

# Bayesian Statistics - Homework 2

Gabry et al., *Visualization in Bayesian workflow* (2018)

Roberto Corti

May 15, 2021

# Visualization in Bayesian workflow

A pipeline for bayesian modelling

Visualization is an invaluable way of justifying and criticize a statistical model.

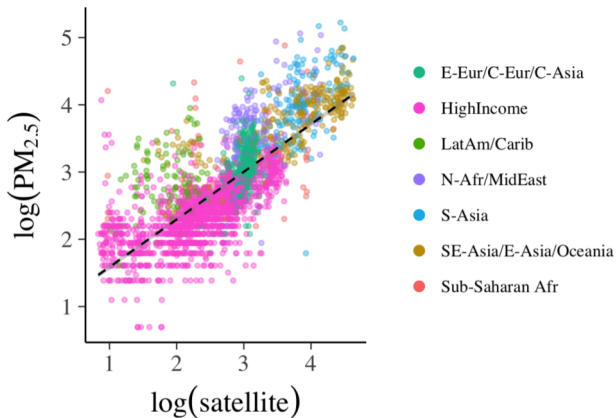
Phases of statistical workflow:

- ▶ Set up an initial model
- ▶ Model check
- ▶ Computational checks for the inference algorithm
- ▶ Posterior predictive checks
- ▶ Model comparison

**Example:** Estimate  $\text{PM}_{2.5}$  concentration through satellite data

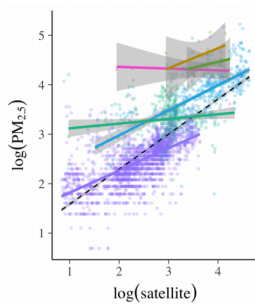
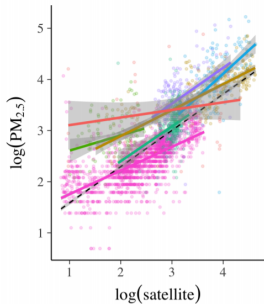
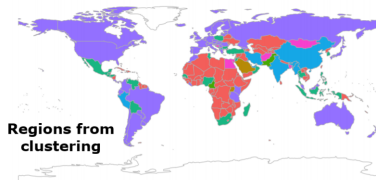
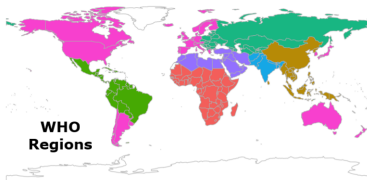
# Exploratory data analysis

More than just plotting the data



# Exploratory data analysis

More than just plotting the data



# Exploratory data analysis

More than just plotting the data

- ▶ **Model 1:** simple linear regression

$$\log(\text{PM}_{2.5}) \sim \mathcal{N}(\alpha + \beta \log(\text{satellite}), \sigma)$$

- ▶ **Model 2:** multilevel model

$$\log(\text{PM}_{2.5,j}) \sim \mathcal{N}(\mu_j, \sigma), \quad \mu_j = \alpha_0 + \alpha_j + (\beta_0 + \beta_j) \log(\text{satellite}_j),$$

where observations are stratified by WHO super-regions.

- ▶ **Model 3:** multilevel model

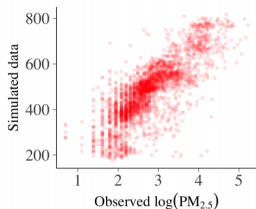
$$\log(\text{PM}_{2.5,j}) \sim \mathcal{N}(\mu_j, \sigma), \quad \mu_j = \alpha_0 + \alpha_j + (\beta_0 + \beta_j) \log(\text{satellite}_j),$$

where observations are stratified by clustered super-region.

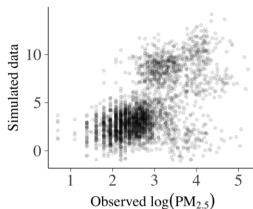
# Prior predictive checking

Fake data can be almost as valuable as real data

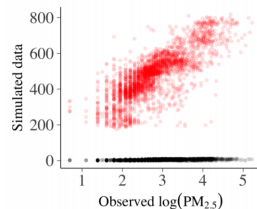
**Generative model:**  $\theta^* \sim p(\theta) \longrightarrow y^* \sim p(y|\theta^*) \iff y^* \sim p(y)$



(a) Vague priors



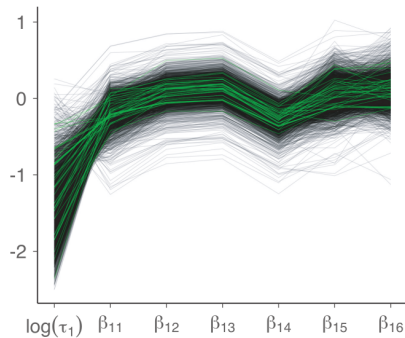
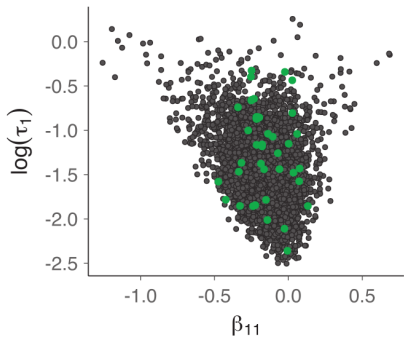
(b) Weakly informative priors



(c) Comparison

# MCMC diagnostics

Moving beyond trace plots

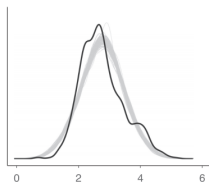


# Posterior predictive checks

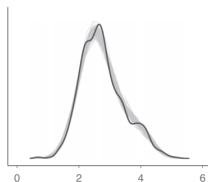
How did we do?

**Posterior predictive distribution:**  $p(\tilde{y}|y) = \int d\theta \, p(\tilde{y}|\theta)p(\theta|y)$

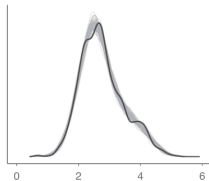
$$\theta^* \sim p(\theta|y) \longrightarrow \tilde{y}^* \sim p(\tilde{y}|\theta^*) \iff \tilde{y}^* \sim p(\tilde{y}|y)$$



**Model 1**



**Model 2**



**Model 3**

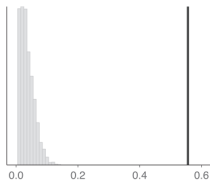


# Posterior predictive checks

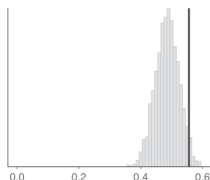
How did we do?

**Posterior predictive distribution:**  $p(\tilde{y}|y) = \int d\theta \, p(\tilde{y}|\theta)p(\theta|y)$

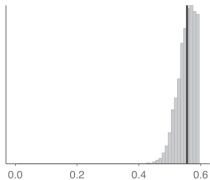
$$T(\tilde{y}) = \text{skew}(\tilde{y})$$



**Model 1**



**Model 2**



**Model 3**

# Pointwise plots for predictive model comparison

Looking *where* a model is better than another

**LOO predictive distribution:**  $p(y_i|y_{-i})$

