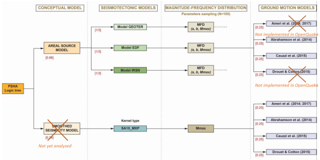
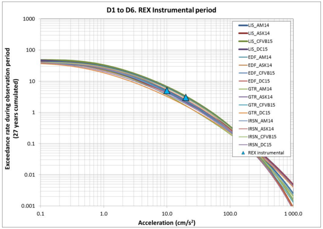
Reconstructing past earthquakes from written testimonies using Bayesian inference

Merlin Keller & Jessie Mayor & Joseph Muré
Paola Traversa, Kévin Manchuel, Guillaume Daniel, Pierre Arroucau

AppliBUGS - December 10, 2021



Drouet et al., (2018) in prep.

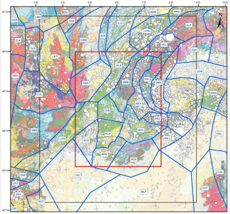


- **Main objectives**: Reduce the uncertainties tainting the seismic hazard curve, by computing in a Bayes-optimal way the weights attributed to the different branches of the PSHA logic tree;
- **Application to France**: Low sismicity, hence few observed data and high uncertainties, with high safety/regulatory constraints to meet
- **Industrial stakes**: Reducing uncertainties on hazard curves and ground spectrum, two key inputs of seismic PSAs, used to assess the capacity of a given structure to whithstand earthquakes.

Overview of PSHA models

SM-model (seismotectonic model)

Subdivision of the region under study in I homogeneous seismicity areas, described by recurrence parameters:

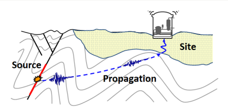


- λ_i : parmeter of the Poisson distribution of annual earthquake counts in i-th area
- β_i : exponential law parameter of the magnitude distribution (Gutenberg-Richter (1944) law).

$$\lambda = \lambda_0 \frac{e^{-\beta(m-M_{max})} - e^{-\beta(m-M_{min})}}{e^{-\beta(M_{max}-M_{min})} - e^{-\beta(M_{max}-M_{min})}}$$

Ground-Motion model (through GMPE)

Relation between an earthquake scenario (M, R) and the ground motion intensity :



$$\log_{10}(a) = G(m, r, \nu) + \epsilon,$$

with a = PGA (peak ground acceleration) or other ;

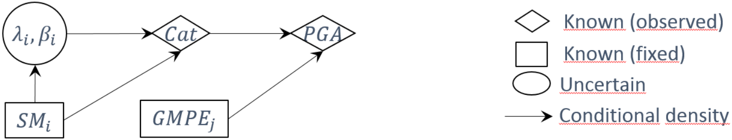
Bayesian updating of model weights (1/2)

- Uncertainty on the choice of a sismotectonical model SM and ground motion prediction equation $GMPE$, described through prior weights $\pi(\mathcal{M})$ for $\mathcal{M} = (SM, GMPE)$ with $SM \in (SM_1, \ldots, SM_K)$ et $GMPE \in (GMPE_1, \ldots, GMPE_L)$
- Bayesian update of model weights given available data according to:

$$\pi(\mathcal{M}|obs) = \frac{m(obs|\mathcal{M})\pi(\mathcal{M})}{\sum_{\mathcal{M}'} m(obs|\mathcal{M}')\pi(\mathcal{M}')},$$

  where $obs = (CAT, PGA)$ represents the data (earthquake catalog and accelerometric records)
- Two possible uses of the updated weights:
  - selection: by maximizing $\pi(\mathcal{M}|obs)$
  - agregation: by averaging predictions within each model $\mathcal{M}$, weighted by $\pi(\mathcal{M}|obs)$ (Bayesian model averaging, ou BMA)

Bayesian updating of model weights (2/2)



- Boils down to compute likelihood $m(obs|\mathcal{M})$, with $obs = (Cat, PGA)$;
- Can be factored into:

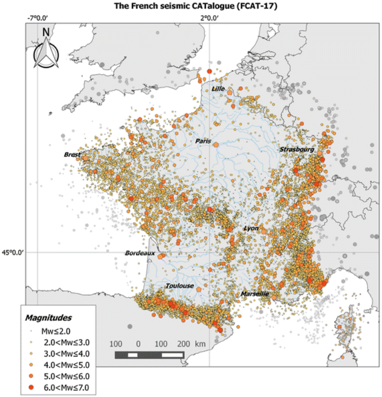
$$
\begin{aligned}
m(obs|\mathcal{M}) &= \mathcal{L}(PGA|CAT, GMPE) \times \int \mathcal{L}(CAT|\lambda, \beta, SM)\pi(\lambda, \beta)d\lambda d\beta \\
&= \mathcal{L}(PGA|CAT, GMPE) \\
&\quad \times \prod_i \int \mathcal{L}(CAT_i|\lambda_i, \beta_i, SM)\pi(\lambda_i, \beta_i)d\lambda_i d\beta_i,
\end{aligned}
$$

with CAT_i the seismic catalog for zone i of SM, with recurrence parameters λ_i, β_i.

- GMPE model weights: **see next talk!**
- Remaining factors obtained by importance sampling (Keller et al. (2021)).

The FCAT-17 seismic catalog

- **Recent earthquakes:** location, magnitude and depth estimated from instrumental measures
  - From 1900 to 1960: first European sismographs
  - Since 1960: first French accelerometric networks
- **ancient earthquakes:** location, magnitude and depth estimated from reported / measured damages

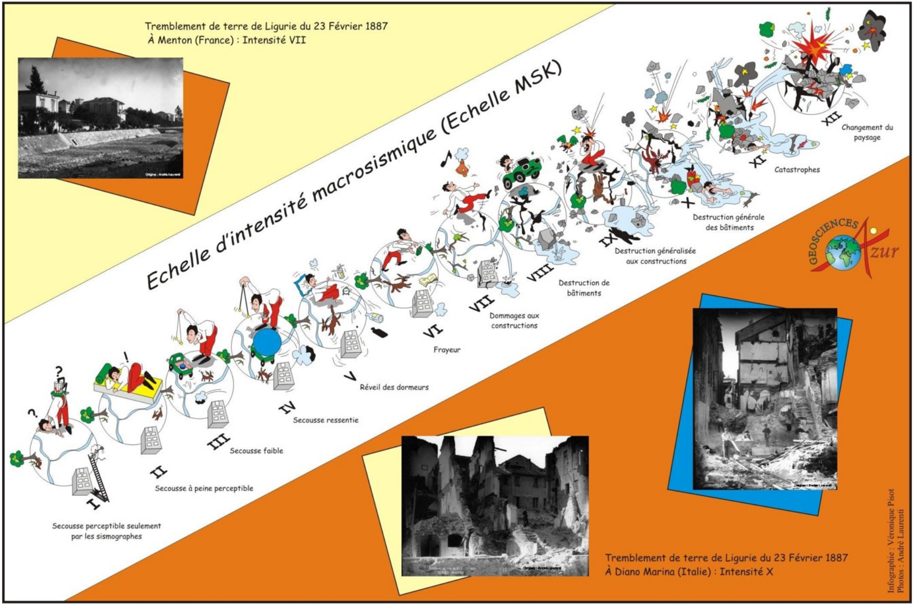


- Overall, $\sim 40\ 000$ earthquakes, from year 463 to $2\ 009$

Objectives of this work

Improve existing approach for ancient earthquake estimation

Macroseismic intensity index: the MSK scale



Modeling the macroseismic intensities

- In FCAT-17, the following intensity prediction equation (IPE) is used (following Baumont et al., 2018):

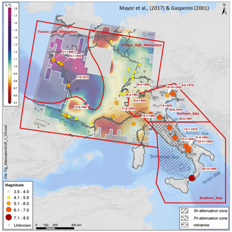
$$I_i = C_1 + C_2 M_w + \beta \log_{10}(R_i) + \gamma R_i,$$

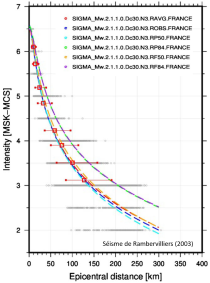
with

  - I_i macroseismic intensity at location i;
  - $R_i = \sqrt{r_i^2 + H^2}$ hypocentral distance, given r_i (epicentral distance) and H (depth);
  - C_1, C_2 source terms, β geometric (elastic) and γ (anelastic) attenuation.

- Current IPE methodology accounts for epistemic uncertainties:
  - regionalized γ properties
  - metric for R_i to adjust to discrete intensities

$\Rightarrow$ 6 regional IPEs used in FCAT-17





Proposed improvements

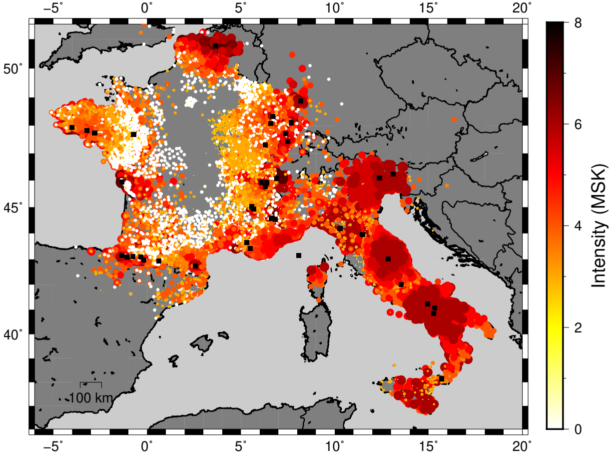
- Avoid having to define a single hypocentral distance for all observations sharing the same discrete intensity by data augmention techniques;
- Use a data-driven clustering method to learn the regional dependance of the apparent attenuation (using a simplified IPE)
- Add prior information on interest parameters through a Bayesian approach

$\Rightarrow$ we have applied the above methodological contribution to the 3 main steps involved in historical seismicity:

1. Clusterize the data;
2. Calibrate IPE within each identified cluster;
3. Invert IPE for a new event, selecting between calibrated IPEs based on model selection criterion

Sommaire

Calibration database (Baumont et al., 2018)



TOTAL : 41 events ; $M_w = [3.6 - 7.1]$; $I_0 = [\text{III - VIII}]$; $H = [2 - 30]$ Km
$+$ associated macroseismic fields (I_i, D_i, Q_i values, over 13000 datapoints)

IPE modeling: beyond hypocentral metrics

- FCAT-17 model (Baumont et al., 2018) + added site effect

$$\begin{aligned} I_i &= C_1 + C_2 M_w + \beta \log_{10}(R_i) + \gamma R_i + \eta_i \\ \eta_i &\overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \end{aligned}$$

- Observation I_i^{obs} modeled as a noisy censored version of I_i

$$\begin{aligned} I_i^{obs} &= \delta k, \quad \text{s.t.} \quad I_i + \varepsilon_i \in [(k-1/2)\delta, (k+1/2)\delta[\\ \varepsilon_i &\overset{ind.}{\sim} \mathcal{N}(0, 1/q_i) \end{aligned} \tag{1}$$

- Bayesian modeling: uniform priors chosen for all model parameters to enforce physically plausible bounds
- Goals: in the following, we want to determine
  - the posterior distribution of $\theta = (C_1, C_2, \beta, \gamma, \sigma)$ (calibration) given the 41 calibration events
  - the posterior distribution of $\eta = (Mw, H)$ (inversion) given a new event and a "well-chosen" θ value (here, the calibration posterior median)

,

Bayesian updating algorithm

- The Laplace approximation can be used exploiting the likelihood integrated over latent variables I_i :

$$\mathcal{L}(I_i^{obs}|d_i, q_i, \theta) = \Phi\left(\frac{I_i^{obs} - G(d_i;\theta) + \delta/2}{\sqrt{\sigma^2 + 1/q_i}}\right) - \Phi\left(\frac{I_i^{obs} - G(d_i;\theta) - \delta/2}{\sqrt{\sigma^2 + 1/q_i}}\right)$$

however, this leads to recurrent numerical instabilities

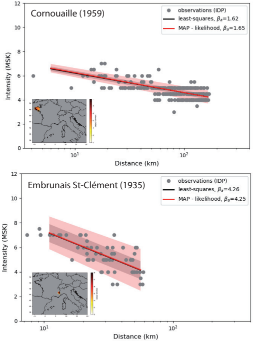
- Instead, we used MCMC algorithms to sample from the high-dimensional posterior distribution of $(\theta, (I_i)_{1\leqslant i\leqslant n})$ (calibration) or $(\eta, (I_i)_{1\leqslant i\leqslant n})$ (inversion)
- To select the right calibrated IPE to invert for a new event, we used the DIC criterion, which boils down in our case to compute the averaged posterior likelihood

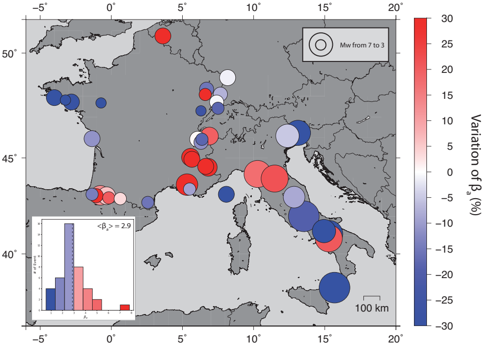
Technical details

- Metropolis within Gibbs used to benefit from partial conjugacy properties: in the calibration (resp. inversion) step, σ (resp. H) is the only variable whose conditional posterior distribution is not available in closed-form, and is sampled using random-walk Metropolis-Hastings;
- Python implementation based on a prototype for new MCMC classes in Open-TURNS `https://openturns.github.io/`;
- Currently, the random-walk proposal standard step must be set manually; in future versions it will adapt automatically.

- Goal: group calibration earthquakes with similar apparent attenuation β_a
- Bayesian estimation of β_a (Laplace approx.) using MAP (Median A Posteriori) using simplified IPE:

$$I_i = C_1 - \beta_a \times \log_{10}(R_i) + \eta_i$$

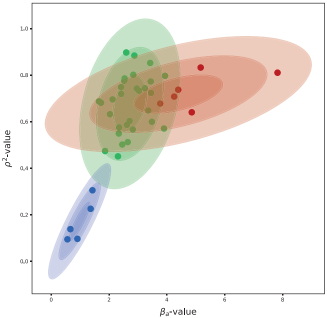




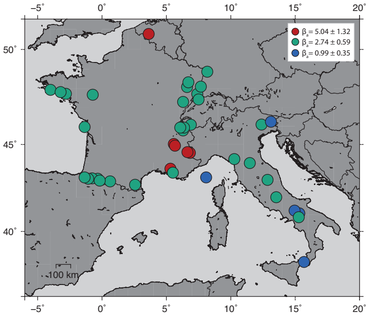
3 clusters for β_a

- Gaussian Mixture Modeling (GMM) of joint features (β_a, ρ^2) with ρ^2 the linear correlation coefficient
- Principal component analysis to identify relevant clusters (β_a, ρ^2)
- Bayesian information criterion (BIC) used to find optimal number k of clusters ($\Rightarrow$ k = 3)

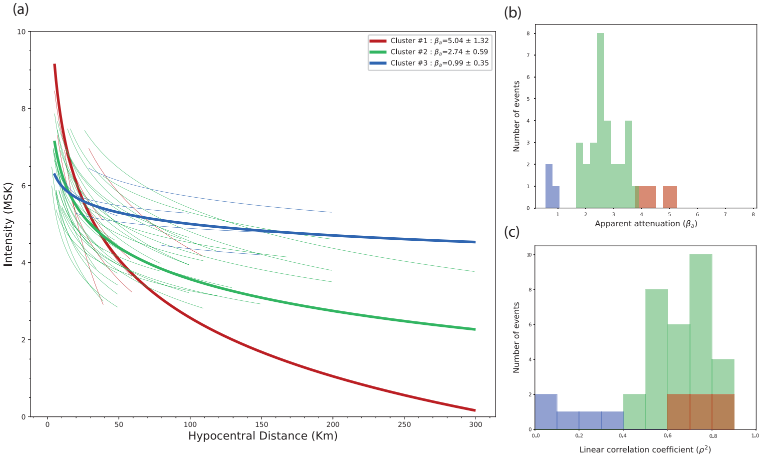
(a)



(b)

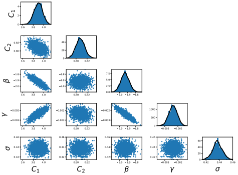


⇒ strong - intermediate - weak apparent attenuation



IPE calibration for whole dataset

- Recall that: $I_i = C_1 + C_2 M_w + \beta \log_{10}(R_i) + \gamma R_i + \eta_i$
- Posterior scatterplots for $\theta = \{C_1; C_2; \beta; \gamma; \sigma\}$

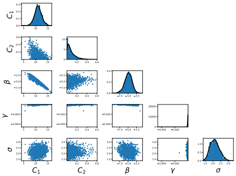


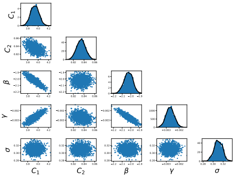
- MAP (Median A Posteriori) values

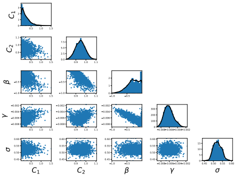
Cluster	C_1	C_2	β	γ	σ	σ_{C_1}	σ_{C_2}	σ_β	σ_γ	σ_σ
whole dataset	3.90	0.81	-1.93	-0.0024	0.44	0.08	0.008	0.05	0.0003	0.007

- posterior correlations between C_1 and C_2, and between C_1, β and γ

- Recall that: $I_i = C_1 + C_2 M_w + \beta \log_{10}(R_i) + \gamma R_i + \eta_i$
- Posterior scatterplots for $\theta = \{C_1; C_2; \beta; \gamma; \sigma\}$







- MAP (Median A Posteriori) values

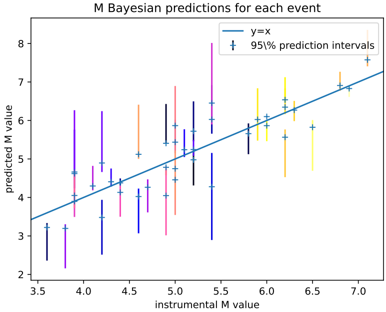
Cluster	C_1	C_2	β	γ	σ	σ_{C_1}	σ_{C_2}	σ_β	σ_γ	σ_σ
high-Q_a	10.97	0.06	-4.91	-0.0000193	2.10	1.50	0.09	1.10	0.0001	0.000112
medium-Q_a	3.94	0.83	-2.03	-0.0027	0.31	0.08	0.0084	0.05	0.00032	0.0084
low-Q_a	0.16	0.95	-0.27	-0.0062	0.52	0.21	0.048	0.20	0.0011	0.023

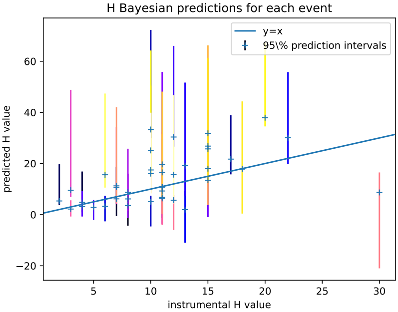
- Parameter estimation difficult in clusters 1 & 3 due to reduced dataset
- posterior correlations between C_1 and C_2, and between C_1, β and γ

Cross-validation study

- Idea: compare (M_w, H) instrumental values to estimates obtained by IPE inversion
- Calibration parameters θ set to their posterior median
- use DIC criterion to allocate each event to a cluster

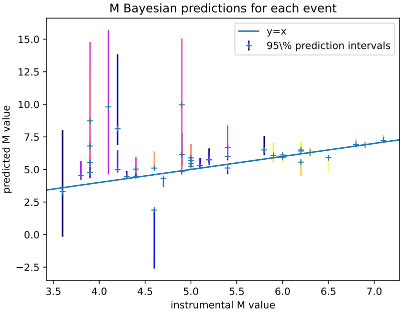
Cross-validation: whole dataset

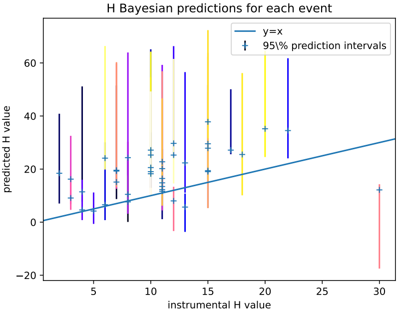




- Significant prediction errors
- Prediction uncertainties are under-estimated

Cross-validation: clusters





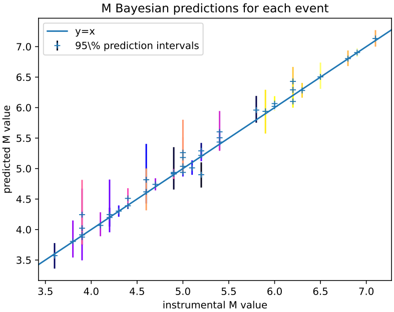
- Clustering reduces prediction errors (especially for M values)
- Increased prediction uncertainties due to smaller datasets

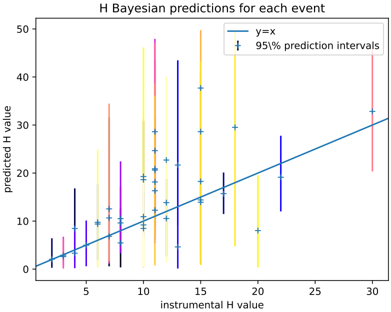
Simulation study

To validate and better interpret the above results, we have conducted a simulation study wherein:

- I^{obs} synthetic values are simulated according to our hierarchical IPE model, setting calibration parameters to their MAP
- For each calibration event, the IPE is inverted having set calibration parameters to their "true" value (used to simulate the data)
- The results represent an <u>oracle</u>, that is, the ideal results we would obtain if macroseismic intensities where perfectly explained by the IPE model, and calibration parameters where known in advance

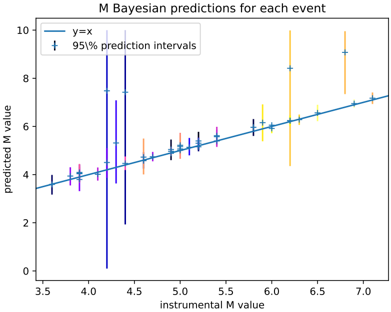
Cross-validation: whole dataset (simulations)

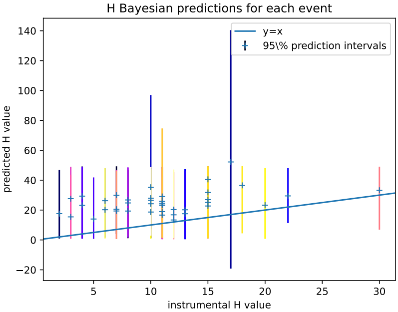




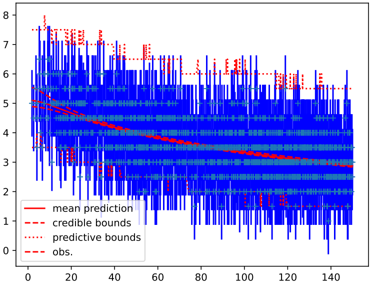
- Good prediction accuracy for M
- H remains difficult to predict

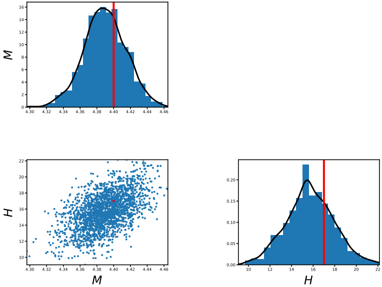
Cross-validation: clusters (simulations)





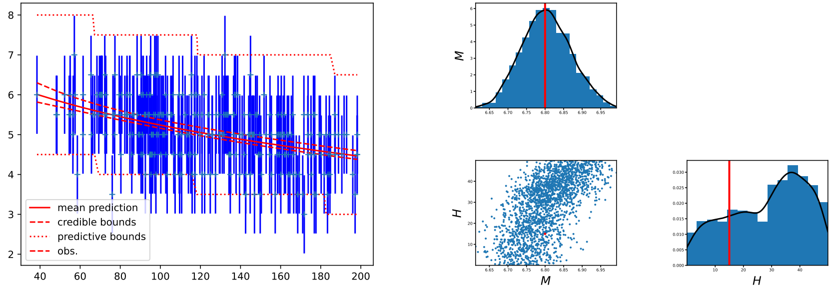
- Similar results to whole dataset simulations
- Locally large prediction uncertainties due to cluster selection





- IPE model fits the data well, in spite of slight bias
- (M_w, H) well identified, but with large uncertainties

Inversion results for event 10001675 (simulations)



- IPE model fits the data well, in spite of slight bias
- multiple (Mw, H) explain the data equally well: identifiability issue

Discussion

- mature statistical methodology to account for:
  1. expert information, through Bayesian inference
  2. discrete macroseismic intensities, avoiding the choice of hypocentral distance metrics
  3. intrinsic (apparent) attenuation properties for each calibration event
- Cross-validation results point to identifiability issues concerning H and M_w

Perspectives

- Need to reduce uncertainties / improve identifability
- several promising avenues:
  - merge clustering and calibration steps in a mixture modeling approach to better fit the data
  - variable selection to reduce model complexity and enhance identifiability
  - integrate calibration parameters uncertainty
- More detailed comparison to current methodology may provide complementary leads