Università degli Studi di Napoli Federico II
Scuola politecnica e delle scienze di Base
Dipartimento di Strutture per l'Ingegneria e l'Architettura (DiSt)
Consorzio Interuniversitario ReLuis

WORKSHOP: Approcci per la valutazione dei modelli di pericolosità sismica in Italia

Villa Orlandi, Anacapri, 7-8 settembre 2023

Warner Marzocchi, Università degli Studi di Napoli Federico II

La fase di test di un modello di pericolosità: problemi e possibili soluzioni

(con applicazioni a MPS19, MPS04 e ESHM20)

Some thoughts on the testing of seismic hazard models

Prof. Warner Marzocchi
(warner.marzocchi@unina.it)

Some general considerations...

About terminology: Consistency, Calibration, Validation

All tests compare model forecasts with data.

Consistency: check if the forecasts of the model are consistent with the data. In this case the data used are the same used to set up the model.

Calibration: check if the forecasts of the model are consistent with independent data. Here the epistemic uncertainty is not considered (conservative tests)

Validation: check if the forecasts of the model are consistent with independent data. Here the epistemic uncertainty is considered. Testing the ontological hypothesis

Some general considerations...

About terminology and the probabilistic framework

- If we consider the “subjective” (sometimes inappropriately called “Bayesian”) interpretation of the probability (e.g., D’Amico and Albarello, 2008), testing one model against the data has a very limited interest (“all models are wrong”, so why wasting time and energy in testing any model?). In fact, in this framework, it makes much more sense to compare the performances of models (Lindley, 2000; Gelmann et al., 2003).

- In this framework, the “significance” of a test does not make sense. Ssubjectivists never test their model according to the Neumann-Pearson approach to model testing, where the significance level has been defined.

- If we consider the “subjective” interpretation of the probability, the uncertainty is only epistemic (bunch of literature about this). Aleatory variability does not exist, and only the mean hazard makes sense (no value at the distribution of the branches).

Some general considerations...

In PSHA, as well as in most of similar enterprises, one model produces the forecasts that are compared with data

This implies that, if the model is correct, some of the observations **must** be on the tails of the forecast distributions.
This is unavoidable.
If we have no observations on tails of the distribution, we are **overfitting** the data.

Some general considerations...

A different approach used in PSHA, used also for testing the model: Leaving the data (ground shaking at the sites) speak for themselves (D'Amico and Albarello, 2008)

This approach has important consequences:

- There is not a single model behind that can produce ground shaking in all sites, because we have no observations on the tail of the forecast distribution.

- It is assumed to have enough data to calculate "exactly" the rates of exceedances. This is particularly problematic. Let us assume that we have a catalog of 475 years. The 10%in50 years ground shaking is the maximum observed in the catalog (almost) no matter what has been observed in the same site in the same period of time.

- Even more important: according to the model, there is (mostly) zero probability to exceed in the next 50 years what has been observed in the catalog.

The MPS19 approach

Fase di test 1: consistenza dei modelli di sismicità. Si verifica la consistenza dei modelli di sismicità utilizzando il catalogo storico rilasciato dal Tavolo 2 (CPTI15). In particolare, si verifica se i modelli di sismicità forniti per il Tavolo 3 sono in grado di spiegare in maniera soddisfacente la sismicità osservata negli ultimi secoli, sia in termini di numero di terremoti, che nella loro distribuzione spaziale. Lo scopo finale è di verificare quale modello non sia adeguato per contribuire al modello finale MPS19.

Fase di test 2: *scoring* dei modelli di sismicità e dello scuotimento del terreno. Si attribuisce un punteggio (*score*) ad ogni modello di sismicità e di scuotimento del terreno in base ad un confronto delle performance predittive di tutti i modelli. Questo tipo di informazione è successivamente utilizzato per la pesatura di ogni componente all'interno del modello finale MPS19.

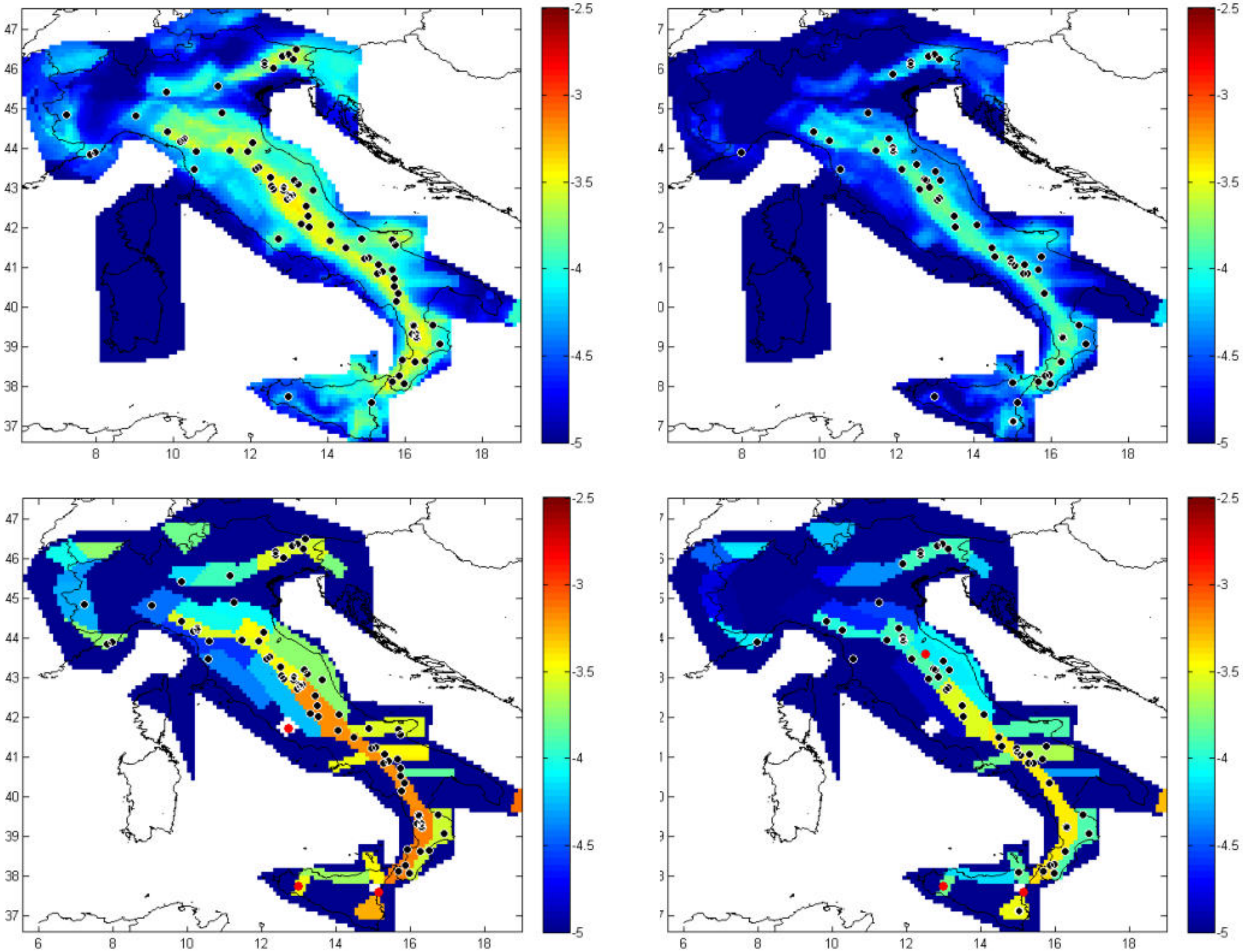
Fase di test 3. La consistenza del modello finale MPS19. Si valuta la coerenza del modello finale MPS19 tramite un confronto con i dati accelerometrici registrati negli ultimi decenni in Italia e contenuti nel database ITACA. In particolare, si vede se il numero di eccedenze in PGA registrate negli ultimi decenni sono compatibili con il modello MPS19. La stessa modalità potrà essere utilizzata per una validazione del modello utilizzando i dati futuri.

Fase di test 4. La consistenza del modello finale MPS19 in termini di intensità macrosismica. Per utilizzare un set di dati più ricco e indipendente da quello utilizzato nella fase di test 3, si considerano le intensità macrosismiche contenute nel database DBMI15 per verificare se le intensità previste dal modello MPS19 sono compatibili con quanto osservato negli ultimi secoli. In questo modo si utilizza un dataset più corposo per la fase di test, ma questa fase di test necessita il passaggio da accelerazioni ad intensità. Questo passaggio non è parte integrante del modello MPS19, per cui eventuali risultati negativi dovrebbero essere valutati con attenzione, poiché potrebbero indicare sia una inadeguatezza del modello MSP18, che un'errata legge di trasformazione da accelerazioni ad intensità.

FASE 1: test sui tassi di sismicità e confronto con il modello MPS04

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

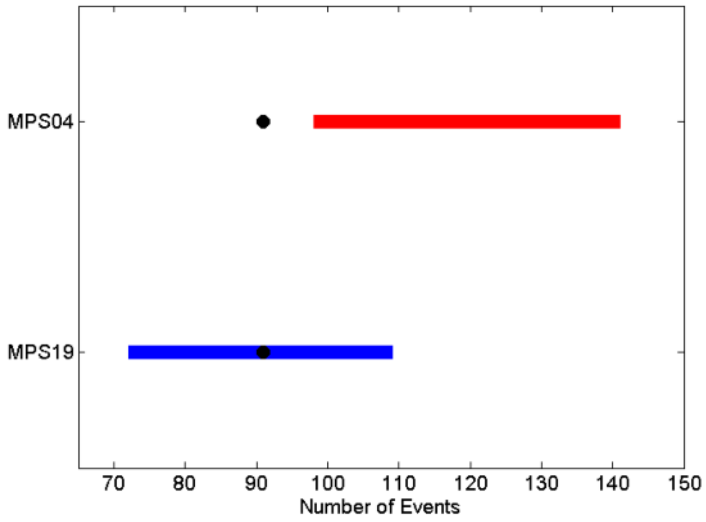
M 5.6+ | M 6.0+



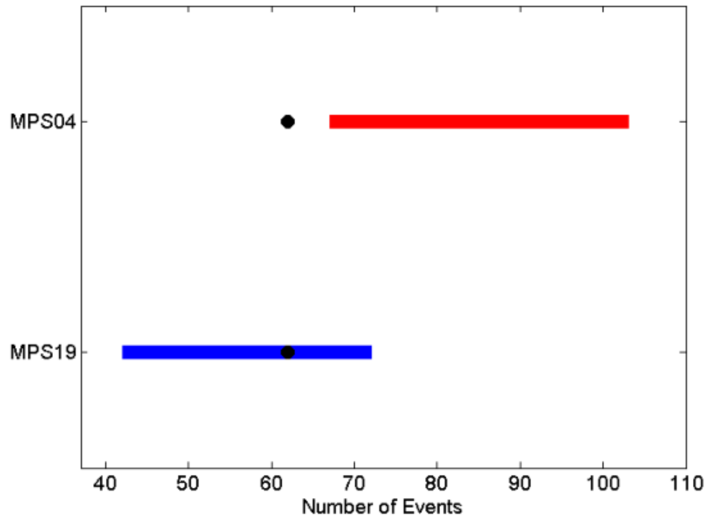
MPS19

MPS04

M 5.6+



M 6.0+



Catalogo: CPTI15 Mw 5.55+ 1787 – 2014 (91 eventi)
Eventi Italia in terra (75 eventi)
Pgain = exp(2095)
Eventi Italia in terra meno i tre peggiori di MPS04 (72 eventi):
Pgain = 6.6

Catalogo: CPTI15 Mw 5.95+ 1660 – 2014 (62 eventi)
Eventi Italia in terra (54 eventi)
Pgain = exp(2092)
Eventi Italia in terra meno i tre peggiori di MPS04 (51 eventi):
Pgain = 0.2

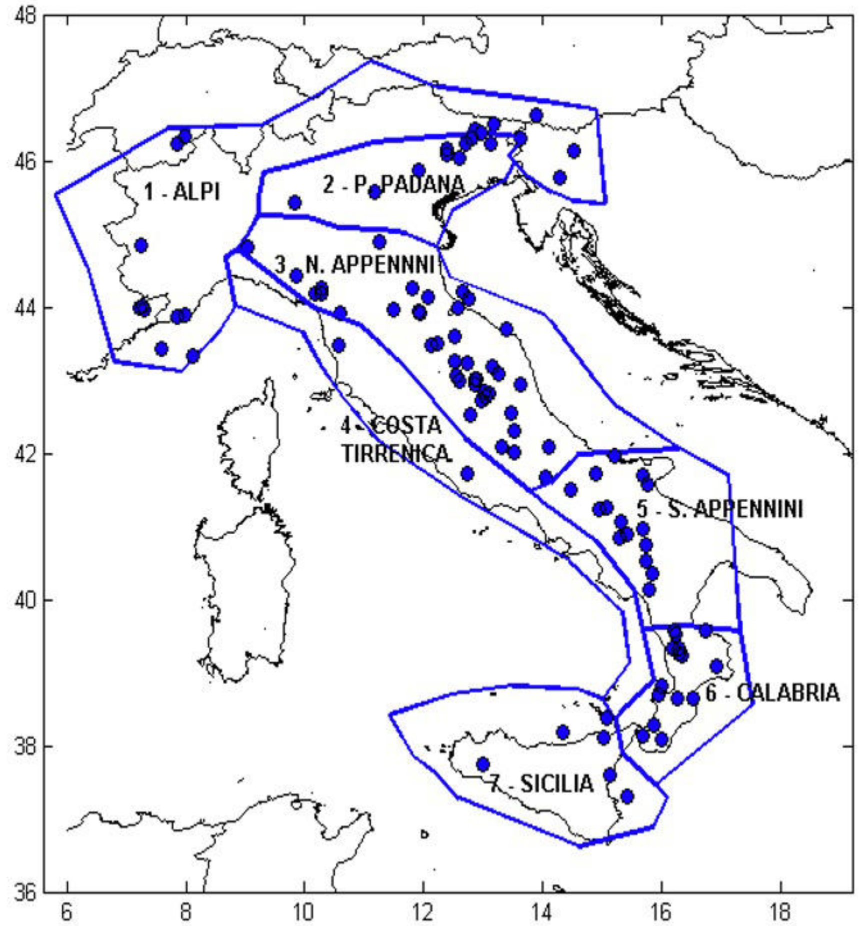
Some thoughts on the MPS19 testing phase

- We check the consistency of M5.6+ and M6+ in historical catalog (a few centuries).

- We check the **number of earthquakes** and their **spatial occurrences**, subdividing Italy in regions (*see figure aside; we cannot use the CSEP grid because it does not match the epicentral uncertainties for historical earthquakes*)

- We include epistemic uncertainty in the consistency testing; using only the **mean hazard, or mean rate,** the tests are overly restrictive (e.g., **rejecting** a model does not mean that the model is bad). Including epistemic uncertainty in testing requires the adoption of a coherent probabilistic framework.

Zona	1 ALPI	2 P.PADANA	3 N-APPENNINI	4 COSTA TIRRENICA	5 S-APPENNINI	6 CALABRIA	7 SICILIA	8 ITALIA
Mw 5.6+	1530	1690	1750	1750	1787	1787	1700	1787
Mw 6.0+	1490	1450	1580	1580	1660	1660	1530	1660



M5.6+ target earthquakes

Testing the rates of each ERF model

Zone and # events	1 (16)	2 (10)	3 (40)	4 (2)	5 (15)	6 (15)	7 (6)	Italia (91)
MA1	0.10	0.31	0.69	0.13	0.54	0.36	0.67	0.39
MA2	0.06	0.82	0.87	0.09	0.54	0.01	0.88	0.22
MA3	0.66	0.96	0.98	0.29	0.52	0.03	0.66	0.56
MA4	0.55	0.69	0.69	0.04	0.52	0.23	0.59	0.58
MA5	0.06	0.40	0.23	0.01	0.90	0.07	0.74	0.23
MF1	0.55	0.80	0.29	0.13	0.21	0.01	0.32	0.21
MF2	0.01	0.78	0.91	0.02	0.33	0.54	0.41	0.15
MS1	0.59	0.25	0.12	0.04	0.48	0.008	0.40	0.12
MS2	0.68	0.16	0.88	**0.005**	0.87	0.12	0.30	0.90
MG1	0.47	0.43	0.25	**0.001**	0.66	0.35	0.11	0.94
MG2	0.20	0.35	**0.0009**	**0.0055**	0.11	**0.0009**	0.04	0.40

The p-values in **bold** are “significant” accounting for multiple testing

Testing some key hypotheses of the ensemble ERF model

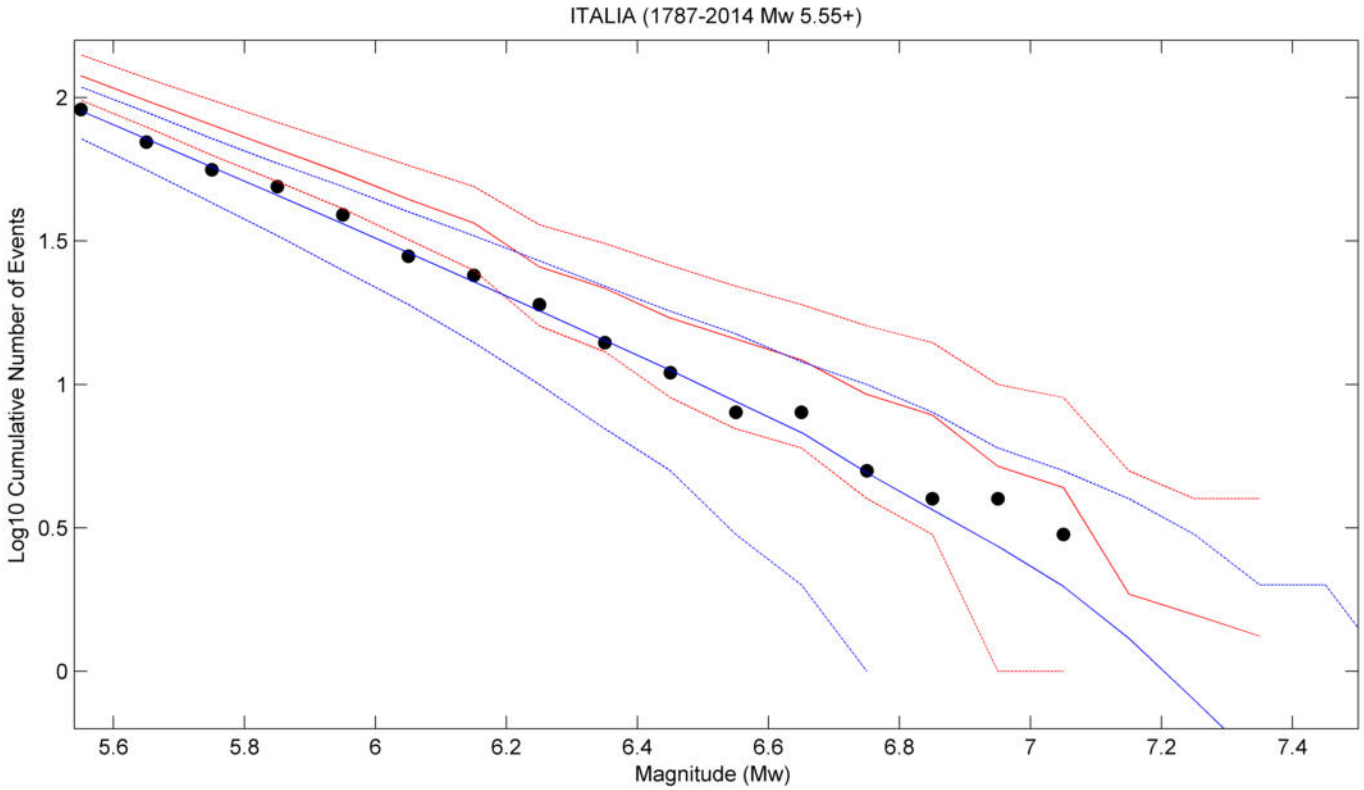
A. H_0: the MFD is exponential (Lilliefors test)
B. H_0: the process is stationary (Runs test)

Zone	1 ALPI	2 P.PADANA	3 N-APPENNINI	4 COSTA TIRRENICA	5 S-APPENNINI	6 CALABRIA	7 SICILIA	8 ITALIA
Mw 5.6+	A 0.16 B 0.77	A 0.28 B 1.0	A 0.50 B 0.63	A N/A B 1.0	A 0.5 B 1.0	A 0.5 B 0.77	A 0.37 B 0.67	A 0.5 B 0.91
Mw 6.0+	A 0.31 B 0.67	A 0.25 B 0.10	A 0.5 B 0.92	A N/A B 1.0	A 0.01 B 0.40	A 0.5 B 0.5	A 0.5 B 1.0	A 0.5 B 0.64

P-value for the test of DEPARTURE FROM A TAPERED-GR LAW (A), and NONSTATIONARITY (B)

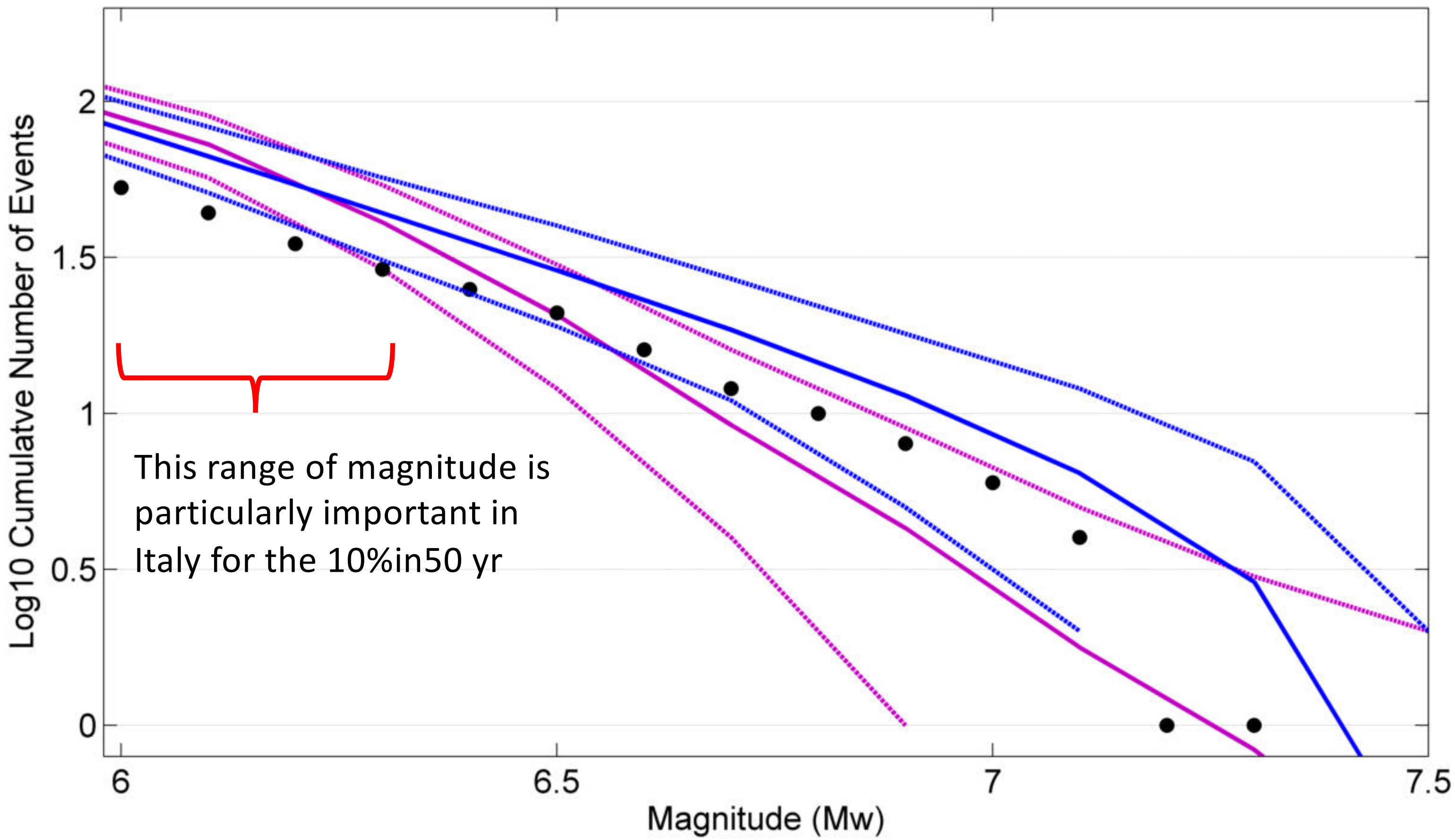
Assumptions A and B are not rejected by the data, but we have a few data and any kind of test cannot be very powerful.
These results indicate only that there are not blatant departures from the assumptions.

Consistency testing of the ensemble MPS19-ERF model



MPS19 in BLUE, 95 cl.

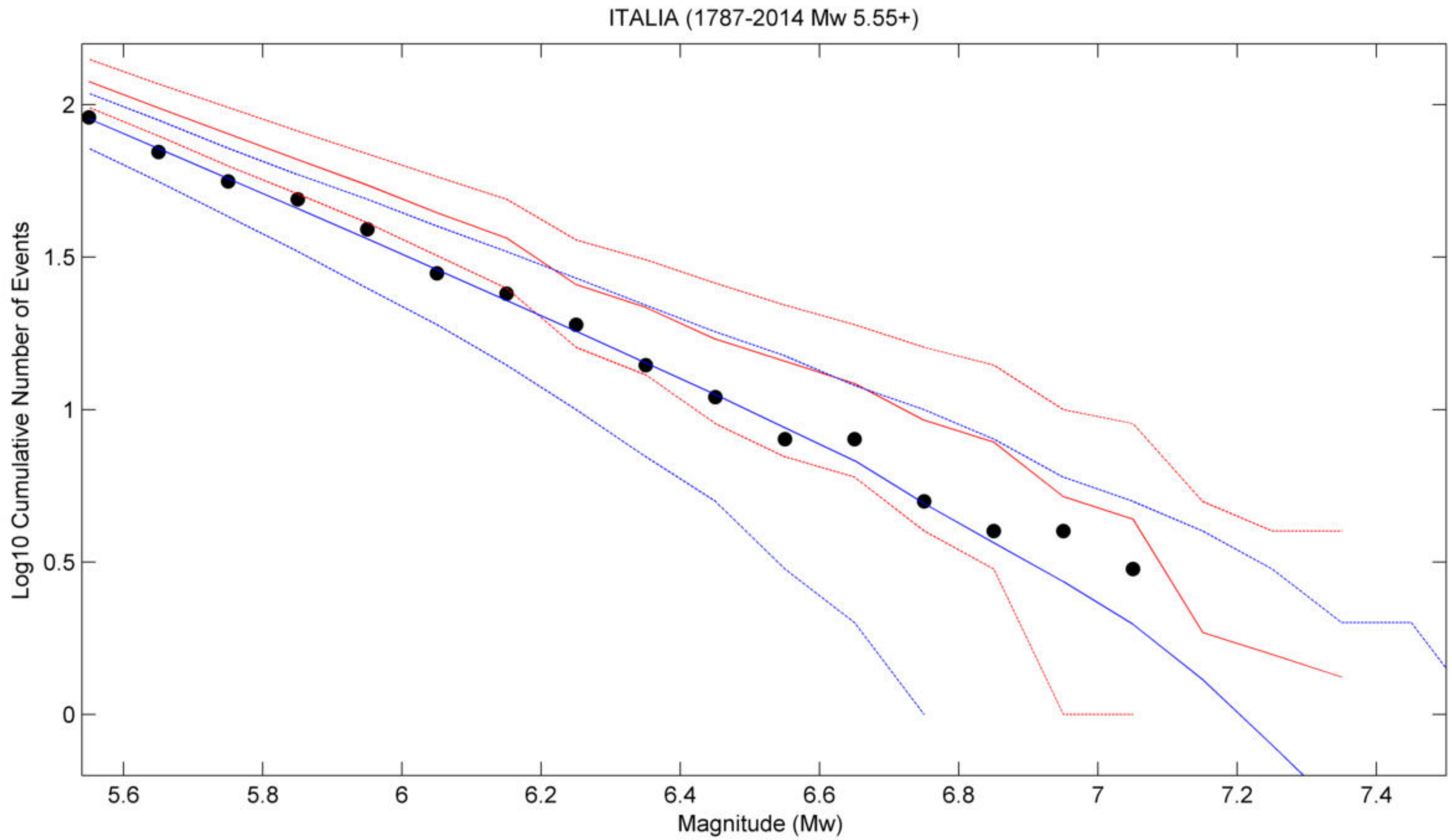
ESHM20 Seismicity rates



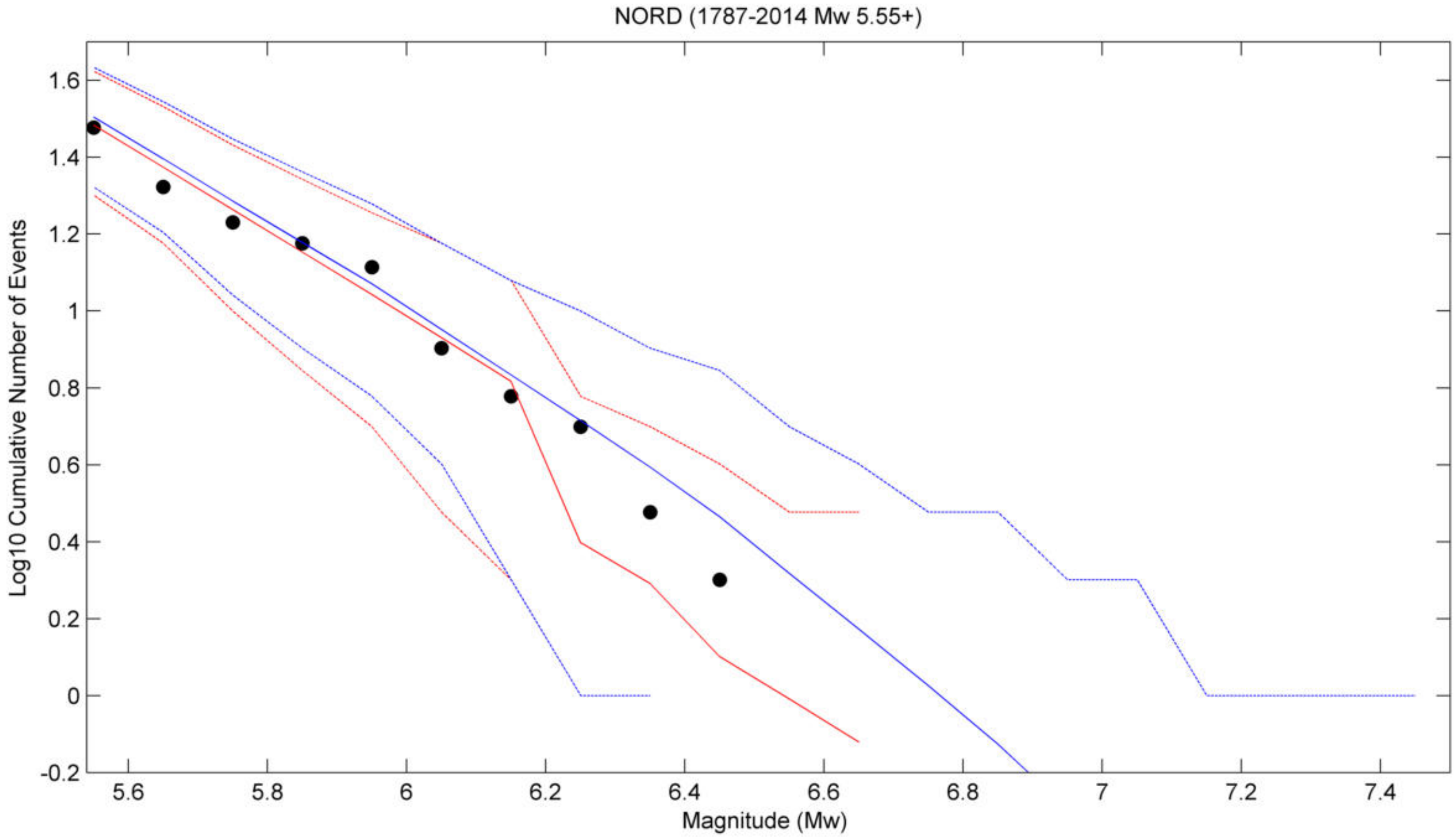
Seismotectonic zonation in PURPLE

Smoothed seismicity and faults in BLUE

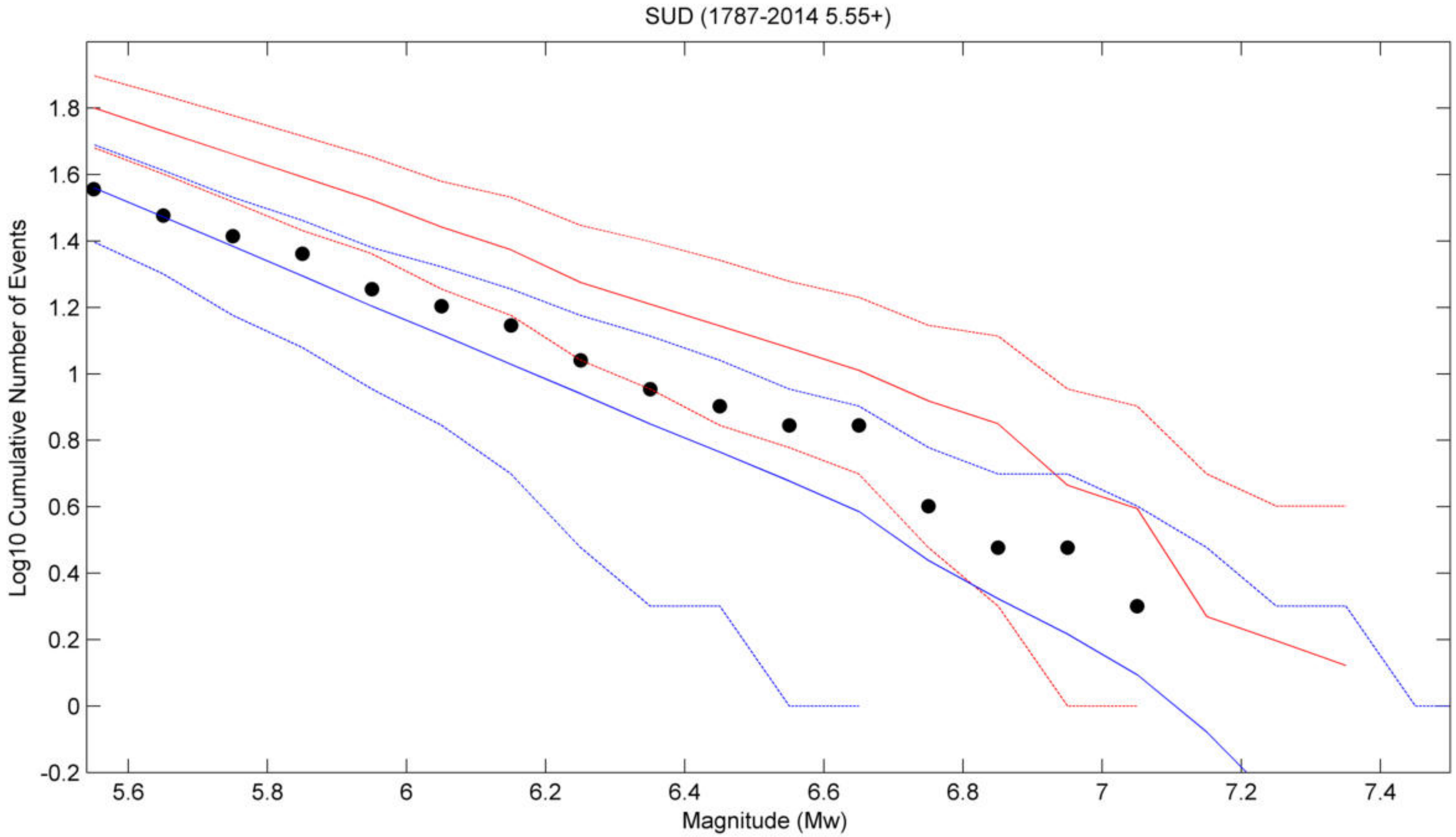
RED – MPS04
BLUE – MPS19



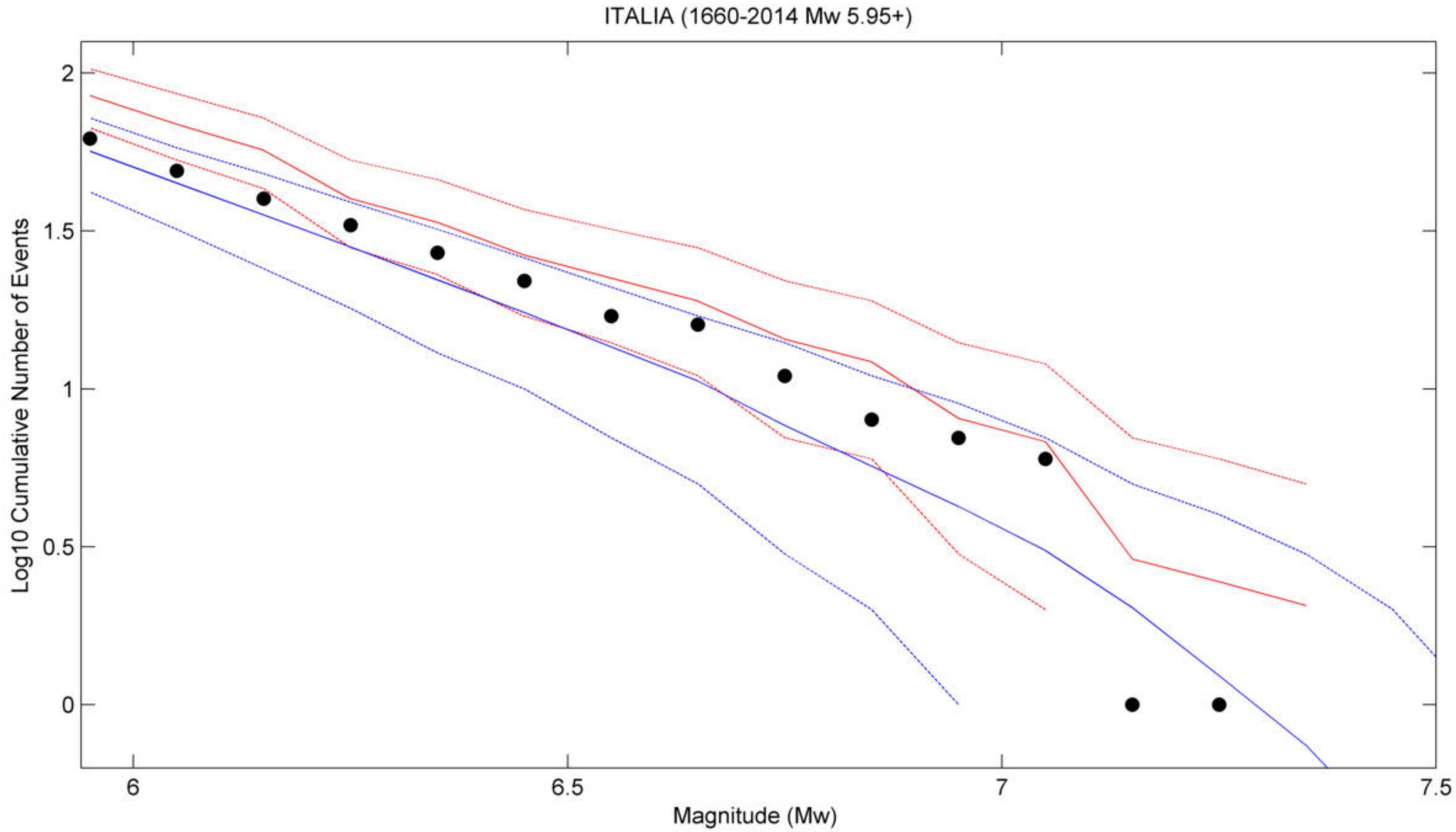
RED – MPS04
BLUE – MPS19



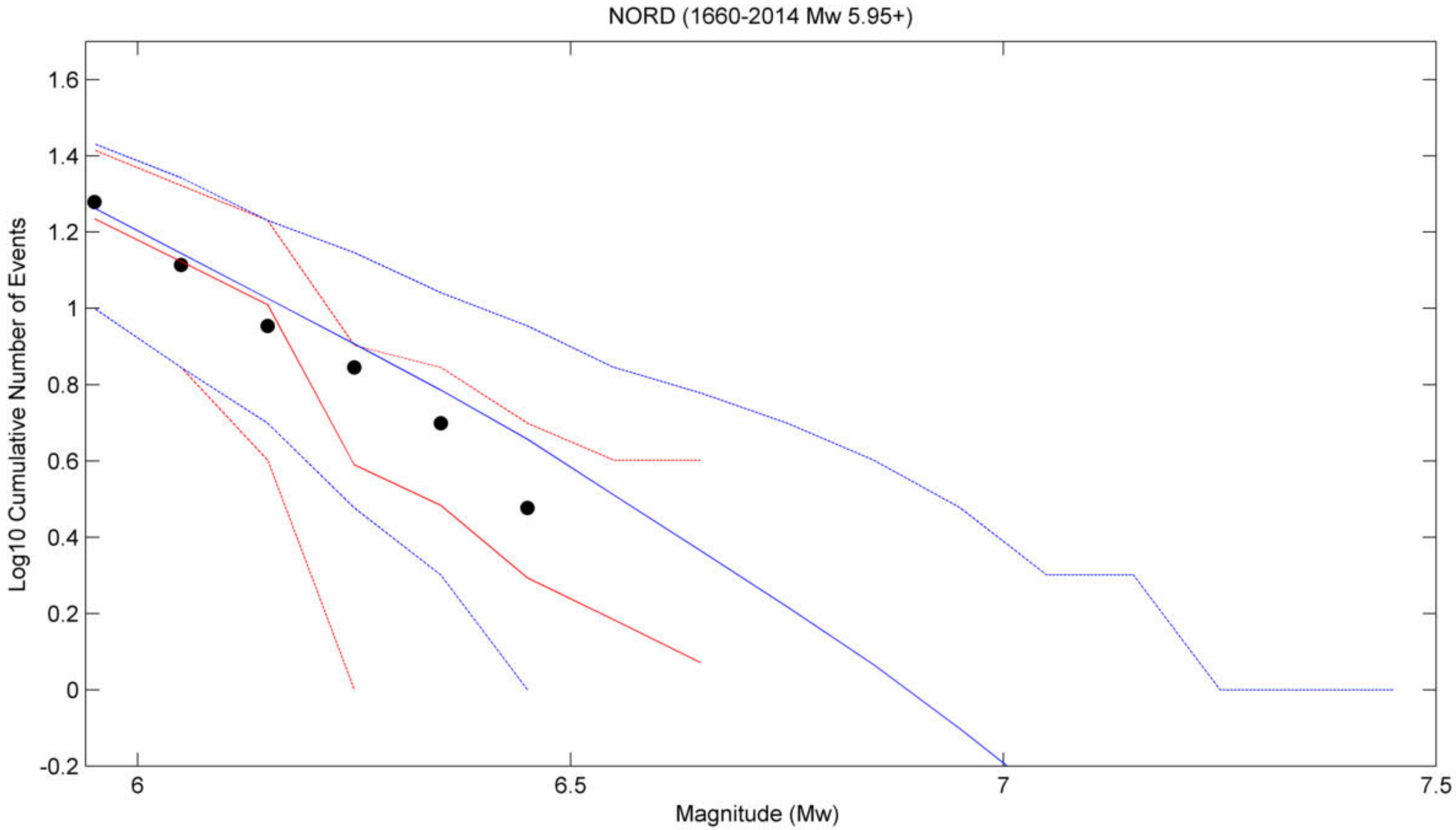
RED – MPS04
BLUE – MPS19



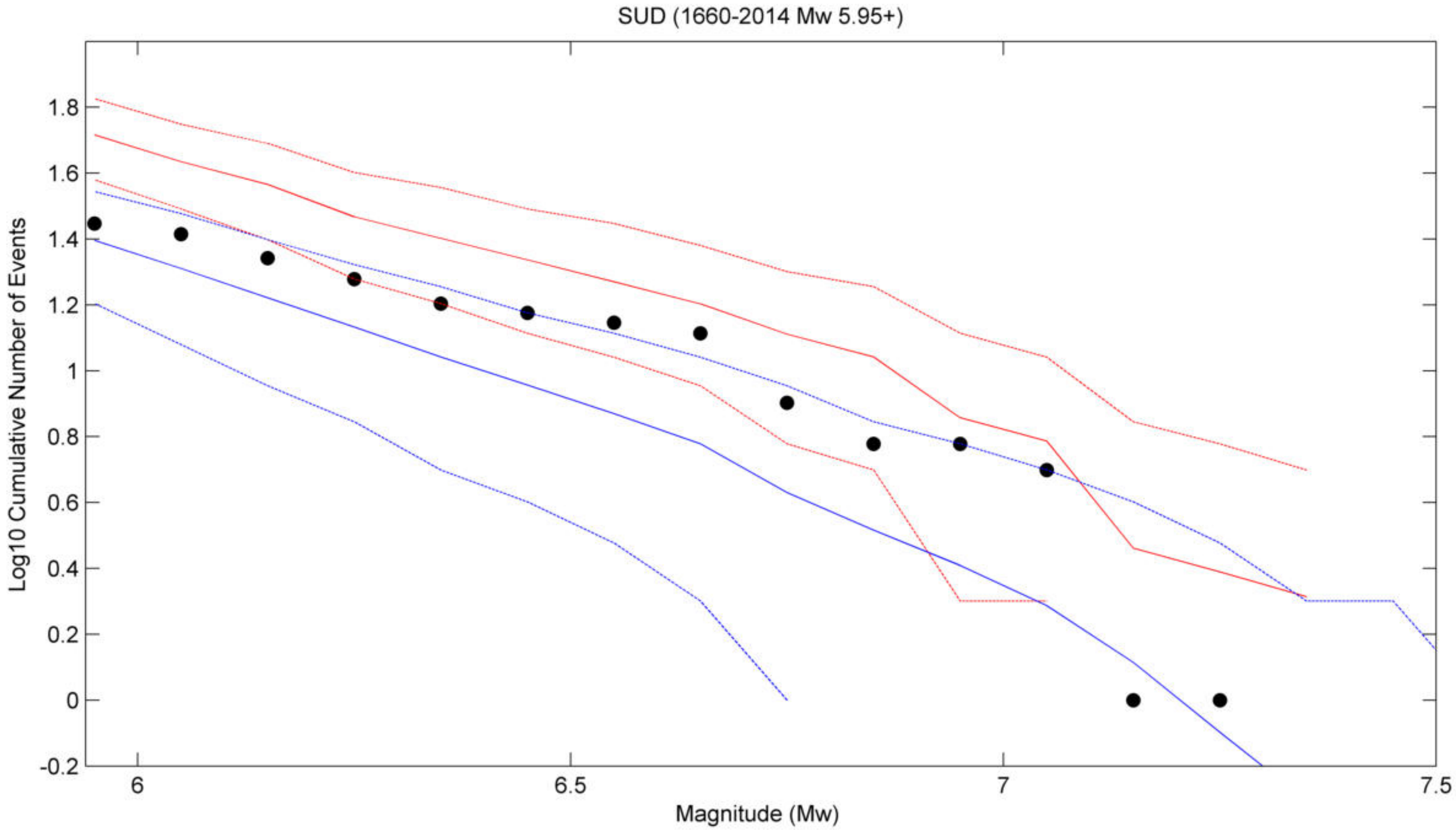
RED – MPS04
BLUE – MPS19



RED – MPS04
BLUE – MPS19



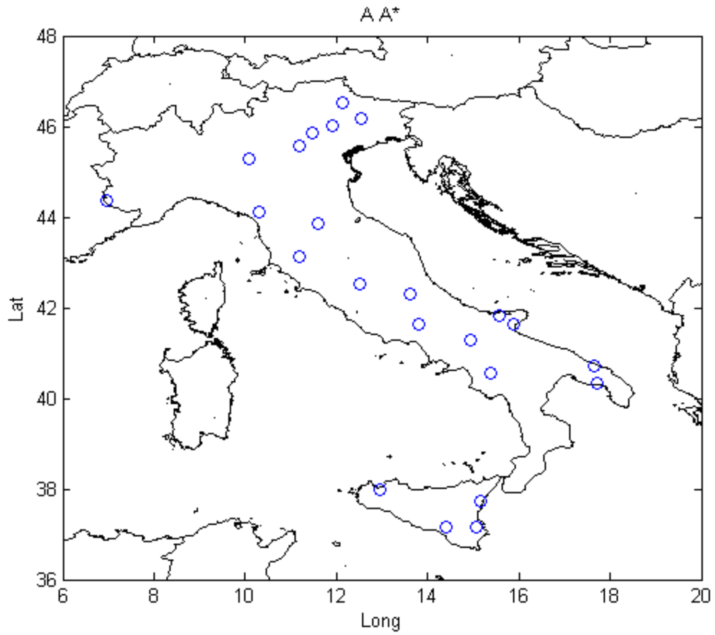
RED – MPS04
BLUE – MPS19

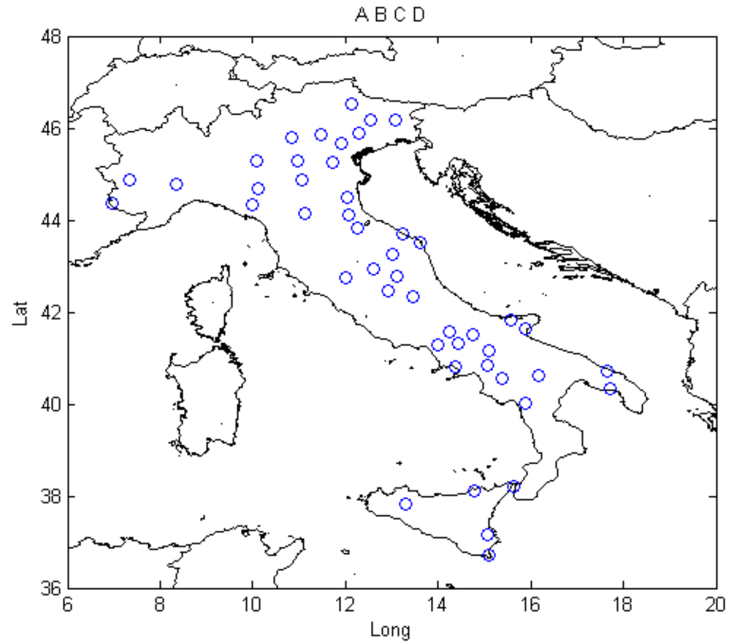


FASE 3: test di MPS19 sui valori di accelerazione strumentali

Stations

Distance between stations > 30 Km





A A*

A B C D

The IM observed at different sites is correlated [Park et al., 2007], and this may be an issue for testing [Iervolino et al., 2017]. We adopt a pragmatic approach choosing sites that have a minimum distance of 30 km from one another.

In this case, although part of the IM correlation remains independently from the distance [Park et al., 2007], we may assume that, for small PoE, exceedances are independent, i.e., the observation of an exceedance E_Y in one site Y does not modify the probability to observe an exceedance E_X in the nearby site X, $P(E_X \mid E_Y) = P(E_X) \equiv \text{PoE}$. Hence, the exceedances follow a binomial distribution with the parameter equal to the selected PoE.

Worthy of note, this assumption makes the test of PGA conservative: if the null hypothesis (the exceedances are generated by a process which is well described by MPS19) is not rejected, it will not rejected also relaxing the assumption, i.e., using a more exact distribution for the number of exceedances, which is overdispersed with respect to the binomial distribution [see Figure 2 in Iervolino et al., 2019].

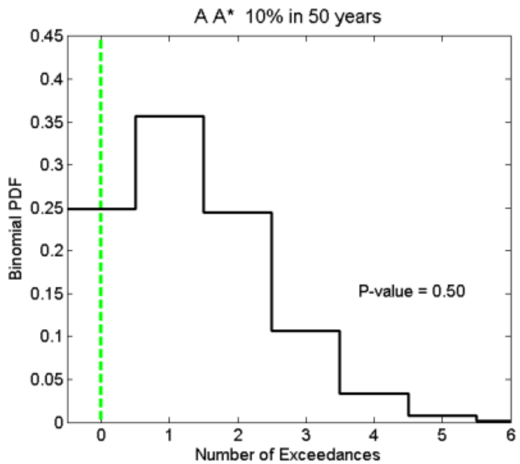
Spatial Correlation of ground shaking ≠ spatial correlation of exceedances

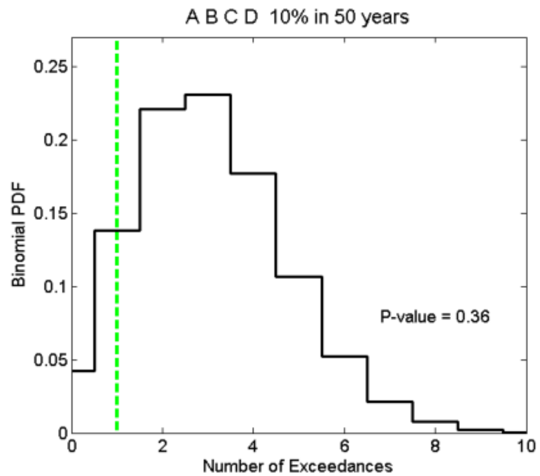
Università degli Studi di Napoli Federico II

PGA

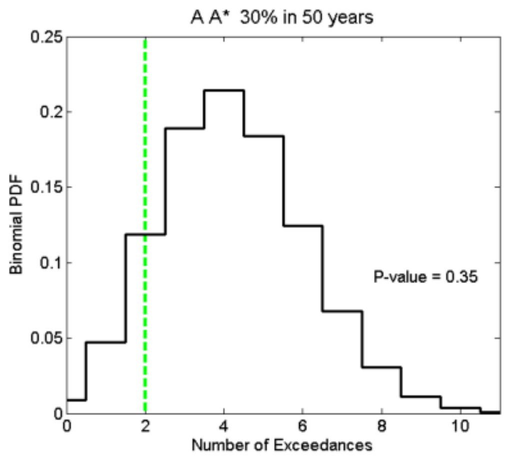
A A* stations | **A B C D stations**

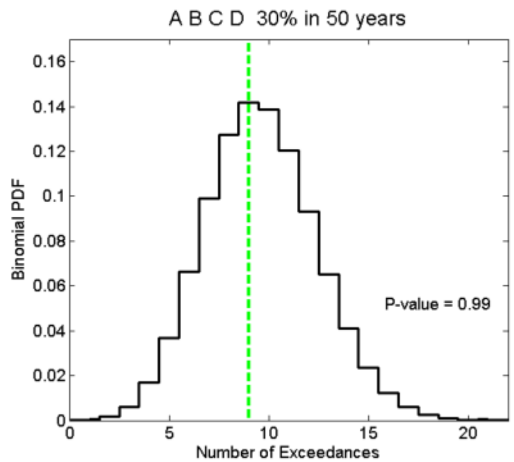
10% in 50 years





30% in 50 years





PGA

Probability gain MPS19 vs MPS04: ratio of the probability to observe the exact number of exceedances according to the two different models (green when MPS19 is better; red when MPS04 is better)

	A A* stations	A B C D stations
10% in 50 years (return period: 475 years)	$P_{gain} = 1$	$P_{gain} = 3.6$
63% in 50 years (return period: 50 years)	$P_{gain} = 65$	$P_{gain} = 111$

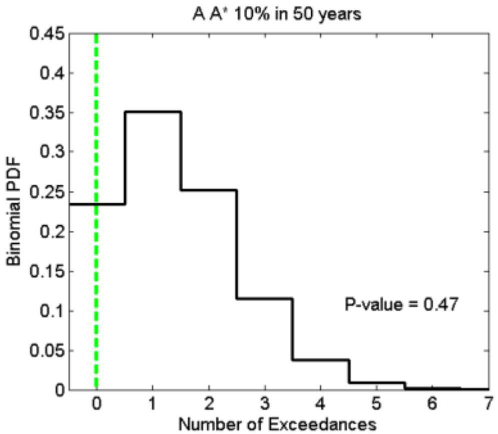
A guidance to P_{gain}
(it requires some assumptions)

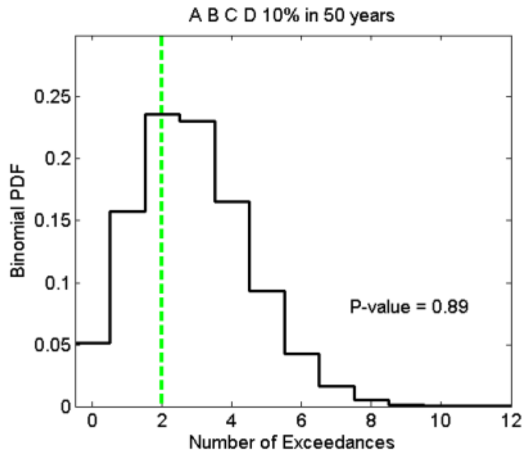
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
>100	Decisive

SA (0.2 s)

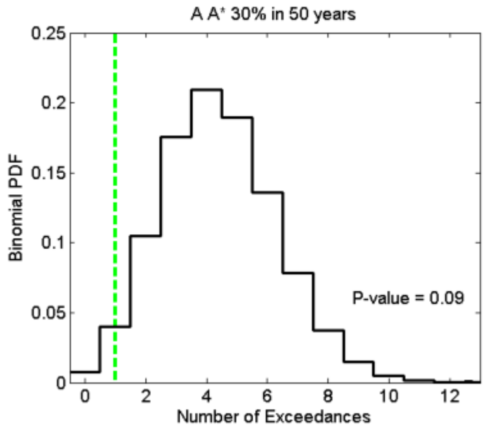
A A* stations | **A B C D stations**

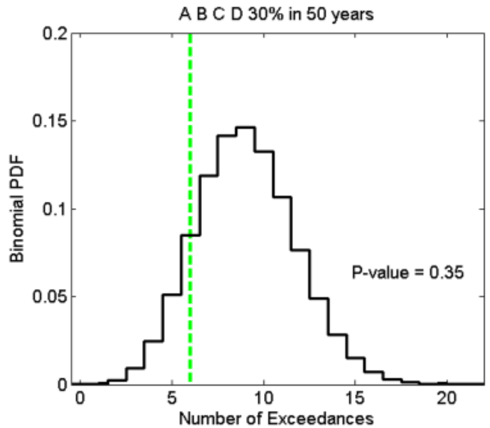
10% in 50 years





30% in 50 years





SA (0.2 s)

Probability gain MPS19 vs MPS04: ratio of the probability to observe the exact number of exceedances according to the two different models (green when MPS19 is better; red when MPS04 is better)

	A A* stations	A B C D stations
10% in 50 years (return period: 475 years)	$P_{gain} = 1$	$P_{gain} = 1$
63% in 50 years (return period: 50 years)	$P_{gain} = 1429$	$P_{gain} = 474$

A guidance to P_{gain}
(it requires some assumptions)

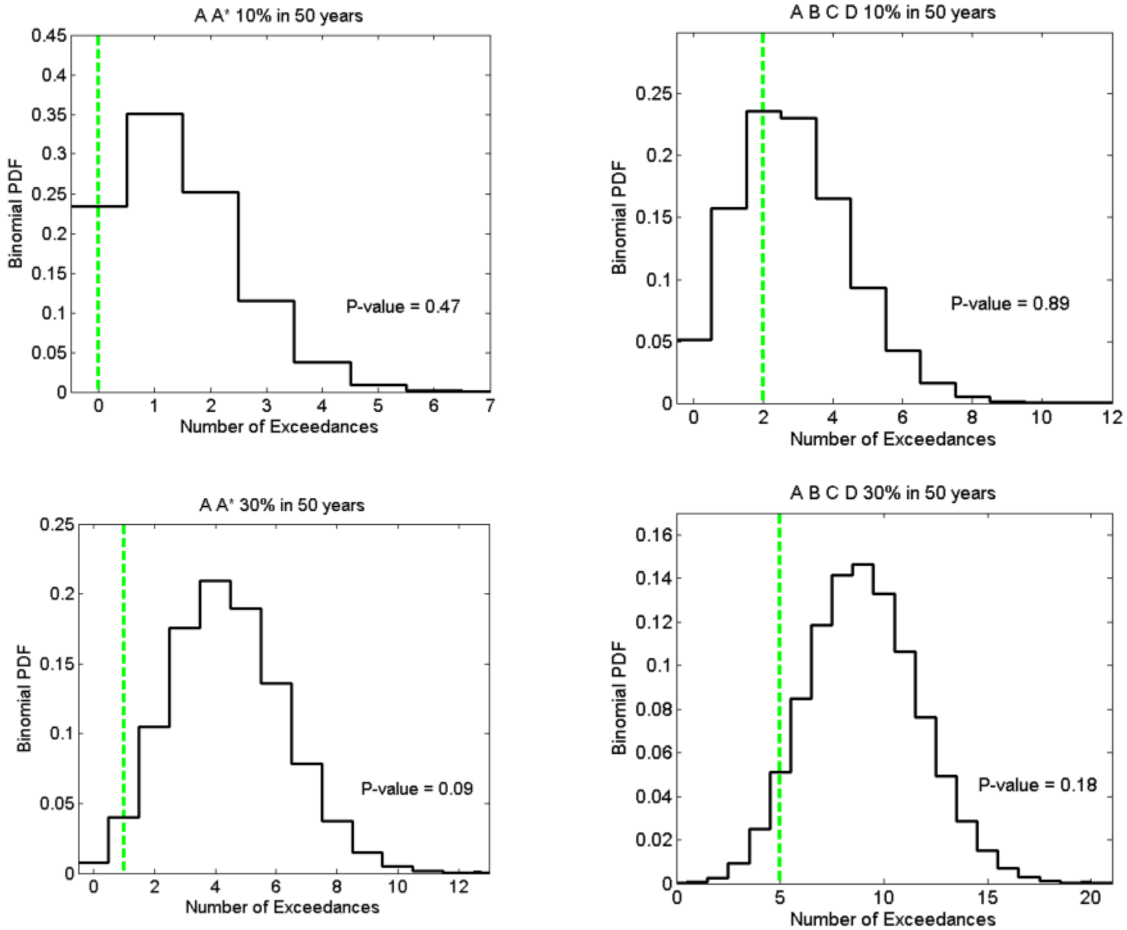
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
>100	Decisive

SA (1.0 s)

A A* stations | **A B C D stations**

10% in 50 years

30% in 50 years

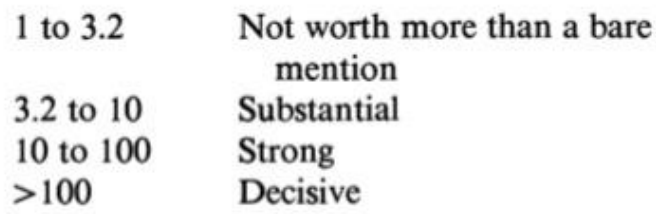


SA (1.0 s)

Probability gain MPS19 vs MPS04: ratio of the probability to observe the exact number of exceedances according to the two different models (green when MPS19 is better; red when MPS04 is better)

	A A* stations	A B C D stations
10% in 50 years (return period: 475 years)	P_{gain} = 1	P_{gain} = 1
63% in 50 years (return period: 50 years)	P_{gain} = 5.4	P_{gain} = 3

A guidance to P_{gain}
(it requires some assumptions)



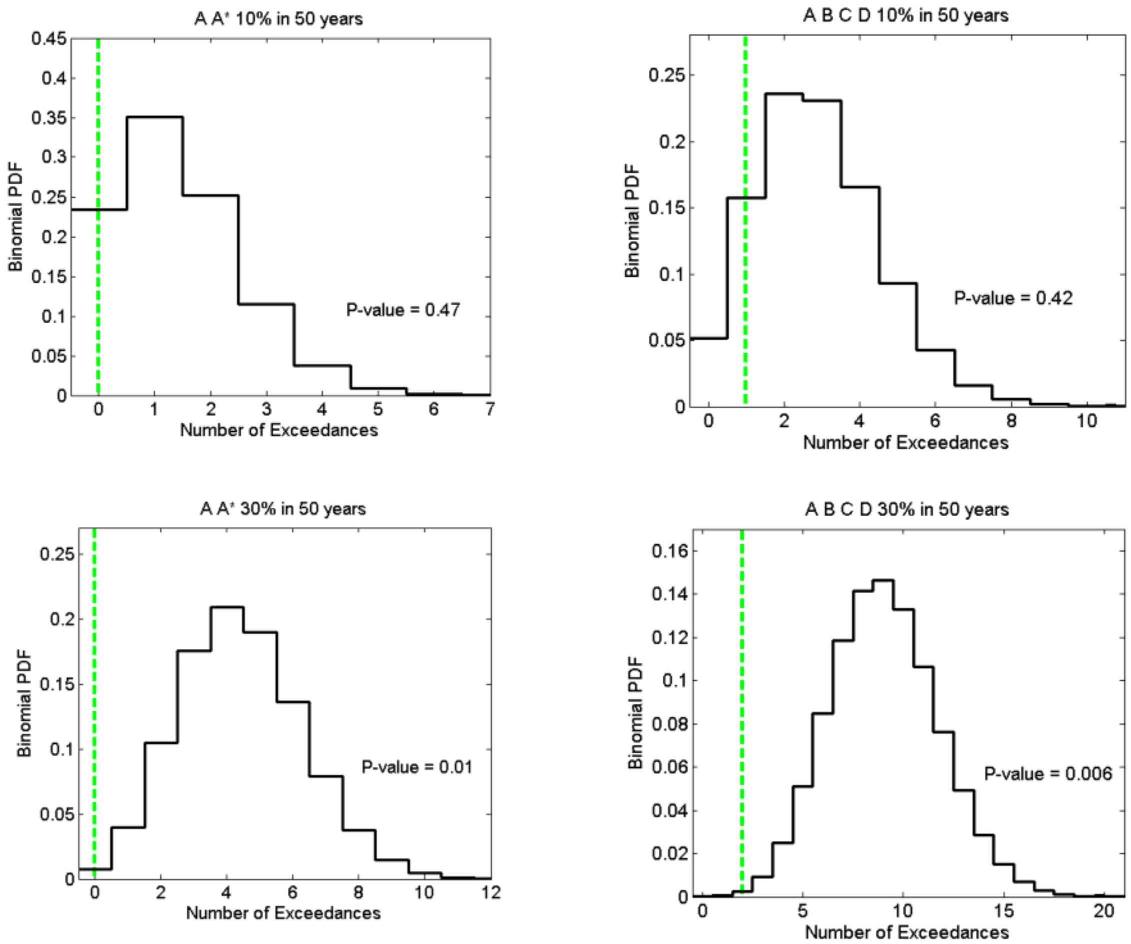
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
>100	Decisive

SA (2.0 s)

A A* stations | **A B C D stations**

10% in 50 years

30% in 50 years



SA (2.0 s)

Probability gain MPS19 vs MPS04: ratio of the probability to observe the exact number of exceedances according to the two different models (green when MPS19 is better; red when MPS04 is better)

	A A* stations	A B C D stations
10% in 50 years (return period: 475 years)	$P_{gain} = 1$	$P_{gain} = 1$
63% in 50 years (return period: 50 years)	$P_{gain} = 1$	$P_{gain} = 3$

A guidance to P_{gain}
(it requires some assumptions)

1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
>100	Decisive

Points to take home

1. MPS19 fits well the exceedances observed in the last 30 years, for different type of soil (A A*, and A B C D) and different exceedance thresholds (10% in 50 and 30% in 50)

2. MPS19 and MPS04 do not have different skill in forecasting the ground shaking exceedances in the last 30 years for return periods of 475 years (10% in 50). Conversely, the skill of MPS19 is definitely superior to the skill of MPS04 in forecasting the ground shaking exceedances for different spectral ordinates for return periods of 50 years (63% in 50).

FASE 4: test di MPS19 sui valori di intensità macrosismica

8.2.4.1. Conversione di MPS19 in termini di intensità macrosismica

Poiché MPS19 non produce direttamente stime in intensità macrosismica, ma in termini di accelerazione per diverse ordinate spettrali, il primo passo è stato quello di convertire i tassi di superamento della PGA di MPS19, $\lambda(x)$ dove x è la PGA, in termini di tassi di superamento delle intensità macrosismiche $\lambda(I)$. Ciò è stato fatto utilizzando la procedura suggerita da D'Amico e Albarello (2008)

$$\lambda(\mathrm{I}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \lambda(\mathrm{x_i}) \mathrm{P}(\mathrm{I}|\mathrm{x_i}) \qquad (8.7)$$

dove $P(I|x_i)$ è la funzione empirica di conversione delle PGA in intensità macrosismica, e *N* è il numero di livelli di PGA considerate nelle curve di hazard del modello. In questo modo è possibile stimare i tassi in termini di intensità. Ovviamente il passaggio introduce un'ulteriore fonte di possibile errore, perché la funzione empirica di trasformazione delle PGA in intensità macrosismica è soggetta ad un'elevata incertezza. Per questa fase di test si utilizzeranno due funzioni stimate da diversi gruppi di ricerca, Faenza e Michelini (2010), e Gomez Capera et al (2018).

8.2.4.2. Selezione dei siti per il test

I 27 siti selezionati per l'analisi (vedi tabella 8.11) sono quelli con una storia sismica più lunga e completa (Catania che ha una storia molto lunga non è stata considerata perché il modello in fase di test non include le aree vulcaniche). Il numero di eccedenze osservate per ogni sito tiene conto della completezza del sito e viene ripulito dalle intensità più basse (I<5.5) e da quelle generate da foreshocks e aftershocks che non sono considerati nel catalogo declusterato.

Una volta applicati i criteri di selezione, il numero di osservazioni si riduce a 202 complessivamente.

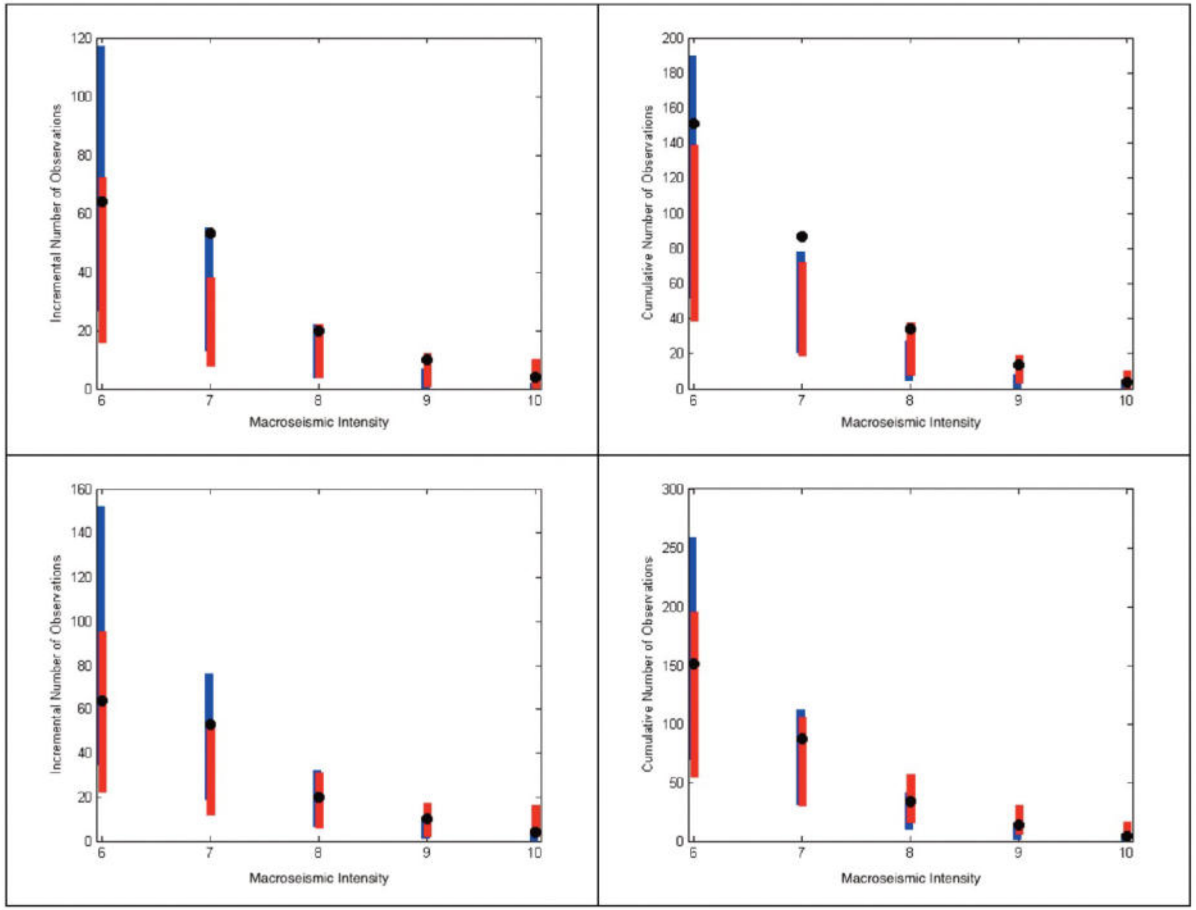
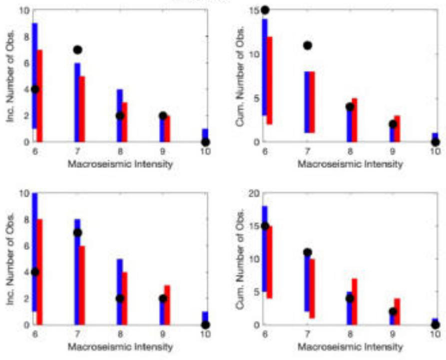
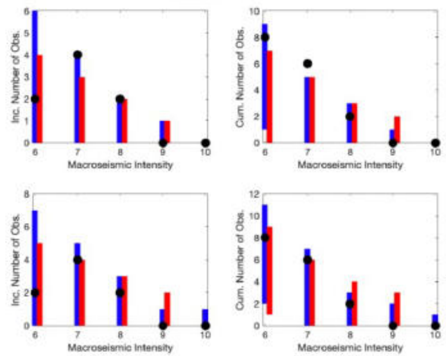


Figure 10. Results of the consistency test of MPS19 with the observed macroseismic intensities (black dots) of the last centuries for different soils. The upper row is for A sites; the lower row is for C sites. The left panels report the incremental number of observations for different macroseismic intensity, and the right panels the cumulative versions; the blue [Faenza and Michelini, 2010] and red bars [Gomez Capera et al., 2018] are the forecasts of MPS19 using two different transformation laws between PGA and macroseismic intensities (see text for more details).

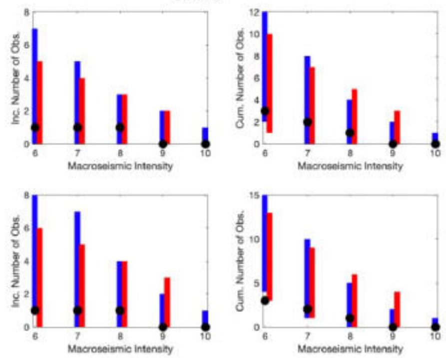
Amatrice



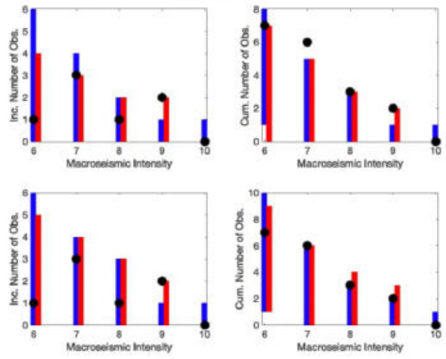
Ancona



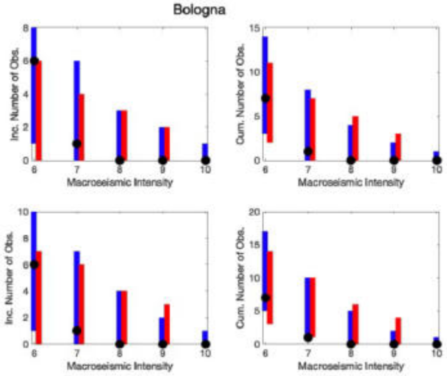
Belluno



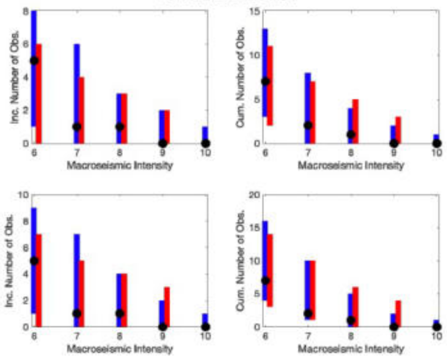
Benevento



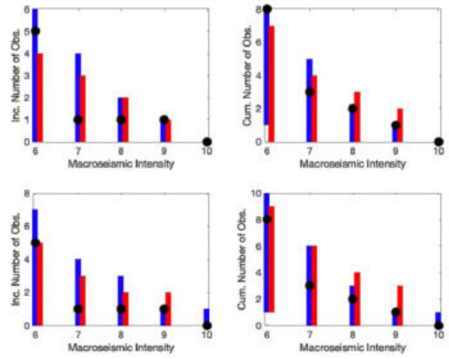
Bologna



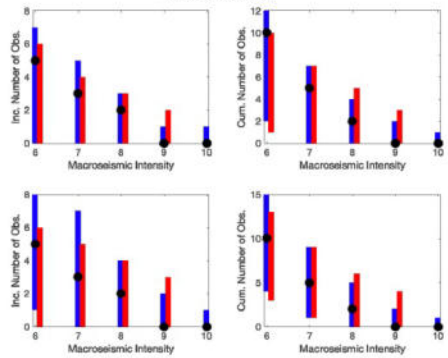
Castelnuovo di Garf.



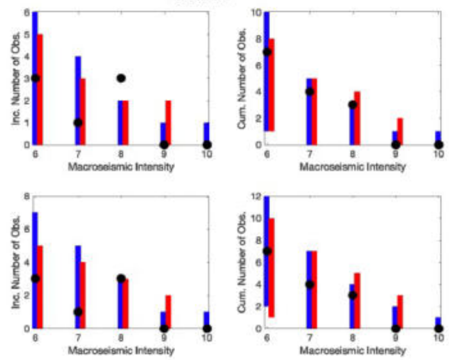
Catanzaro



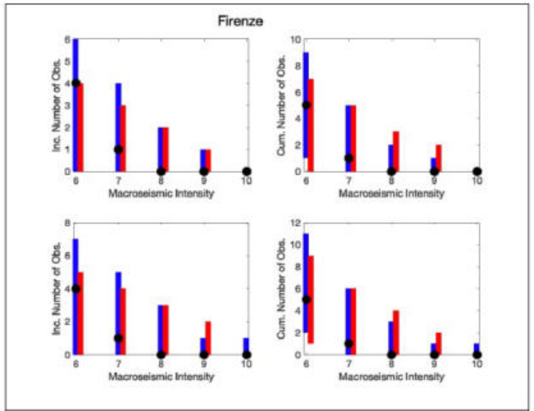
Città di Castello

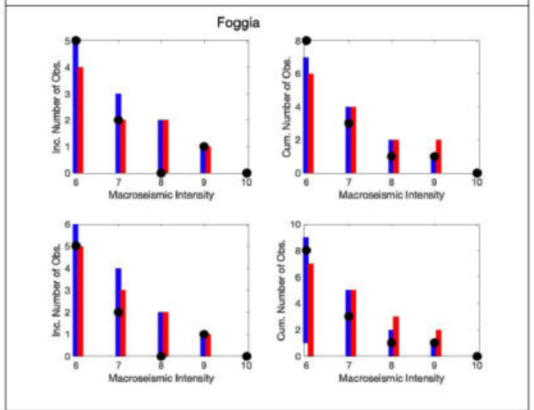


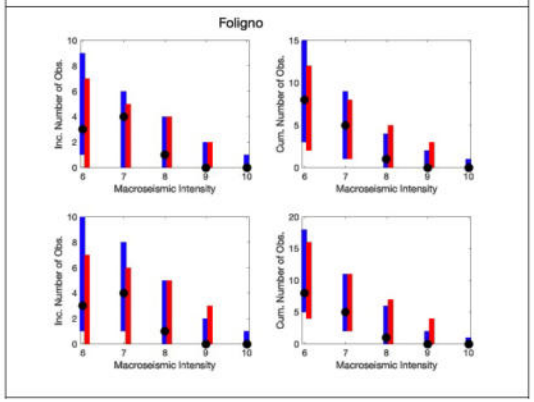
Cosenza

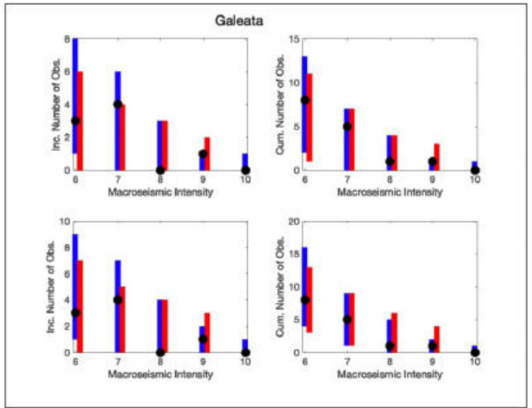


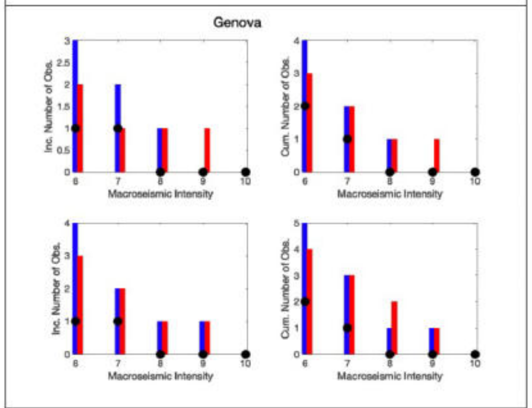
UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

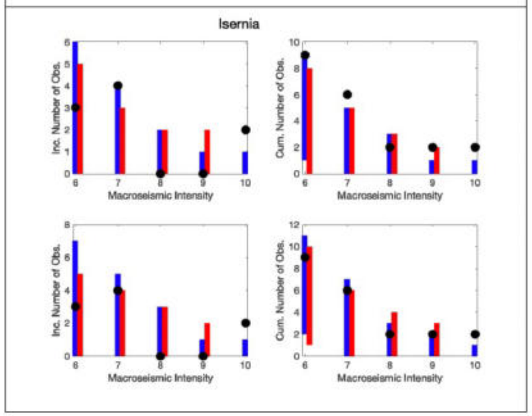


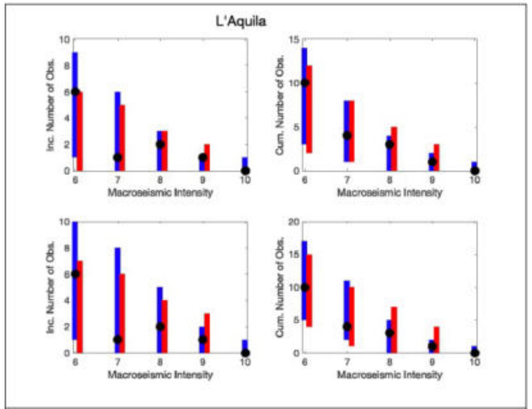


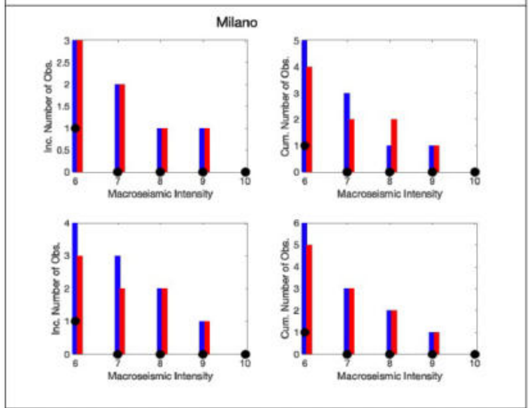


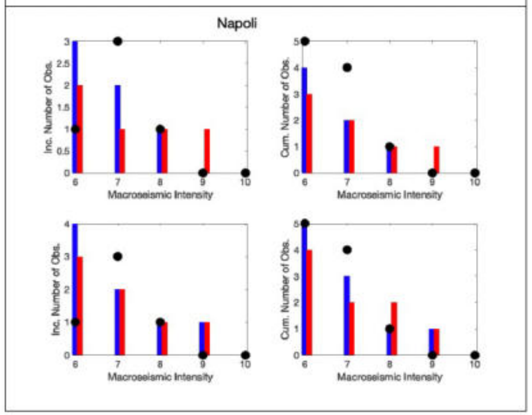




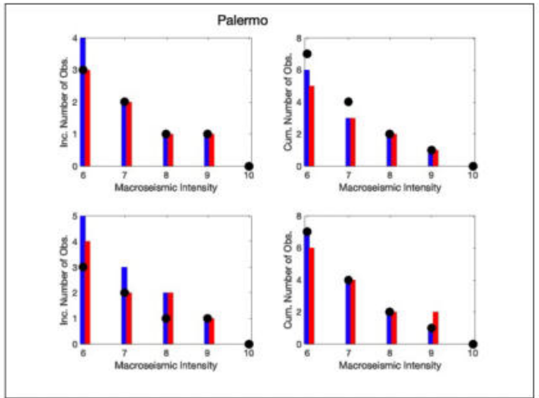


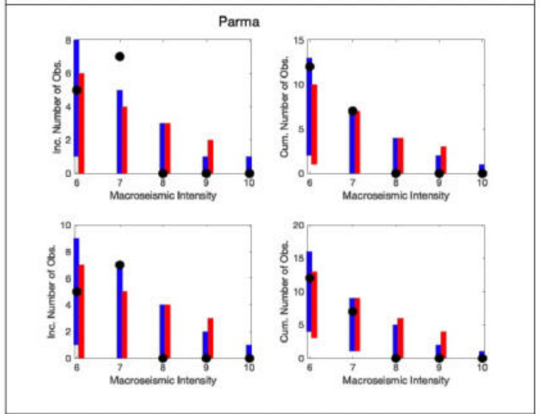


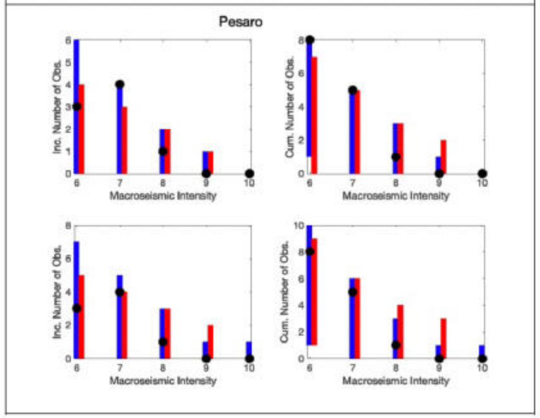


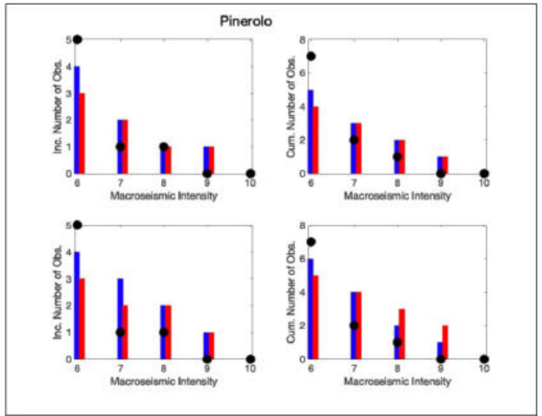


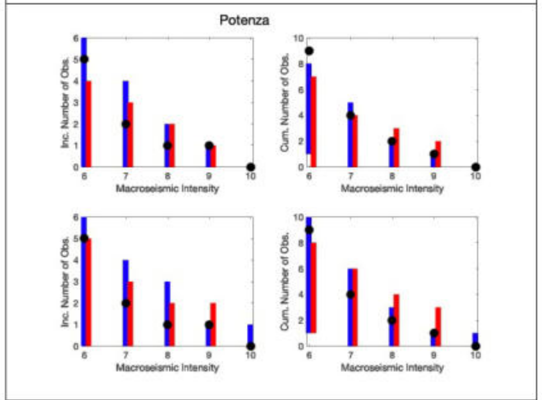
UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

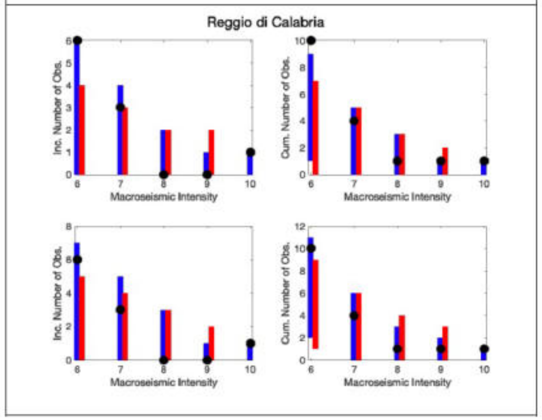


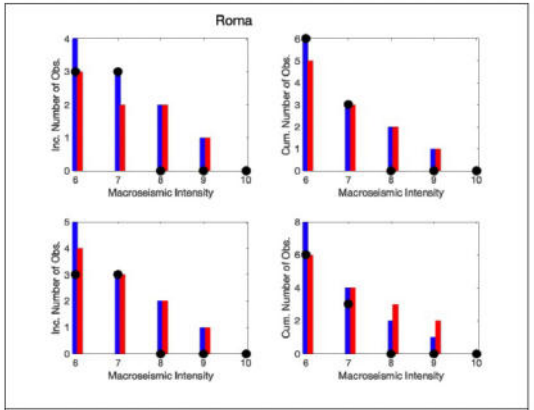


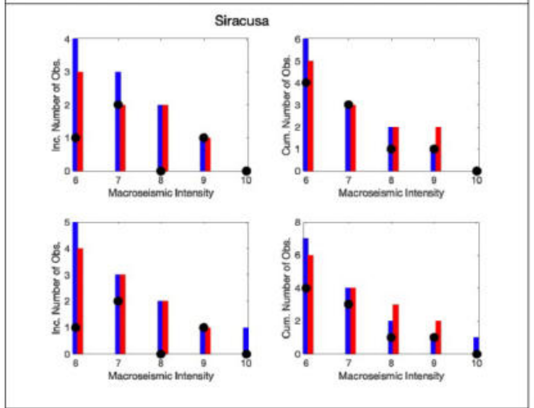


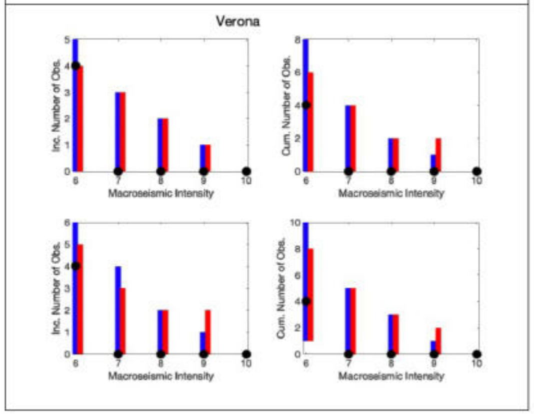








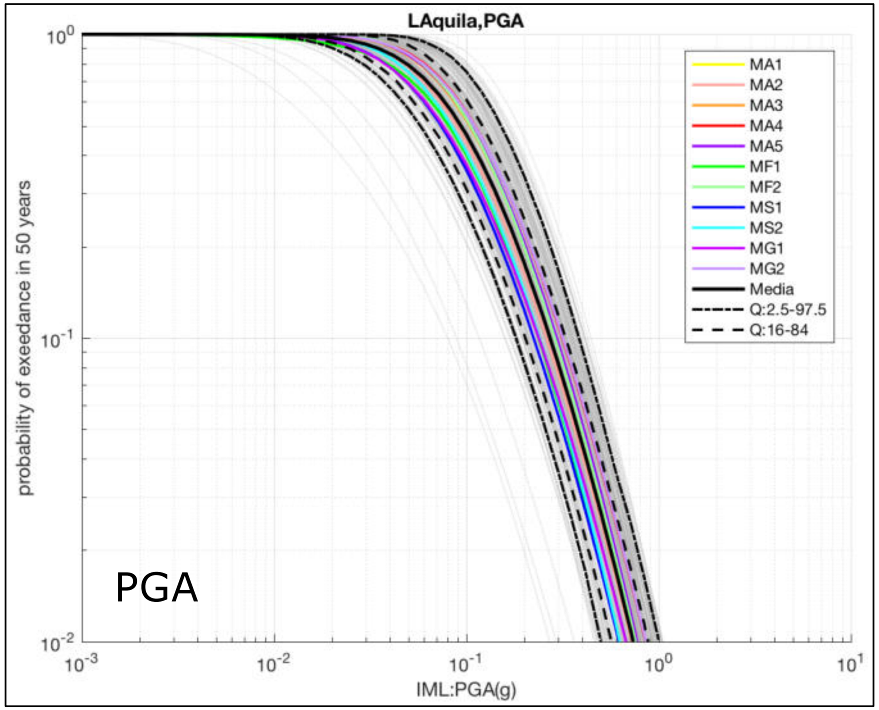


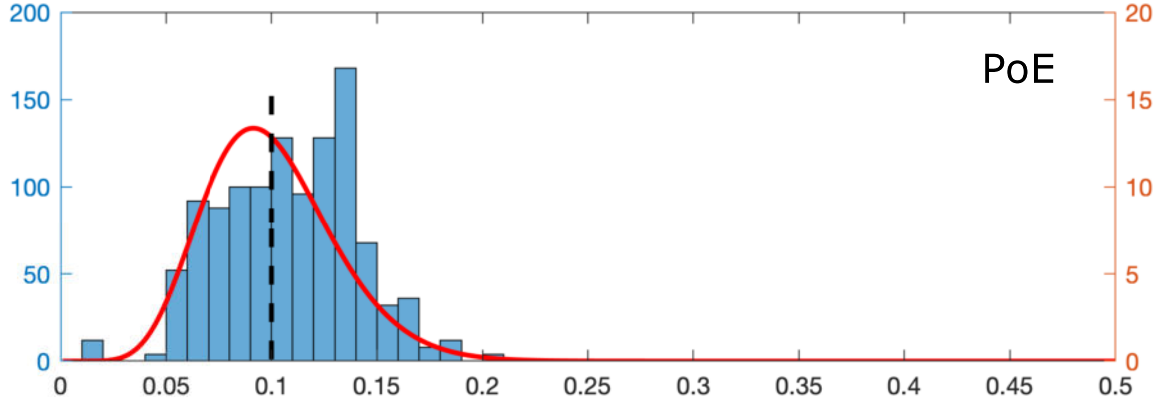


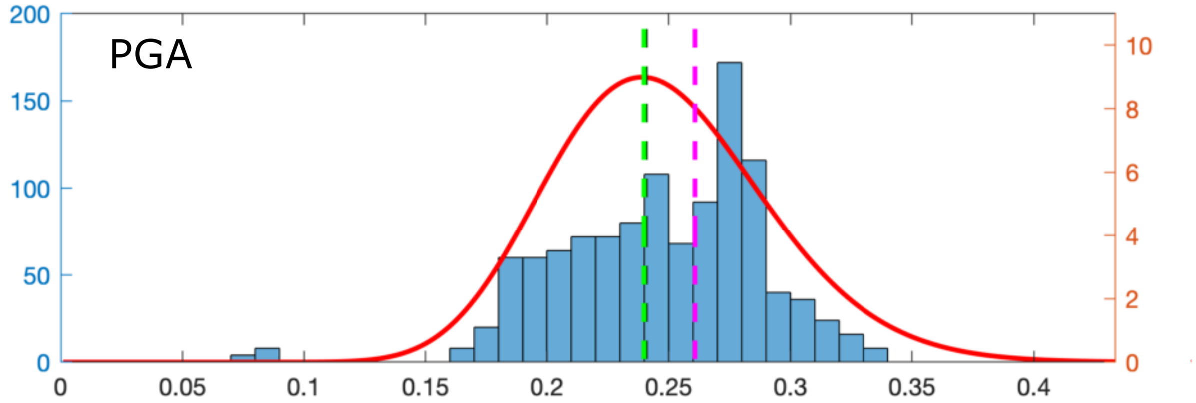
How MPS19 handles different uncertainties

Quantifying uncertainties: L’Aquila

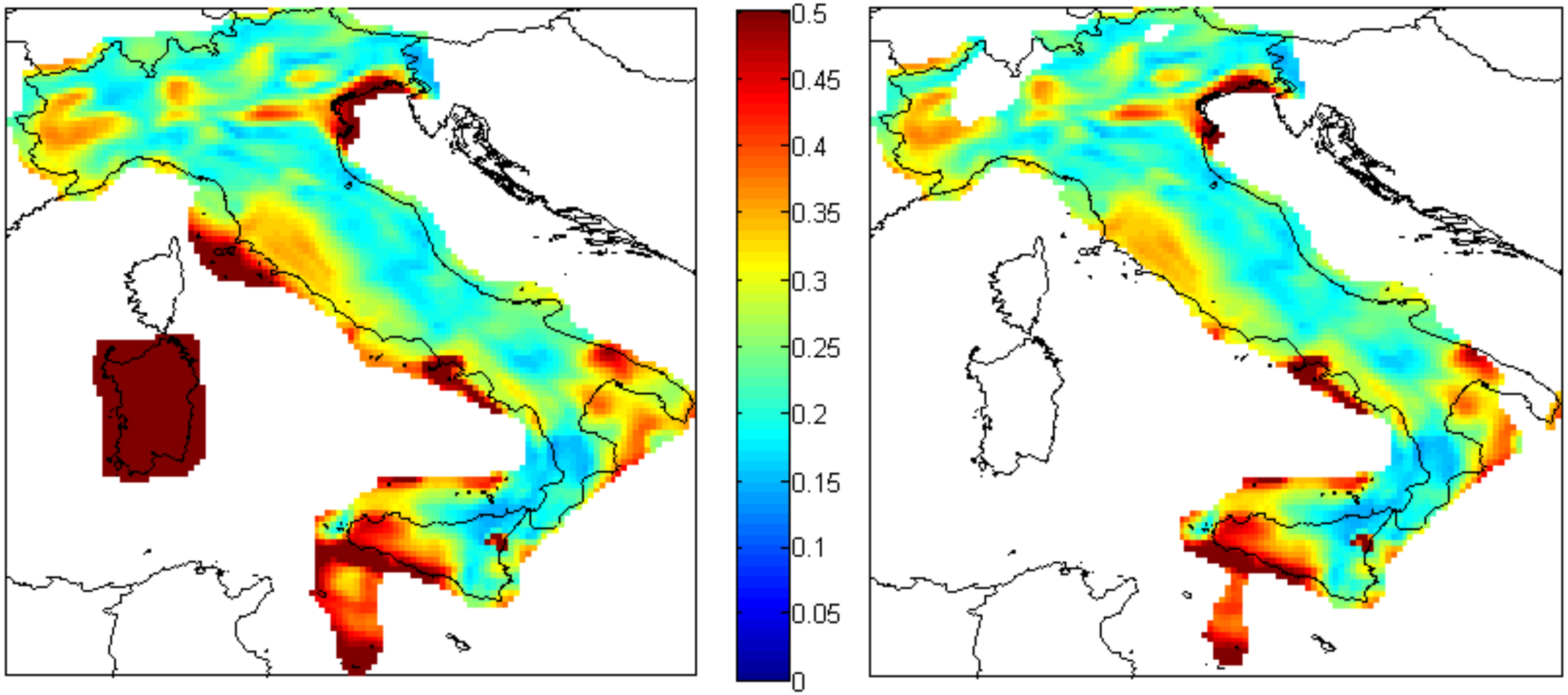
10% in 50





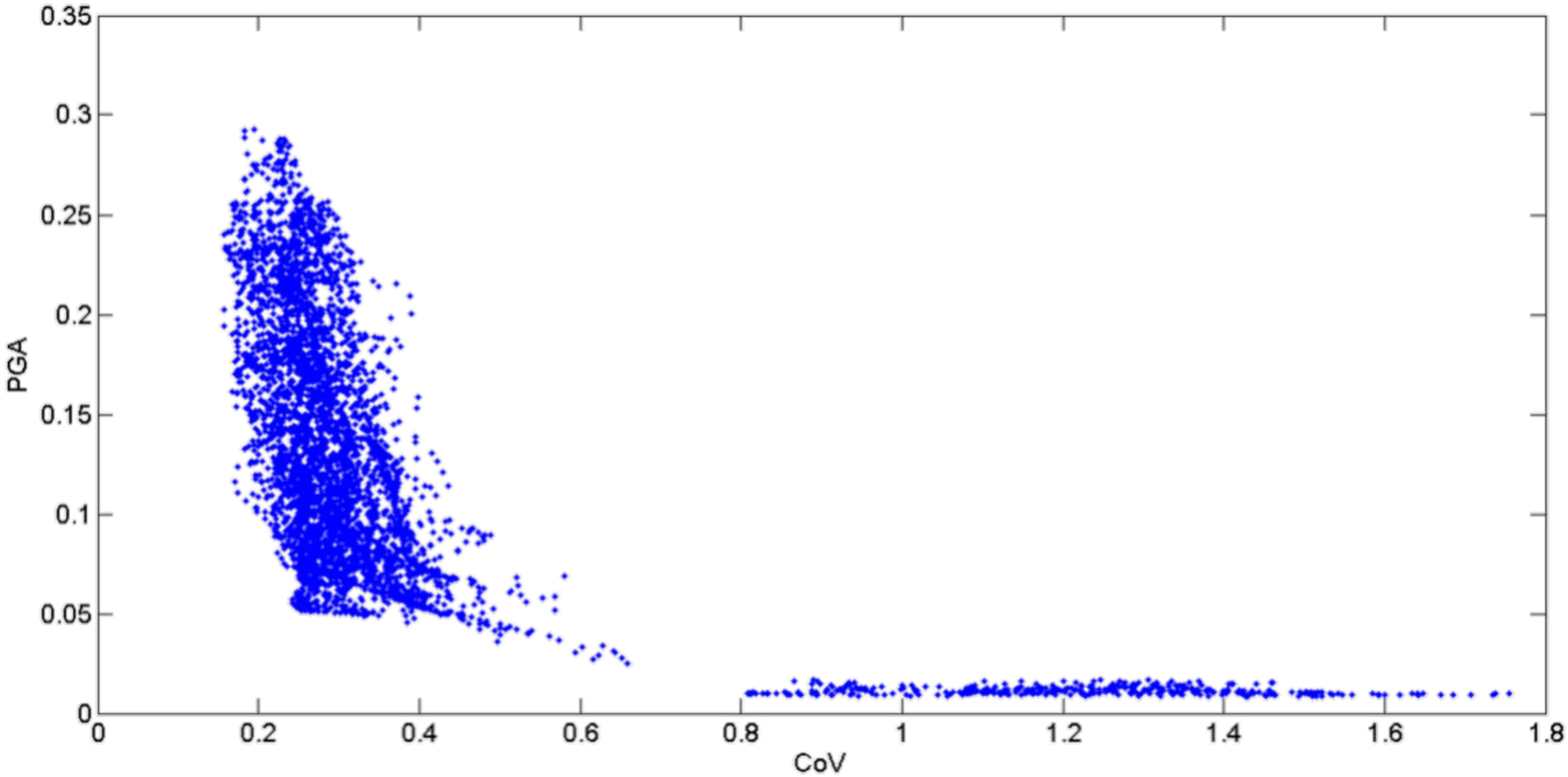


Quantifying uncertainties: the spatial distribution



All sites **Sites with PGA value greater than 20th percentile**

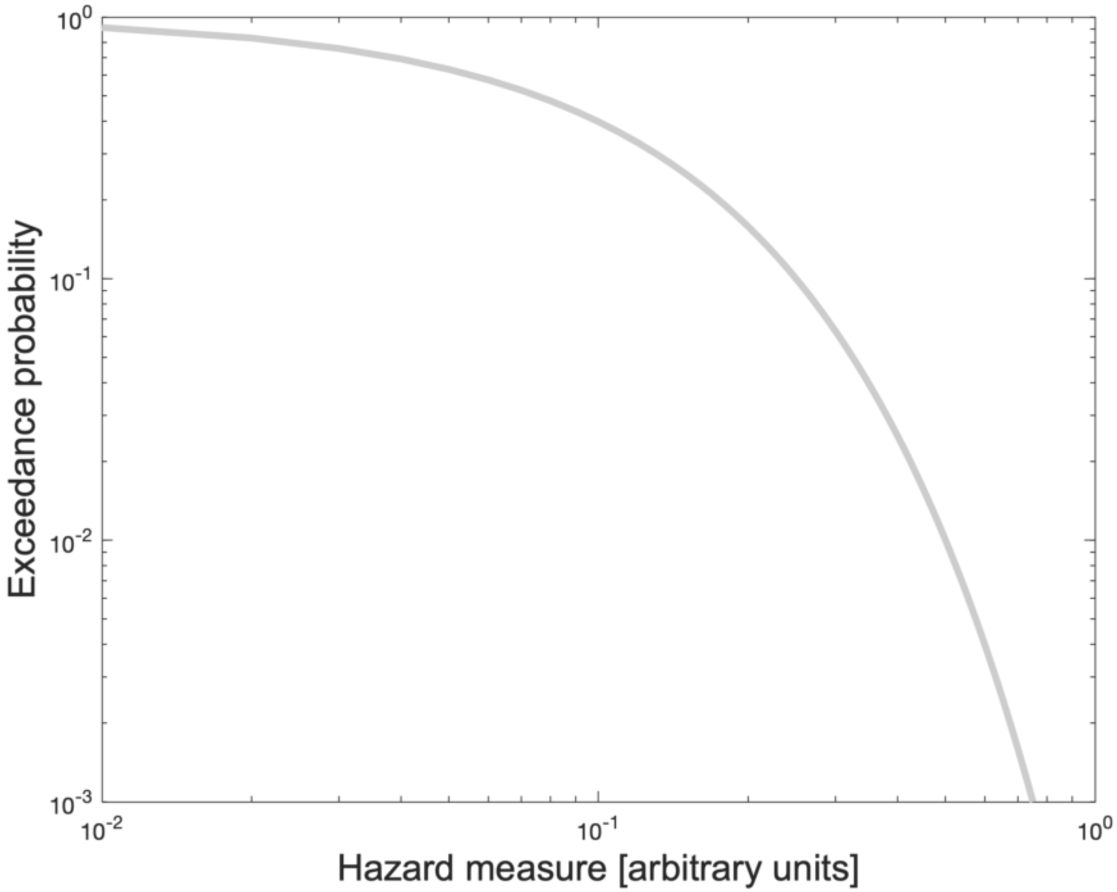
Coefficients of variation (CoV) for PGA values with 10% prob. of exceed. in 50 years.
Higher values of CoV are in areas with low hazard, where generally the available information is poorer and the uncertainty is higher.



Comparison between CoV and PGA.

CoV greater than 0.5 is for grid points with PGA (10% in 50 years) lower than 0.05g.

A common story...



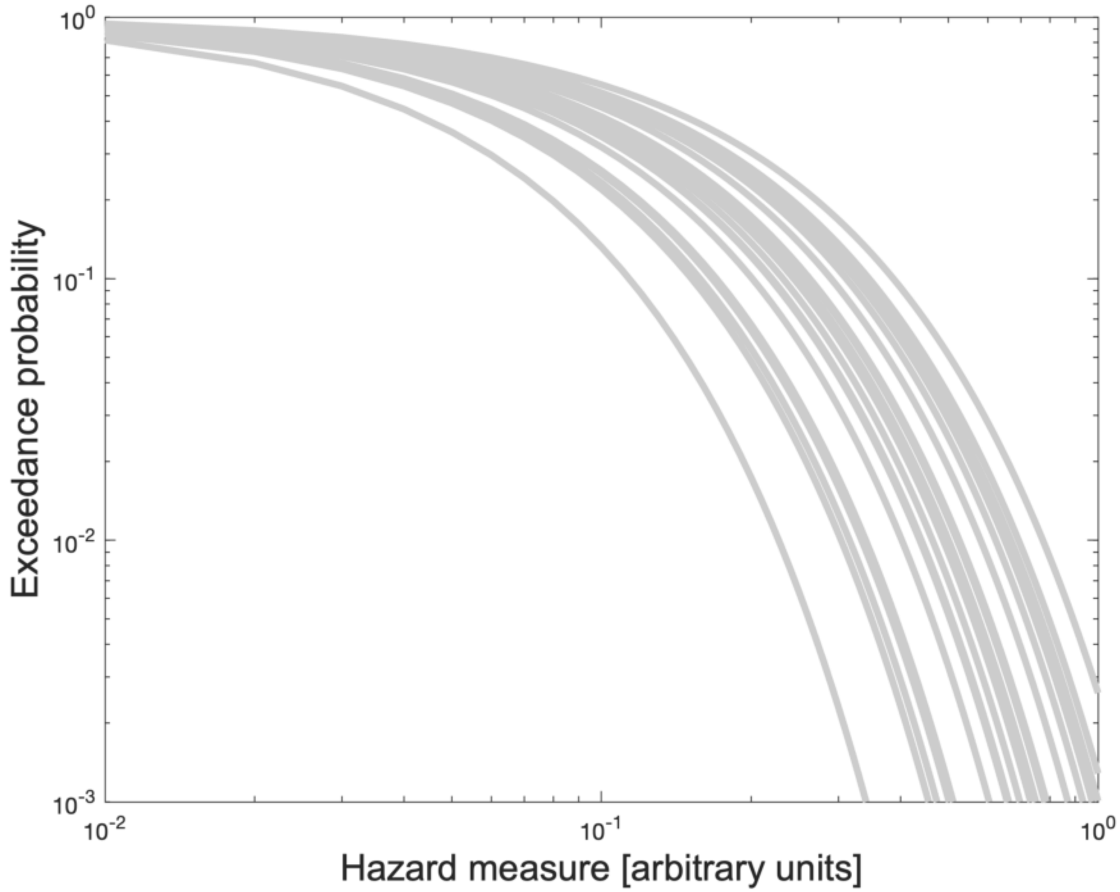
Hazard intensities in natural systems cannot be predicted deterministically, so we may assign a probability distribution to the hazard intensity (e.g., ground shaking, velocity of the wind, amount of rain, etc..).

Here we use a *survival distribution to characterize the hazard intensity*,. i.e., an exceedance probability (or *hazard*) curve

The probability distribution is aimed at describing the intrinsic (natural) randomness of the system (sometimes called ***aleatory variability***)

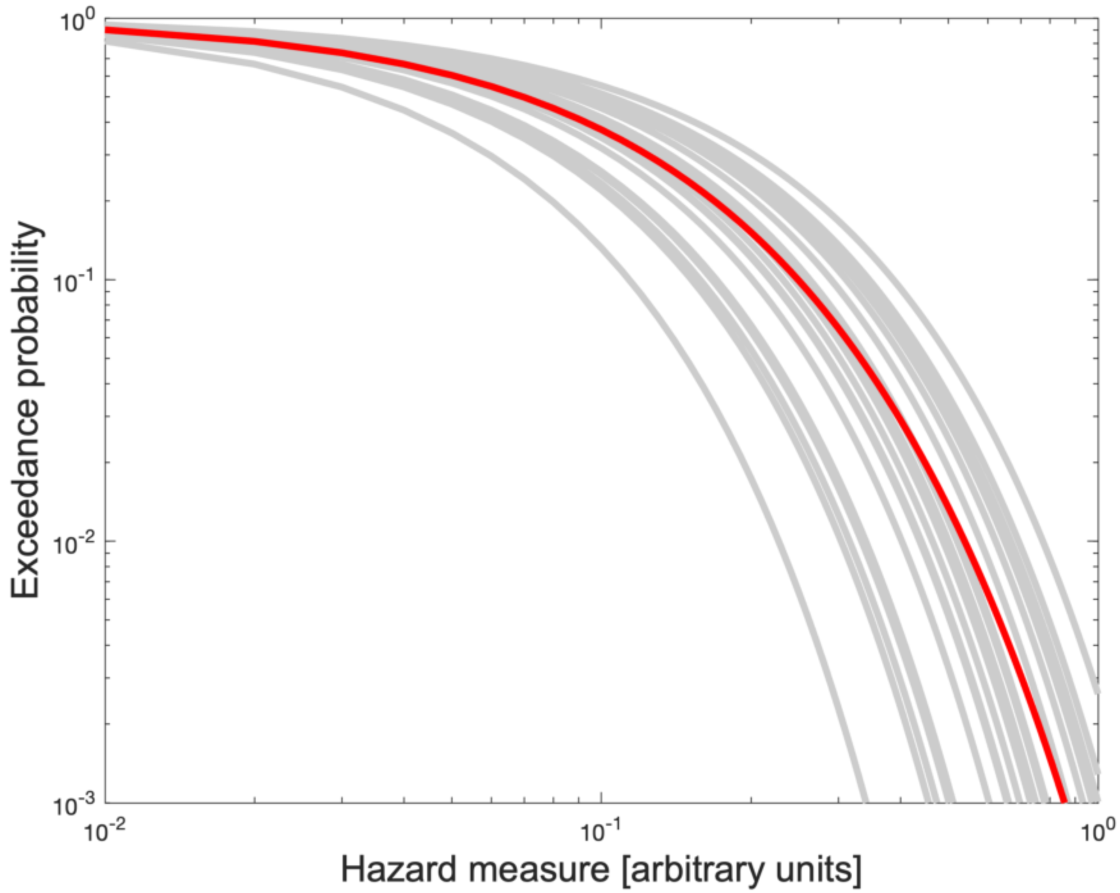
[the original ensemble model in weather forecasting]

A common story...



Often (always?) we do not know the ***true*** model, or its ***true*** parametrization, and we describe this additional and different kind of uncertainty through the use of multiple models (i.e., multiple distributions). This uncertainty is often named ***epistemic uncertainty***.

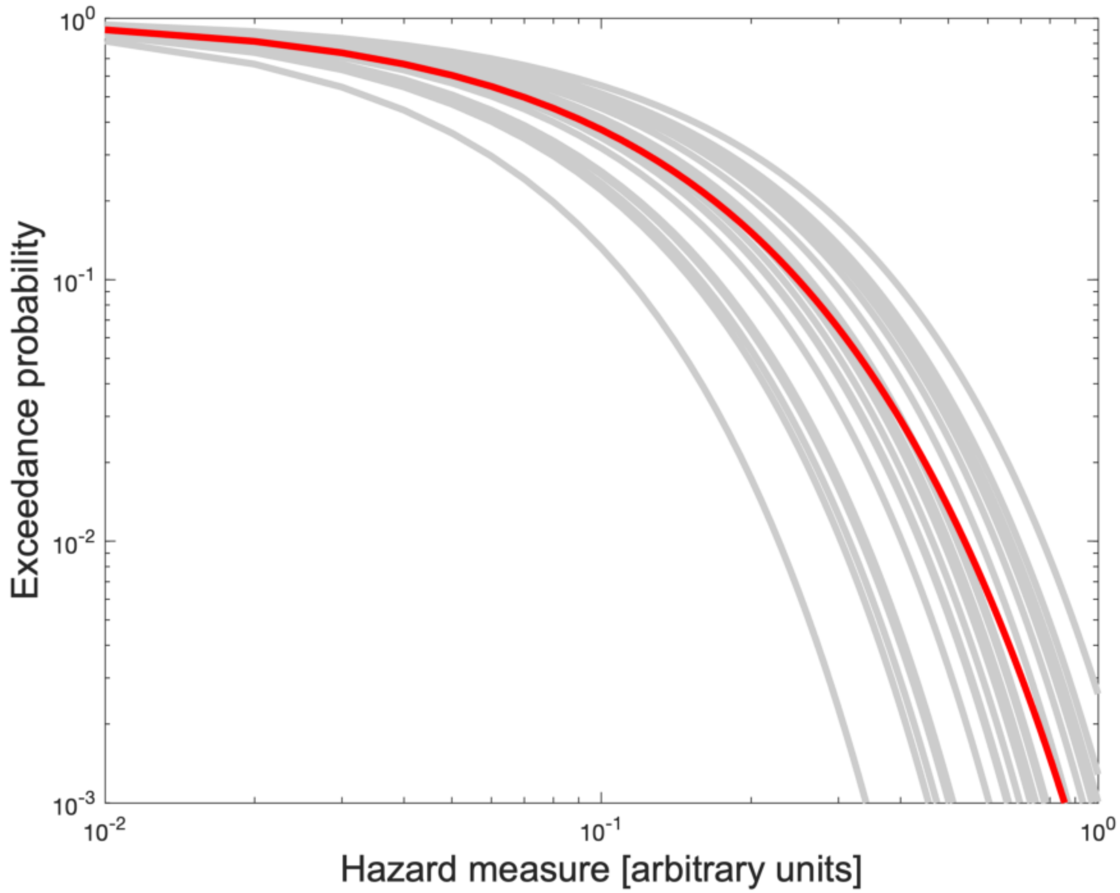
A common story...



The usual approach is to collapse all distributions into one (often, but not always, the “mean hazard”).

We name this distribution ensemble distribution (sometimes called multiensemble in weather forecasting and climate)

A common story...

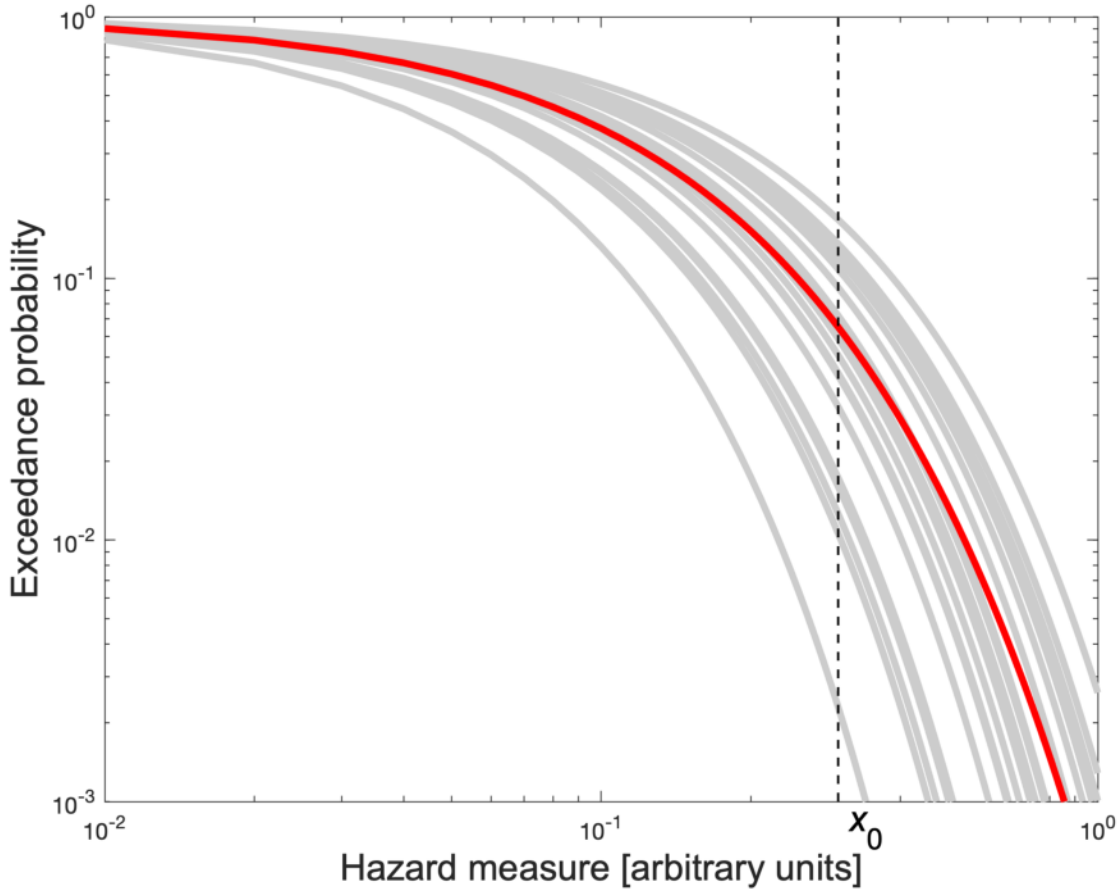


The usual approach is to collapse all distributions into one (often, but not always, the "mean hazard").

The mean hazard corresponds to a mixed distribution, which is the convolution of all distributions. The mean hazard can be used as testing distribution

We name this distribution ensemble distribution (sometimes called multiensemble in weather forecasting and climate)

Where the problems come from...



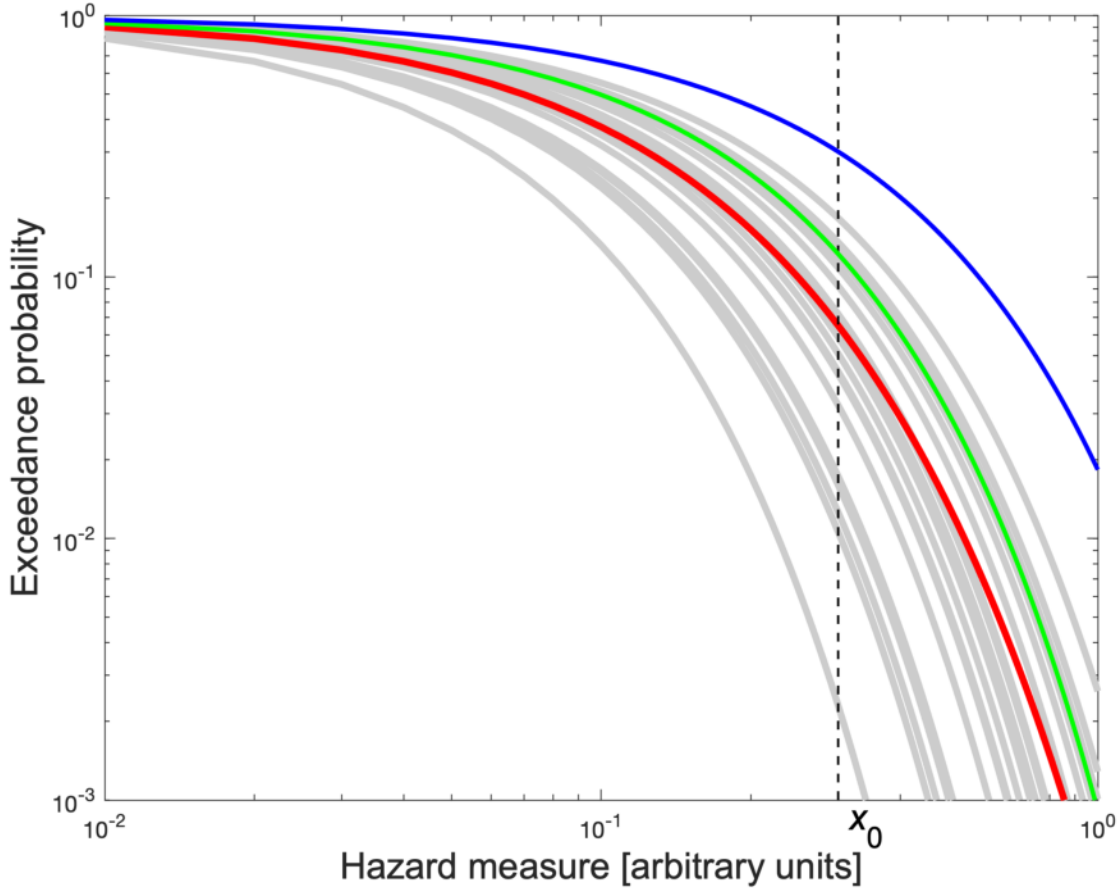
What is the EP for one specific hazard intensity x_0 ? Usually, the probability given by the ensemble distribution (red curve) is taken,

but...

this means to throw away all other distributions. For example, we can get the same mean hazard from widely dispersed or very close by distributions.
Is this information irrelevant?

If you think so, give a look at the next slides that describes the implications for testing the model.

The problem for testing ...



Let as consider two cases, (A) and (B), where we know the *true* distributions that we assume to exist:

(A) Green hazard curve

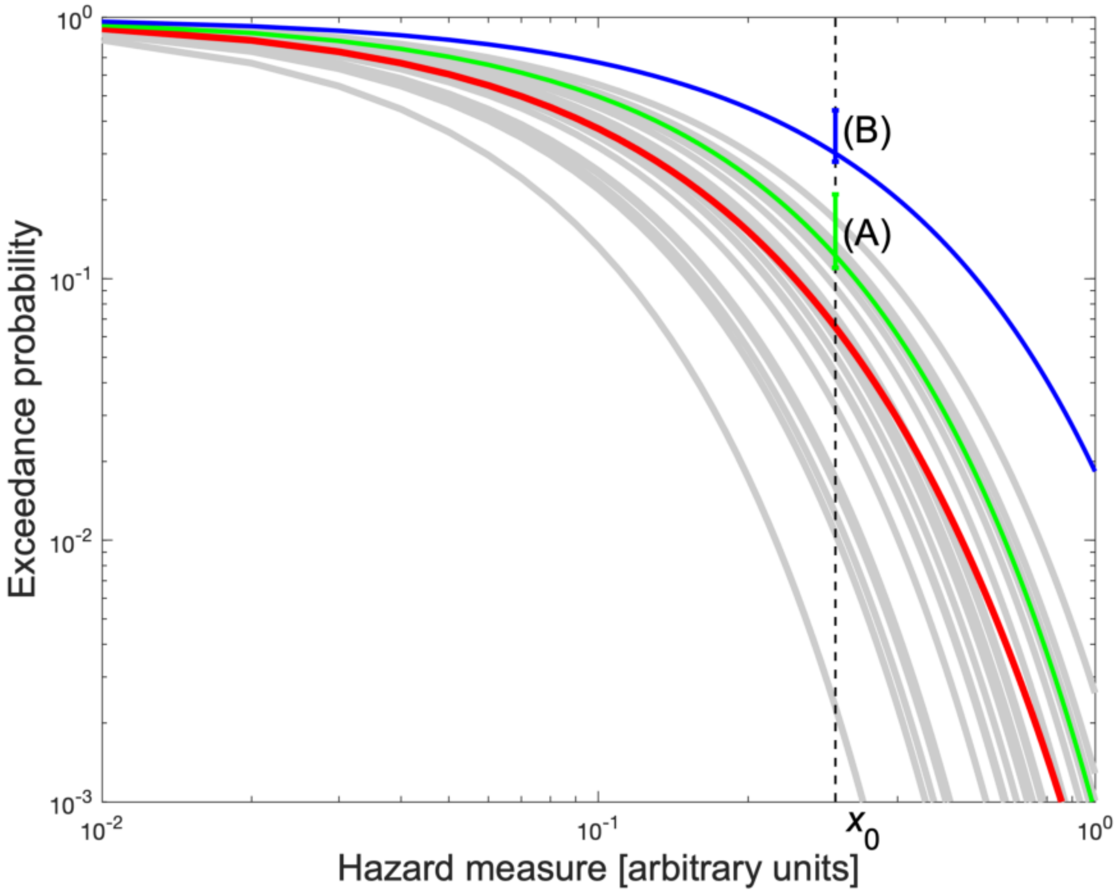
(B) Blue hazard curve.

They are different from the ensemble distribution.

Q: Are they compatible with our model?

If we give a heuristic meaning to the distribution of all models, the green line (case(A)) seems to be coherent with our model, whereas the blue line (case(B)) is not.

The problem for testing...

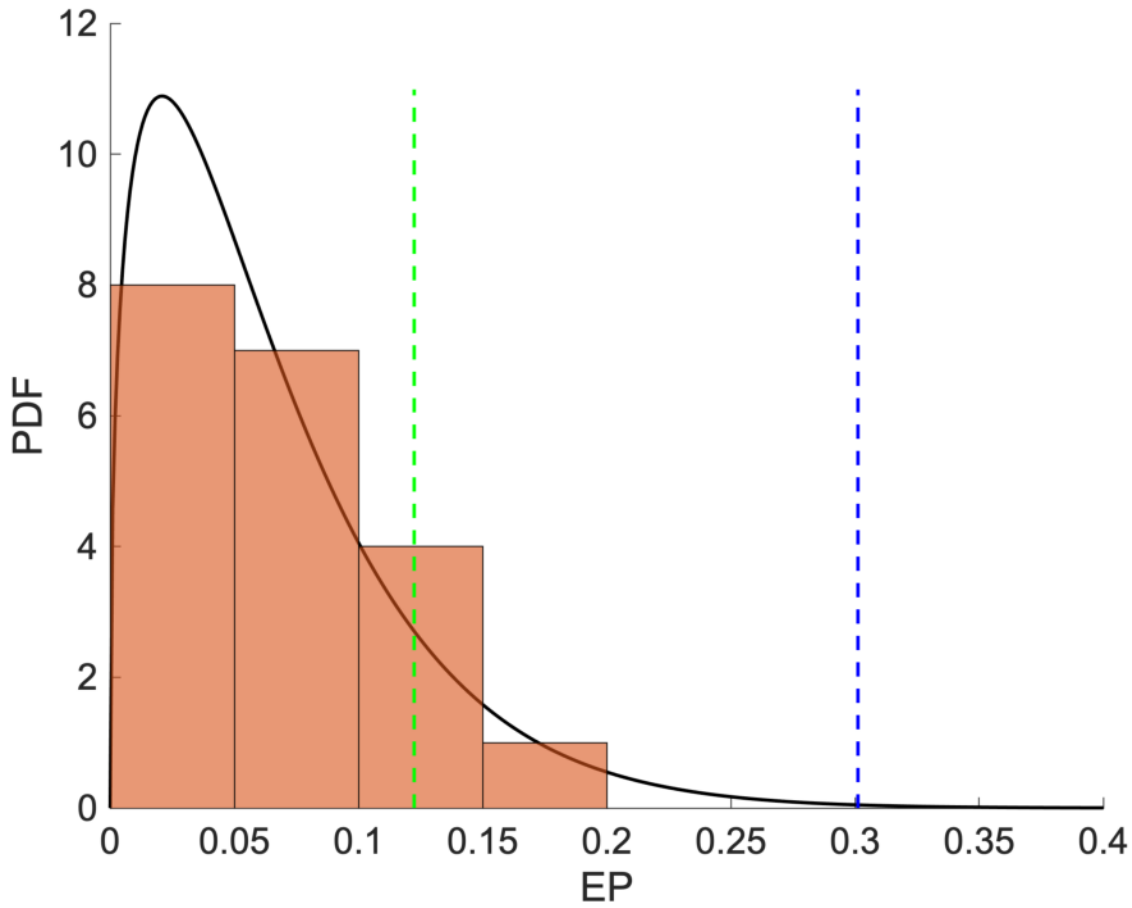


Let us collect some measurement of the exceedance frequencies for case (A) and (B) (green and blue bar respectively).

In this case, we can see (still heuristically) that

- the observations of case (A) are inside the range of distributions;
- the observation of case (B) are outside the range of distributions;
- for both cases, the ensemble distribution is not appropriate to describe the observations

Some conceptual problem...



Q: which is the EP of the hazard intensity x_0 ?

If we consider all models (i.e., we consider the epistemic uncertainty) we get a distribution of probability (the histogram and the interpolating black curve) instead of one single EP.
From this representation it is easy to see that the true probability of case (A) is coherent with the model, whereas the blue line of case (B) is not.

but...

the two most common probabilistic frameworks (***frequentist*** and ***subjectivist***) consider the probability as a **single number, not as a distribution**

We need a different framework to handle a set of probabilities keeping separated uncertainties of different kind.

The following slides introduce one possible solution:
the unified probabilistic framework

- The unified probabilistic framework is rooted in the “experimental concept”, which defines what we want to describe with our model (related to the usefulness of the model).
- Once an experimental concept is defined, the unified framework derives smoothly from it.

Definition of the "Experimental Concept"

- Specifies collections of data, observed and not yet observed, that are judged to be **exchangeable** when conditioned on a set of explanatory variables
  - *Definition:* A sequence of random variables $\{E_n : n = 1, 2, \dots, N\}$ is exchangeable if it can be embedded in an infinite sequence that has a joint probability distribution invariant with respect to permutations in the data ordering

- Exchangeable events can be modeled as identical and conditionally independent random variables with a well-defined frequency of occurrence (De Finetti's theorem)

$$p(e_1, e_2, \dots, e_N) = \int_0^1 \prod_{n=1}^{N} p(e_n | \phi)\, dP(\phi) = \int_0^1 \prod_{n=1}^{N} p(e_n | \phi)\, p(\phi)\, d\phi$$

  - **Exchangeability judgments** allow us to test Bayesian models using the Frequentist concept of experimental repeatability through identical trials

The unified probabilistic framework: definition of the "Experimental Concept"

Example of an exchangeable sequence of ***squares*** and ***circles***; we may define a long-term frequency of ***squares*** even though we know that they have different colors. Colors are considered irrelevant and/or their relevance not known.

The experimental concept creates a bridge between the *subjective world* of **unique events that cannot be described by a frequency,** and the *frequentist world* **in which probability is a frequency**.

Experimental concept is linked to the data-generating process *that is external to the model*

An example of two experimental concepts (but the same process) in PSHA having different aleatory–epistemic–ontological uncertainties

1. Collection of the ground shaking exceedance every year (**one** annual exceedance frequency, $f^{(I)}$)

2. Suppose to measure a binomial variable *A* that indicates years in which earthquakes are more or less likely. In this case we collect **two** series of yearly ground shaking exceedances, one when A=0, and the other when A=1 (**two** different annual frequencies, $f_1^{(II)}$ and $f_2^{(II)}$)

$$f^{(I)} \neq f_1^{(II)} \neq f_2^{(II)}$$

All of them are correct, but they refer to different experimental concepts (and the same process)

The definition of the **experimental concept allows ontological testing** of a complete probabilistic forecasting model through a hierarchy of uncertainties

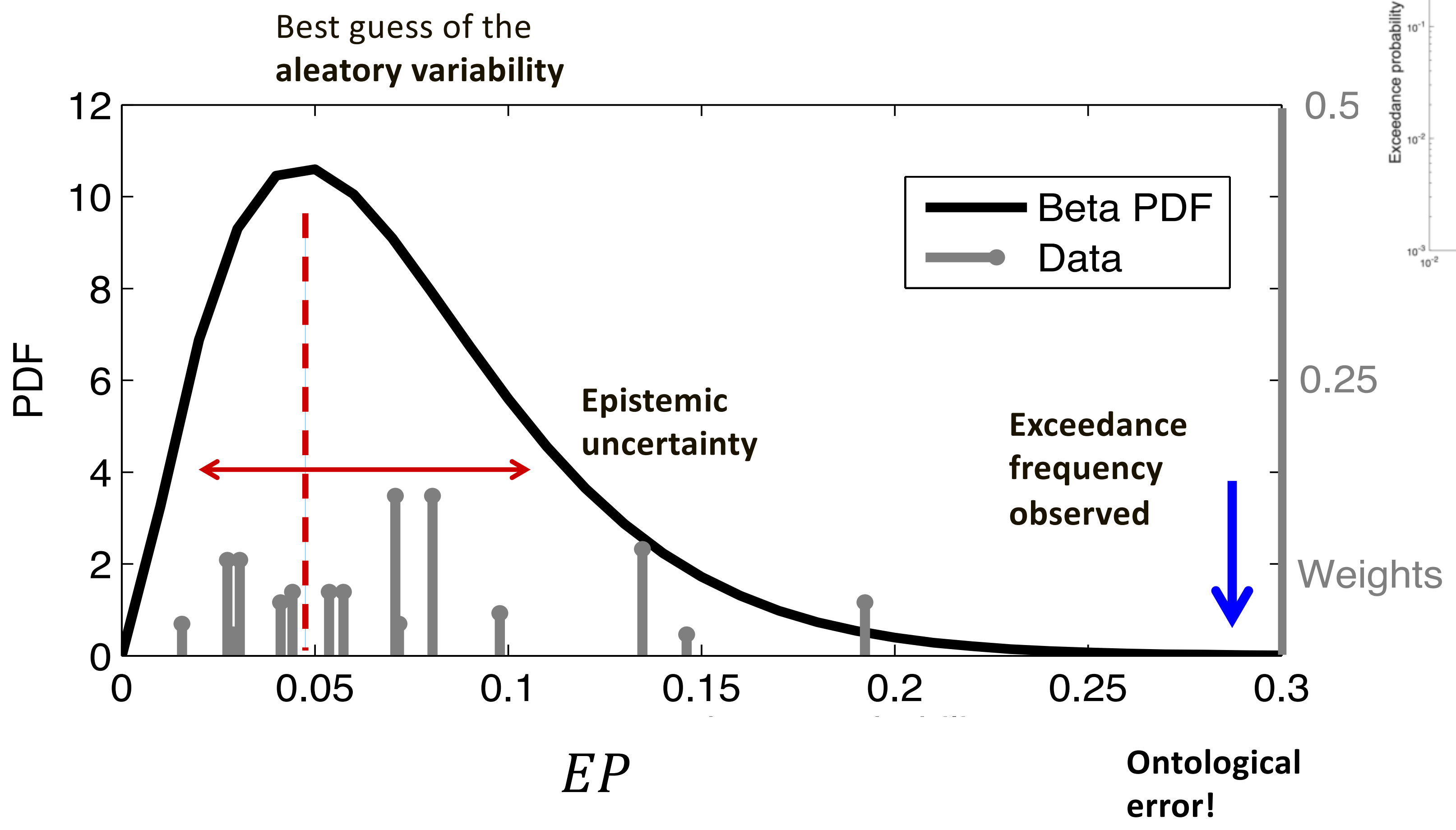
A hierarchy of uncertainties is necessary for testing

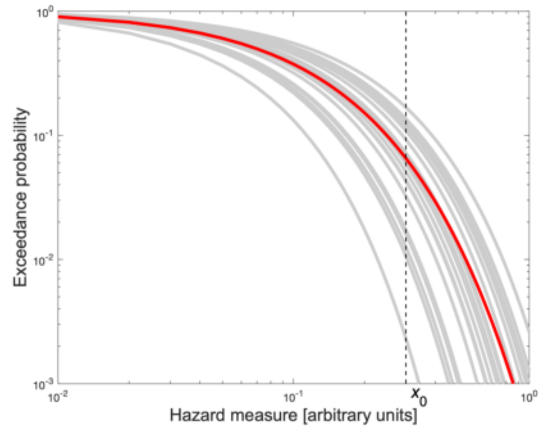
- **Aleatory variability** is quantified by the **expected (long-run) frequency** of events belonging to the data-generating process (defined by the experimental concept). Hypotheses about aleatory variability can be tested against observations by **frequentist (error-statistical) methods**.

- **Epistemic uncertainty** measure **lack of knowledge in the estimation of such frequency**; it implies a **distribution over the probability**. Bayesian methods are appropriate for reducing epistemic uncertainties as new knowledge is gained through observation.

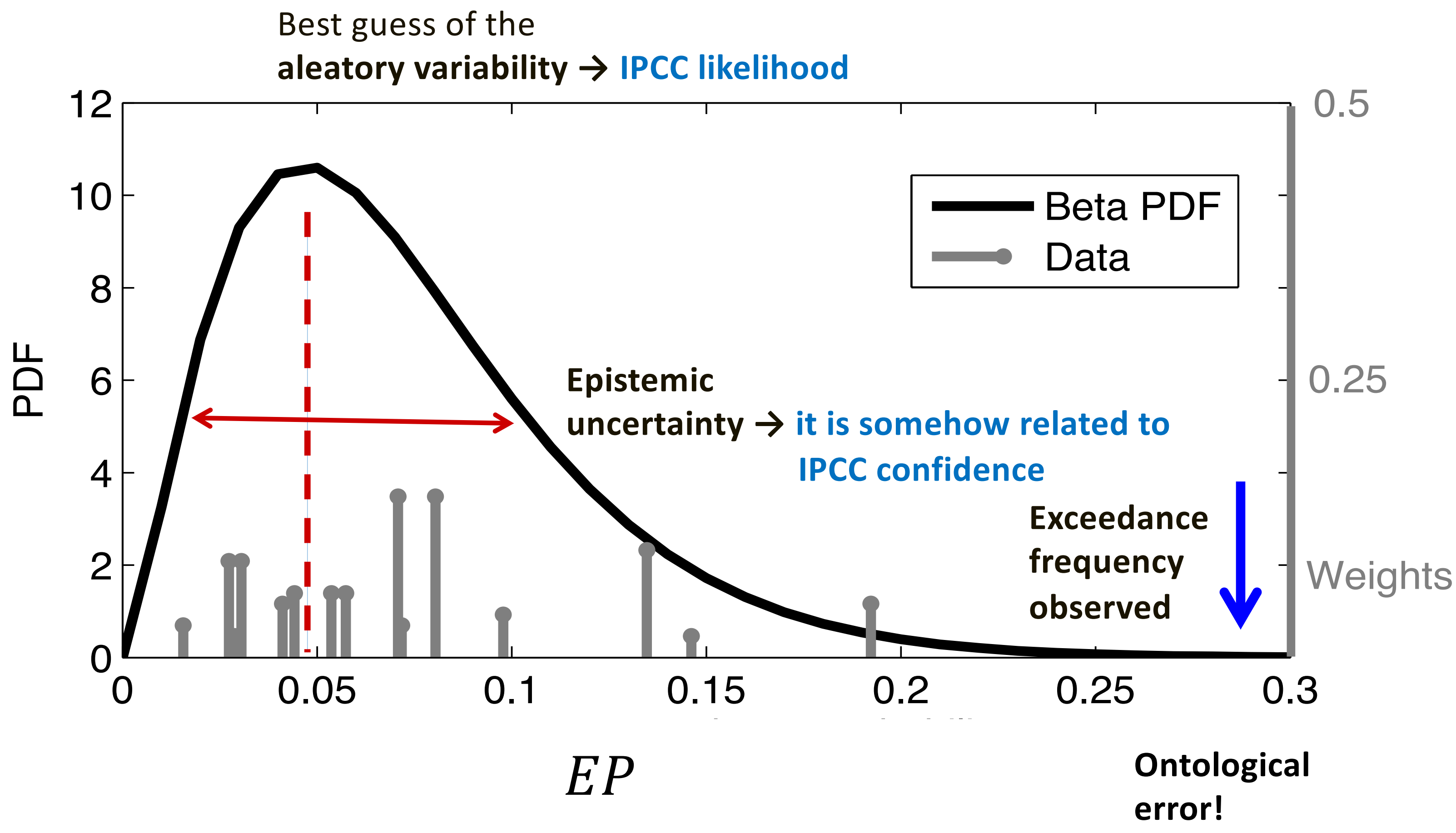
- **Ontological error** is identified by the rejection of a null hypothesis, here called the "**ontological null hypothesis**", which states that the *true frequency of the random events is a sample from the (joint) probability distribution describing the epistemic uncertainties.*

Different kinds of uncertainty and the *Extended Experts Distribution* EED





The link with IPCC framework ....



Here the aleatory variability (center of the distribution) is the likelihood given by IPCC. The epistemic uncertainty, i.e., the variability of the distribution, may be related to the subjective confidence given by IPCC, but a comparison is **not** adequate because the confidence **cannot** be interpreted as a probability.

Uncertainty Hierarchy of Earthquake Forecasting

Aleatory variability

"There are known knowns; there are things we know that we know.

Epistemic uncertainty

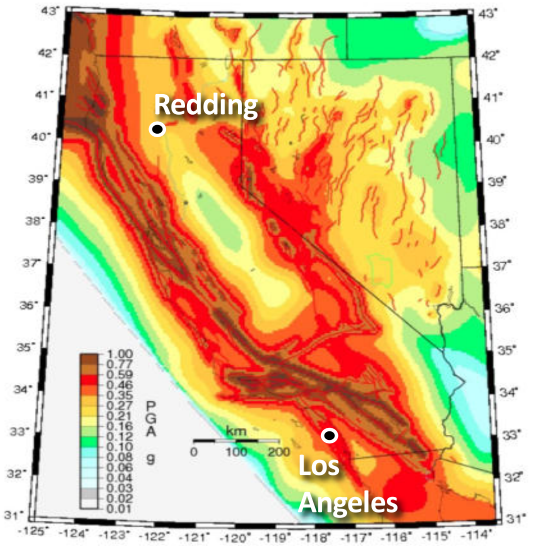
There are known unknowns; that is to say there are things that, we now know we don't know.

Ontological errors

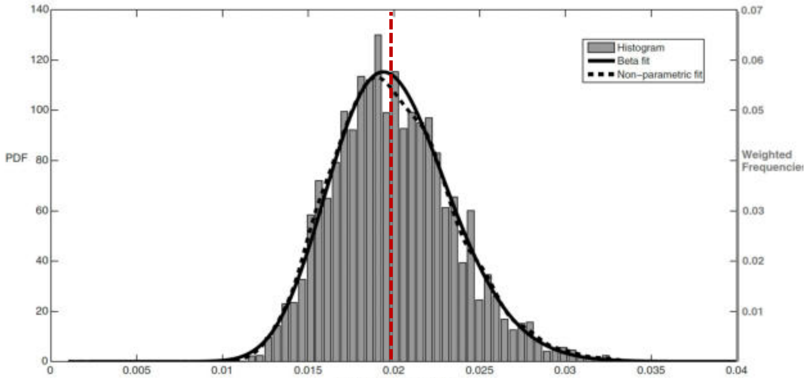
But there are also unknown unknowns – there are things we do not know we don't know."

Why is so important to keep separate uncertainties?

"This is not simply semantics: distinguishing between the two types of uncertainty is fundamental to the way that they are dealt with in the hazard calculations and how uncertainty is handled in decision making on the basis of the hazard analysis." (Abrahamson & Bommer, 2005)

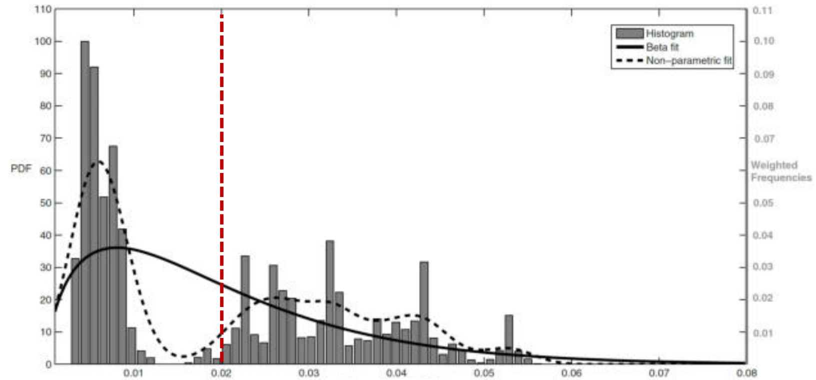


Los Angeles $\phi_0 = 2\%/50\ \mathrm{yr}$ $\bar{x}_0 = \bar{f}^{-1}(\phi_0) = 0.82\ \mathrm{g}$



PoE

Redding $\phi_0 = 2\%/50\ \mathrm{yr}$ $\bar{x}_0 = \bar{f}^{-1}(\phi_0) = 0.36\ \mathrm{g}$



PoE

Marzocchi et al., BSSA 2015