A Bayesian approach to estimate weights of ground motion prediction equations

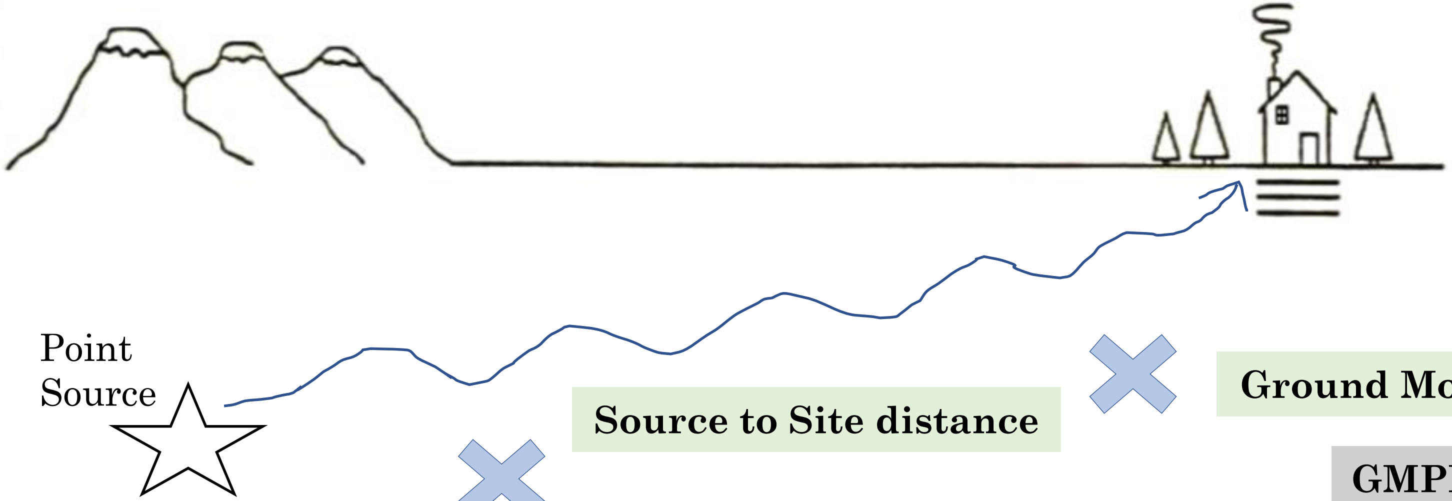
AppliBUGS – Dec 10th, 2021

Joint Post-doctoral Research Project EDF / NCSU

Saran Bodda, Abhinav Gupta (NCSU)

Merlin Keller, Gloria Senfaute (EDF)

Probabilistic Seismic Hazard Analysis (PSHA)

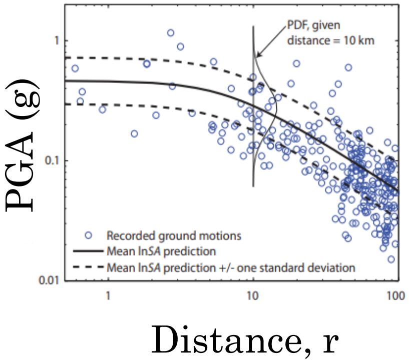


Point Source

Source to Site distance

×

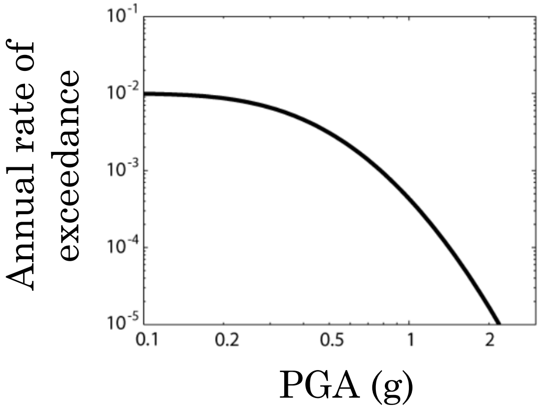
Ground Motion Prediction



$$\text{Risk} = \int P_{f|\lambda} \cdot \left| \frac{dH(\lambda)}{d\lambda} \right| d\lambda$$

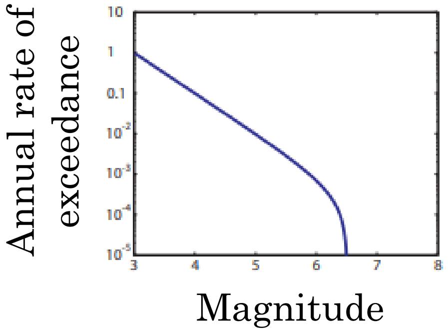
=

Hazard Curve



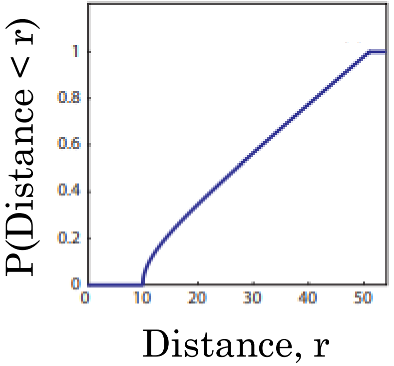
×

Distribution of EQ Magnitude



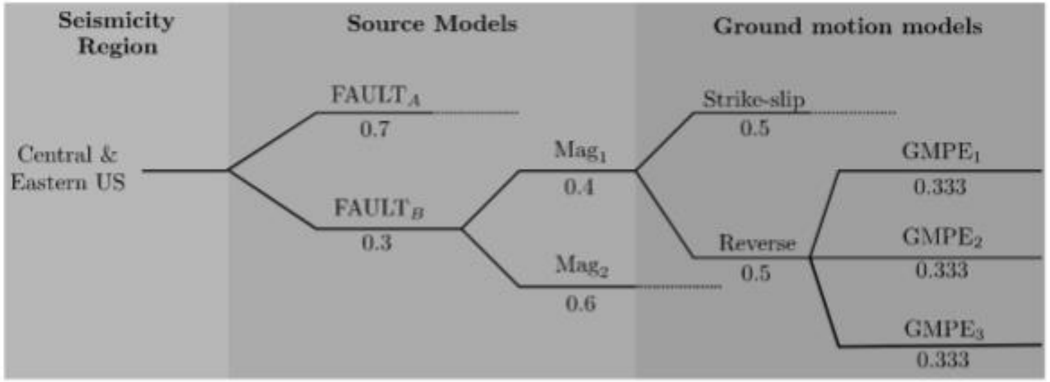
Source Models

$M_{S,1}$
$M_{S,2}$
⋮
⋮



GMPE Models

$M_{G,1}$
$M_{G,2}$
⋮
⋮
⋮



Model Uncertainty

- ***"Essentially all models are wrong, but some are useful."*** *– George E. P. Box*

- It depends on how accurately a mathematical model describes the true system for a real-life situation, since models are almost always only approximations to reality.

- Model Selection – process of selecting one candidate model from among a collection of candidate models.

- Model Averaging (ensemble modeling).

Objectives

Step 1: Methodology Development

- Bayesian model averaging (BMA) approach.
- Estimate weights associated to a list of fixed GMPEs with added bias/variance terms.
    - Bayesian linear models (conjugate prior approach)

Step 2: Application

- Apply the methodology to: (i) simple case study and (ii) a set of GMPEs and pan-European Engineering Strong Motion (ESM) database.

➢ Update GMPE internal parameters and *then calculate the weights.*

- Adaptive Metropolis (AM), Automated Factor Slice Sampling (AFSS), Laplace approximation and Sampling importance resampling (SIR), Bayesian linear models (BLM)

Bayesian Model Averaging (BMA)

- BMA approach refers to the process of estimating a desired quantity of interest (y) under each candidate model (M_k) and then averaging the estimates based on how likely each model is, in a list of candidate models.

$$p(y|D)=\sum_{k=1}^{K} p(y|M_k,D)\,p(M_k|D)$$

- Posterior Probability (or weights) of each model

*Model prior probability: **uniform***

$$p(M_k|D)=w_k=\frac{p(D|M_k)\,p(M_k)}{\sum_{k=1}^{K} p(D|M_k)\,p(M_k)}$$

- Marginal Likelihood of candidate model

$$p(D|M_k)=\int_{\theta_k} p(D|\theta_k,M_k)\,p(\theta_k|M_k)\,d\theta_k$$

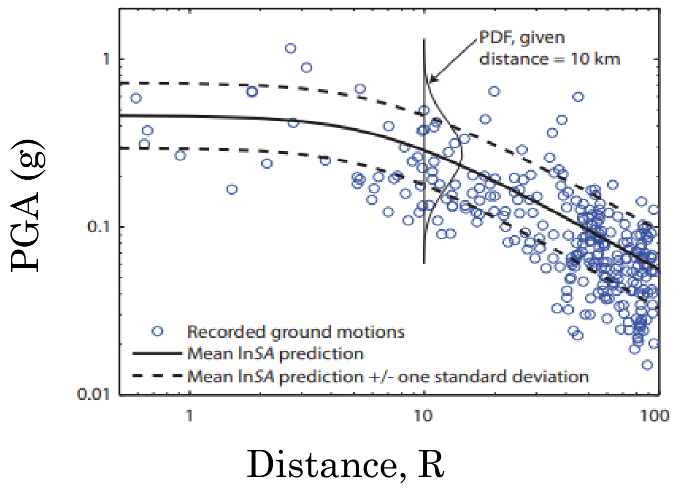
Statistical Model 1

General model

Stochastic error: $\varepsilon_{k,i} \sim N\left(0, \sigma_k^2\right)$

$$y_i = f_k(x_i, \beta) + \varepsilon_{k,i}$$

$$f_k(x_i, \beta) = \beta_0 + (\beta_1 + \beta_2 M_W)\log\left(\sqrt{R^2 + \beta_5^{\,2}}\right) + \beta_3 M_W + \beta_4 M_W^2$$



y_i is the QoI of the considered database

$f_k(x_i, \beta)$ is the QoI predicted by the k-th candidate model, given a vector x_i (M_W, R) of regressors

$$p(y \mid D) = \sum_{k=1}^{K} p(y \mid M_k, D)\, p(M_k \mid D)$$

Fixed parameter setting

$$y_i = f_k(x_i, \beta) + \varepsilon_{k,i}$$

$$\sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} w_k f_k(y_i \mid x_i, \beta)\right)$$

$f_k(x_i, \beta)$ is fixed. Weights can be estimated using Maximum likelihood approach

Statistical Model 2

Fixed parameter setting with added bias term

Bias Term

$$y_i = f_k(x_i, \beta) + \mu_k + \varepsilon_{k,i} \qquad \varepsilon_{k,i} \sim N\left(0, \sigma_k^2\right)$$

Since $f_k(x_i, \beta)$ is fixed, the unknown parameters are $\theta_k = (\mu_k, \sigma_k^2)$

- Parameters and weights can be estimated using
  - Bayesian Linear models (conjugate prior approach)
  - MCMC methods

Statistical Model 3

Uncertain parameter setting

$$y_i = f_k(x_i, \beta) + \varepsilon_{k,i} \qquad \varepsilon_{k,i} \sim N\left(0, \sigma_k^2\right)$$

The unknown parameters are $\theta_k = (\beta_k, \sigma_k^2)$

- Parameters and weights can be estimated using
  - Bayesian **Linear** models (conjugate prior approach)
  - MCMC methods

Estimation Method: Bayesian Linear Model (BLM)

$$y = X\beta + \varepsilon \qquad \theta = (\beta, \sigma^2)$$

Prior

$$p(\beta,\sigma^2) = p(\beta \mid \sigma^2)\, p(\sigma^2) = N(\mu_\beta, \sigma^2 V_\beta) \times IG(a,b) = NIG(\mu_\beta, V_\beta, a, b)$$

Banerjee (2008)

Likelihood

$$p(D \mid \beta, \sigma^2) \sim N(X\beta, \sigma^2 I)$$

Posterior

$$p(\beta, \sigma^2 \mid D) \sim NIV(\mu^*, V^*, a^*, b^*)$$

Marginal Likelihood

$$p(D) \sim MVSt_{2a}\left(X\mu, \frac{b}{a}\left(I + XVX^T\right)\right)$$

Model probabilities

$$w_k(D) = \frac{p_k(D)}{\sum_{k=1}^{K} p_k(D)}$$

BMA Prediction

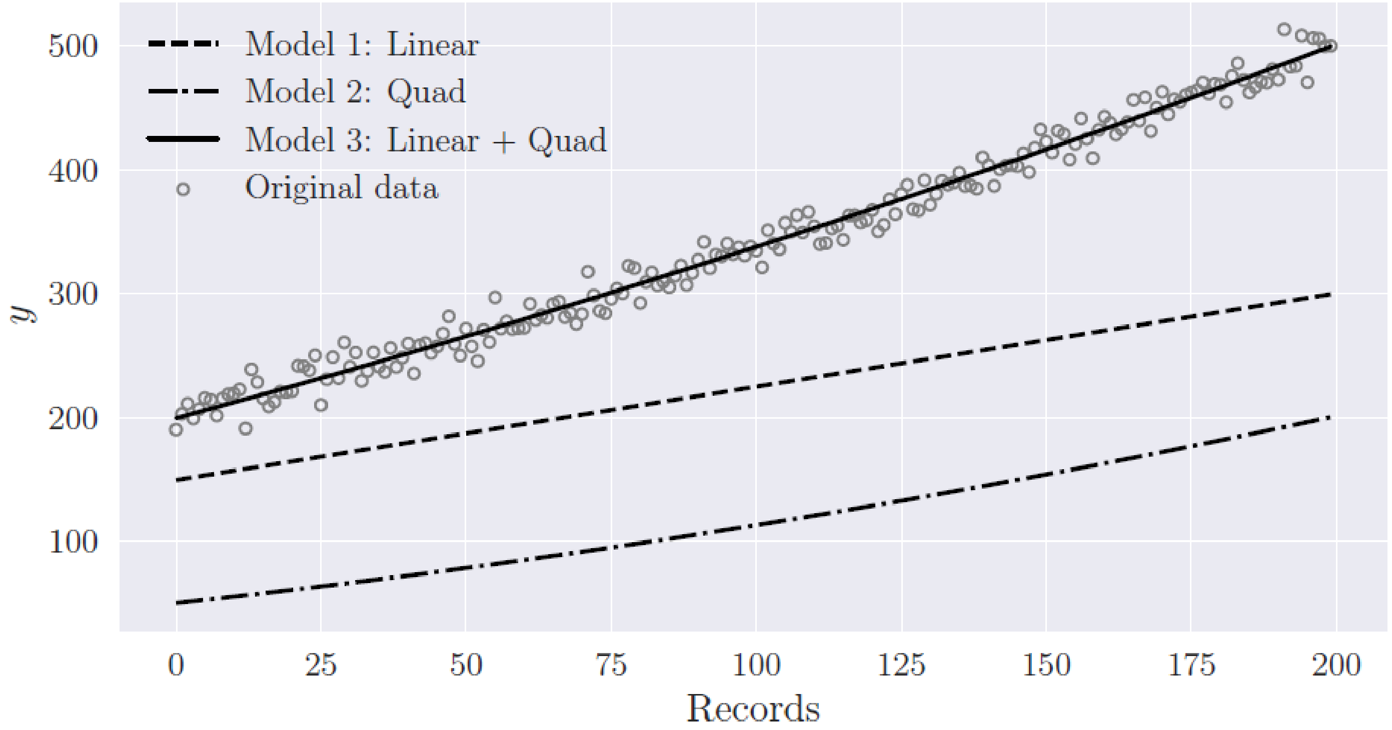
$$p(\tilde{y} \mid D, \tilde{x}) = \sum_{k=1}^{K} w_k(D)\, p_k(\tilde{y} \mid D, \tilde{x})$$

Simple Case Study

True model

$$y_i = f(x_i) + \varepsilon_i$$
$$= -0.7 + 30x_i + 2x_i^2 + \varepsilon_i$$

$$\varepsilon_i \sim N\left(0, 10^2\right)$$



Candidate models

$$f(x_i) = f_1(x_i) + f_2(x_i)$$

$$f_1(x_i) = -0.7 + 30x_i$$
$$f_2(x_i) = 2x_i^2$$
$$f_3(x_i) = f_1(x_i) + f_2(x_i)$$

Model residual

$$r_{k,i} = y_i - f_k(x_i)$$

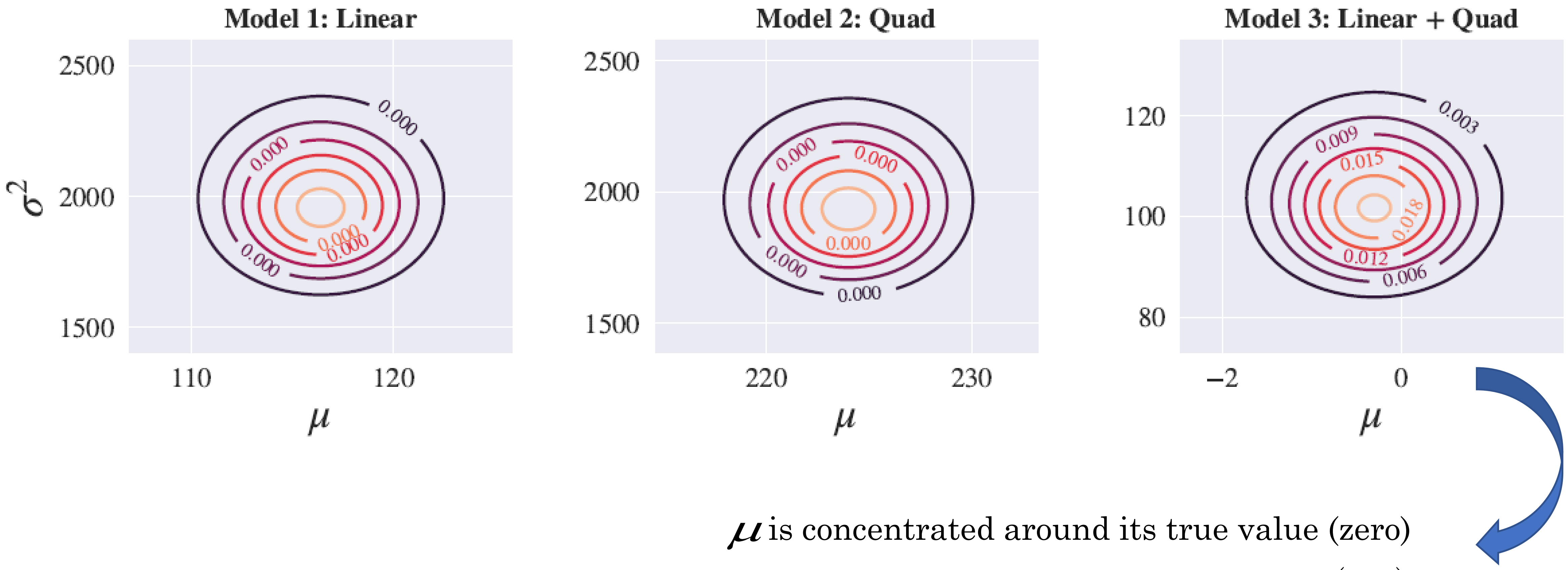
$$y_i = f_k(x_i) + \mu_k + \varepsilon_{k,i}$$
$$r_{k,i} = \mu_k + \varepsilon_{k,i}$$
$$r_k = 1_n \mu_k + \varepsilon_k$$

$$r_k = 1_n \mu_k + \varepsilon_k$$

$$y = X\beta + \varepsilon$$

$$\theta = (\beta, \sigma^2)$$

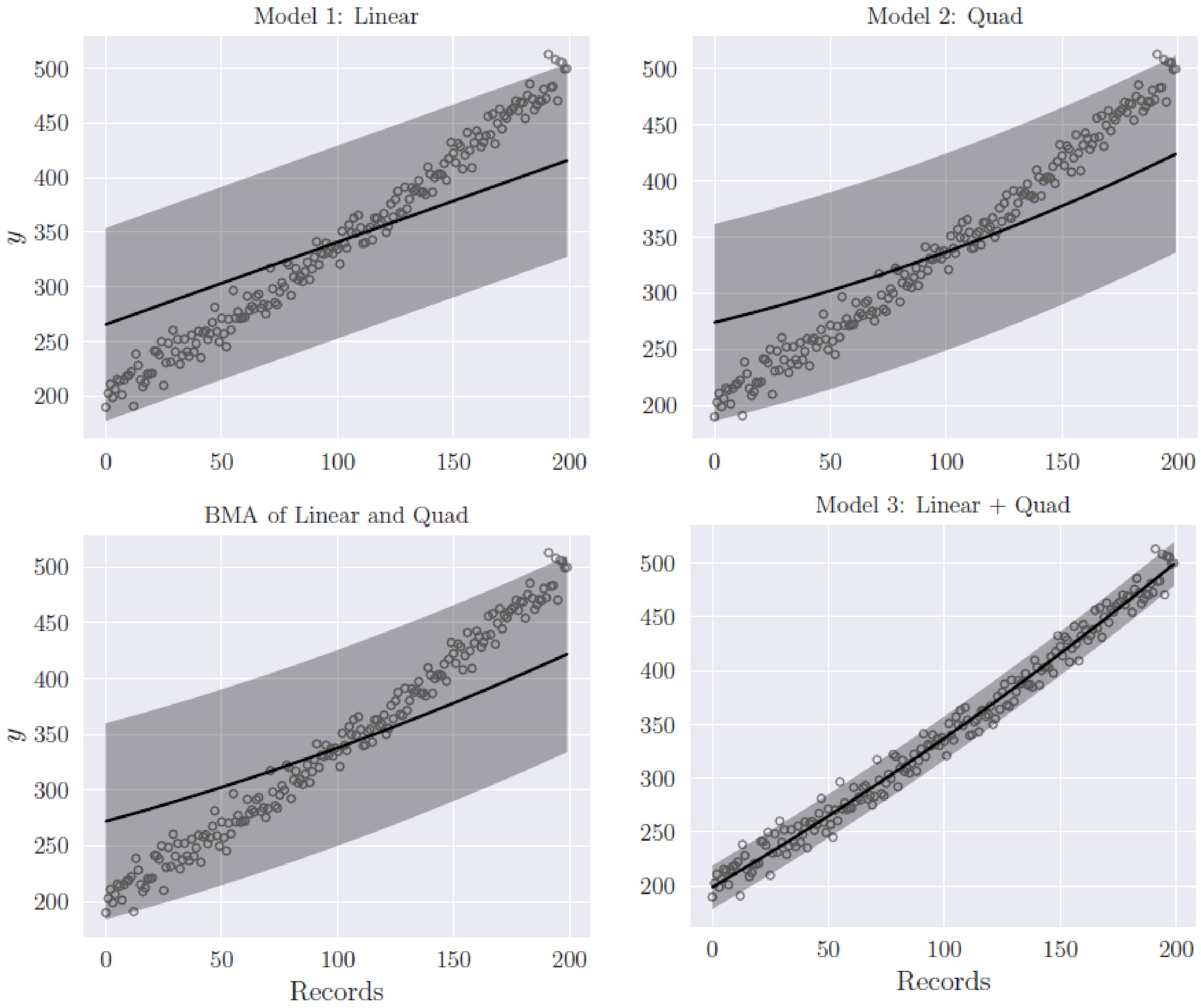
Joint Posterior Distribution



μ is concentrated around its true value (zero)

σ^2 is concentrated around its true value $\left(10^2\right)$

BMA Predictions



Model	Weights
	Case 1
f_1 : Linear	0.2425
f_2 : Quad	0.7575
f_3 : Linear + Quad	–

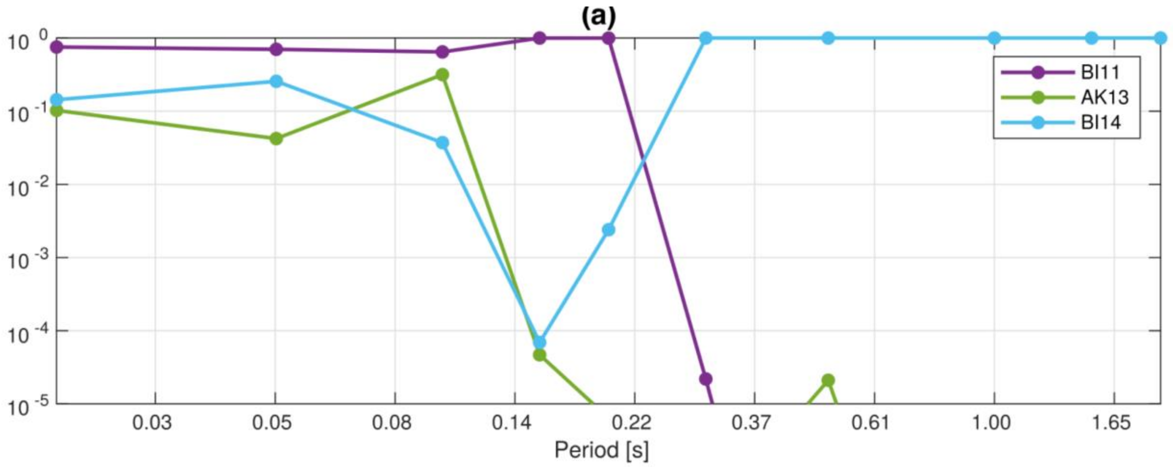
- BMA boils down to Bayesian model selection

Bertin et al. (2019)

$$y_i = f_k(x_i, \beta) + \mu_k + \varepsilon_{k,i}$$

Using Bayesian model averaging to improve ground motion predictions

- Combines 9 GMPE models using logic-tree
- RESORCE-2013 database
- Estimate weights using Bayesian model averaging (BMA) approach
- BMA is implemented using two techniques: Markov chain Monte Carlo (MCMC) method and Maximum Likelihood estimation
- Marginal Likelihood of candidate model



$$\theta_k = (\mu_k, \sigma_k^2)$$

$$p(D \mid M_k) = \int_{\theta_k} p(D \mid \theta_k, M_k)\, p(\theta_k \mid M_k)\, d\theta_k$$

$$\mu_k^* = \frac{1}{N}\sum_{i=1}^{N} d_i - f_k(x_i)$$

$$\log p(D \mid M_k) \approx -N\left(\frac{\log 2\pi}{2} + \log \sigma_k^*\right) - \frac{N}{2} - \log(\mu_b - \mu_a) - \log(\sigma_b - \sigma_a)$$

$$\sigma_k^* = \left[\frac{1}{N}\sum_{i=1}^{N}\left(d_i - f_k(x_i) - \mu_k^*\right)^2\right]^{1/2}$$

GMPE Application

- 3 GMPE models from Bertin et al. (2019)
- pan-European Engineering Strong Motion (ESM)-2018 database
- Openquake and GMPE-smtk toolkit
- BMA is implemented using conjugate prior approach – Bayesian Linear Models

- Bindi11: https://github.com/gem/oq-engine/blob/master/openquake/hazardlib/gsim/bindi_2011.py
- Akkar14: https://github.com/gem/oq-engine/blob/master/openquake/hazardlib/gsim/akkar_2014.py
- Bindi14: https://github.com/gem/oq-engine/blob/master/openquake/hazardlib/gsim/bindi_2014.py

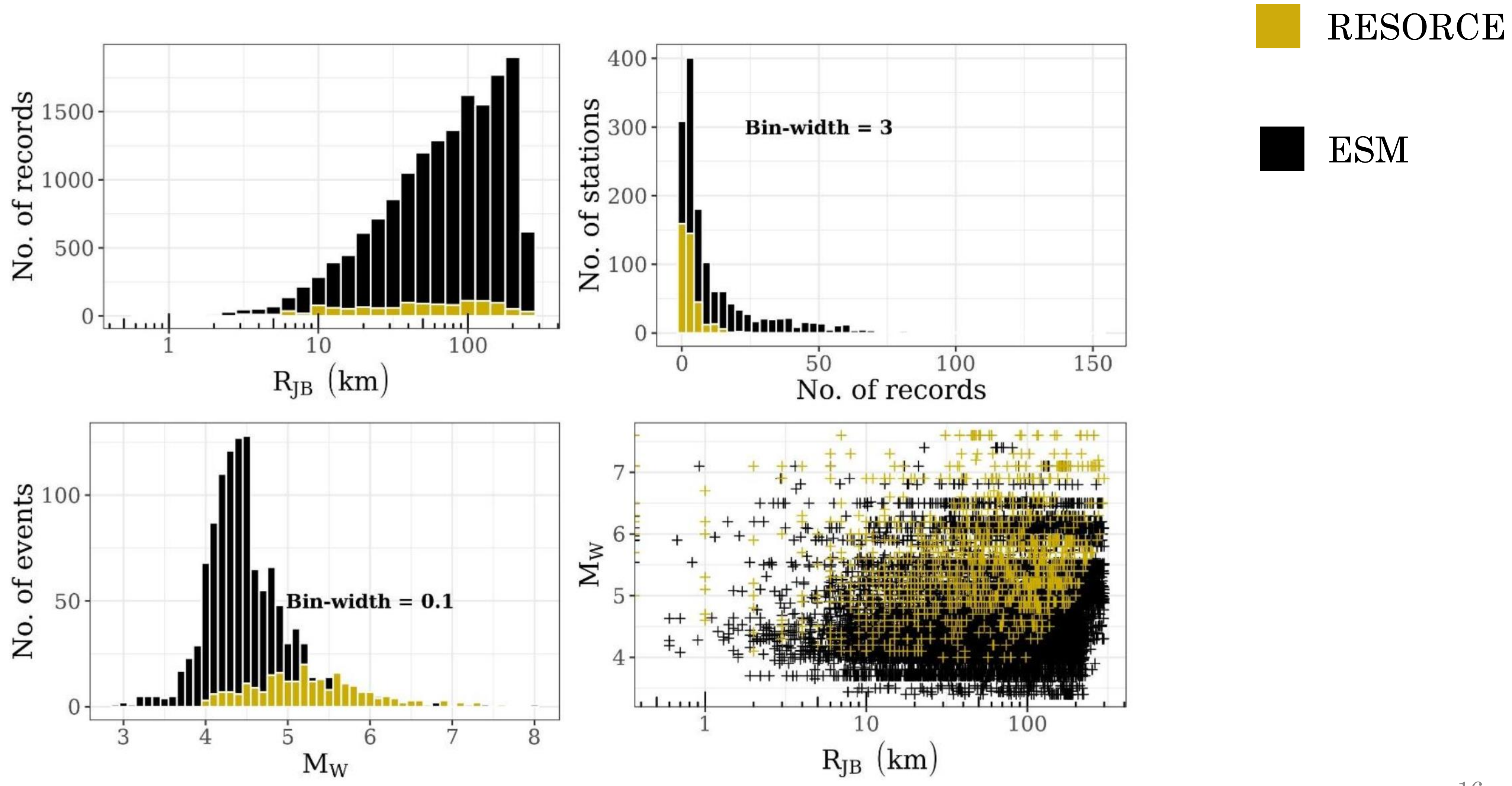
ESM Comprehensive Reference Table

The flat file consists of (Lanzano et al., 2019):

- 23,014 recordings from 2179 earthquakes and 2080 stations from Europe and Middle-East (1969 – 2018).
- Magnitudes range from 3.5 – 8.0 (includes shallow active crustal and subduction zones).
- Moment magnitude, focal depth, several distance metrics, style of faulting and parameters for site characterization.
- ***QoI:*** Spectral amplitudes (5% damping, acceleration and displacement response) are provided for 36 periods, in the interval 0.01–10 s.

Comparison with RESORCE database

- ESM is an updated database of RESORCE



Bertin et al. (2019) – Selection Criteria

- RESORCE Database (5882 records)
- **Selection Criteria: 939 records – training (739), testing (200)**
- Moment magnitude [5 – 7.3]
- Distance [4 – 150 km]
- V_{S30} velocity [300 – 1200 m/s]
- Fault mechanism – *Normal, Strike-Slip and Reverse*
- Geometric mean of horizontal spectral Acceleration is computed for 10 periods

	1	2	3	4	5	6	7	8	9	10
Period [s]	0.02	0.05	0.1	0.15	0.2	0.3	0.5	1	1.5	2

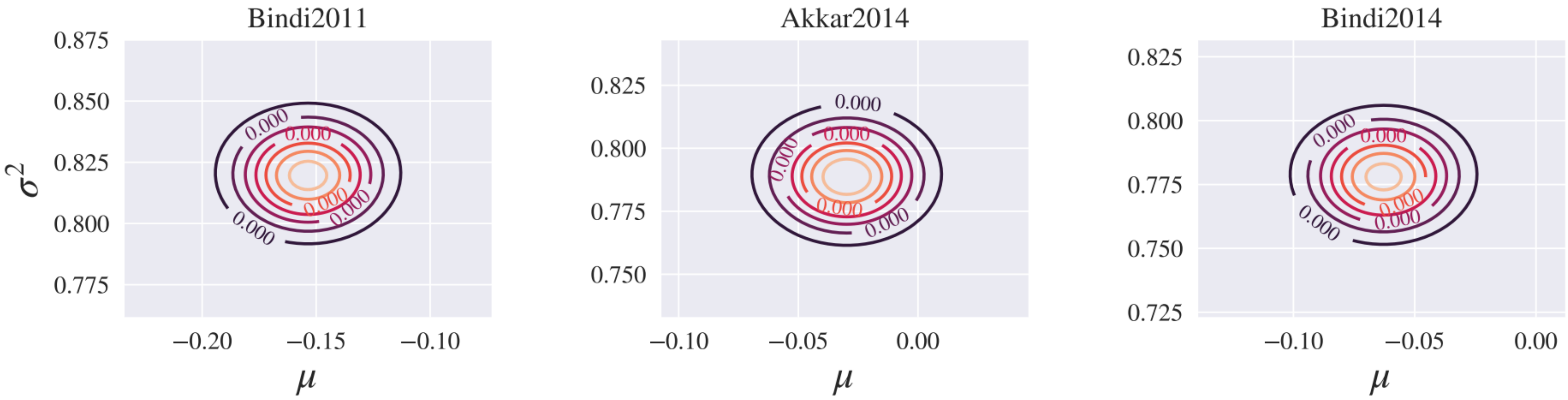
ESM Database (19197 records)

Selection Criteria: 2356 records – training (1650), testing (706)

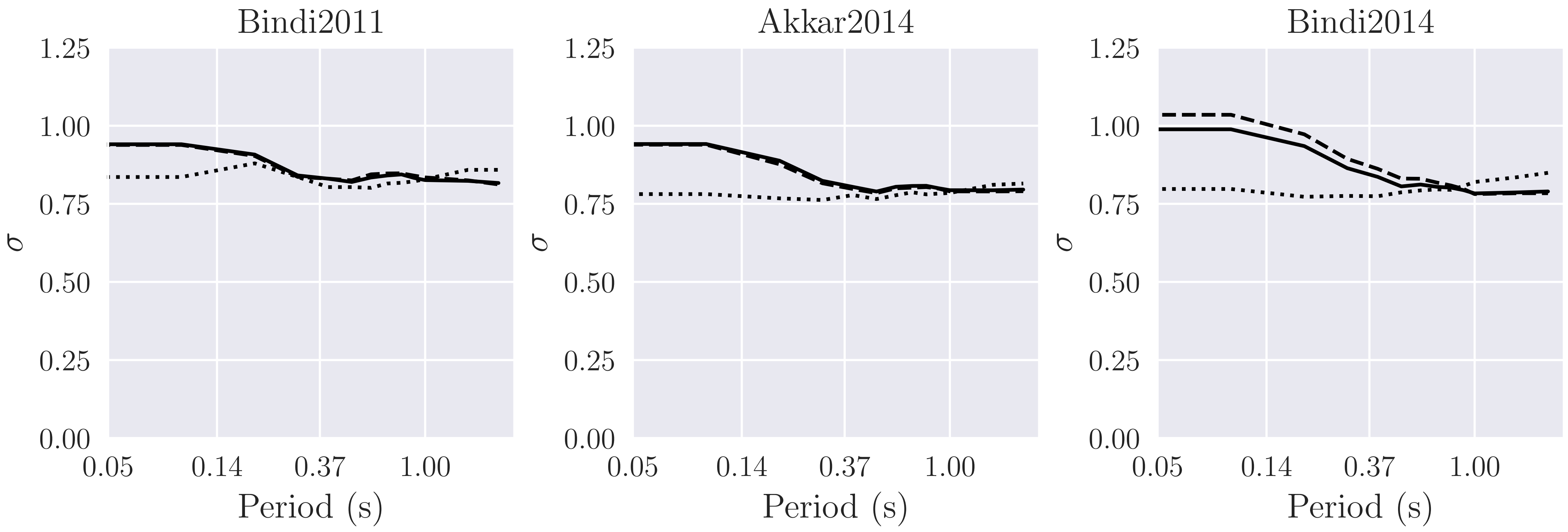
K-fold cross-validation

Joint Posterior Distribution (T = 1 sec)

- $\mu \neq 0$: Discrepancy between the GMPE predictions and ESM training database



Comparison of Total σ



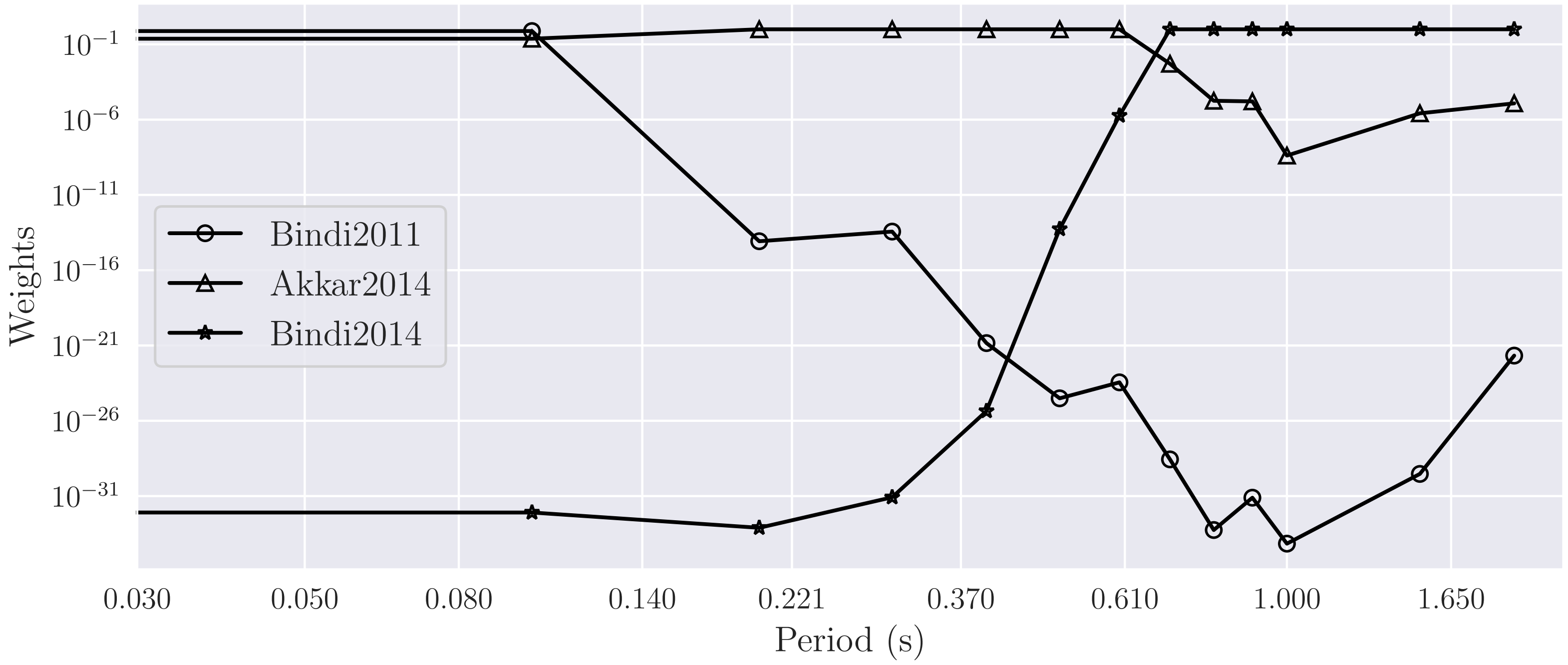
Dotted: σ of original GMPEs on their own data set

Dashed: σ of original GMPEs on ESM training dataset – increase due to wide range of records

Solid: σ of original GMPEs with added bias term on ESM training dataset

Model Weights

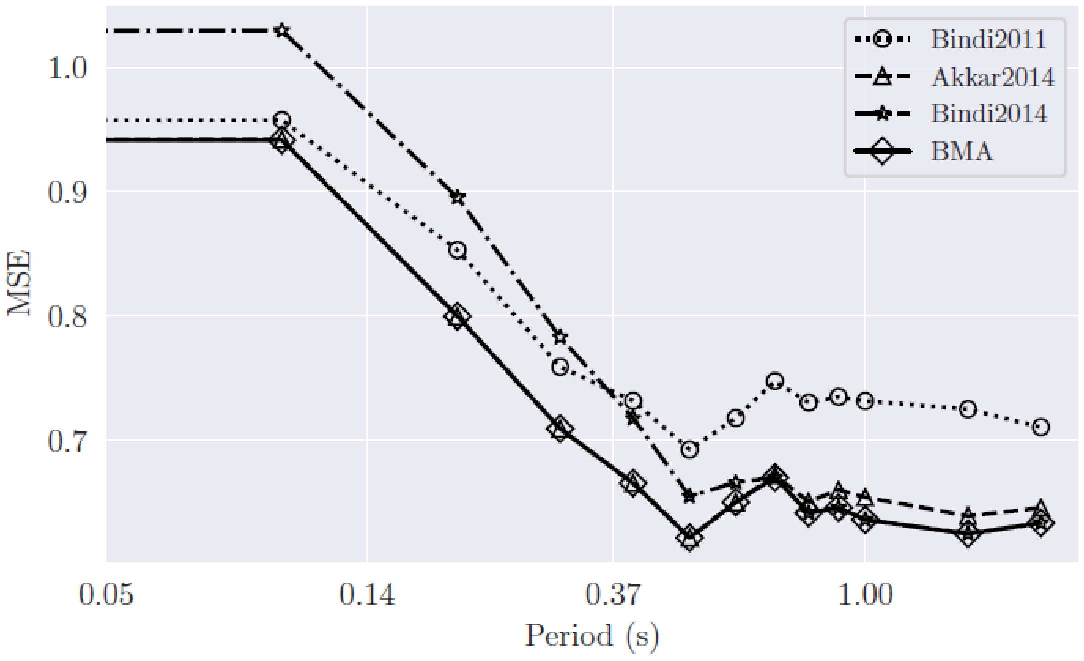
- At a given period, one GMPE completely takes over ($w \approx 1$) the other GMPEs.
- No single GMPE dominates over the entire range of period.
- The response predicted by the combined model using BMA may perform better than a single GMPE at all periods.



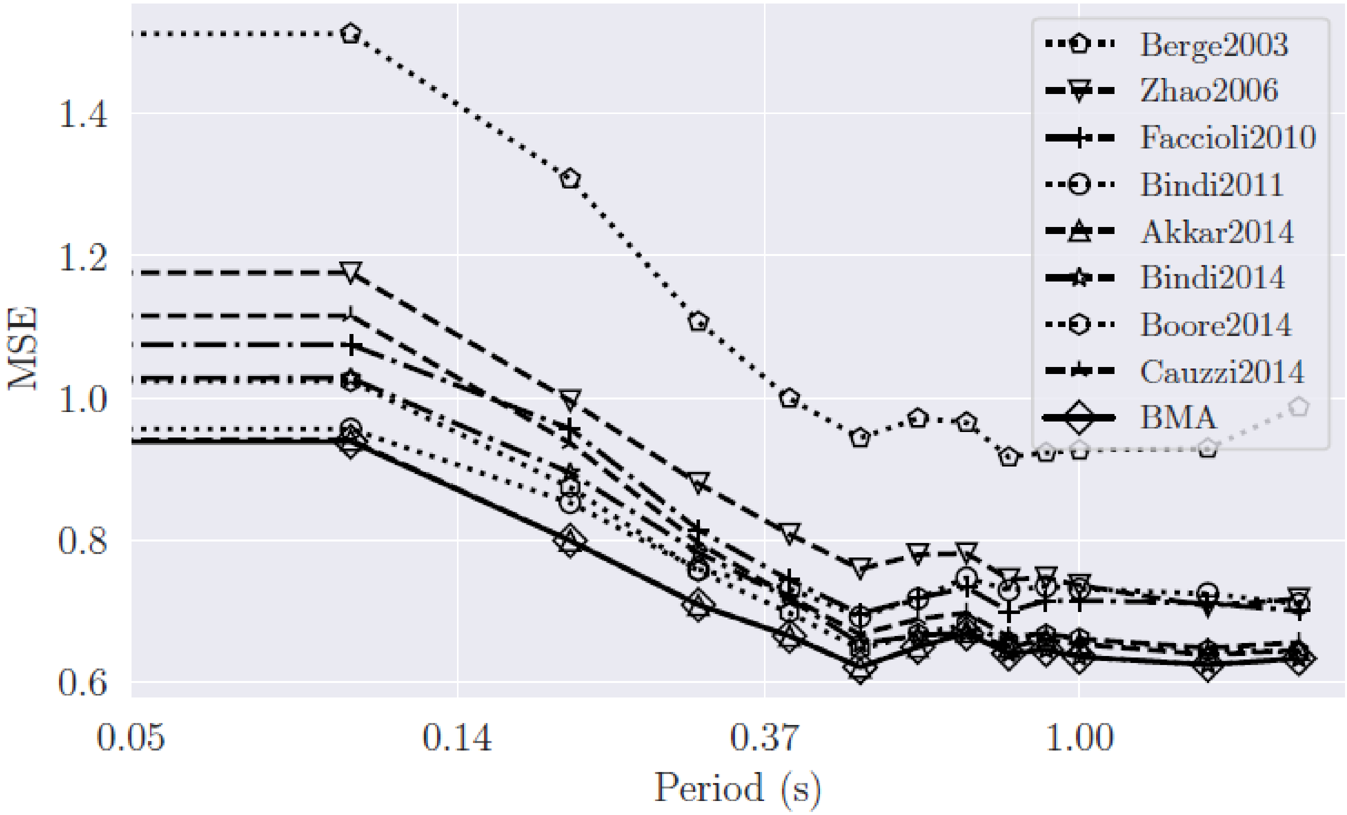
GMPE Performance

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - E\big[p(\tilde{y}\,|\,D,\tilde{x})\big]\Big)^2$$

3 GMPEs



8 GMPEs



- The BMA approach yields the most optimal predictions at all time periods.

Conclusions

- The GMPE model weights can be estimated using analytical formulations and without any approximations.
- The proposed methodology can assist experts to make a better judgment by going from a heuristic approach to a quantified or semi-quantified approach.
- There is an increase in sigma when the original GMPEs are estimated on ESM data set compared to the estimation on their own data set.
- The BMA model has a lower mean squared error than any of the GMPEs at all periods for the testing data set indicating that the BMA model yields the most optimal predictions.

Bayesian Update of GMPE coefficients

- Motivation: can we have even better predictions by updating the GMPE parameters?

Bodda et al. (2021)

Functional Form of Bindi et al. (2014)

$$f_{BI14}(x) = e_1 + f_D(R,M) + f_M(M) + f_S(EC8, V_{s,30}) + f_{SOF}(FC)$$

$$f_D(R,M) = [c_1 + c_2(M - M_{ref})]\log\left(\sqrt{R^2 + h^2} / R_{ref}\right) - c_3\left(\sqrt{R^2 + h^2} - R_{ref}\right)$$

$$f_M(M) = \begin{cases} b_1(M - M_h) + b_2(M - M_h)^2 & \text{for } M \le M_h = 6.75 \\ b_3(M - M_h) & \text{otherwise} \end{cases}$$

$$f_S(EC8, V_{s,30}) = \gamma \log(V_{s,30} / V_{ref})$$

$$f_{SOF}(FC) = f_1 F_N + f_2 F_R + f_3 F_S$$

- Linear functional form: *h* is constant
- Moment magnitude [3 – 5.2], Distance [1 – 350 km], V_{S30} velocity [800 – 1200 m/s], Fault mechanism – *Normal, Strike-Slip and Reverse*
- 3154 records – *2200 (updating), 954 (testing)*

Statistical Models

M0: Original

$$\log y_i = \log\left(10^{f_{BI14}(x_i)-2} / g\right) + \varepsilon_i$$
$$= f_{BI14}(x_i)\log 10 - 2\log 10 - \log g + \varepsilon_i$$

Correction Term

$$\text{M1:} \log y_i = \log\left(10^{f_{BI14}(x_i)+\mu_{BI14}-2} / g\right) + \varepsilon_{BI14,i}, \quad \theta_{M1} = \left(\mu_{BI14}, \sigma_{BI14}\right)$$

Linear

$$\text{M2:} \log y_i = \log\left(10^{f_{BI14}(x_i,\eta)-2} / g\right) + \varepsilon_{BI14,i}, \quad \theta_{M2} = \left(\eta, \sigma_{BI14}\right)$$

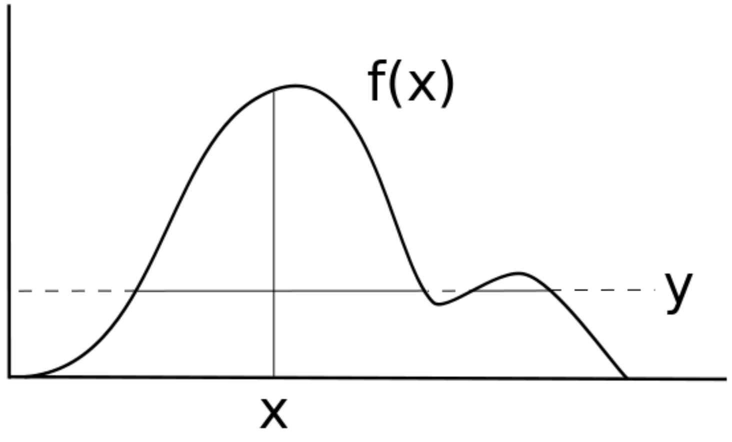
Nonlinear

$$\text{M3:} \log y_i = \log\left(10^{f_{BI14}(x_i,\eta,h)-2} / g\right) + \varepsilon_{BI14,i}, \quad \theta_{M3} = \left(\eta, h, \sigma_{BI14}\right)$$

$$\eta = (e_1, c_1, c_2, c_3, b_1, b_2, \gamma, f_1, f_2)$$

- Bayesian linear models (BLM), Adaptive Metropolis (AM) algorithm, Automated Factor Slice Sampling (AFSS), Laplace approximation and Sampling importance resampling (SIR).
- Implemented in R using LaplacesDemon package (Statisticat and LLC., 2020).

Slice Sampling



- Given a sample x we choose y uniformly at random from the interval $[0, f(x)]$
- Given y figure out all the line segments under the curve.
- From all the line segments, draw a value of x uniformly.

Disadvantages

- Probably slower than Metropolis Hastings.
- Finding the roots of the intersection between horizontal line and distribution is tricky.

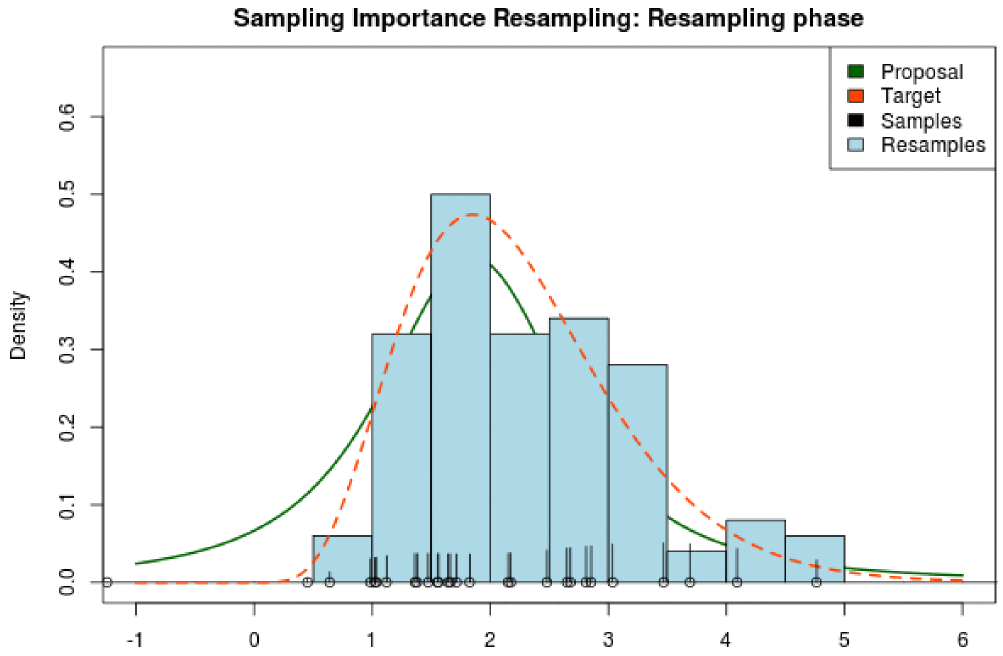
Laplace + Sampling Importance Resampling

Normal Approximation to the Posterior Distribution

Proposal density: $q(\theta \mid D) \sim \mathcal{MVN}\left(\tilde{\theta}, -H^{-1}\right)$

Sample weights: $w_i = \dfrac{p(\theta_i \mid D) / q(\theta_i \mid D)}{\sum_{\theta} p(\theta_i \mid D) / q(\theta_i \mid D)}$

Effective sample size: $ESS = \dfrac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2}$



$\tilde{\theta}$ is the mode of $p(\theta \mid D)$, $H = \nabla^2 \log p(\theta \mid D)|_{\theta=\tilde{\theta}}$ is a Hessian matrix.

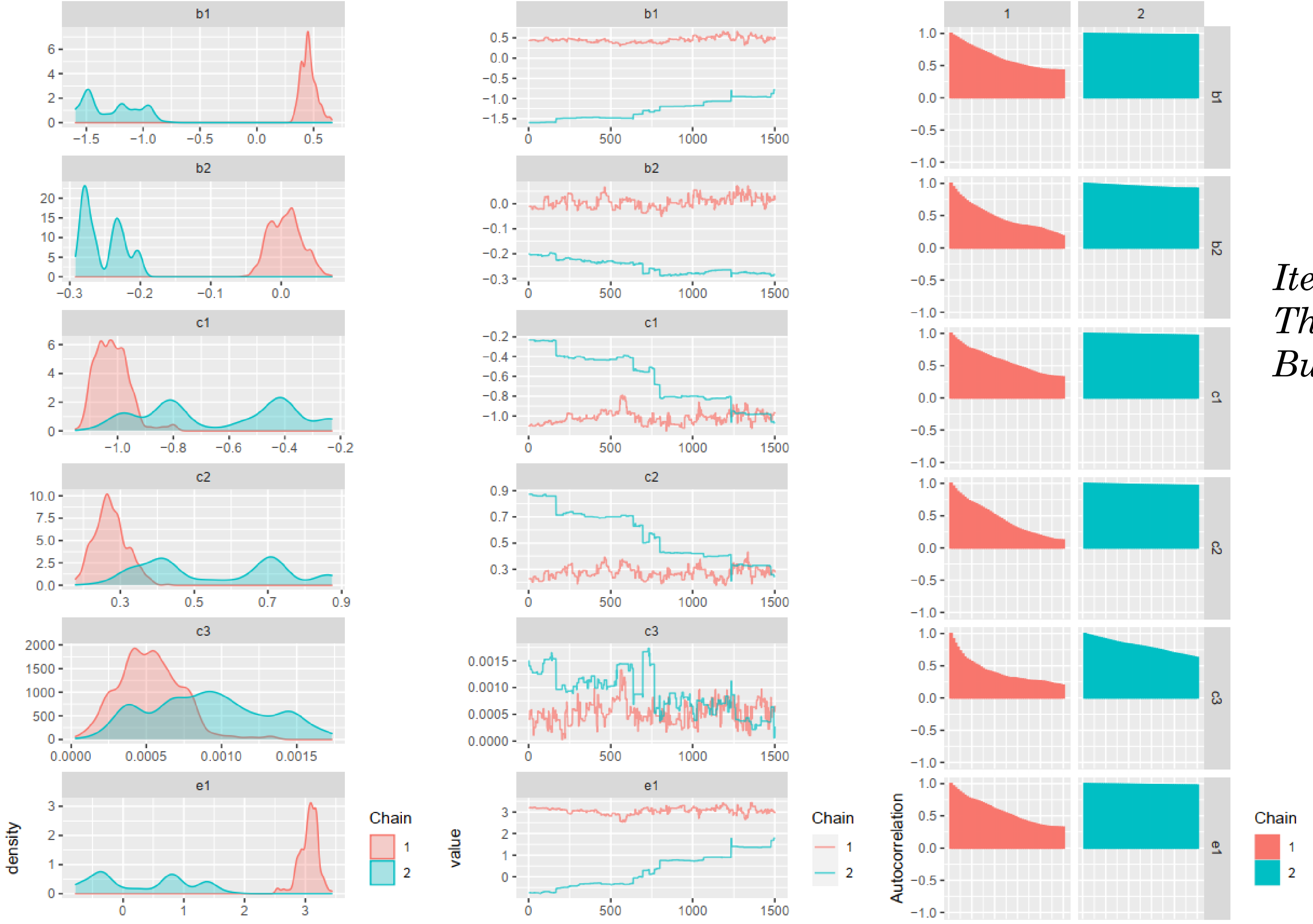
- Quality of SIR is measured by the effective sample size (ESS)
- $1 < ESS < S \equiv$ the size of an equivalent iid posterior sample

Bayesian Update of Parameters

Comparison of original BI14 coefficients with the M3 recalibrated coefficients for PSA ($T = 1$ s) (the standard deviation of coefficients are given in brackets)

Coefficients	Org.	*M2: Linear*	textitM3: Non-Linear		
		BLM	SIR	AM	AFSS
e_1	3.1247				
Uninformative		2.63 (0.224)	3.7075 (0.469)	3.19 (0.277)	3.2189 (0.243)
Informative		3.5769 (0.148)	3.2299 (0.253)	3.1023 (0.14)	3.2383 (0.248)
c_1	− 1.0527				
Uninformative		− 1.0379 (0.095)	− 1.039 (0.111)	− 0.9946 (0.071)	− 0.9877 (0.076)
Informative		− 0.7448 (0.082)	− 0.9902 (0.079)	− 1.0104 (0.052)	− 0.9904 (0.078)
c_2	0.1035				
Uninformative		0.2357 (0.061)	0.2377 (0.063)	0.2589 (0.083)	0.2764 (0.045)
Informative		0.4395 (0.051)	0.2751 (0.047)	0.2775 (0.044)	0.274 (0.048)
c_3	0				
Uninformative		7e−4 (2.8e−4)	6.5e−4 (3.3e−4)	7.5e−4 (4.2e−4)	7e−4 (2.7e−4)
Informative		0.001 (2.8e−4)	6.8e−4 (2.8e−4)	5.5e−4 (2e−4)	6.9e−4 (2.8e−4)
b_1	0.3066				
Uninformative		0.9674 (0.295)	0.9628 (0.302)	0.5825 (0.193)	0.6252 (0.174)
Informative		− 0.1853 (0.194)	0.6328 (0.176)	0.5032 (0.063)	0.6385 (0.172)
b_2	− 0.1476				
Uninformative		0.1012 (0.054)	0.1011 (0.055)	0.0271 (0.047)	0.0437 (0.039)
Informative		− 0.0643 (0.039)	0.0449 (0.039)	0.0185 (0.019)	0.0455 (0.038)
γ	− 0.8266				
Uninformative		− 1.2179 (0.191)	− 1.2146 (0.191)	− 1.1872 (0.192)	− 1.1275 (0.168)
Informative		− 1.0783 (0.166)	− 1.1209 (0.166)	− 1.1216 (0.189)	− 1.1236 (0.168)
f_1	0.0263				
Uninformative		− 0.0034 (0.021)	− 0.003 (0.021)	0.0067 (0.05)	− 0.0037 (0.022)
Informative		0.0004 (0.022)	− 0.0026 (0.021)	− 0.0018 (0.024)	− 0.0025 (0.022)
f_2	0.0186				
Uninformative		− 0.081 (0.027)	− 0.0808 (0.027)	− 0.0622 (0.08)	− 0.081 (0.027)
Informative		− 0.0794 (0.027)	− 0.0811 (0.027)	− 0.0811 (0.03)	− 0.0801 (0.027)
h	4.4161				
Uninformative		–	4.5108 (1.205)	4.5272 (0.999)	4.4313 (0.339)
Informative		–	4.4215 (0.337)	4.4644 (0.316)	4.4102 (0.332)
σ_T	0.3561				
Uninformative		0.8672 (0.151)	0.87 (0.013)	0.8644 (0.054)	0.8697 (0.013)
Informative		0.8842 (0.154)	0.8698 (0.013)	0.8699 (0.013)	0.87 (0.013)

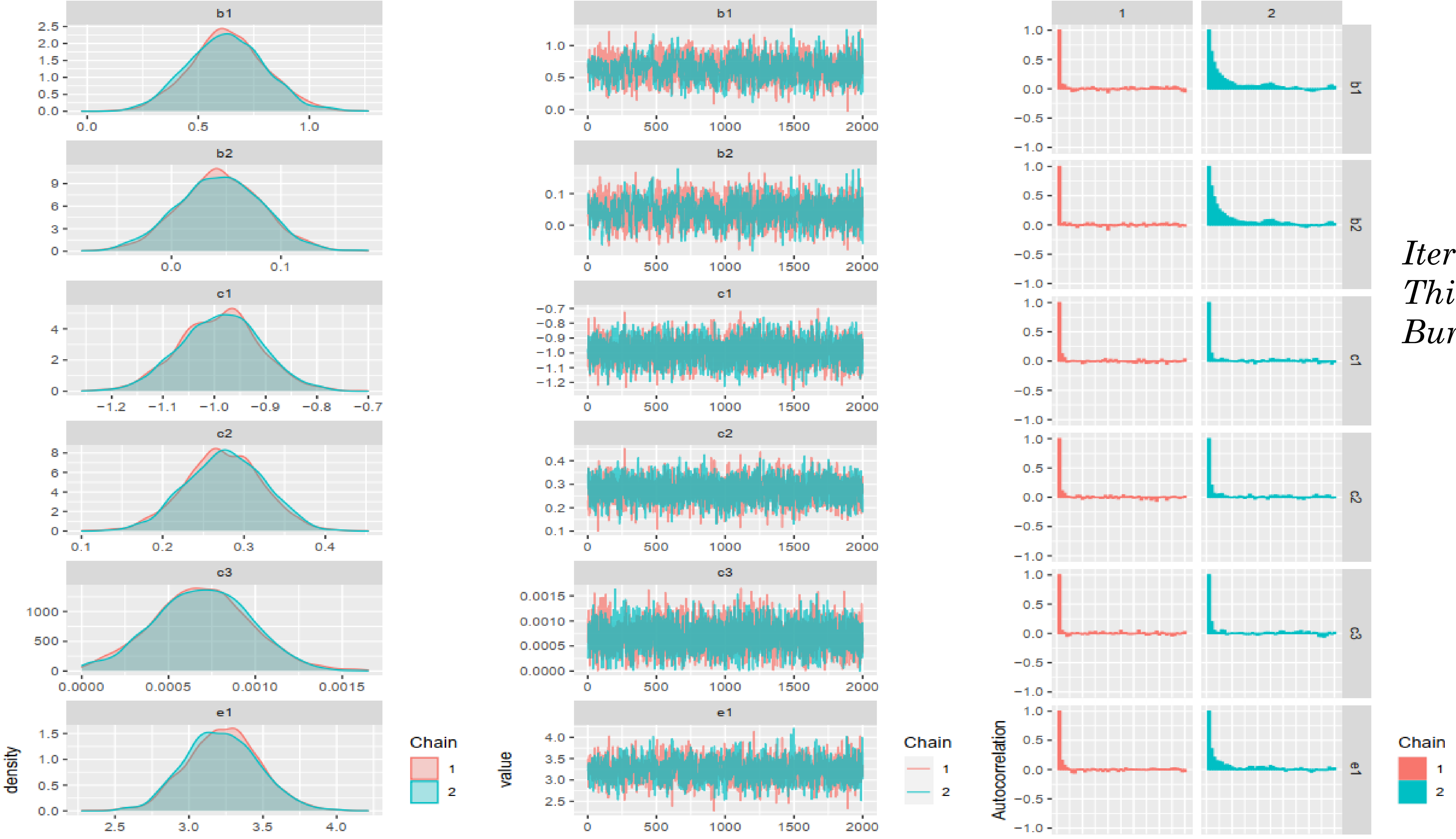
Adaptive Metropolis – Convergence Diagnostics



Marginal Density **Trace Plots** **Autocorrelation**

Iterations: 100,000
Thinning: 50
Burn-in: 500

AFSS – Convergence Diagnostics



Marginal Density **Trace Plots** **Autocorrelation**

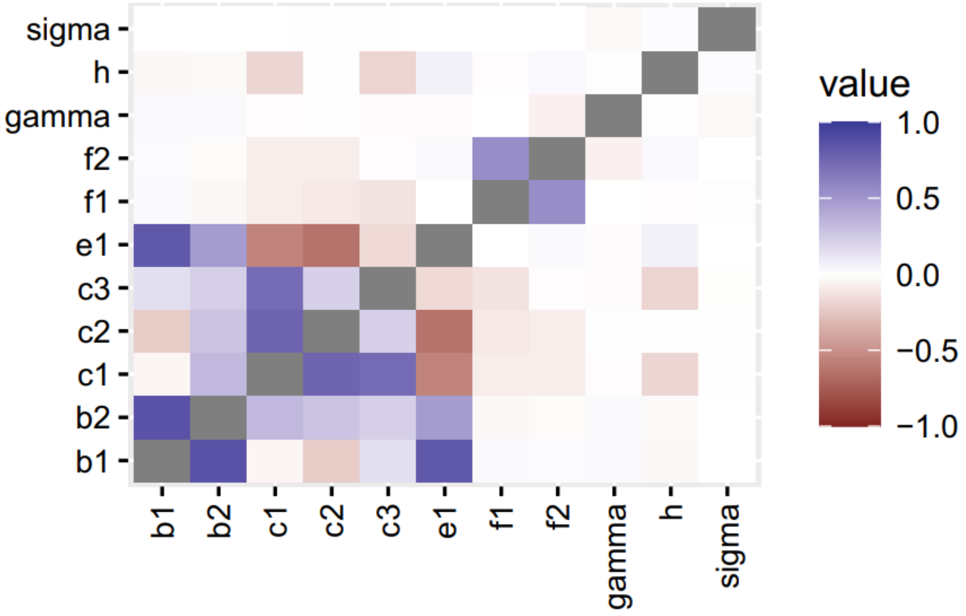
Iterations: 5000
Thinning: 2
Burn-in: 500

Convergence Diagnostics

SIR Convergence

Iterations: 10,000

ESS: 8300-9600



Pearson’s correlation matrix for posterior samples

Comparison of original BI14 coefficients with the M3 recalibrated coefficients for PSA (T = 1 s) (the standard deviation of coefficients are given in brackets)

Coefficients	Org.	*M2: Linear*	textitM3: Non-Linear		
		BLM	SIR	AM	AFSS
e_1	3.1247				
Uninformative		2.63 (0.224)	3.7075 (0.469)	3.19 (0.277)	3.2189 (0.243)
Informative		3.5769 (0.148)	3.2299 (0.253)	3.1023 (0.14)	3.2383 (0.248)
c_1	−1.0527				
Uninformative		−1.0379 (0.095)	−1.039 (0.111)	−0.9946 (0.071)	−0.9877 (0.076)
Informative		−0.7448 (0.082)	−0.9902 (0.079)	−1.0104 (0.052)	−0.9904 (0.078)
c_2	0.1035				
Uninformative		0.2357 (0.061)	0.2377 (0.063)	0.2589 (0.083)	0.2764 (0.045)
Informative		0.4395 (0.051)	0.2751 (0.047)	0.2775 (0.044)	0.274 (0.048)
c_3	0				
Uninformative		7e−4 (2.8e−4)	6.5e−4 (3.3e−4)	7.5e−4 (4.2e−4)	7e−4 (2.7e−4)
Informative		0.001 (2.8e−4)	6.8e−4 (2.8e−4)	5.5e−4 (2e−4)	6.9e−4 (2.8e−4)
b_1	0.3066				
Uninformative		0.9674 (0.295)	0.9628 (0.302)	0.5825 (0.193)	0.6252 (0.174)
Informative		−0.1853 (0.194)	0.6328 (0.176)	0.5032 (0.063)	0.6385 (0.172)
b_2	−0.1476				
Uninformative		0.1012 (0.054)	0.1011 (0.055)	0.0271 (0.047)	0.0437 (0.039)
Informative		−0.0643 (0.039)	0.0449 (0.039)	0.0185 (0.019)	0.0455 (0.038)
γ	−0.8266				
Uninformative		−1.2179 (0.191)	−1.2146 (0.191)	−1.1872 (0.192)	−1.1275 (0.168)
Informative		−1.0783 (0.166)	−1.1209 (0.166)	−1.1216 (0.189)	−1.1236 (0.168)
f_1	0.0263				
Uninformative		−0.0034 (0.021)	−0.003 (0.021)	0.0067 (0.05)	−0.0037 (0.022)
Informative		0.0004 (0.022)	−0.0026 (0.021)	−0.0018 (0.024)	−0.0025 (0.022)
f_2	0.0186				
Uninformative		−0.081 (0.027)	−0.0808 (0.027)	−0.0622 (0.08)	−0.081 (0.027)
Informative		−0.0794 (0.027)	−0.0811 (0.027)	−0.0811 (0.03)	−0.0801 (0.027)
h	4.4161				
Uninformative		–	4.5108 (1.205)	4.5272 (0.999)	4.4313 (0.339)
Informative		–	4.4215 (0.337)	4.4644 (0.316)	4.4102 (0.332)
σ_T	0.3561				
Uninformative		0.8672 (0.151)	0.87 (0.013)	0.8644 (0.054)	0.8697 (0.013)
Informative		0.8842 (0.154)	0.8698 (0.013)	0.8699 (0.013)	0.87 (0.013)

Model Selection Criteria

- Deviance

$$D(Y \mid \theta) = -2\log\left[f(Y \mid \theta)\right]$$

- Deviance Information Criteria

Posterior mean of deviance

$$DIC = \bar{D} + p_D = D(Y \mid \hat{\theta}) + 2p_D$$

$$p_D = \bar{D} - \hat{D}$$

Deviance at posterior mean of θ

- Watanabe-Akaike Information Criteria

$$WAIC = -2\sum_{i=1} \log\left\{\bar{f}_i\right\} + 2p_W$$

$$\bar{f}_i = \mathrm{E}\left[f(Y_i \mid \theta) \mid Y\right]$$

$$p_W = \sum_{i=1} \mathrm{Var}\left[\log\left(f(Y_i \mid \theta)\right) \mid Y\right]$$

Model Selection

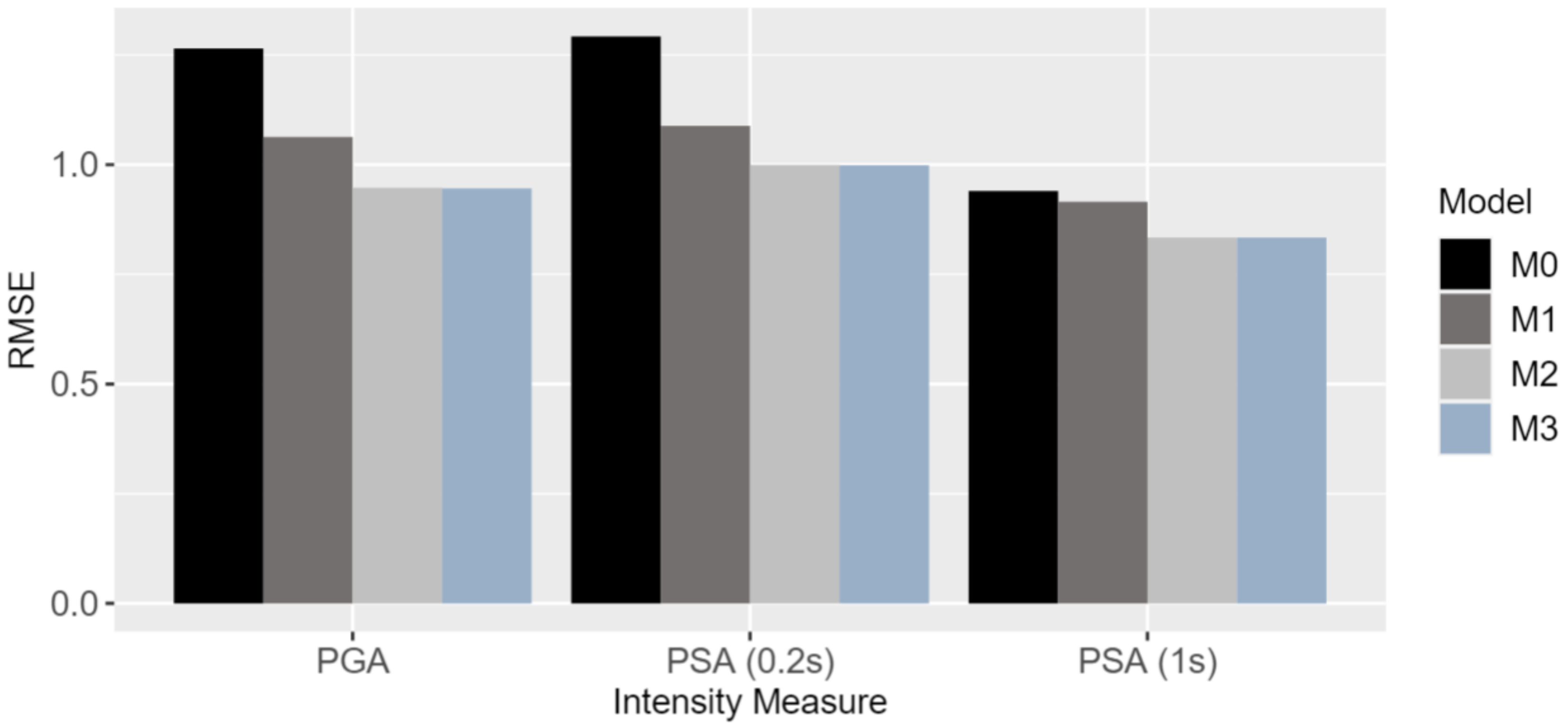
- WAIC scores for M2 and M3 are greater than their DIC scores when the parameters are estimated using AM algorithm.
- WAIC and DIC scores are close to each other for M2 and M3 when the parameters are estimated using BLM, SIR, and AFSS algorithms.
- Maybe select linear model (M2) rather than non-linear model (M3) for Bayesian recalibration of parameters.
- For non-linear functional form SIR algorithm can be employed for better computational efficiency.

Model	Estimation Method	PGA		PSA (0.2s)		PSA (1s)	
		DIC	*WAIC*	*DIC*	*WAIC*	*DIC*	*WAIC*
M1: Added-Bias	BLM	6712.30	6713.18	6742.08	6743.01	5987.16	5988.64
	SIR	6716.35	6716.25	6746.15	6746.08	5991.22	5991.71
	AM	6722.23	6721.19	6749.61	6750.59	5992.52	5993.08
	AFSS	6716.59	6716.43	6746.07	6746.00	5991.18	5991.69
M2: Linear	BLM	6139.22	6140.53	6322.09	6323.29	5633.46	5635.42
	SIR	6140.92	6141.92	6324.39	6325.11	5633.72	5635.73
	AM	6168.59	6171.24	6669.41	9611.83	5793.69	6608.39
	AFSS	6150.25	6144.02	6326.97	6325.25	5636.17	5630.25
M3: Non-Linear	SIR	6135.53	6136.56	6321.81	6322.55	5633.96	5636.01
	AM	6403.54	8584.04	6704.76	9917.72	5768.62	6378.53
	AFSS	6142.98	6136.03	6323.77	6323.94	5638.72	5608.87

Testing

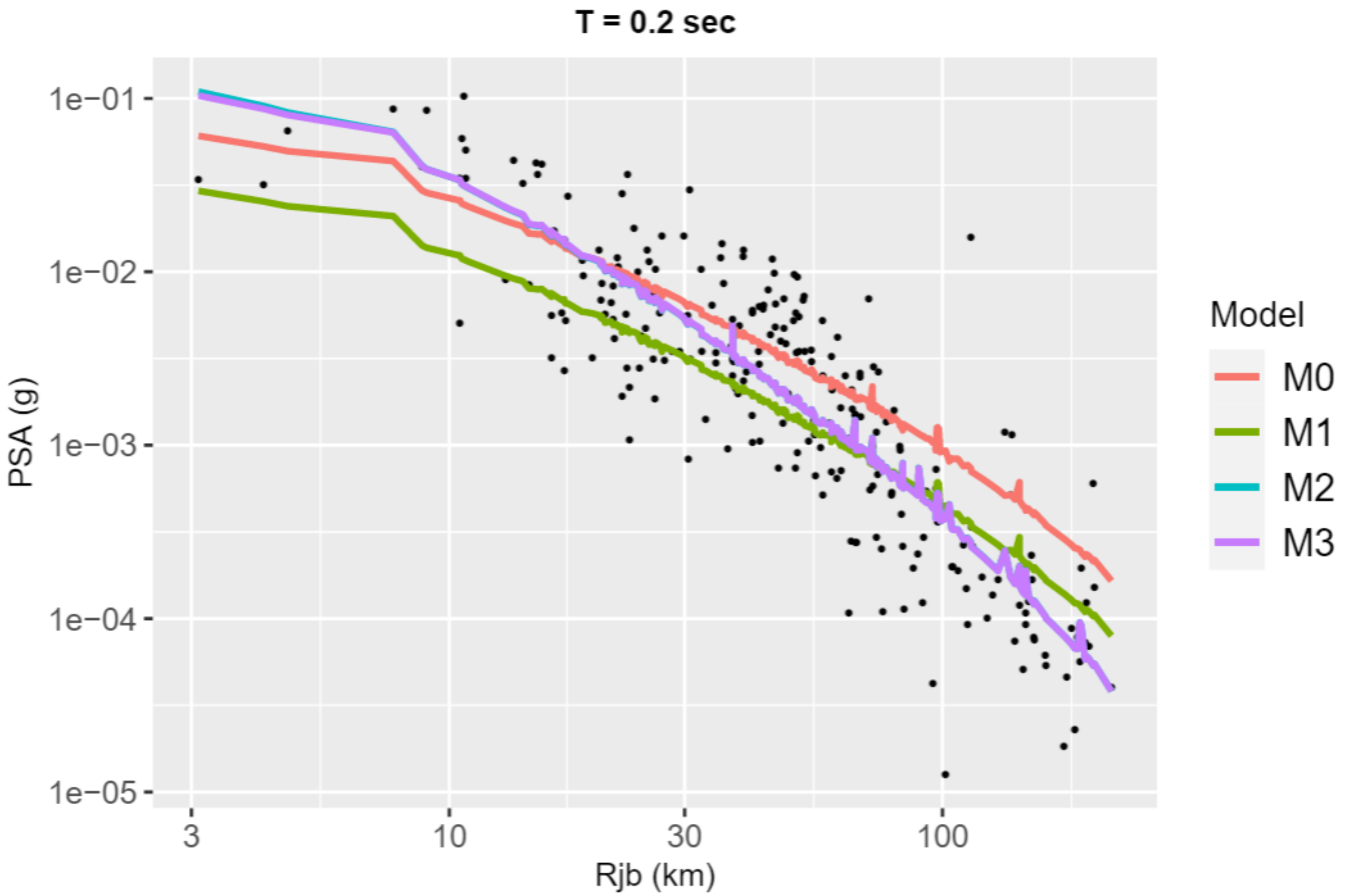
- Single-split

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - E\left[p(y_i \mid x_i, y_{train})\right]\right)^2}{N}}$$



Model Comparison

- Predictions estimated using non-linear model M3 are almost similar to the predictions estimated using linear model M2.



Conclusions

- The linear statistical model (M2) based on the linear functional form of the GMPE can be considered for Bayesian update of parameters and to reduce the RMSE when the GMPE is tested against a new data set.

- The parameters in M2 can be recalibrated using conjugate priors (analytical formulation) approach for Bayesian linear models (BLM) for the best computational efficiency and accuracy.

- The parameters in the non-linear functional form (M3) of the GMPE model can be recalibrated using sampling importance resampling.

Perspectives

- Evaluate the applicability of this study across a wider set of GMPEs and databases in the future.

- Extend this methodology for updating the coefficients in GMPE models and then integrate the model uncertainty by averaging predictions over all the GMPE's using BMA approach.

- Develop scaled backbone GMPE models.

- Integrate Source models and GMPE models to do complete PSHA calculations.

References

- Akkar, S., Sandıkkaya, M., Şenyurt, M., Sisi, A. A., Ay, B., Traversa, P., Douglas, J., Cotton, F., Luzi,L., Hernandez, B. et al. (2014). "Reference database for seismic ground-motion in Europe (RESORCE)." Bulletin of earthquake engineering, 12(1), pp. 311–339.
- Banerjee, S. (2008). "Bayesian linear model: Gory details." Downloaded from http://www. biostat. umn. edu/˜ph7440.
- Bertin, M., Marin, S., Millet, C. and Berge-Thierry, C. (2020). "Using Bayesian model averaging to improve ground motion predictions." Geophysical Journal International, 220(2), pp. 1368–1378.
- Bindi, D., Massa, M., Luzi, L., Ameri, G., Pacor, F., Puglia, R. and Augliera, P. (2014). "Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods up to 3.0 s using the RESORCE dataset." Bulletin of Earthquake Engineering, 12(1), pp. 391–430.
- Bodda, S. S., Keller, M., Gupta, A. and Senfaute, G. (2021). "A Methodological Approach to Update Ground Motion Prediction Models Using Bayesian Inference." Pure Appl. Geophys.
- Lanzano, G., Sgobba, S., Luzi, L., Puglia, R., Pacor, F., Felicetta, C., D'Amico, M., Cotton, F. and Bindi,D. (2019). "The pan-European Engineering Strong Motion (ESM) flatfile: compilation criteria and data statistics." Bulletin of Earthquake Engineering, 17(2), pp. 561–582.
- Statisticat and LLC. (2020). LaplacesDemon: Complete Environment for Bayesian Inference. R package version 16.1.4

Thank you!

Any Questions?