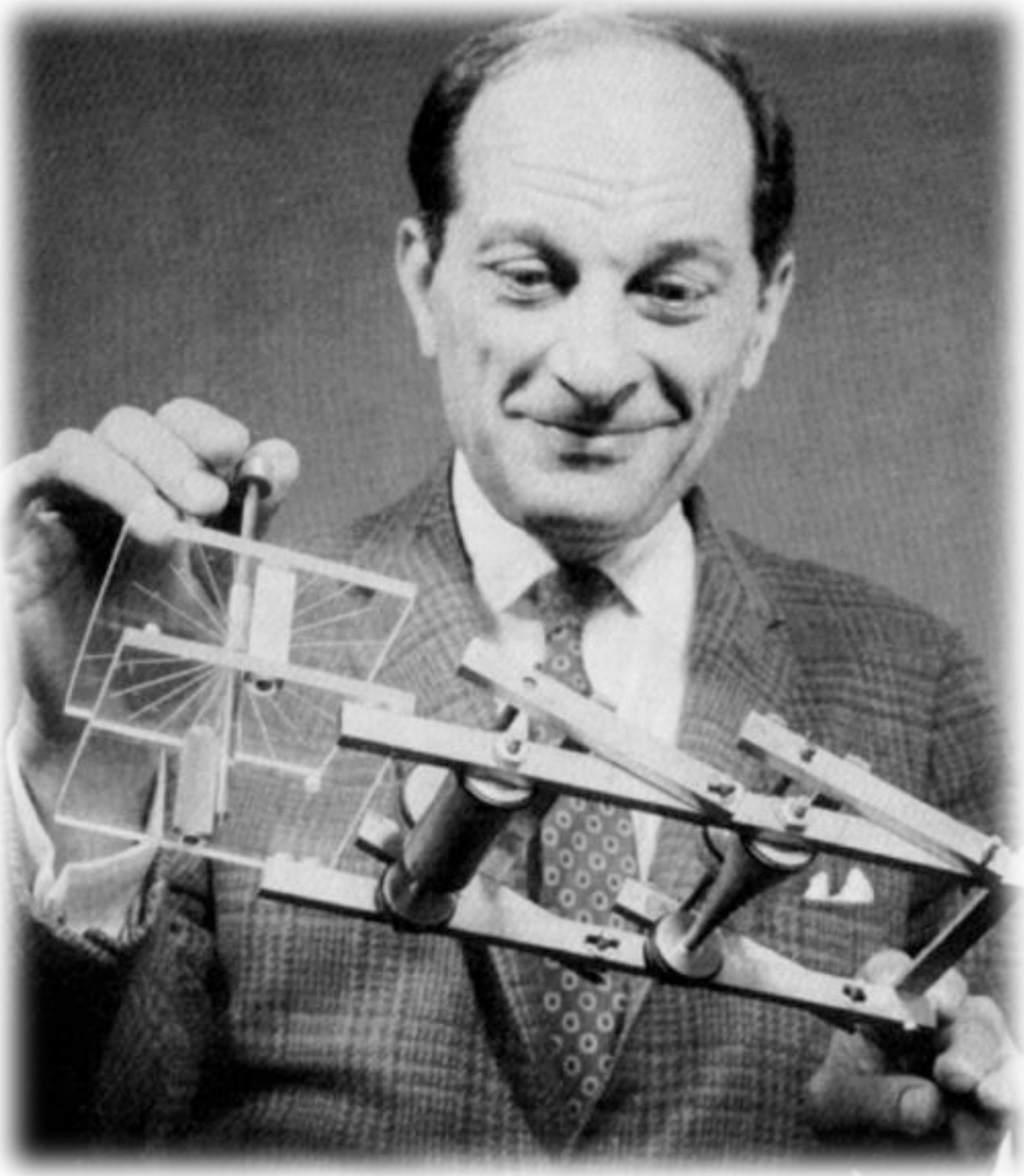
Jonah Gabry & Lauren Kennedy
Columbia University

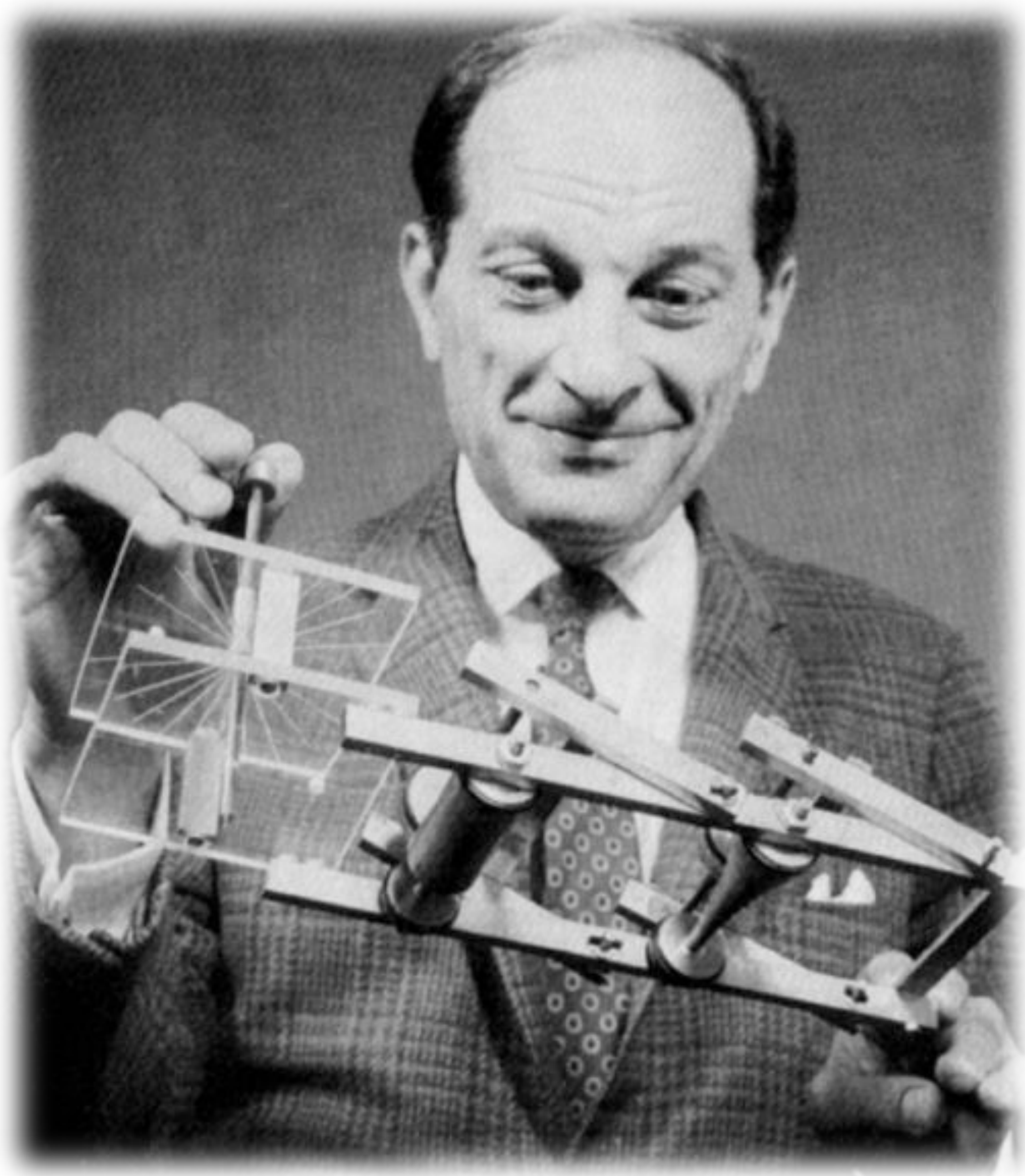
github.com/jgabry/stancon2018helsinki_intro

Why “Stan”?

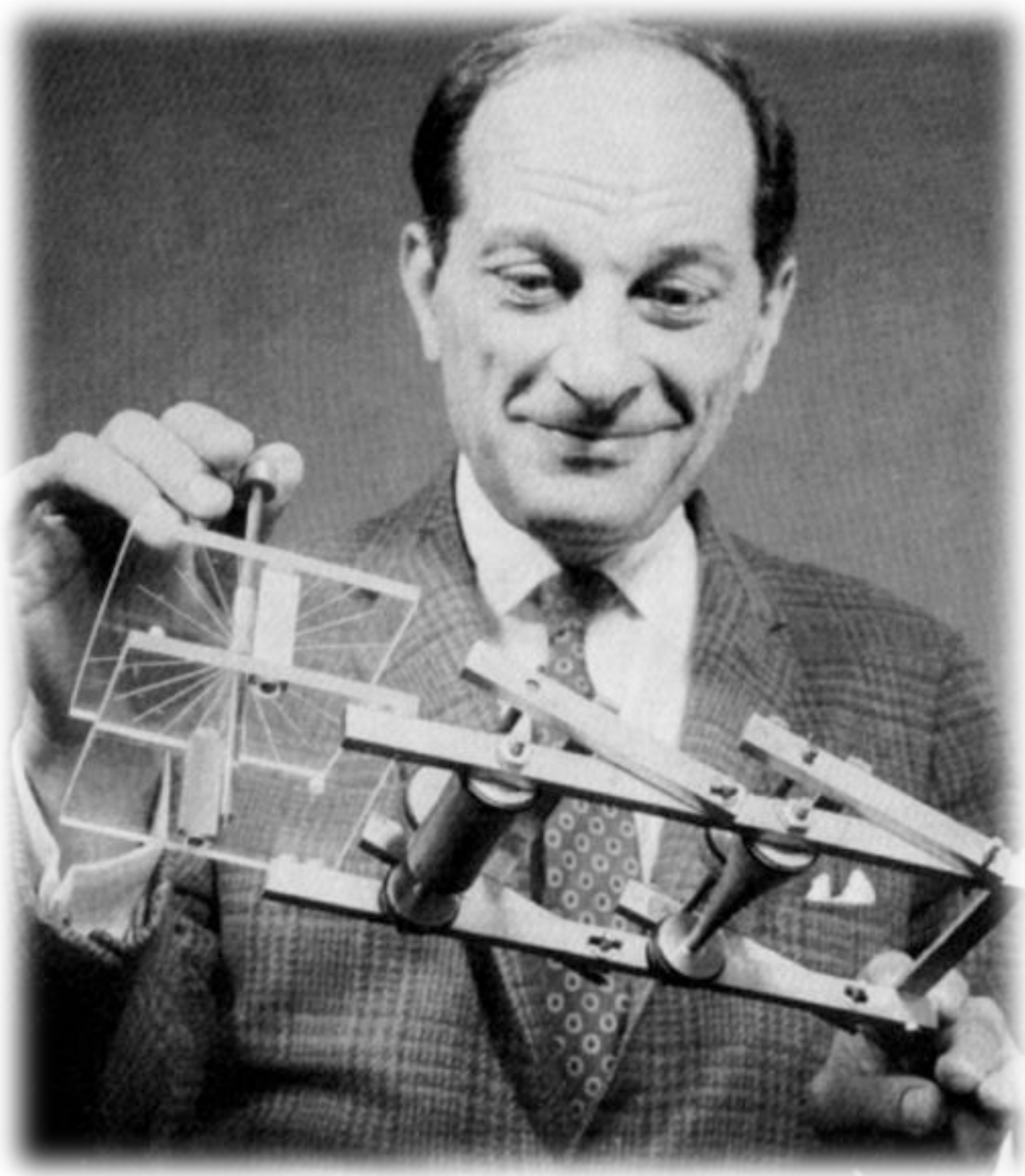
suboptimal SEO





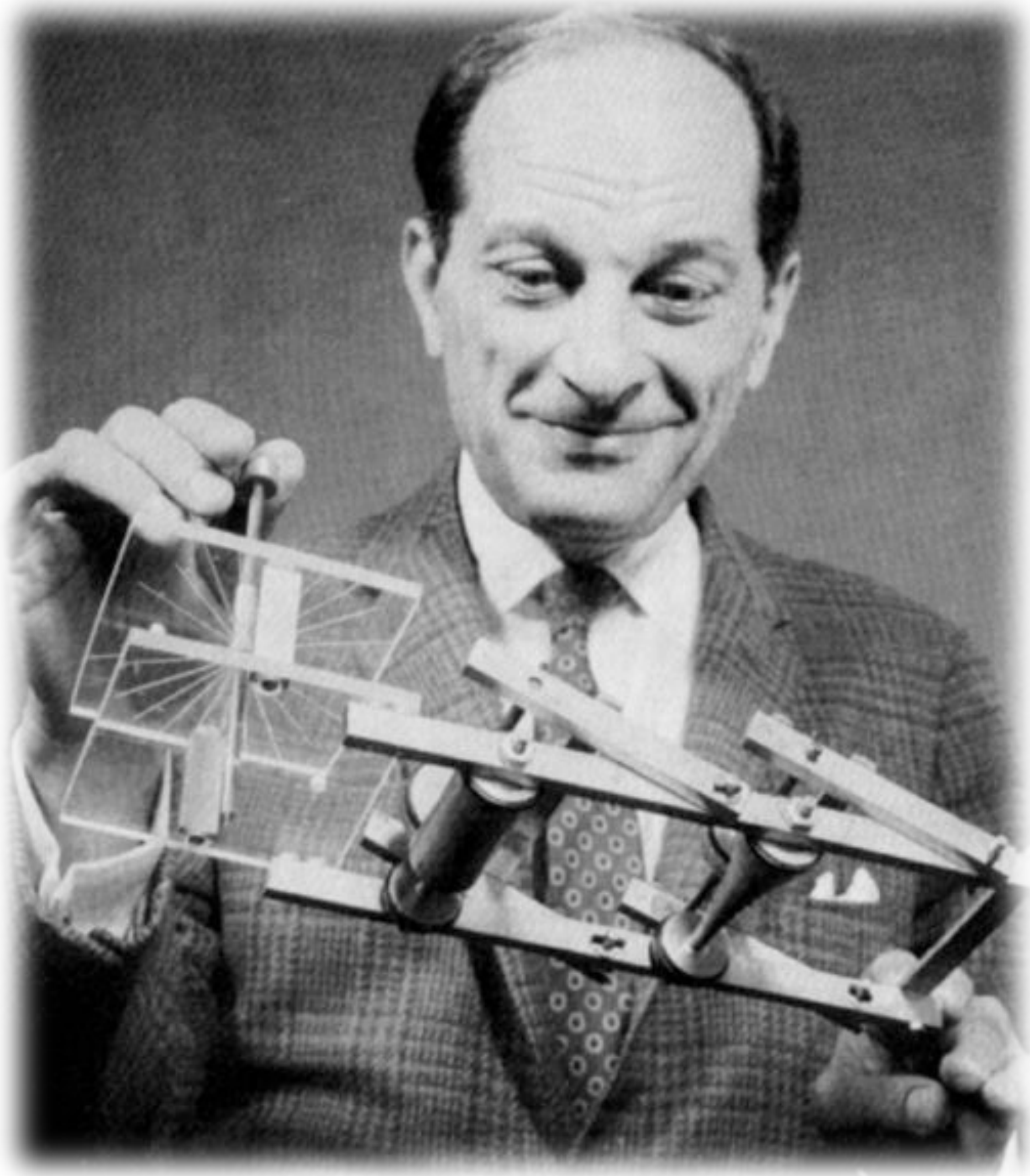


Stanislaw Ulam
(1909–1984)



Stanislaw Ulam
(1909–1984)

Monte Carlo
Method



Stanislaw Ulam
(1909–1984)

H-Bomb

Monte Carlo
Method

What is Stan?

What is Stan?

- Open source probabilistic **programming language**, **inference algorithms**

What is Stan?

- Open source probabilistic **programming language**, **inference algorithms**
- Stan **program**
  - declares data and (constrained) parameter variables
  - defines log posterior (or penalized likelihood)

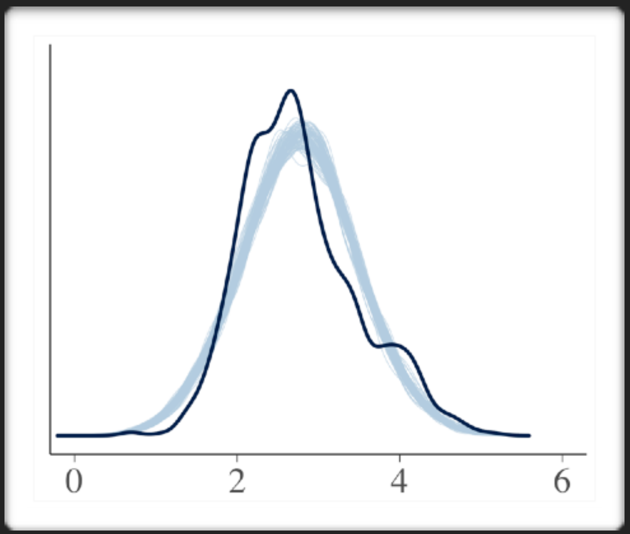
What is Stan?

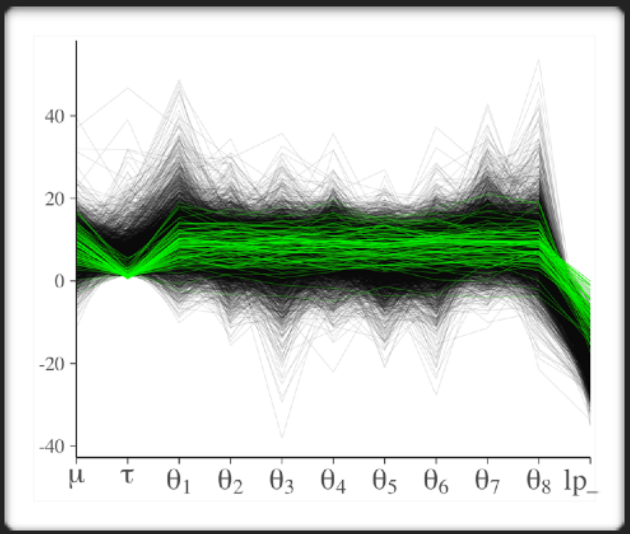
- Open source probabilistic **programming language**, **inference algorithms**
- Stan **program**
  - declares data and (constrained) parameter variables
  - defines log posterior (or penalized likelihood)
- Stan **inference**
  - MCMC for full Bayes
  - VB for approximate Bayes
  - Optimization for (penalized) MLE

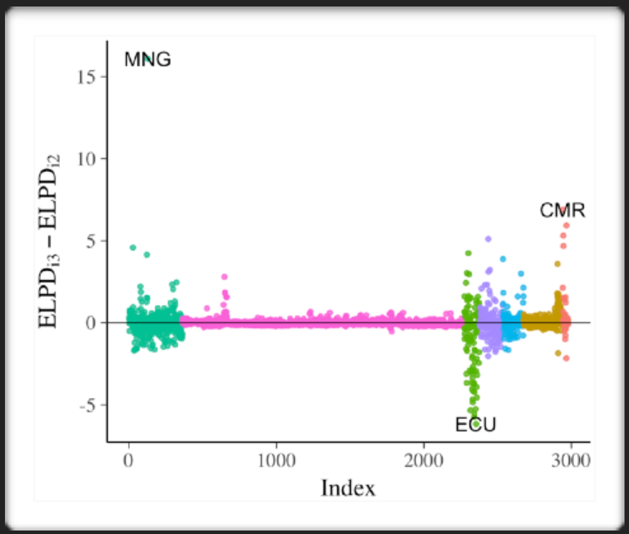
What is Stan?

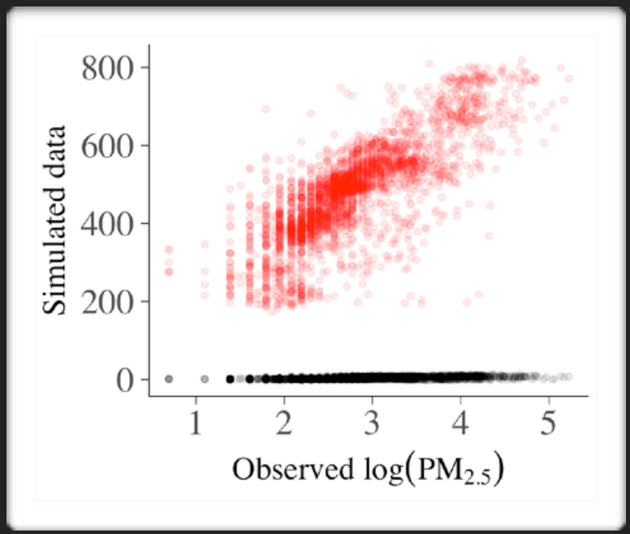
- Open source probabilistic **programming language**, **inference algorithms**
- Stan **program**
  - declares data and (constrained) parameter variables
  - defines log posterior (or penalized likelihood)
- Stan **inference**
  - MCMC for full Bayes
  - VB for approximate Bayes
  - Optimization for (penalized) MLE
- Stan **ecosystem**
  - lang, math library (C++)
  - interfaces and tools (R, Python, many more)
  - documentation (example model repo, user guide & reference manual, case studies, R package vignettes)
  - online community (Stan Forums on Discourse)

Visualization in Bayesian workflow

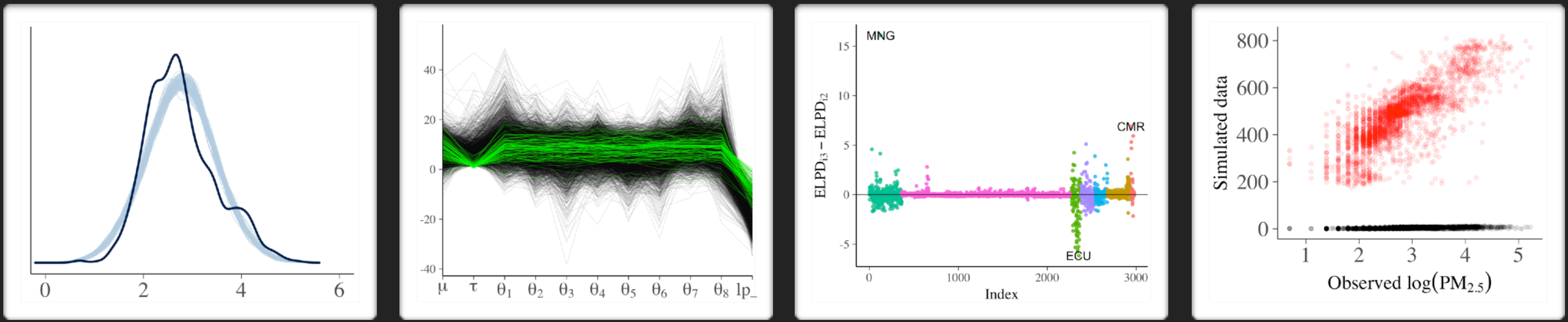








Visualization in Bayesian workflow



Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2018).
Visualization in Bayesian workflow.
Journal of the Royal Statistical Society Series A, accepted for publication.
arXiv preprint: arxiv.org/abs/1709.01449
Code: github.com/jgabry/bayes-vis-paper

Workflow

Bayesian data analysis

Workflow

Bayesian data analysis

- Exploratory data analysis

Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking

Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking
- Model fitting and algorithm diagnostics

Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking
- Model fitting and algorithm diagnostics
- *Posterior* predictive checking

Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking
- Model fitting and algorithm diagnostics
- *Posterior* predictive checking
- Model comparison (e.g., via cross-validation)

Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking
- Model fitting and algorithm diagnostics
- *Posterior* predictive checking
- Model comparison (e.g., via cross-validation)

Example

Example

Goal Estimate global PM2.5 concentration

Example

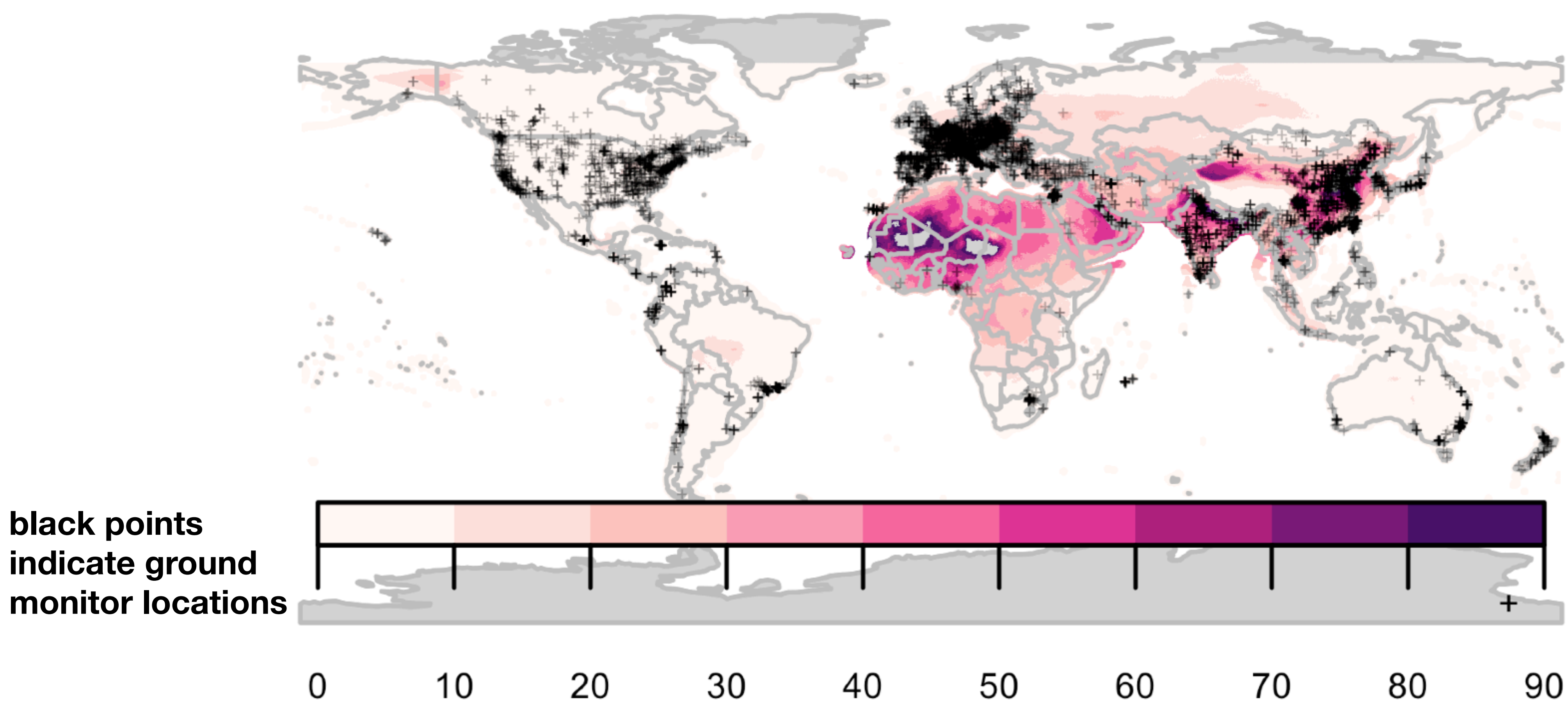
Goal Estimate global PM2.5 concentration

Problem Most data from noisy satellite measurements (ground monitor network provides sparse, heterogeneous coverage)

Example

Goal Estimate global PM2.5 concentration

Problem Most data from noisy satellite measurements (ground monitor network provides sparse, heterogeneous coverage)



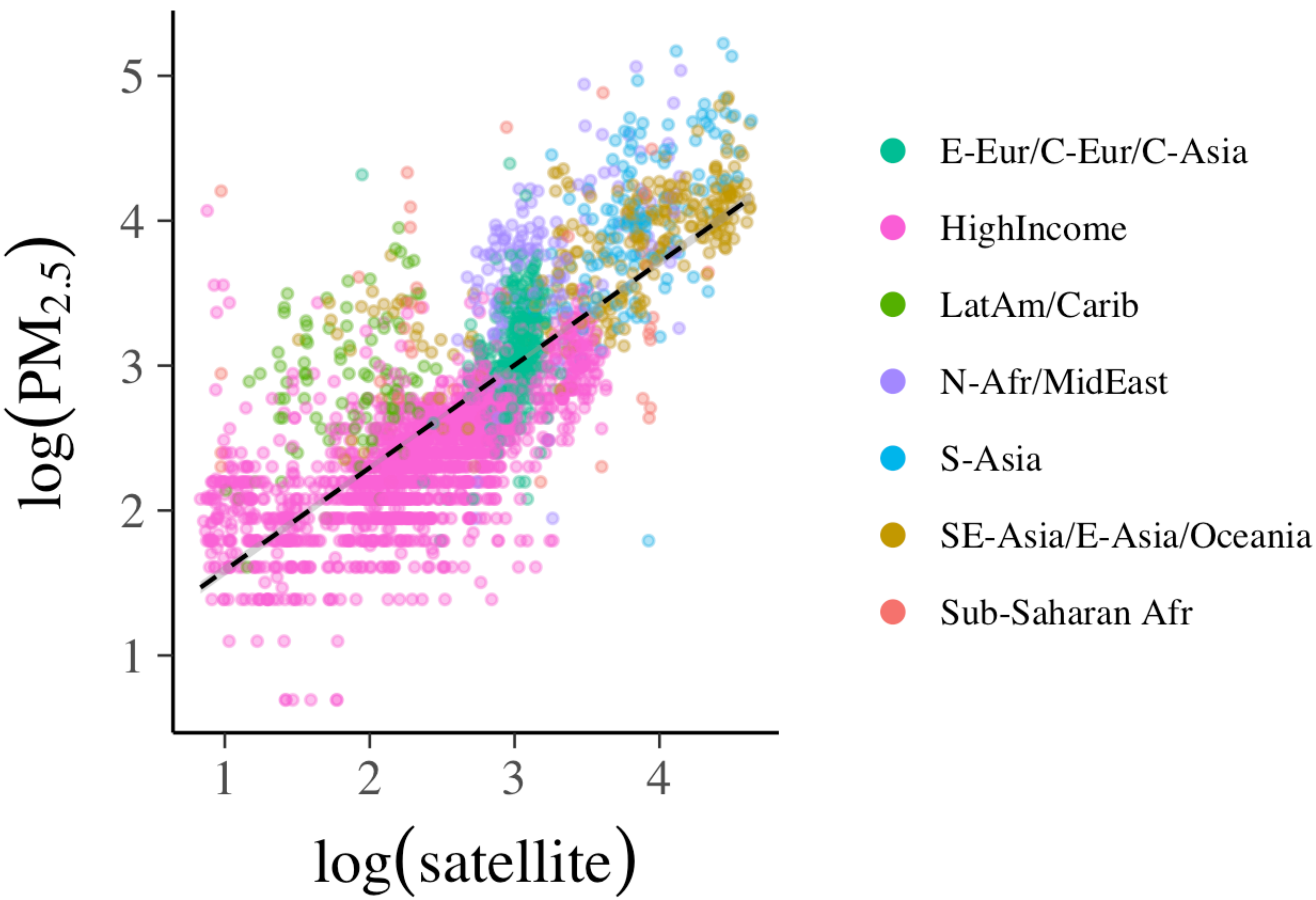
Satellite estimates of PM2.5 and ground monitor locations

Workflow

Bayesian data analysis

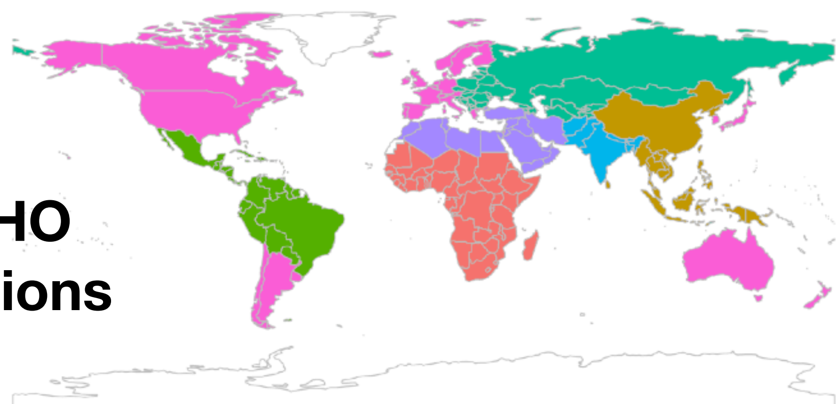
- **Exploratory data analysis**
- *Prior* predictive checking
- Model fitting and algorithm diagnostics
- *Posterior* predictive checking
- Model comparison (e.g., via cross-validation)

Exploratory data analysis
building a network of models

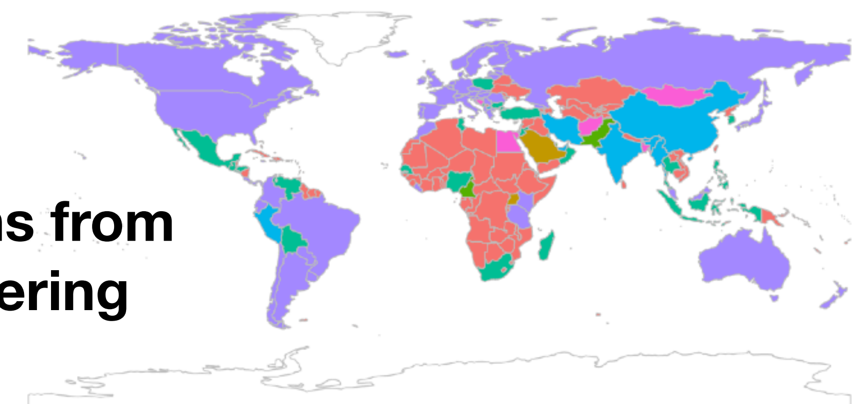


Exploratory data analysis
building a network of models

WHO Regions



Regions from clustering

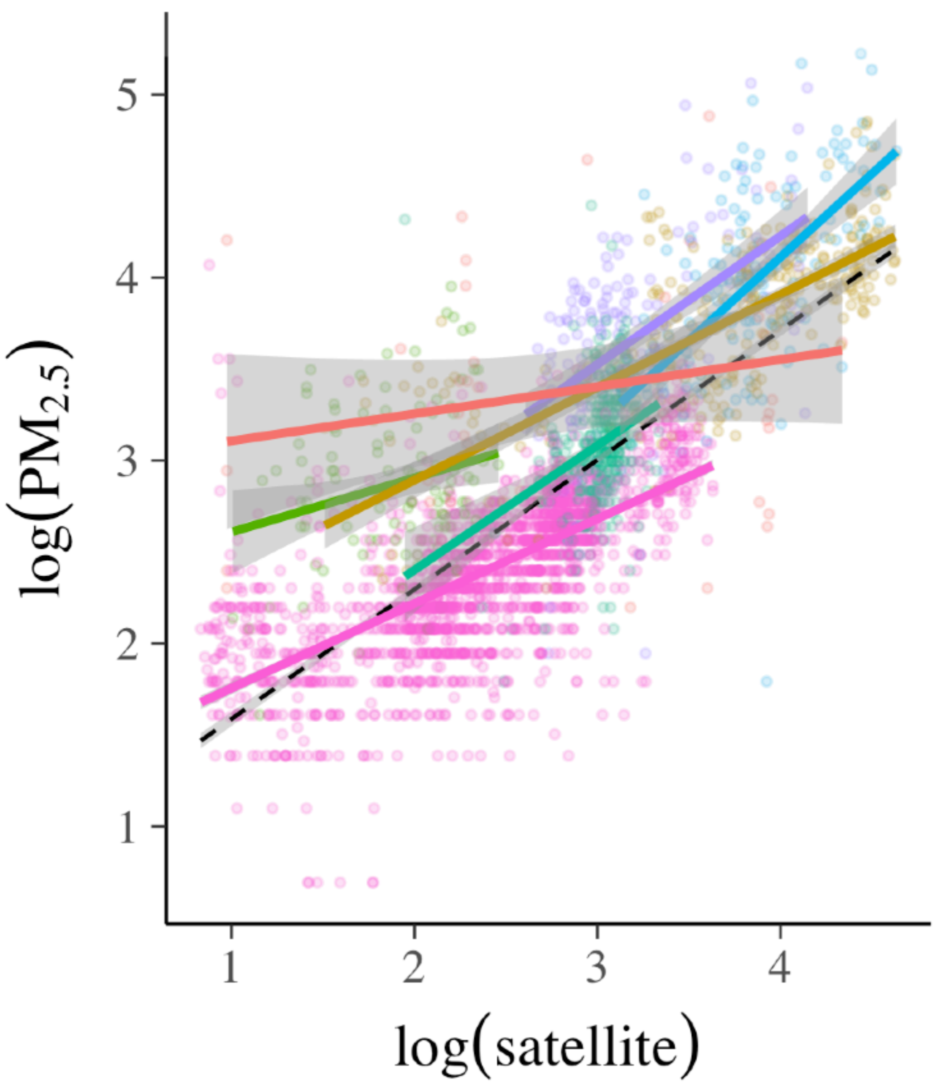


Exploratory data analysis

building a network of models

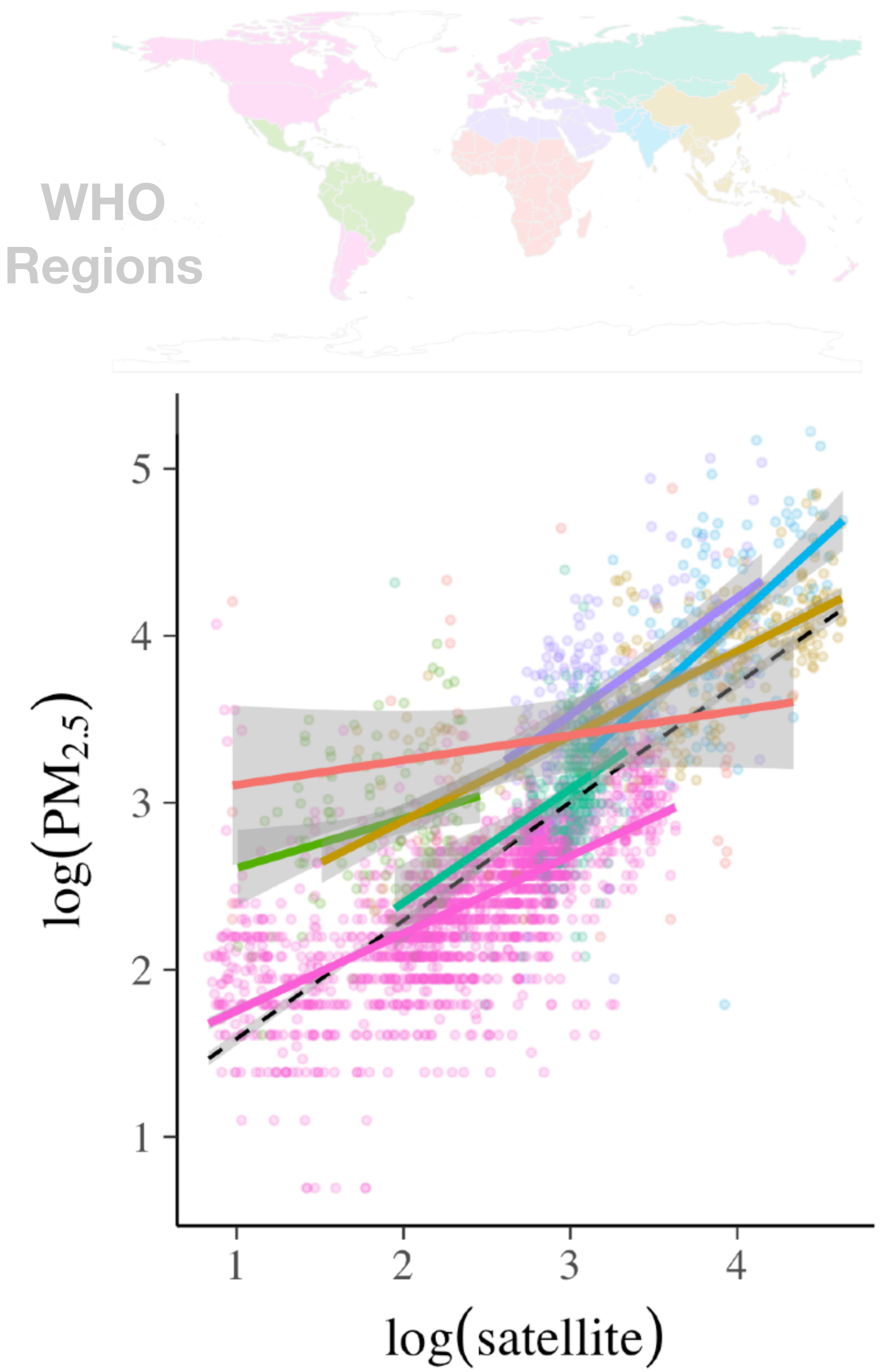
WHO Regions

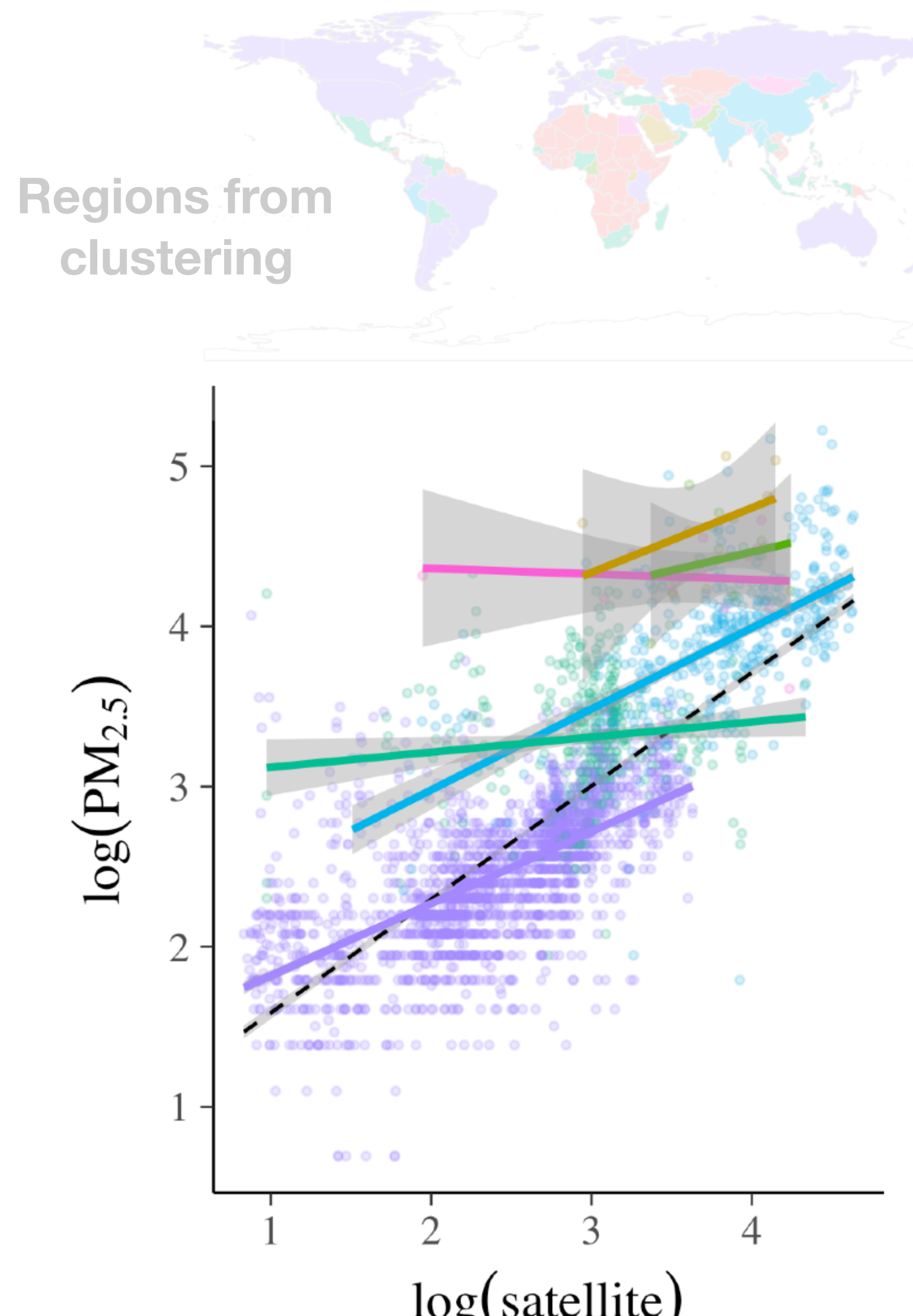
Regions from clustering



Exploratory data analysis

building a network of models





Exploratory data analysis

building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Model 1

$$\log(\mathbf{PM}_{2.5})_{n,j} = \alpha + \beta \log(\mathbf{sat})_{n,j} + \epsilon_{n,j}, \quad \epsilon_{n,j} \sim N(0,\sigma)$$

Exploratory data analysis

building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Model 1

$$\log(\textbf{PM}_{2.5})_{n,j} = \alpha + \beta \log(\textbf{sat})_{n,j} + \epsilon_{n,j}, \quad \epsilon_{n,j} \sim N(0,\sigma)$$

Exploratory data analysis
building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Model 1

$$\log(\mathbf{PM}_{2.5})_{n,j} = \alpha + \beta \log(\mathbf{sat})_{n,j} + \epsilon_{n,j}, \quad \epsilon_{n,j} \sim N(0,\sigma)$$

$$\log(\mathbf{PM}_{2.5})_{n,j} \sim N(\alpha + \beta \log(\mathbf{sat})_{n,j}, \sigma)$$

Exploratory data analysis

building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Models 2 and 3

Exploratory data analysis
building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Models 2 and 3

$$\log\left(\mathrm{PM}_{2.5,nj}\right) \sim N(\mu_{nj}, \sigma)$$

Exploratory data analysis
building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Models 2 and 3

$$\log\left(\mathrm{PM}_{2.5,nj}\right) \sim N(\mu_{nj}, \sigma)$$

$$\mu_{nj} = \alpha_0 + \alpha_j + (\beta_0 + \beta_j)\log\left(\mathrm{sat}_{nj}\right)$$

Exploratory data analysis

building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Models 2 and 3

$$\log\left(\mathrm{PM}_{2.5,nj}\right) \sim N(\mu_{nj}, \sigma)$$

$$\mu_{nj} = \boxed{\alpha_0 + \alpha_j} + \boxed{(\beta_0 + \beta_j)} \log\left(\mathrm{sat}_{nj}\right)$$

Exploratory data analysis

building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Models 2 and 3

$$\log\left(\mathrm{PM}_{2.5,nj}\right) \sim N(\mu_{nj}, \sigma)$$

$$\mu_{nj} = \boxed{\alpha_0 + \alpha_j} + \boxed{(\beta_0 + \beta_j)} \log\left(\mathrm{sat}_{nj}\right)$$

$$\alpha_j \sim N(0, \tau_\alpha) \quad \beta_j \sim N(0, \tau_\beta)$$

Exploratory data analysis
building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Models 2 and 3

$$\log\left(\mathrm{PM}_{2.5,nj}\right) \sim N(\mu_{nj}, \sigma)$$

$$\mu_{nj} = \boxed{\alpha_0 + \alpha_j} + \boxed{(\beta_0 + \beta_j)} \log\left(\mathrm{sat}_{nj}\right)$$

$$\boxed{\alpha_j \sim N(0, \tau_\alpha)} \quad \beta_j \sim N(0, \tau_\beta)$$

Exploratory data analysis

building a network of models

For measurements $n = 1, \ldots, N$

and regions $j = 1, \ldots, J$

Models 2 and 3

$$\log\left(\mathrm{PM}_{2.5,nj}\right) \sim N(\mu_{nj}, \sigma)$$

$$\mu_{nj} = \boxed{\alpha_0 + \alpha_j} + \boxed{(\beta_0 + \beta_j)} \log\left(\mathrm{sat}_{nj}\right)$$

$$\boxed{\alpha_j \sim N(0, \tau_\alpha)} \quad \boxed{\beta_j \sim N(0, \tau_\beta)}$$

Workflow

Bayesian data analysis

- Exploratory data analysis
- ***Prior* predictive checking**
- Model fitting and algorithm diagnostics
- *Posterior* predictive checking
- Model comparison (e.g., via cross-validation)

What is a Bayesian model?

What is a Bayesian model?

- Building a Bayesian model forces us to build a model for how the data is generated

What is a Bayesian model?

- Building a Bayesian model forces us to build a model for how the data is generated
- We often think of this as specifying a **prior** and a **likelihood**, as if these are two separate things

What is a Bayesian model?

- Building a Bayesian model forces us to build a model for how the data is generated
- We often think of this as specifying a **prior** and a **likelihood**, as if these are two separate things
- They are not!

What is a Bayesian model?

- Building a Bayesian model forces us to build a model for how the data is generated
- We often think of this as specifying a **prior** and a **likelihood**, as if these are two separate things
- They are not!

Gelman, A., Simpson, D., and Betancourt, M. (2017).
The prior can often only be understood in the context of the likelihood.
arXiv preprint: arxiv.org/abs/1708.07487

A Bayesian modeler commits to an a priori *joint distribution*

Likelihood x Prior

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \mathbf{y})p(\mathbf{y})$$

Posterior x Marginal Likelihood

Data
(observed)

Parameters
(unobserved)

What is the problem with “vague” priors?

What is the problem with "vague" priors?

- If we use an *improper* prior, then we do not specify a joint model for our data and parameters

What is the problem with "vague" priors?

- If we use an *improper* prior, then we do not specify a joint model for our data and parameters
- More importantly, we do not specify a data generating mechanism ***p(y)***

What is the problem with “vague” priors?

- If we use an *improper* prior, then we do not specify a joint model for our data and parameters

- More importantly, we do not specify a data generating mechanism ***p(y)***

- By construction, these priors do not regularize inferences, which is quite often a bad idea

What is the problem with “vague” priors?

- If we use an *improper* prior, then we do not specify a joint model for our data and parameters
- More importantly, we do not specify a data generating mechanism ***p(y)***
- By construction, these priors do not regularize inferences, which is quite often a bad idea
- Proper but diffuse is better than improper but is still often problematic

Generative models

Generative models

- If we disallow improper priors, then Bayesian modeling is generative

Generative models

- If we disallow improper priors, then Bayesian modeling is generative

- In particular, we have a simple way to simulate from $p(y)$:

Generative models

- If we disallow improper priors, then Bayesian modeling is generative

- In particular, we have a simple way to simulate from *p(y)*:

$$\theta^\star \sim p(\theta)$$

Generative models

- If we disallow improper priors, then Bayesian modeling is generative

- In particular, we have a simple way to simulate from *p(y)*:

$$
\begin{array}{c}
\theta^\star \sim p(\theta) \\
\downarrow \\
y^\star \sim p(y|\theta^\star)
\end{array}
$$

Generative models

- If we disallow improper priors, then Bayesian modeling is generative

- In particular, we have a simple way to simulate from *p(y)*:

$$\theta^\star \sim p(\theta)$$
$$\downarrow$$
$$y^\star \sim p(y|\theta^\star)$$

$$\Longleftrightarrow$$

$$y^\star \sim p(y)$$

Prior predictive checking:

fake data is almost as useful as real data

What do vague/non-informative priors imply about the data our model can generate?

Prior predictive checking:
fake data is almost as useful as real data

What do vague/non-informative priors imply about the data our model can generate?

$$\alpha_0 \sim N(0, 100)$$
$$\beta_0 \sim N(0, 100)$$
$$\tau_\alpha^2 \sim \mathrm{InvGamma}(1, 100)$$
$$\tau_\beta^2 \sim \mathrm{InvGamma}(1, 100)$$

Prior predictive checking:

fake data is almost as useful as real data

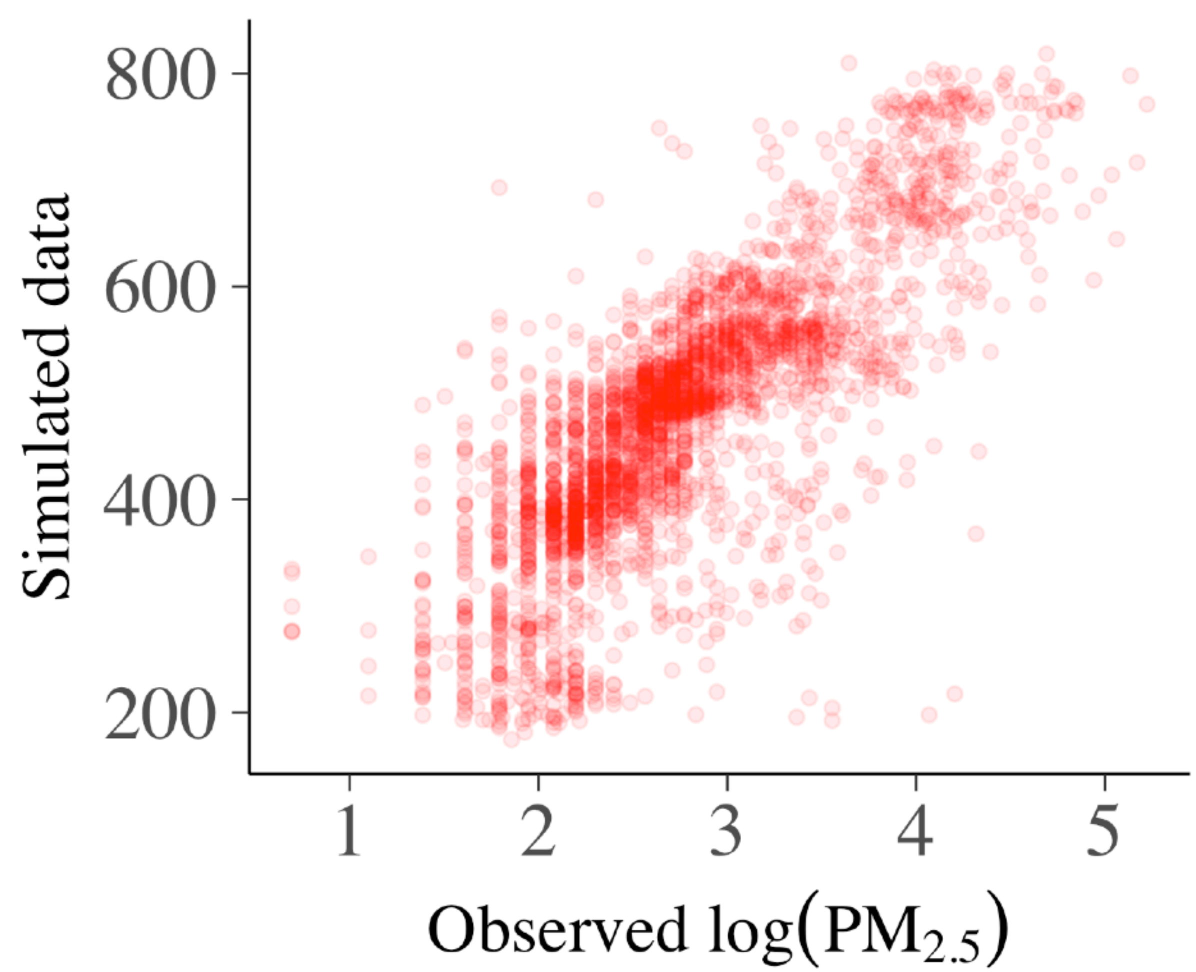
What do vague/non-informative priors imply about the data our model can generate?

$$\alpha_0 \sim N(0, 100)$$

$$\beta_0 \sim N(0, 100)$$

$$\tau_\alpha^2 \sim \text{InvGamma}(1, 100)$$

$$\tau_\beta^2 \sim \text{InvGamma}(1, 100)$$

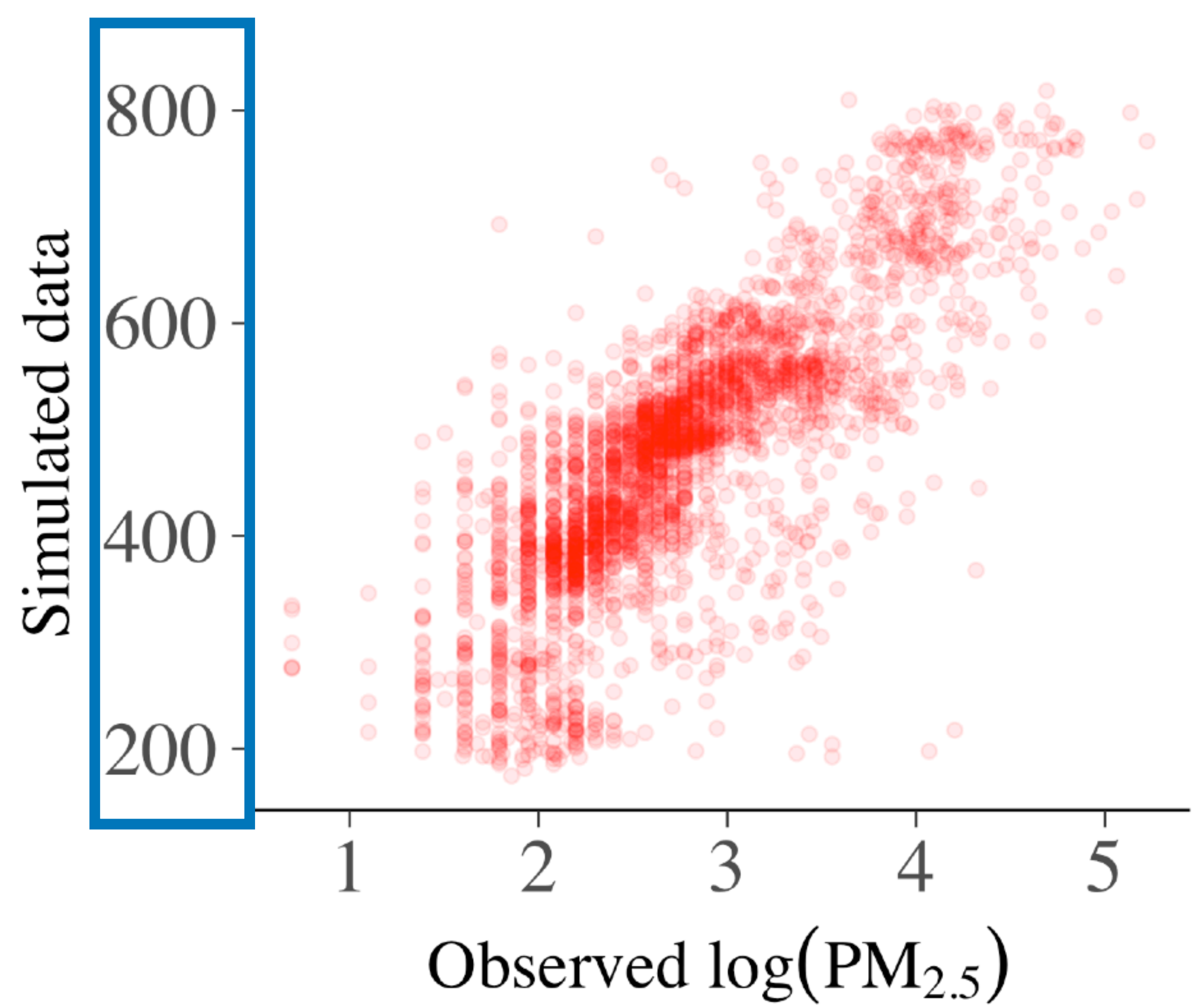


Prior predictive checking:

fake data is almost as useful as real data

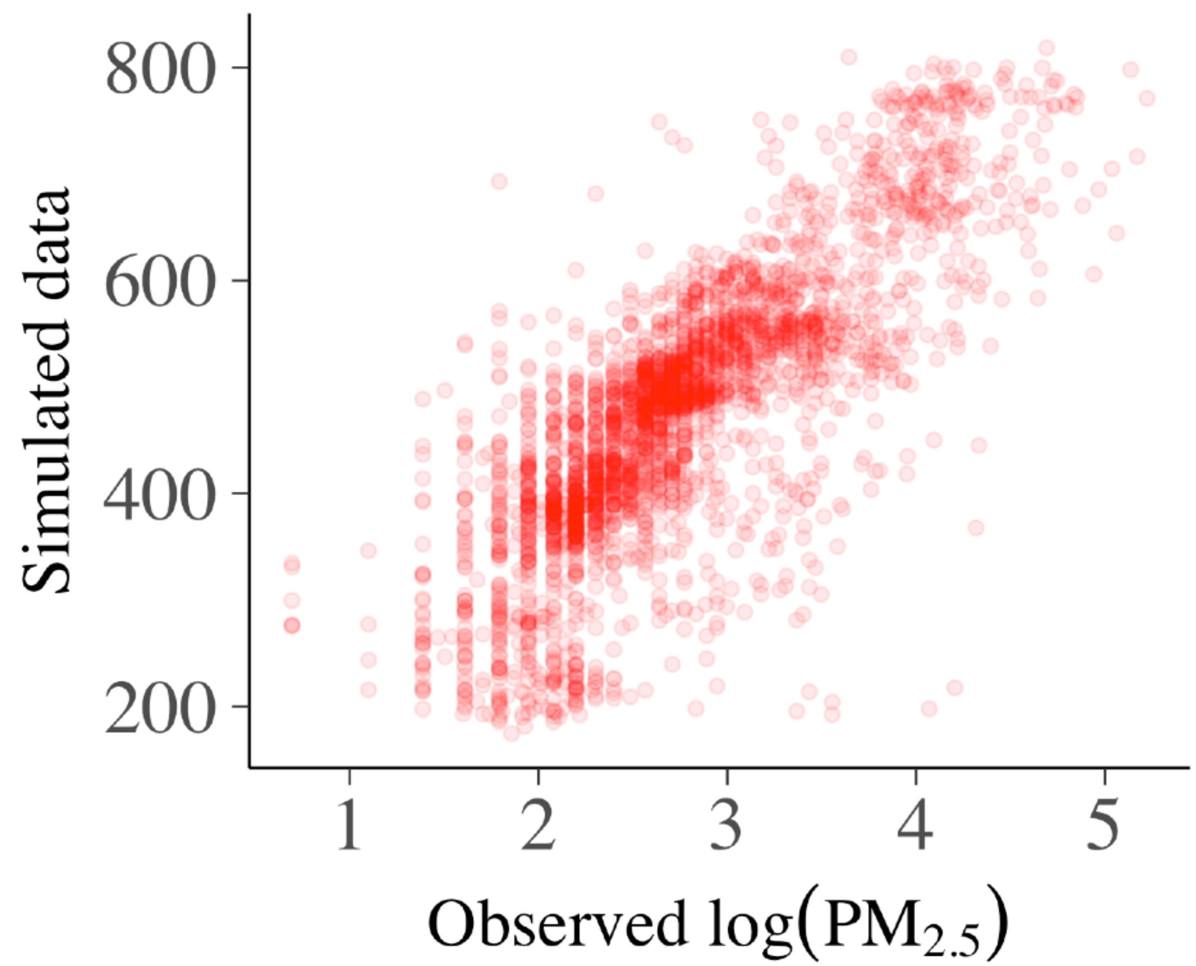
What do vague/non-informative priors imply about the data our model can generate?

$$\alpha_0 \sim N(0, 100)$$
$$\beta_0 \sim N(0, 100)$$
$$\tau_\alpha^2 \sim \mathrm{InvGamma}(1, 100)$$
$$\tau_\beta^2 \sim \mathrm{InvGamma}(1, 100)$$



Prior predictive checking:
fake data is almost as useful as real data

- The prior model is **two orders of magnitude** off the real data
- Two orders of magnitude **on the log scale!**



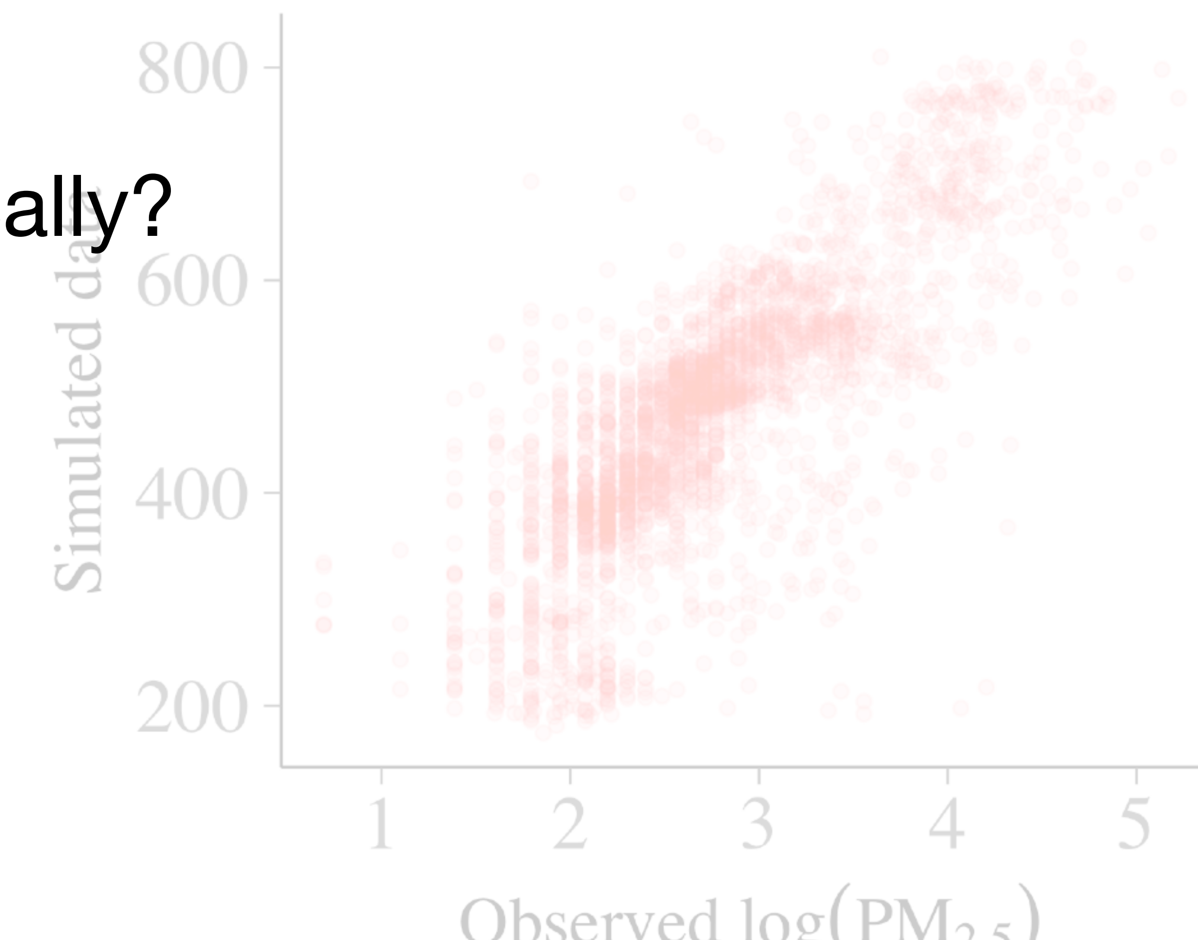
Prior predictive checking:

fake data is almost as useful as real data

- The prior model is **two orders of magnitude** off the real data
- Two orders of magnitude **on the log scale!**

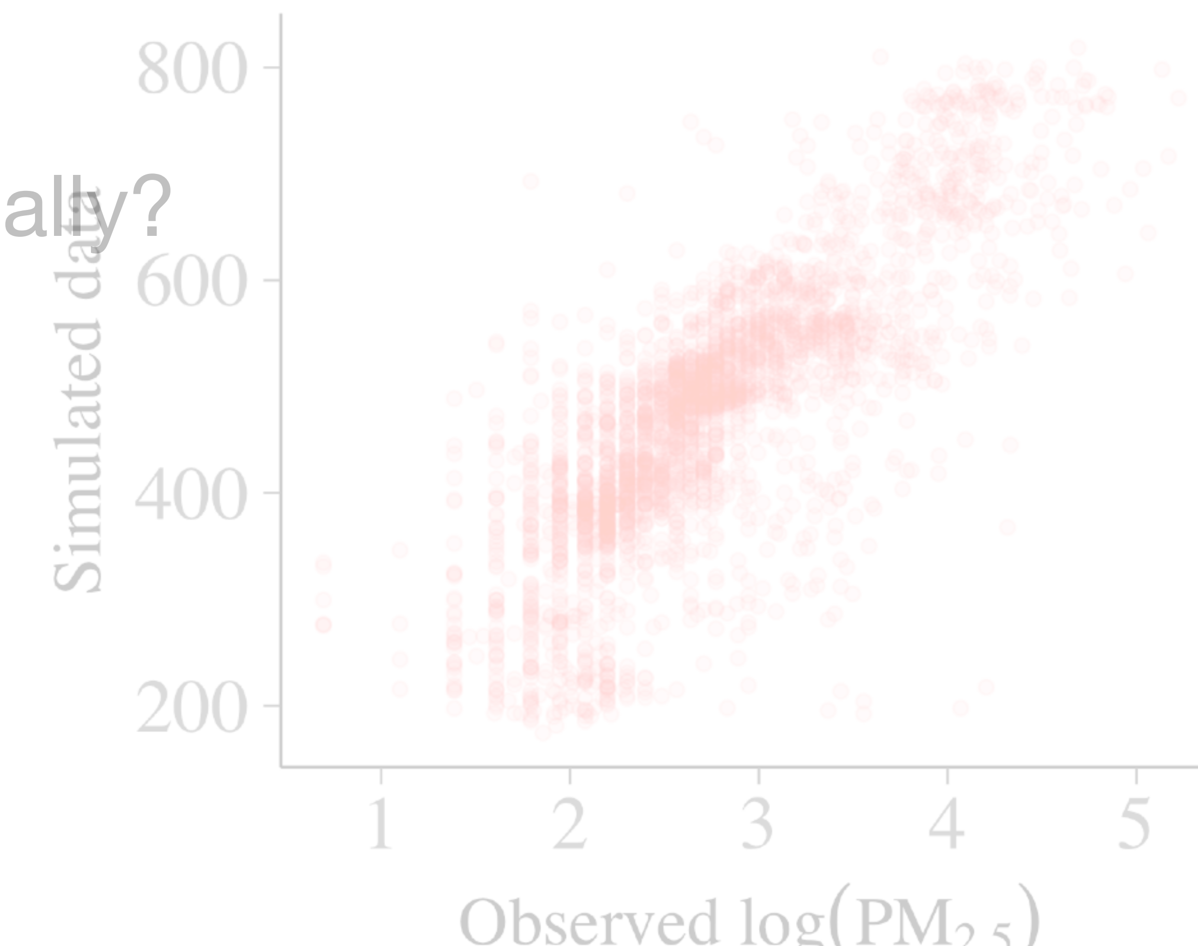
Prior predictive checking:
fake data is almost as useful as real data

- The prior model is **two orders of magnitude** off the real data
- Two orders of magnitude **on the log scale!**
- What does this mean practically?



Prior predictive checking:
fake data is almost as useful as real data

- The prior model is **two orders of magnitude** off the real data
- Two orders of magnitude **on the log scale!**
- What does this mean practically?
- The data will have to overcome the prior…



Prior predictive checking:
fake data is almost as useful as real data

What are better priors for the global intercept and slope and the hierarchical scale parameters?

Prior predictive checking:

fake data is almost as useful as real data

What are better priors for the global intercept and slope and the hierarchical scale parameters?

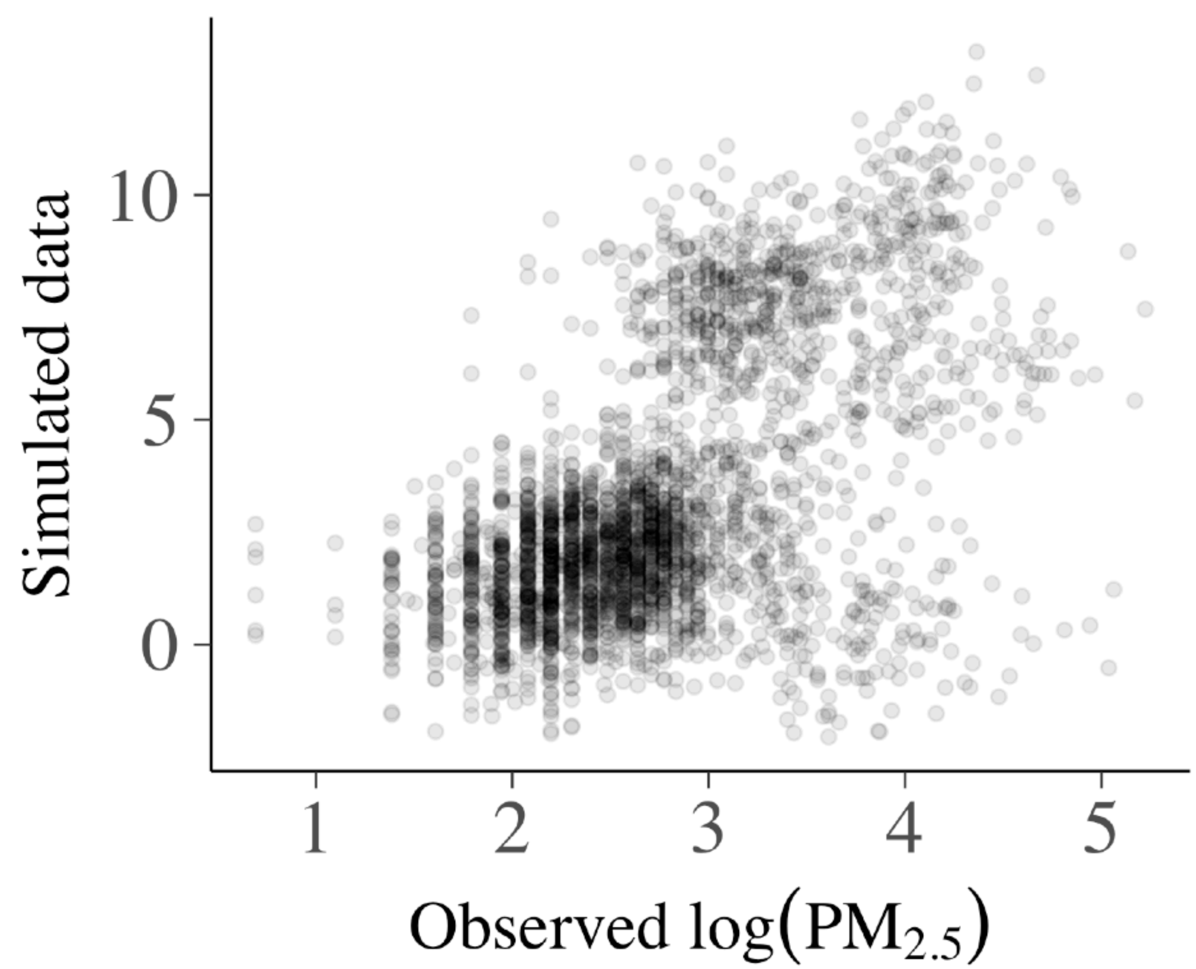
$$
\begin{aligned}
\alpha_0 &\sim N(0,1) \\
\beta_0 &\sim N(1,1) \\
\tau_\alpha &\sim N_+(0,1) \\
\tau_\beta &\sim N_+(0,1)
\end{aligned}
$$

Prior predictive checking:

fake data is almost as useful as real data

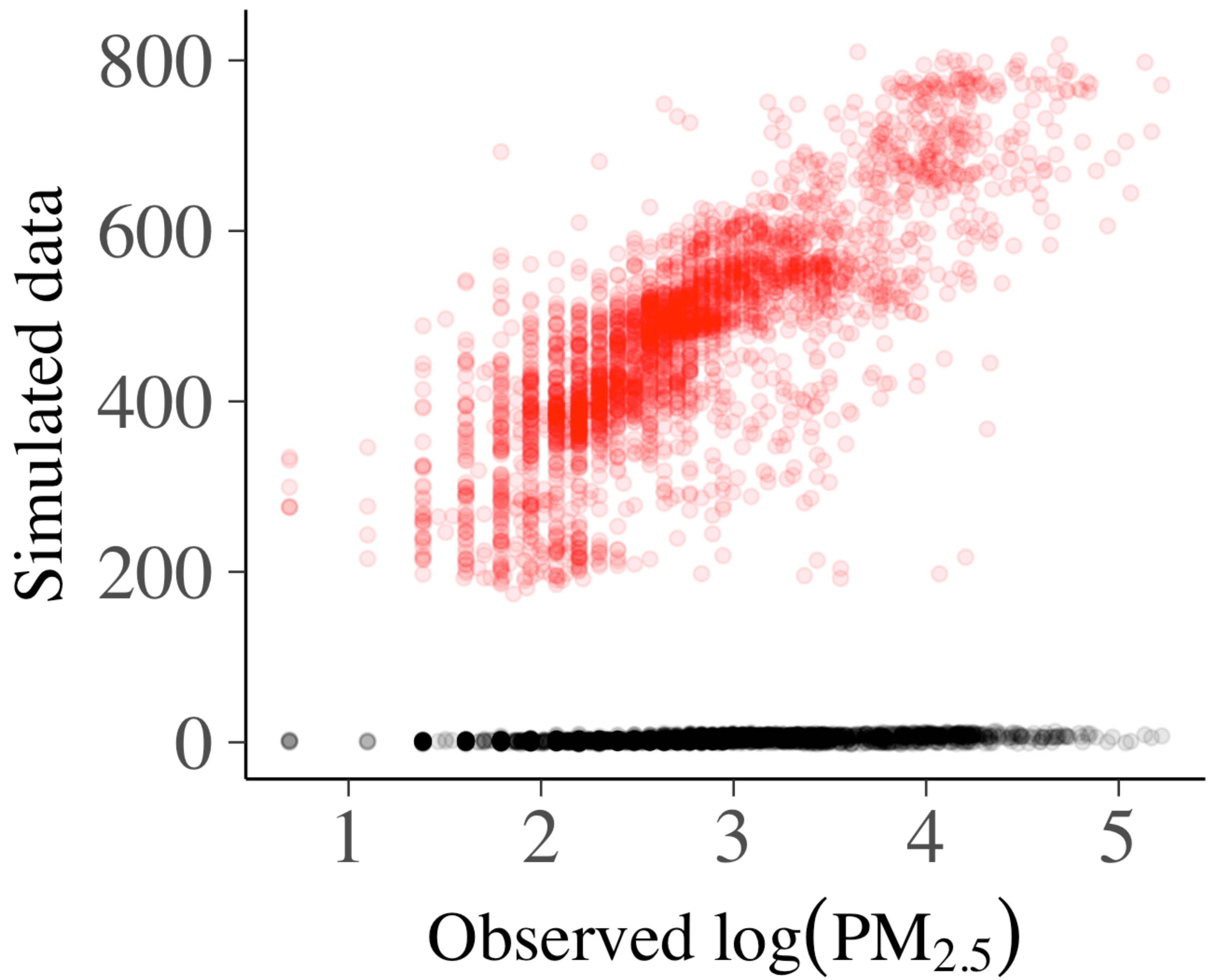
What are better priors for the global intercept and slope and the hierarchical scale parameters?

$$\alpha_0 \sim N(0,1)$$
$$\beta_0 \sim N(1,1)$$
$$\tau_\alpha \sim N_+(0,1)$$
$$\tau_\beta \sim N_+(0,1)$$



Prior predictive checking:

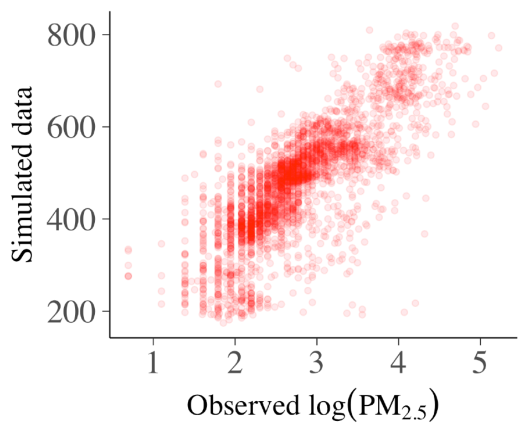
fake data is almost as useful as real data

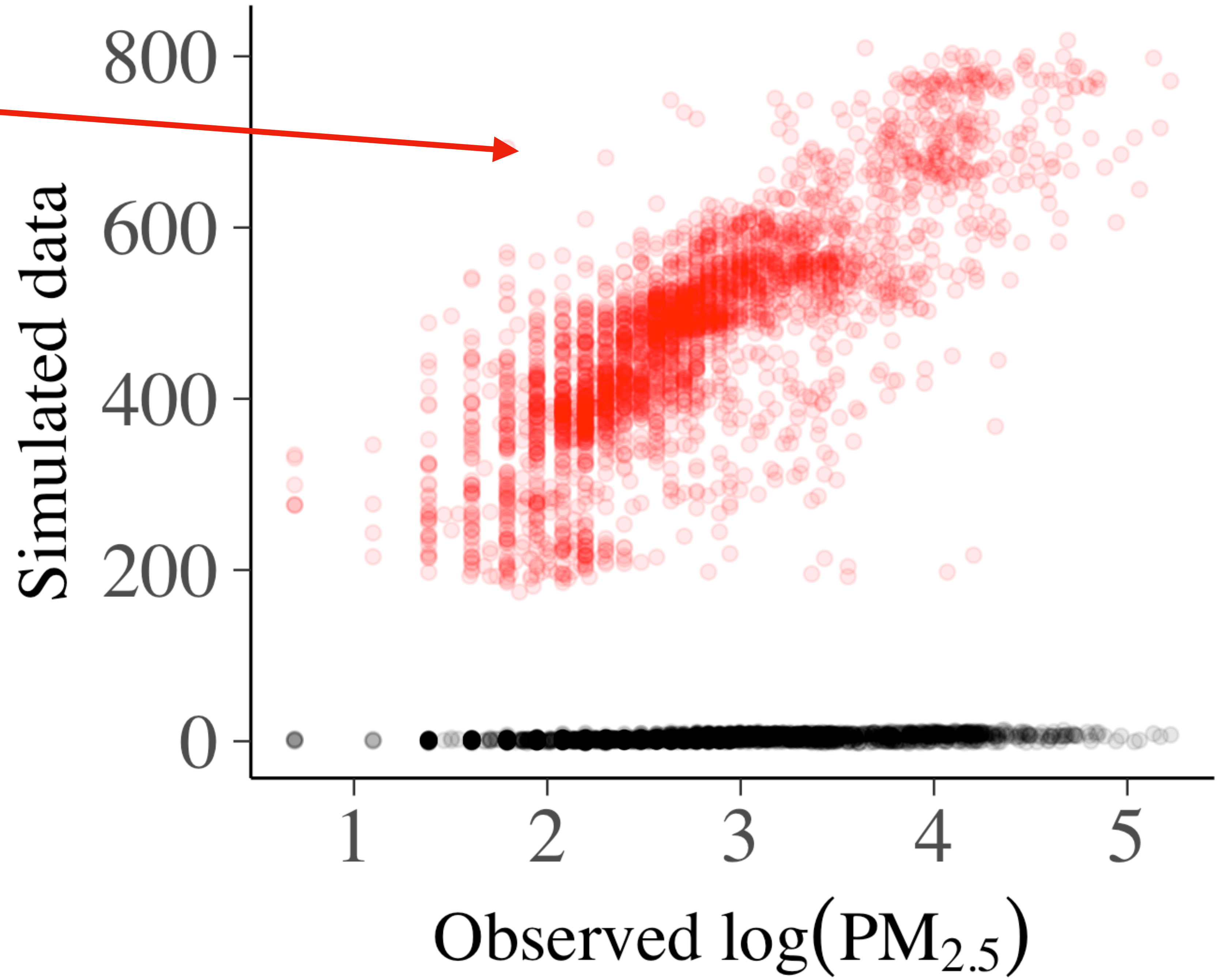


Prior predictive checking:

fake data is almost as useful as real data

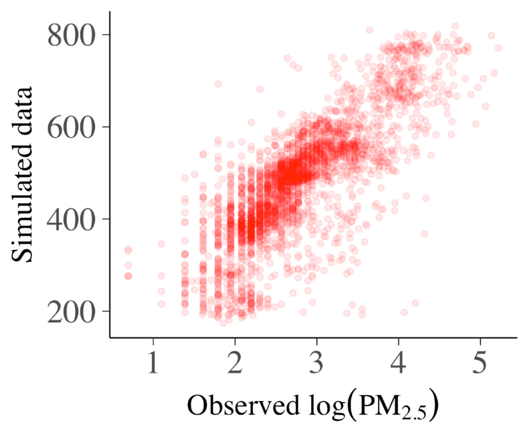
Non-informative



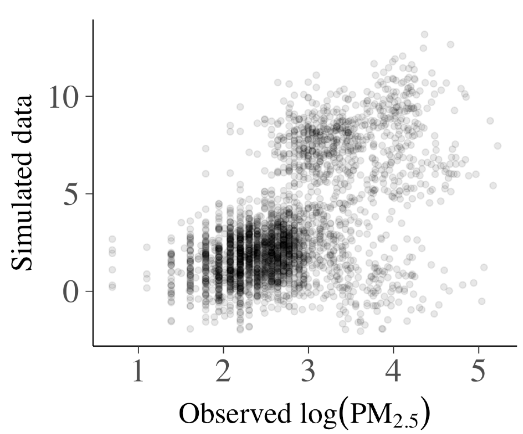


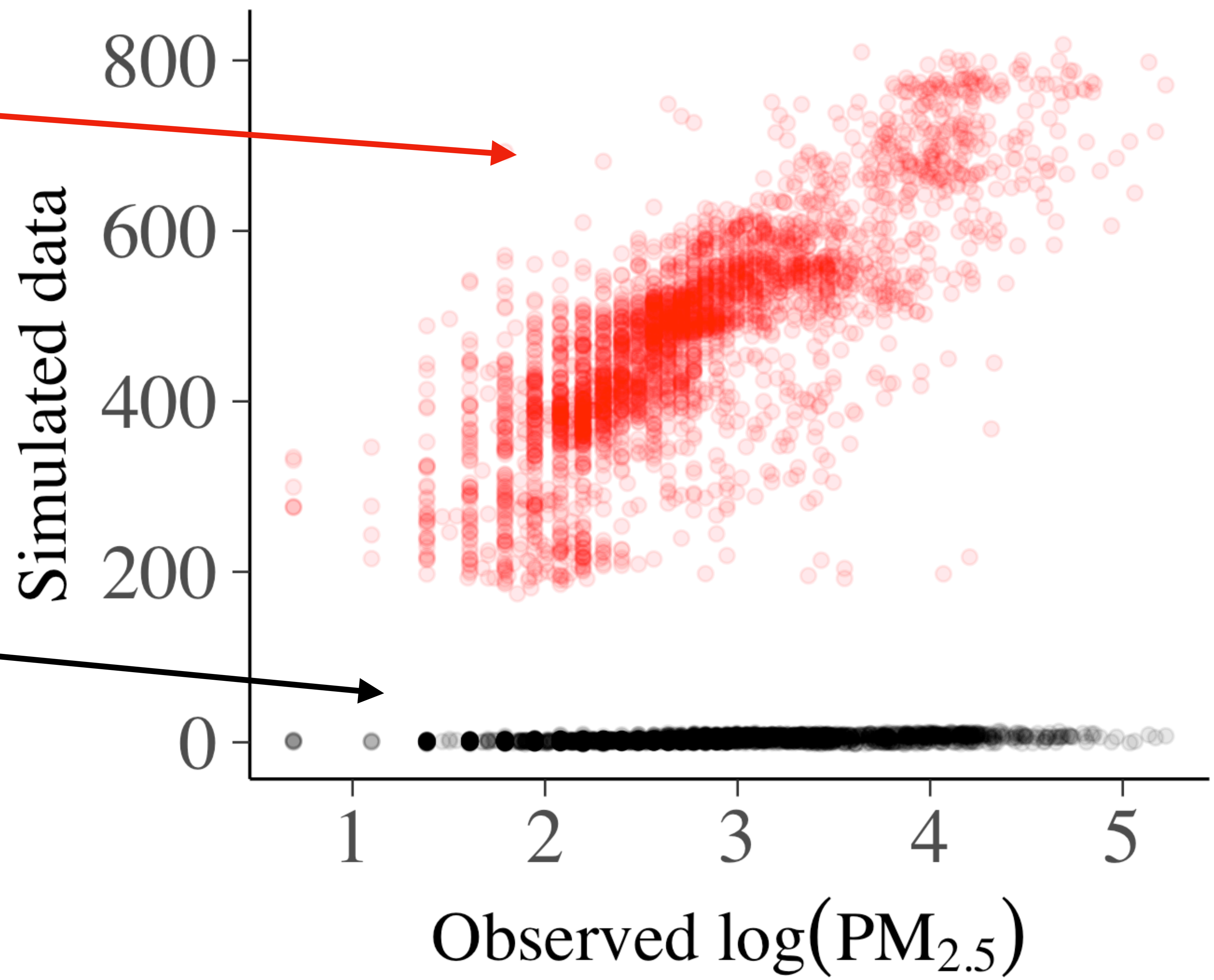
Prior predictive checking:
fake data is almost as useful as real data

Non-informative



Weakly informative





Markov Chain Monte Carlo

MCMC, HMC, NUTS, other fancy acronyms…

Chi Feng's MCMC demos: http://chi-feng.github.io/mcmc-demo/

Betancourt, M. (2017).
A conceptual introduction to Hamiltonian Monte Carlo.
arXiv preprint:
arxiv.org/abs/1701.02434

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2018).
Visualization in Bayesian workflow.
Journal of the Royal Statistical Society Series A, accepted for publication.
arxiv.org/abs/1709.01449 | github.com/jgabry/bayes-vis-paper

Workflow

Bayesian data analysis

- Exploratory data analysis
- Prior predictive checking
- **Model fitting and algorithm diagnostics**
- *Posterior* predictive checking
- Model comparison (e.g., via cross-validation)

Workflow

Bayesian data analysis

- Exploratory data analysis
- Prior predictive checking
- **Model fitting and algorithm diagnostics**

13:00 - 14:30 **Hierarchical models, part 1** Ben Goodrich

13:00 - 14:30 **Ordinary differential equation (ODE) models in Stan** Daniel Lee

14:30 - 15:00 **Coffee break**

15:00 - 16:30 **Hierarchical models, part 2** Ben Goodrich

Workflow

Bayesian data analysis

- Exploratory data analysis
- Prior predictive checking
- **Model fitting and algorithm diagnostics**

13:00 - 14:30 **Hierarchical models, part 1** Ben Goodrich

13:00 - 14:30 **Ordinary differential equation (ODE) models in Stan** Daniel Lee

14:30 - 15:00 **Coffee break**

15:00 - 16:30 **Hierarchical models, part 2** Ben Goodrich

Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking
- Model fitting and algorithm diagnostics
- ***Posterior* predictive checking**
- Model comparison (e.g., via cross-validation)

Posterior predictive checking
visual model evaluation

The *posterior predictive distribution* is the average data generation process over the entire model

Posterior predictive checking

visual model evaluation

The *posterior predictive distribution* is the average data generation process over the entire model

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)\, p(\theta|y)\, d\theta$$

Posterior predictive checking

visual model evaluation

Posterior predictive checking
visual model evaluation

- Misfitting and overfitting both manifest as tension between measurements and predictive distributions

Posterior predictive checking

visual model evaluation

- Misfitting and overfitting both manifest as tension between measurements and predictive distributions
- Graphical posterior predictive checks visually compare the observed data to the predictive distribution

Posterior predictive checking
visual model evaluation

- Misfitting and overfitting both manifest as tension between measurements and predictive distributions

- Graphical posterior predictive checks visually compare the observed data to the predictive distribution

$$\theta^\star \sim p(\theta|y)$$

Posterior predictive checking
visual model evaluation

- Misfitting and overfitting both manifest as tension between measurements and predictive distributions
- Graphical posterior predictive checks visually compare the observed data to the predictive distribution

$$\theta^\star \sim p(\theta|y)$$

$$\downarrow$$

$$\tilde{y} \sim p(y|\theta^\star)$$

Posterior predictive checking
visual model evaluation

- Misfitting and overfitting both manifest as tension between measurements and predictive distributions
- Graphical posterior predictive checks visually compare the observed data to the predictive distribution

$$\theta^\star \sim p(\theta|y)$$

$$\downarrow$$

$$\tilde{y} \sim p(y|\theta^\star)$$

$$\Longleftrightarrow$$

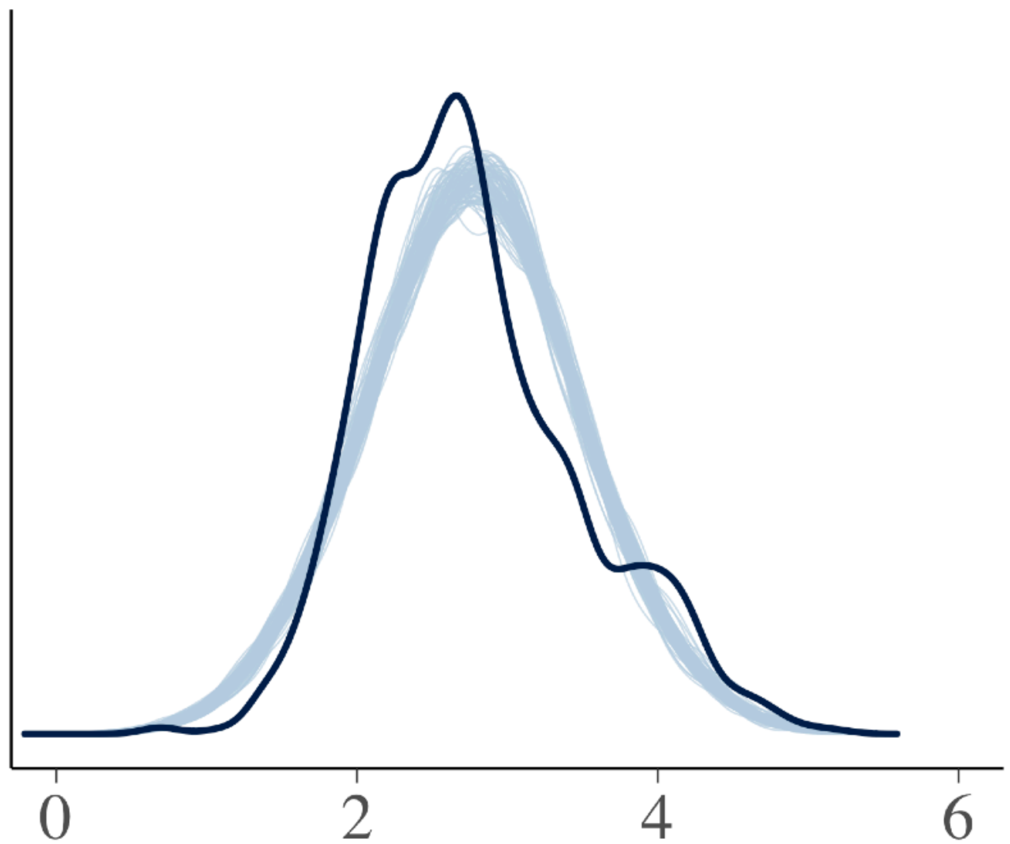
$$\tilde{y} \sim p(\tilde{y}|y)$$

Posterior predictive checking

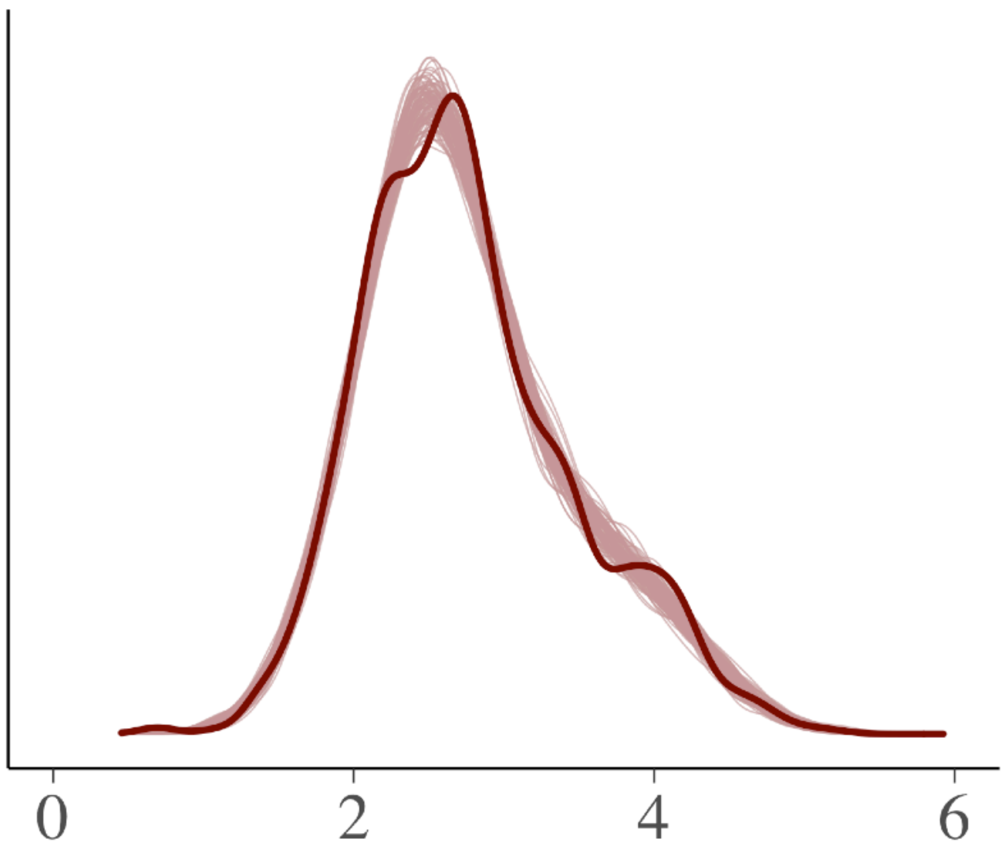
visual model evaluation

Observed data vs posterior predictive simulations

Model 1 (single level)



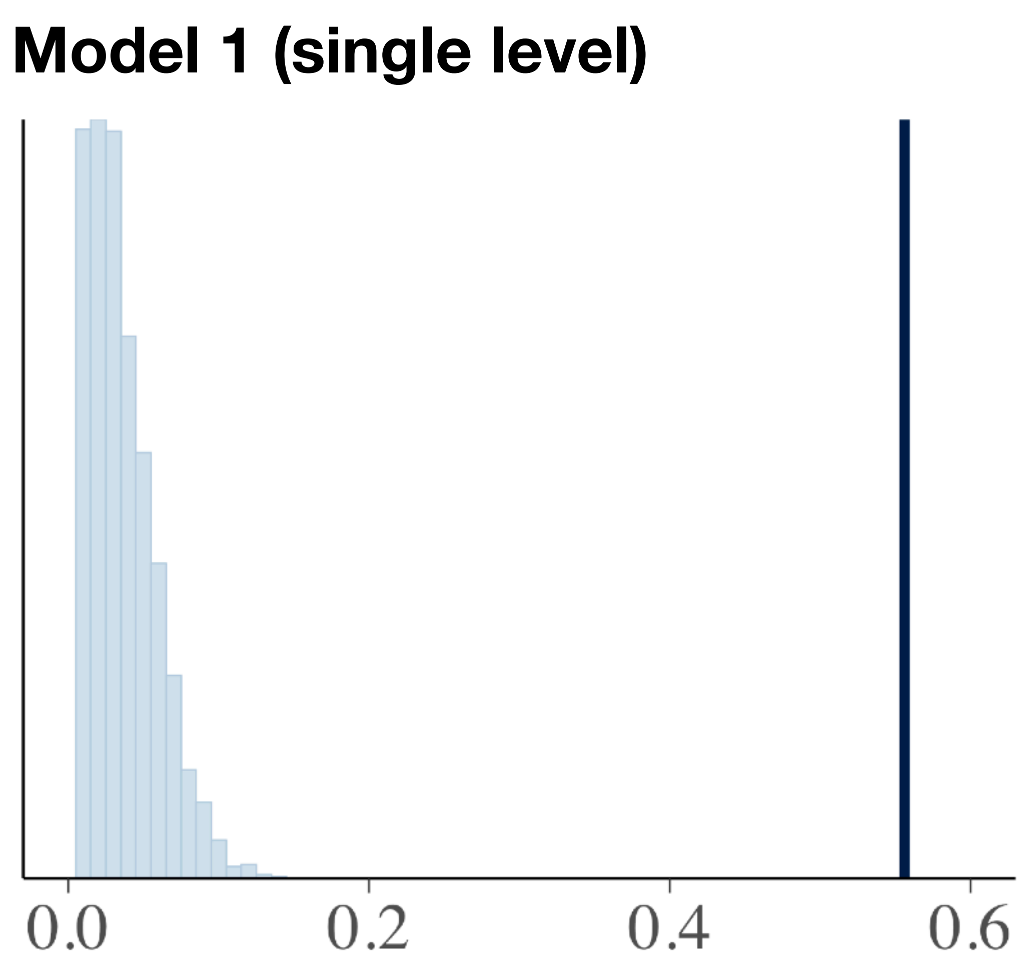
Model 3 (multilevel)

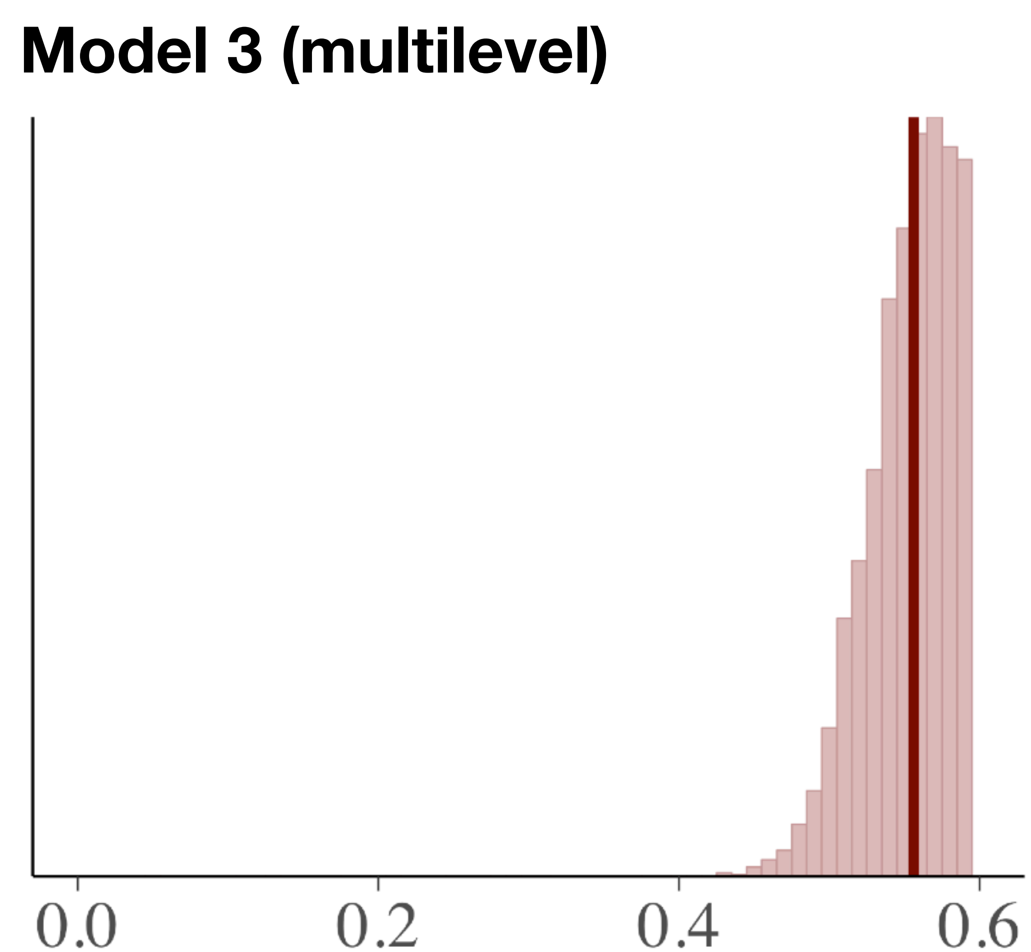


Posterior predictive checking

visual model evaluation

Observed statistics vs posterior predictive statistics

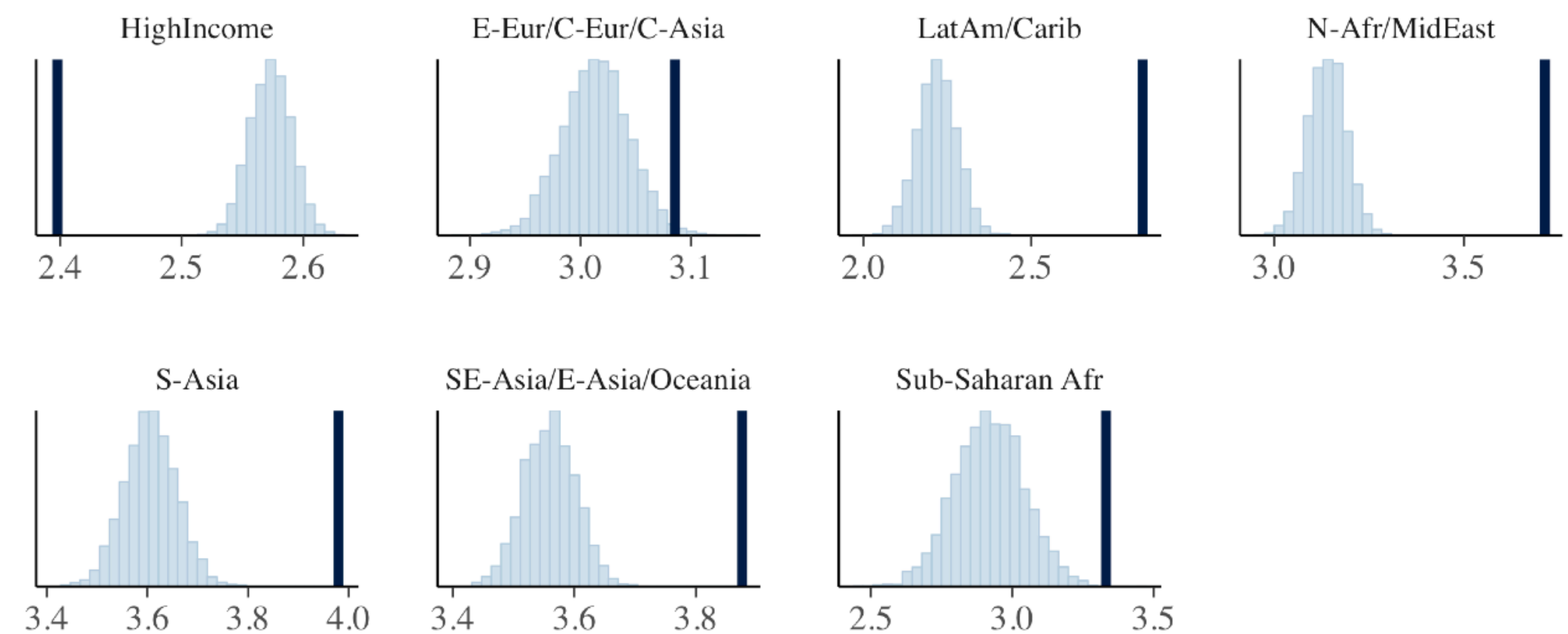




$$T(y) = \mathrm{skew}(y)$$

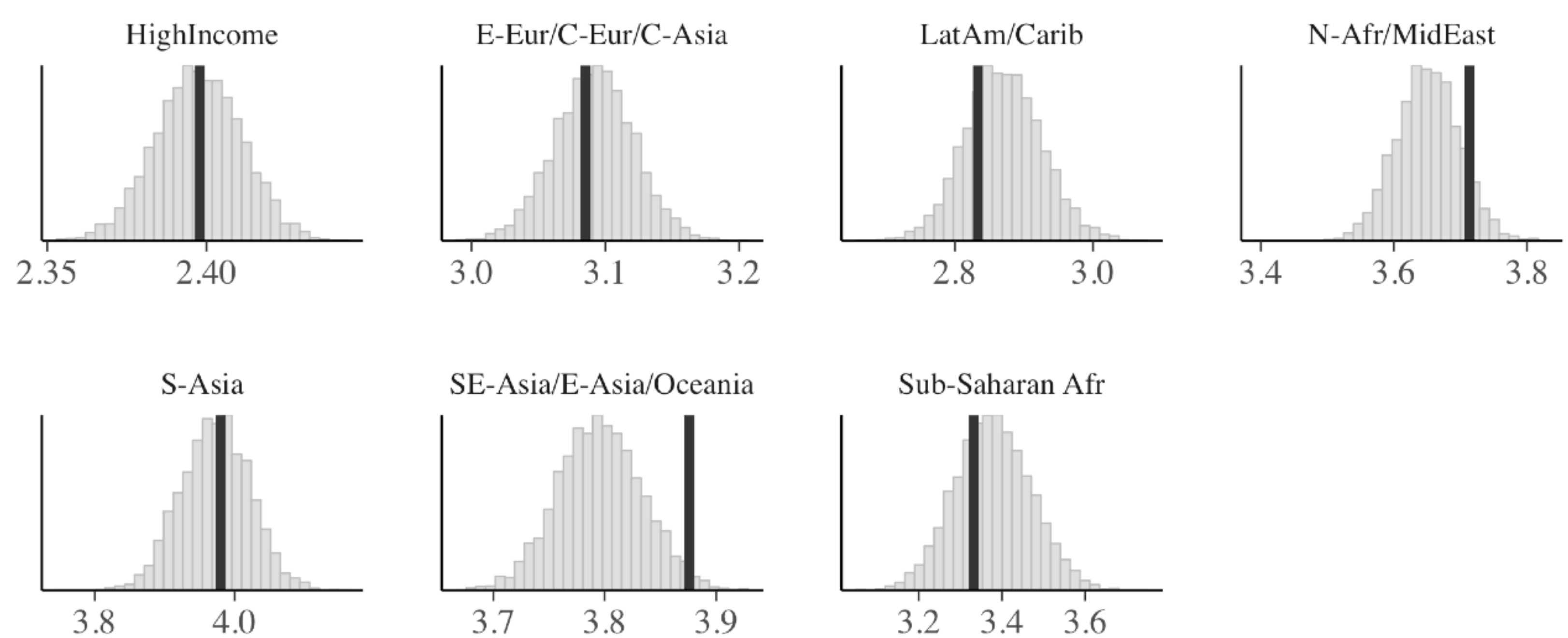
Posterior predictive checking:
visual model evaluation

Model 1 (single level)



$$T(y) = \mathrm{med}(y|\mathrm{region})$$

Model 2 (multilevel)



Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking
- Model fitting and algorithm diagnostics
- *Posterior* predictive checking
- **Model comparison (e.g., via cross-validation)**

Workflow

Bayesian data analysis

- Exploratory data analysis
- *Prior* predictive checking
- Model fitting and algorithm diagnostics
- *Posterior* predictive checking
- **Model comparison (e.g., via cross-validation)**

16:30 - 17:30 **Model assessment and selection** Aki Vehtari