# A Smart Elicitation Technique for Informative Priors in Ground-Motion Mixture Modelling

A. K. Runge, A. Händel, C. Riggelsen, F. Scherbaum — Institute of Earth and Environmental Science, University of Potsdam

Antonia.Runge@geo.uni-potsdam.de, Annabel.Haendel@uni-potsdam.de, Carsten.Riggelsen@geo.uni-potsdam.de Frank.Scherbaum@geo.uni-potsdam.de



## 1. Motivation

**Geoscientific processes and systems** are often described by models making simplifying assumptions. Insufficient/incomplete observations and poor/inadequate understanding of the underlying relationships often result in the development of competing models.

In **probabilistic seismic hazard analysis** (PSHA), e.g., the ground motion at a particular site of interest is typically estimated as an empirical function of source, path and site related parameters.

Following issues arise:

- **Competing ground motion prediction equations** (GMPEs) developed for sites with many available observations (e.g., California, etc.) capture different physical aspects.
- **Foreign GMPEs** must be applied if sparse seismic data is observed or if no dedicated model was developed for a site.
- One **large source of uncertainty** is the selection and judgement of appropriate GMPEs.

## 2. Bayesian Mixture Modelling

**A standard mixture model** aggregates several existing GMPEs (Fig. 1), instead of having a single model that tries to capture the possible ground motion at the site of interest. [a]

GMPE$_{mix}(x) = \sum_{i=1}^{n} w_i \text{GMPE}_i(x)$ , for $n$ component GMPE$_i$ .

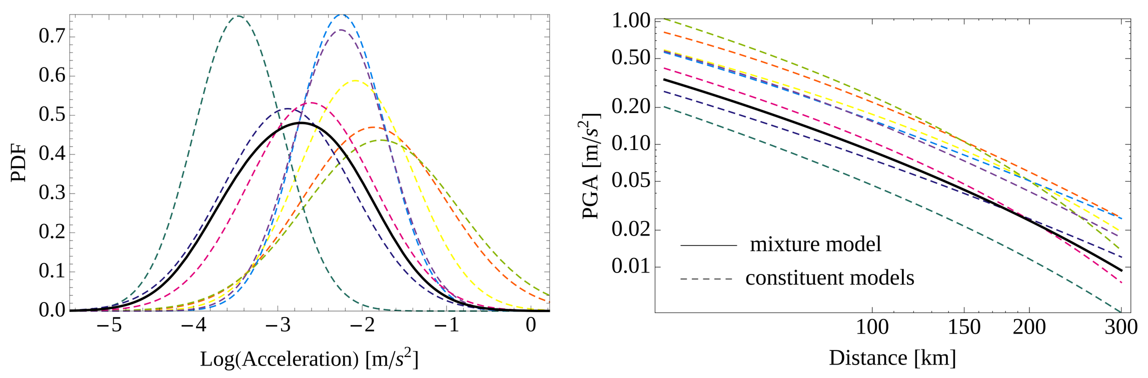

**Fig. 1:** GMPEs and mixture model for an Mw 6 interface event. Left: distribution in 120 km distance. Right: peak ground acceleration (PGA) versus distance.

**The mixture weights** capture appropriateness of GMPEs in the mixture. Their distribution is estimated within a Bayesian statistical framework:

$$f_{\text{posterior}}(w_i) \propto f_{\text{prior}}(w_i) \cdot p(data|w_i)\,,$$

for the likelihood $p(data|w_i)$ of observations $data$ .

**A priori distributions** are typically chosen based on algebraic and/or computational convenience rather than attempting to **capture domain expert's beliefs**. Partly because it is thought to be a non-trivial task to elicit expert knowledge in terms of a distribution.

**A major challenge for experts** in this context is: [b]

- to provide logically consistent probability estimates (in the sense of Kolmogorov's axioms),
- to be aware of the multitude of heuristics, and
- to minimise biases affecting judgement under uncertainty.

[a] Riggelsen et al. (2011)

[b] O'Hagan et al. (2006), Scherbaum and Kühn (2011), Tversky and Kahneman (1974)

## 3. A Smart Elicitation Tool (SmElT) to Quantify Expert Belief

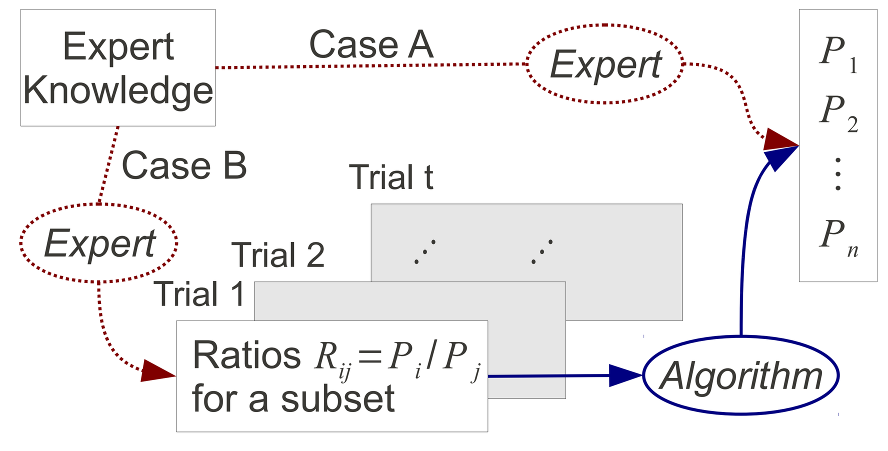


**Fig. 2:** Representation of two different ways to elicit subjective probabilities.

**Simple elicitation task** with the platform-independent, interactive program: [a]

- Quantification, elicitation and transfer of expert knowledge into degree of belief (DOB) distributions $f_{\text{prior}}(w_i)$ .
- Sequential evaluation of relative weights for small model subsets, instead of a single step task (Fig. 2 and 3).

[a] A.K. Runge et al. (2013)

**Optimization** of the model subset presented to the expert in each trial: [a]

- Experimental design theory is applied.
- The design maximizes the expected information based on all previously conducted trials.

[a] Curtis and Wood (2004)

**Results** of the elicitation process:

- A set of logically consistent probabilities: best-fit solution to the set of elicited constraints (Fig. 4).
- A measure of confidence: amount of conflicting information provided by the expert during the relative weighting process (Fig. 5).

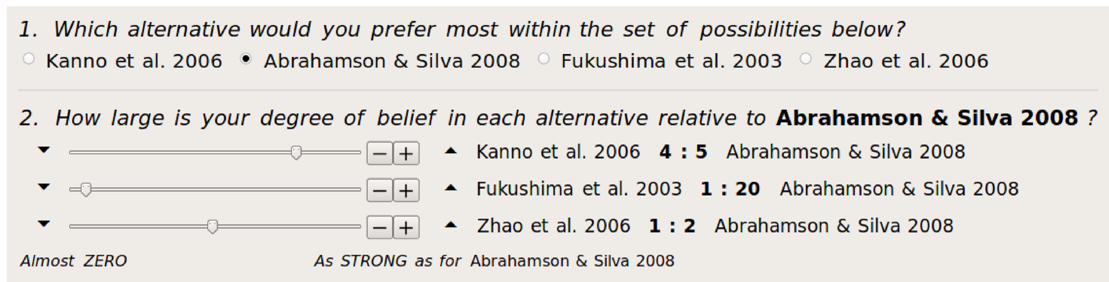

**Fig. 3:** A fictional elicitation trial conducted with our program, where 4 GMPEs are judged relativ to each other.

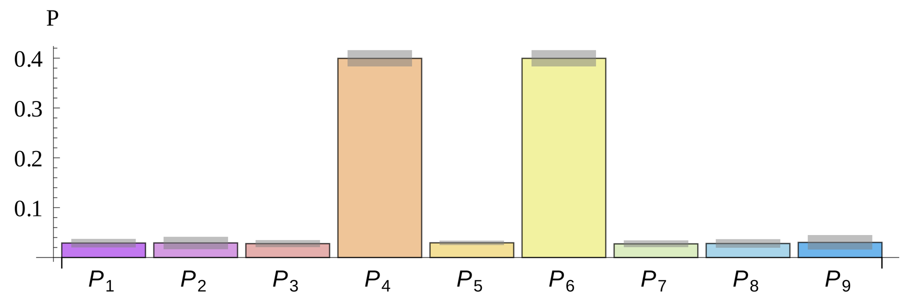

**Fig. 4**: Distribution of an expert's DOB (coloured boxes) and corresponding residual uncertainty (grey boxes).


**Fig. 5:** Residual error (SER) of all elicited statements compared to a zero error ("most consistent") and to the error in the case of "random statements".

## 4. Data and GMPEs

**Data:** 371 interface and 713 intraslab strong-motion records (49 and 90 events respectively) from Northern and Central Chile coming from:

- Arango et al. (2012),
- IPOC network, year 2006 - 2012.

**GMPEs:** 9 GMPEs developed for different subduction zones of the world.

| Model | Short |
|---|---|
| Youngs et al. (1997) | y97 |
| Zhao et al. (2006) | z06 |
| Kanno et al. (2006) | k06 |
| Atkinson and Boore (2003) | ab03 |
| Lin and Lee (2008) | ll08 |
| BC Hydro (2013) | bch13 |
| Mc Verry et al. (2006) | mv06 |
| Arroyo et al. (2010) | a10 |
| Garcia et al. (2005) | g05 |

## 5. The Impact of Priors

**Do large numbers of records overrule the prior distribution?**

- Define a priori distributions $f_{\text{prior}}(w_i)$ based on different scenarios of likely expert behaviors.
- Kullback-Leibler distance (KLD) measures the information difference of two distributions:

$$\text{KLD} = \int \log_2\left(\frac{f_{\text{prior}}(x)}{f_{\text{posterior}}(x)}\right) f_{\text{prior}}(x)\mathrm{d}x\,.$$

- KLD measures information difference in bits: "How many questions need to be asked to return from $f_{\text{posterior}}(x)$ to $f_{\text{prior}}(x)$ ?"
- **Result**: If the prior is too sharp (small variance) the data is not able to shift the posterior into its supported region, even if many records are available (Fig. 6).

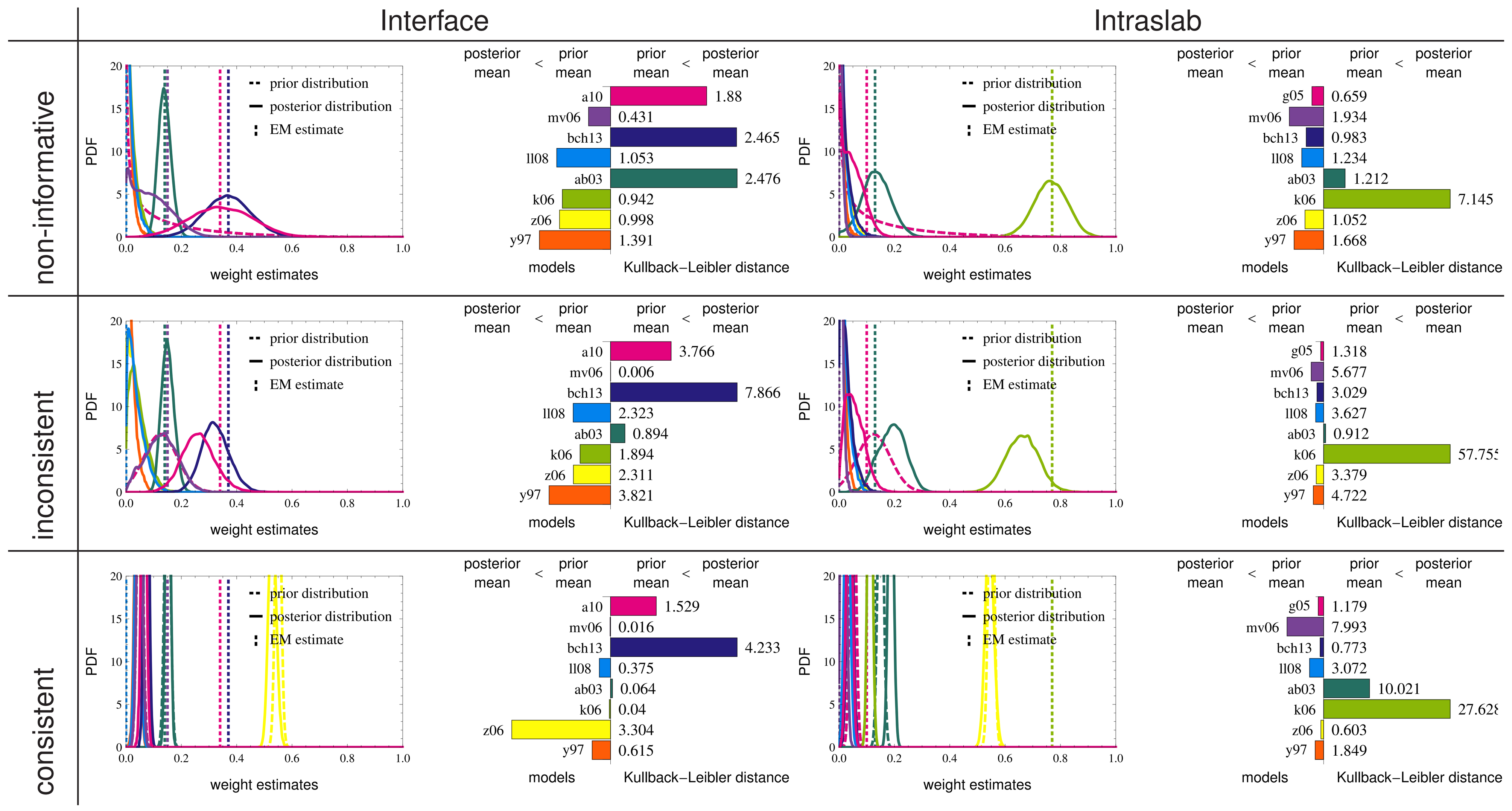

**Fig. 6:** $f_{\text{prior}}(w_i)$ and $f_{\text{posterior}}(w_i)$ after the update with 371 interface or 713 intraslab records and KLDs.

## 6. The Impact of Data

**Does information converge at a particular number of records or events?** - Analysis of the KLDs:

- Apply non-informative priors to measure solely the influence of data.
- Add events as they occur in time, backwards and randomly.
- **Result**: Some data or events shift particular mixture-weight distributions more than others. Distinguish and interpret the reasons for and against the support of different models by the data (Fig. 7).

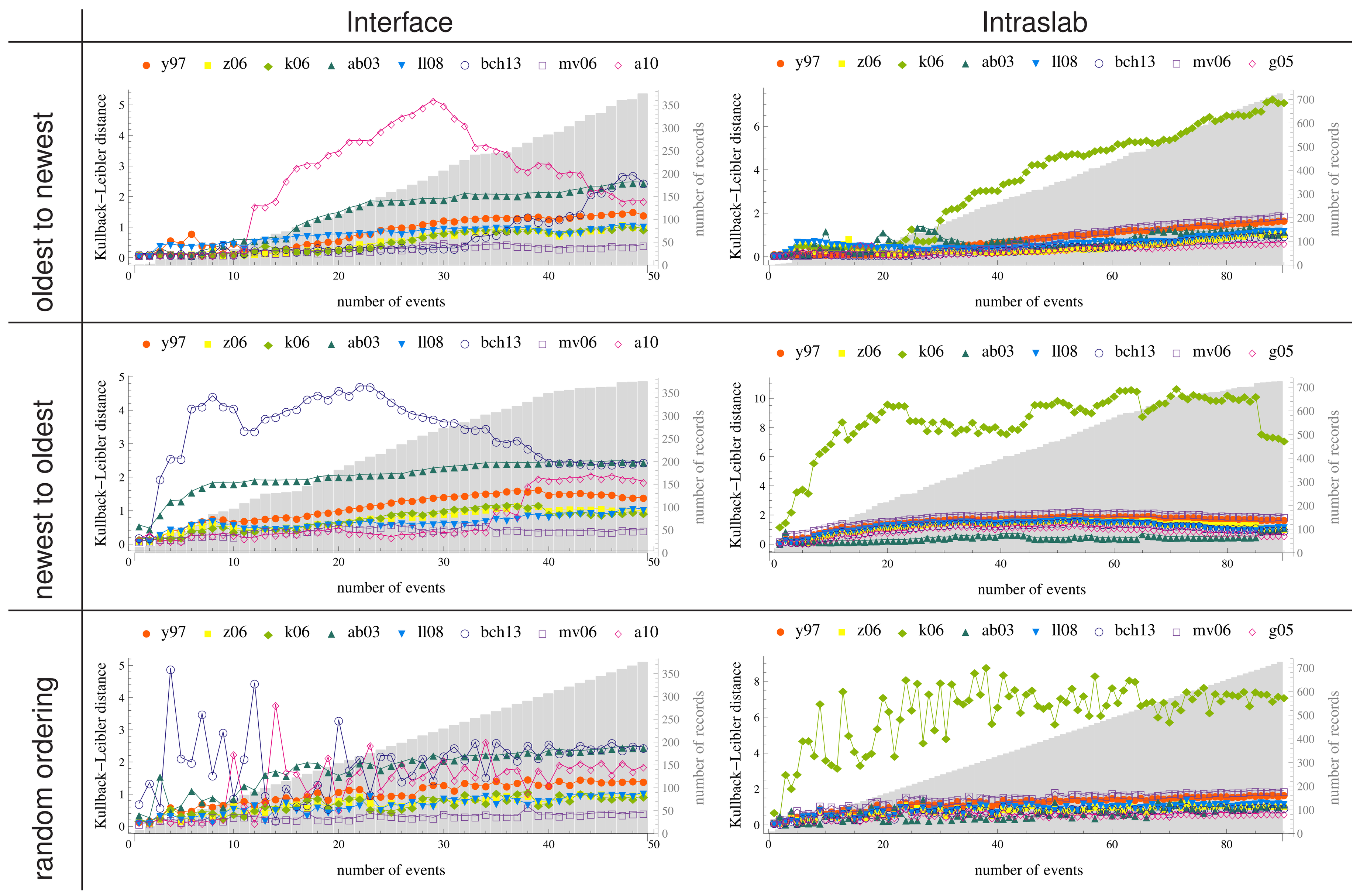

**Fig. 7:** KLDs with an increasing number of interface or intraslab events for 3 different orderings.