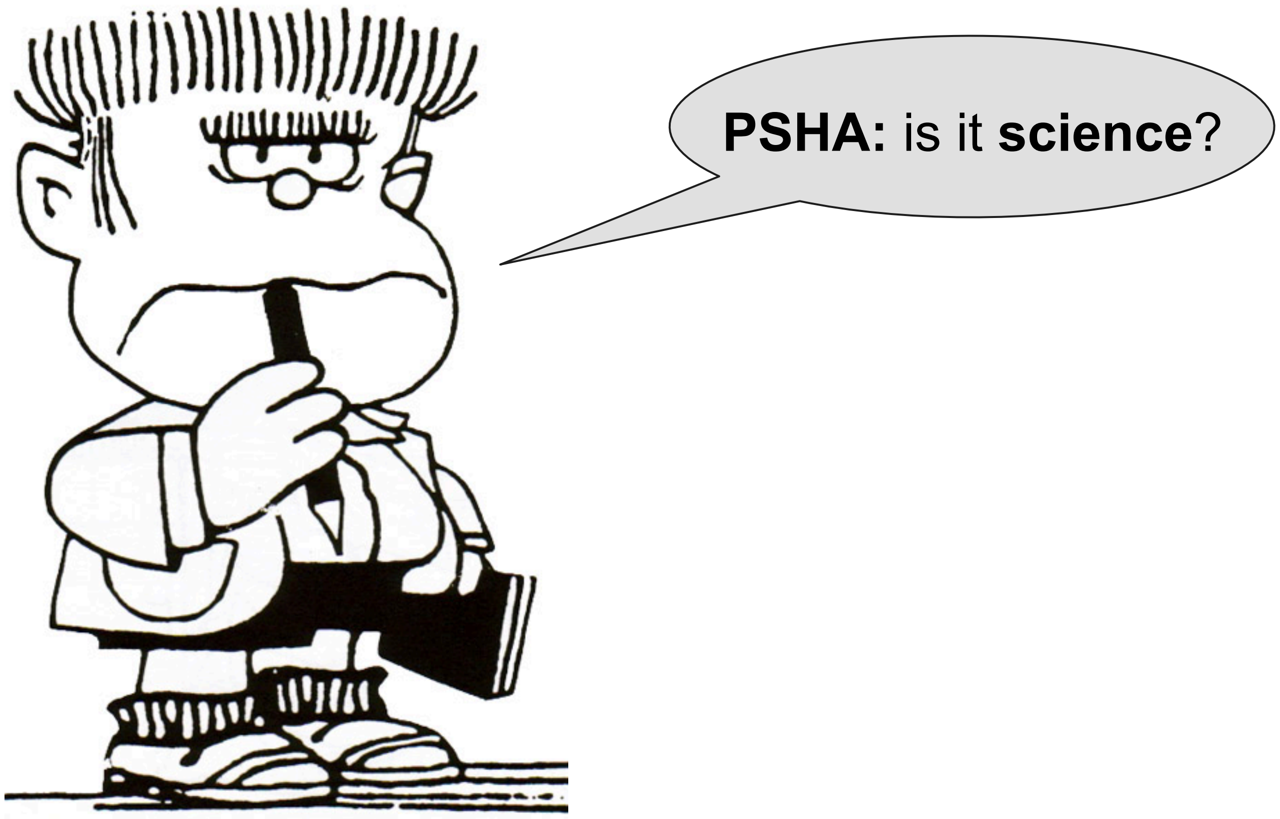
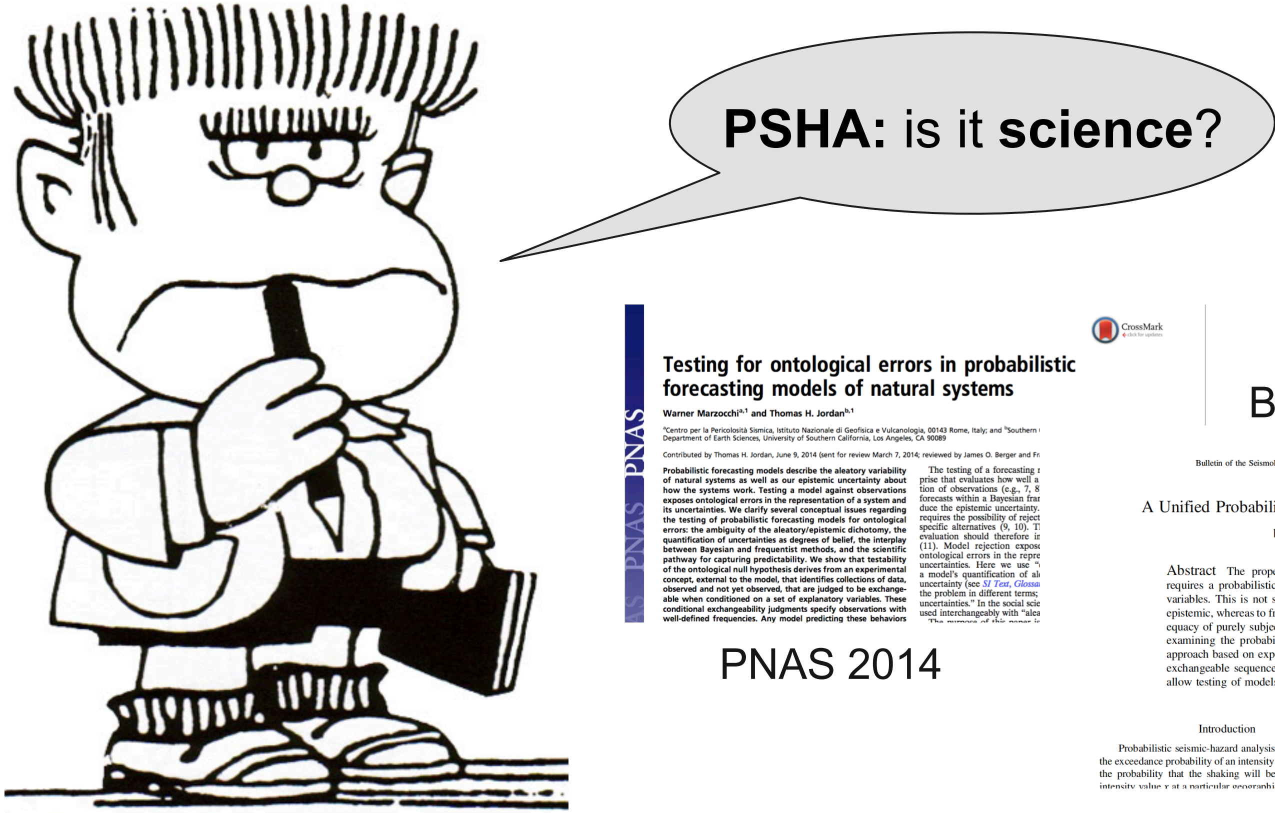
A unified probabilistic framework for seismic hazard analysis

Warner Marzocchi[1] & Thomas H. Jordan[2]

1 - Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

2 - University of Southern California, Los Angeles, USA


PSHA: is it **science**?



Testing for ontological errors in probabilistic forecasting models of natural systems

Warner Marzocchi[a,1] and Thomas H. Jordan[b,1]

[a]Centro per la Pericolosità Sismica, Istituto Nazionale di Geofisica e Vulcanologia, 00143 Rome, Italy; and [b]Southern
Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089

Contributed by Thomas H. Jordan, June 9, 2014 (sent for review March 7, 2014; reviewed by James O. Berger and Fr

Probabilistic forecasting models describe the aleatory variability of natural systems as well as our epistemic uncertainty about how the systems work. Testing a model against observations exposes ontological errors in the representation of a system and its uncertainties. We clarify several conceptual issues regarding the testing of probabilistic forecasting models for ontological errors: the ambiguity of the aleatory/epistemic dichotomy, the quantification of uncertainties as degrees of belief, the interplay between Bayesian and frequentist methods, and the scientific pathway for capturing predictability. We show that testability of the ontological null hypothesis derives from an experimental concept, external to the model, that identifies collections of data, observed and not yet observed, that are judged to be exchangeable when conditioned on a set of explanatory variables. These conditional exchangeability judgments specify observations with well-defined frequencies. Any model predicting these behaviors

The testing of a forecasting
prise that evaluates how well a
tion of observations (e.g., 7, 8
forecasts within a Bayesian fra
duce the epistemic uncertainty.
requires the possibility of reject
specific alternatives (9, 10). T
evaluation should therefore i
(11). Model rejection expose
ontological errors in the repre
uncertainties. Here we use "
a model's quantification of al
uncertainty (see *SI Text, Glossa*
the problem in different terms;
uncertainties." In the social scie
used interchangeably with "alea

PNAS 2014

BSSA 2017, in press

Bulletin of the Seismological Society of America, Vol. , No. , pp. –, , doi: 10.1785/0120170008

A Unified Probabilistic Framework for Seismic-Hazard Analysis

by W. Marzocchi and T. H. Jordan

Abstract The proper scientific interpretation of the seismic-hazard estimates requires a probabilistic framework that admits epistemic uncertainties on aleatory variables. This is not straightforward because, to subjectivists, all probabilities are epistemic, whereas to frequentists, all probabilities are aleatory. We illustrate the inadequacy of purely subjectivist and purely frequentist interpretations of probability by examining the probabilistic meaning of the mean hazard. We advocate a unified approach based on experimental concepts that define aleatory variability in terms of exchangeable sequences of observations, and we show how experimental concepts allow testing of models based on expert opinion by frequentist methods.

Introduction

Probabilistic seismic-hazard analysis (PSHA) estimates the exceedance probability of an intensity measure X; that is, the probability that the shaking will be larger than some

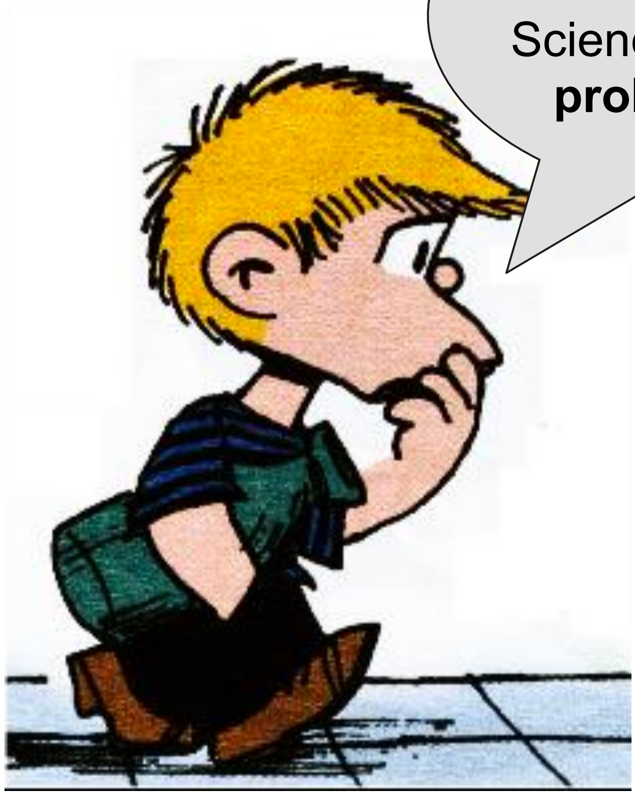
experts' distribution by $P(F(x);\mathcal{E})$. Various inductive methods have been used for estimating such a continuous distribution (e.g., SSHAC, 1997; Bommer, 2012; Goulet *et al.*, The growing ability of scientists to make **accurate predictions** about **natural phenomena** provides convincing evidence that we really are **gaining** in our **understanding** of how the **world works**.

AAAS (1989). Science for All Americans: A Project 2061 Report on Literacy Goals in Science, Mathematics and Technology

The growing ability of scientists to make **accurate predictions** about **natural phenomena** provides convincing evidence that we really are **gaining** in our **understanding** of how the **world works**.

Accurate prediction means that a forecasting model can be tested **against independent observations** and rejected when necessary

AAAS (1989). Science for All Americans: A Project 2061 Report on Literacy Goals in Science, Mathematics and Technology


Ok, I got it! "Forecast" is the essence of Science. But can the **probabilities be tested**?

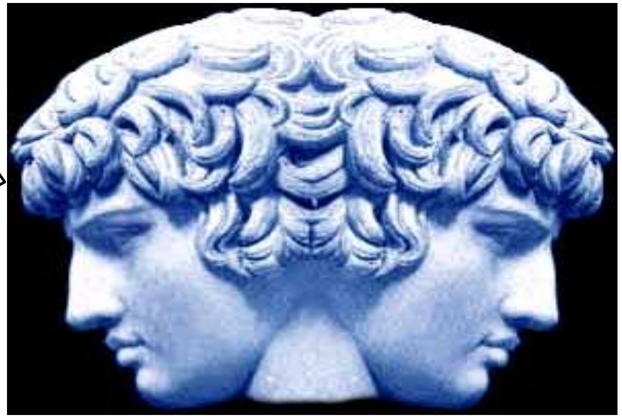

Well, that's not **so easy**!
mafalda

Are probabilities testable?

The Janus face of probability

(two apparently irreconcilable views of probability)

I am a subjective **degree of belief.** I am one single number that measures the **epistemic uncertainty,** that is the only kind of uncertainty.
I describe a **state of knowledge,** and not anything that can be measured in a physical experiment. **Probability is not a frequency and it is intrinsically subjective and untestable**



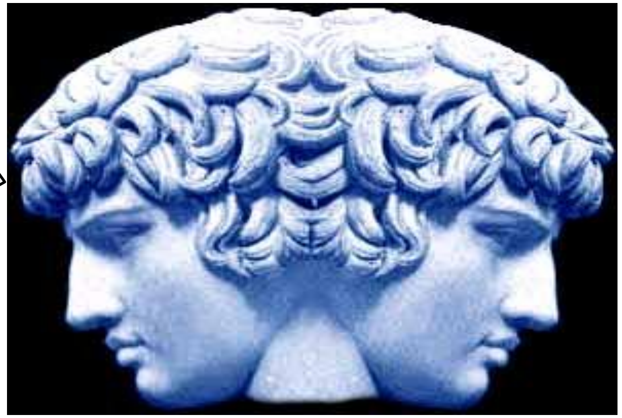
I am one single (unknown) number that reflects the **aleatory variability** of the system.
I am an **objective quantity** associated with a system model, and there is **no room for subjectivity** that cannot be posed at the same level as real measurements. **Subjectivity is fatally unscientific**

Are probabilities testable?

The Janus face of probability

(two apparently irreconcilable views of probability)

I am a subjective **degree of belief.** I am one single number that measures the **epistemic uncertainty,** that is the only kind of uncertainty.
I describe a **state of knowledge,** and not anything that can be measured in a physical experiment. **Probability is not a frequency and it is intrinsically subjective and untestable**



I am one single (unknown) number that reflects the **aleatory variability** of the system.
I am an **objective quantity** associated with a system model, and there is **no room for subjectivity** that cannot be posed at the same level as real measurements. **Subjectivity is fatally unscientific**

A common view in PSHA implies a dichotomy between **SUBJECTIVISM** and **SCIENCE.** **Expert opinion** implies that probability is a **degree of belief** and so **untestable**, and so also **nonscientific**.


Subjectivity is not an issue for Science!!! (pure objectivity is a myth)

Subjectivity is not an issue for Science!!! (pure objectivity is a myth)

“It is **not** unscientific to make a **guess**, although many people who are not in science think it is”

Subjectivity is not an issue for Science!!! (pure objectivity is a myth)

“It is **not** unscientific to make a **guess**, although many people who are not in science think it is”

“We look for new law by the following process: first, we guess it …. “

Probability & testing:
The Bayesian view

The view of an "objective Bayesian" about testing

"**All models are wrong**, and the purpose of model checking (as we see it) **is not to reject a model** but rather to understand the ways in which it does not fit the data. From a Bayesian point of view, the posterior distribution is what is being used to summarize inferences, so this is what we want to check."

A. Gelman

The view of an "objective Bayesian" about testing

"**All models are wrong**, and the purpose of model checking (as we see it) **is not to reject a model** but rather to understand the ways in which it does not fit the data. From a Bayesian point of view, the posterior distribution is what is being used to summarize inferences, so this is what we want to check."
A. Gelman

Models cannot be meaningfully tested against independent data ("**all models are wrong**", so why wasting time to validate them?).
You can only compare different models

Probability describes only the **epistemic uncertainty** and **cannot be tested** (meaningfully) with independent data

(plus some other important by-products...)

- Probability is ***one*** number: the ***mean*** hazard is ***the*** hazard
- **Fractiles** do not have any probabilistic meaning

Probability & testing:
The Frequentist view

- Probability is ***one*** number that estimates the long-term frequency (**aleatory variability**), and so it may be potentially tested against data.
- However, there is no room for **epistemic uncertainty**.

Probability & testing:
The unificationist view
(a univocal hierarchy of uncertainties)

A common view on aleatory variability and epistemic uncertainty

Aleatory variability is related to the inherent complexity of the process generating ground shaking

Epistemic uncertainty comes from our lack of knowledge about the process

A common view on aleatory variability and epistemic uncertainty

Aleatory variability is related to the inherent complexity of the process generating ground shaking

Epistemic uncertainty comes from our lack of knowledge about the process

This definition raises some problems

We cannot define unambiguously what is aleatory and epistemic, and, in the limit, **all uncertainties are epistemic**.

If aleatory variability and epistemic uncertainty **cannot be unequivocally defined** (they are moving targets as long as the knowledge increases), **how can they be helpful for testing a model**?

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 $\{E_n : n = 1, 2, \ldots , N\}$ 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
  - **Exchangeability judgments** allow us to test Bayesian models using the Frequentist concept of experimental repeatability through identical trials
- Definition of the **experimental concept allows ontological testing** of a complete probabilistic forecasting model
  - By modifying the experimental concept to incorporate multiple sets of exchangeable data, we can construct more stringent tests of the model

A hierarchy of uncertainties is necessary for testing

- **Aleatory variability** is quantified by the **expected (long-run) frequency** of events belonging to an 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*.

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."*

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, $f^{(l)}$)

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, $f^{(I)}$)

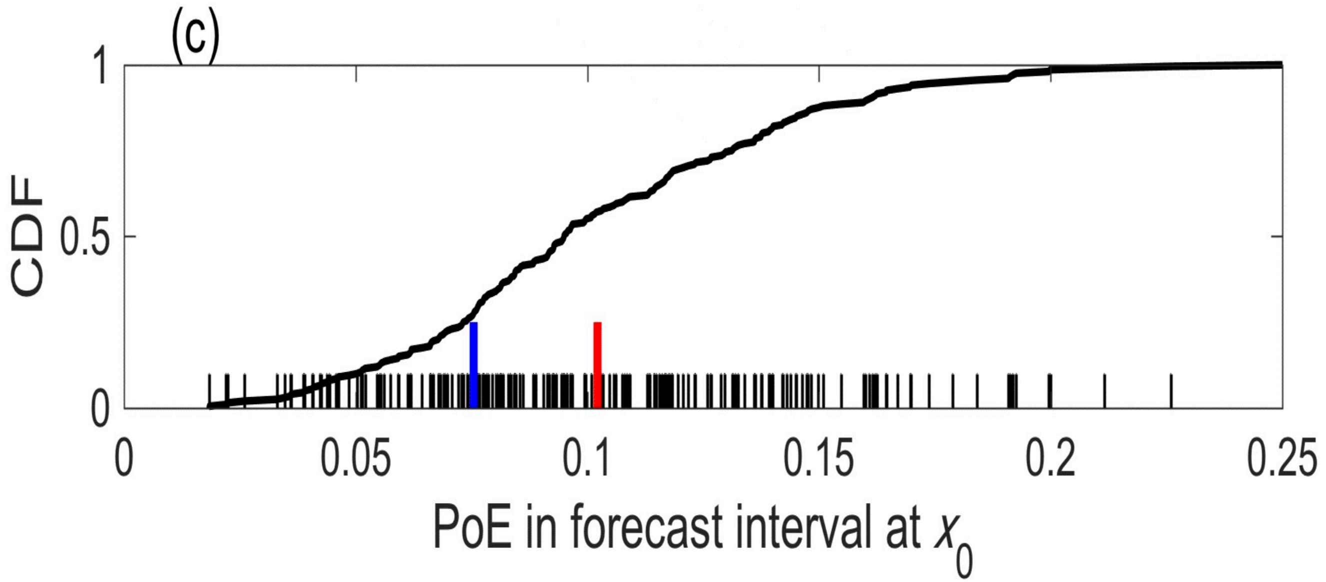
2. Suppose to measure a binomial variable *A* that indicates years in which earthquakes are more or less likely. In this case we collect **two** series of yearly ground shaking exceedances, one when *A*=0, and the other when *A*=1 (**two** different annual frequencies, $f_1^{(II)}$ and $f_2^{(II)}$)

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

The Unificationist view

$$\bar{F}(x) \equiv \sum_{m=1}^{M} F_m(x)\,\pi_m = \sum_{m=1}^{M} F(x \mid H_m)P(H_m).$$

$P(H_m)$ is the weight of the *m*-th model to measure the true frequency



Extended experts' distribution (EED); the *ensemble* modeling

The Ontological Null Hypothesis

$\hat{f}(x) \equiv$ *data-generating hazard curve* ("true hazard")

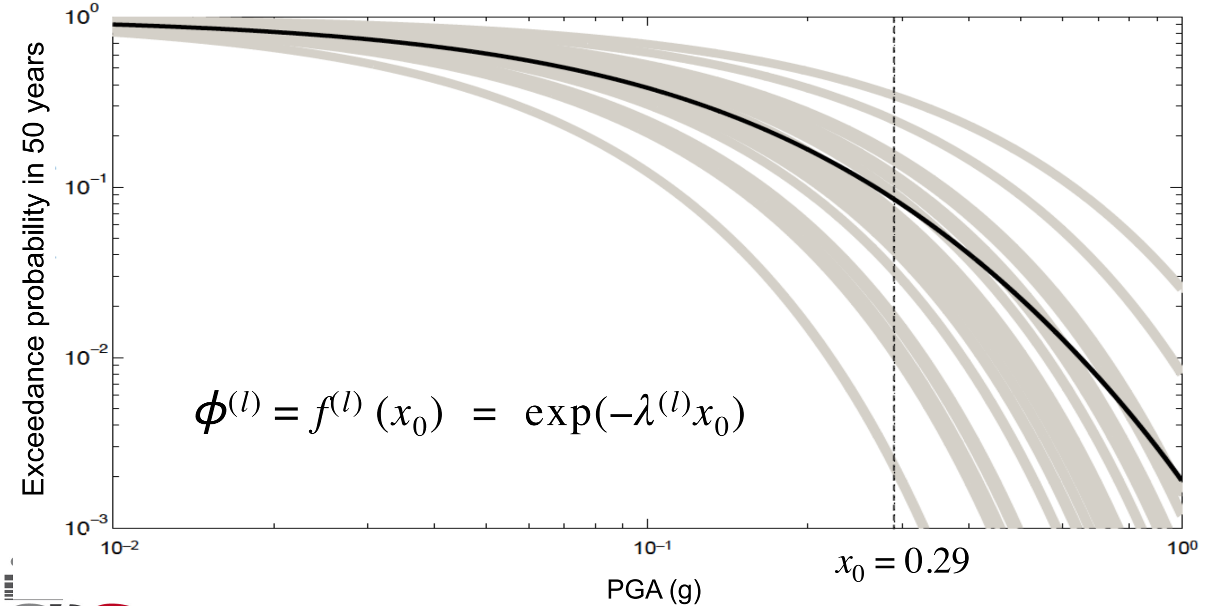
- Testability derives from an ***ontological null hypothesis***, which states that the true hazard curve is a realization of the EED:

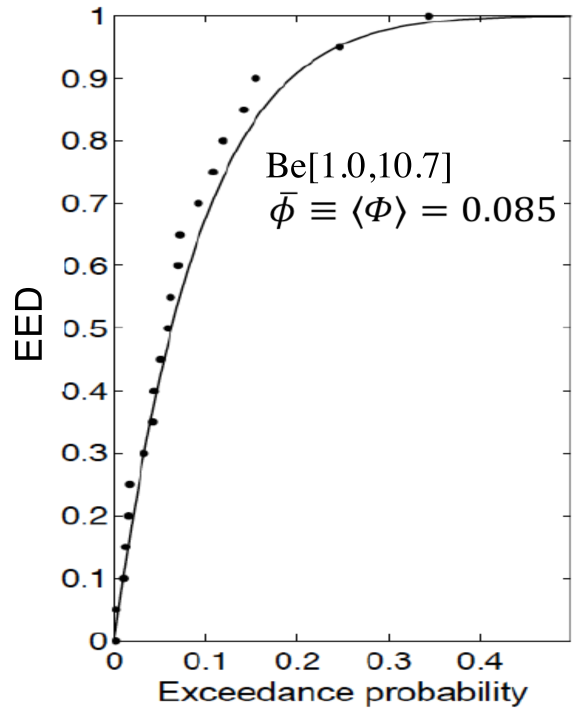
$$\mathbb{O} : \hat{f}(x) \sim P(F(x); \mathcal{E})$$

- Rejection of this null hypothesis implies an ontological error
  - Very different from the much more stringent statement that $\hat{f}(x) = \bar{f}(x)$
- Setting up an ontological test requires an *experimental concept* that appropriately conditions the aleatory variability of the natural system
  - In forecast testing, the most crucial feature of an experimental concept is the judgment that past and future events sample an *exchangeable sequence*

Simple PSHA Example
(Time-independent model for a single location)

Experts' ensemble comprising 20 exponential hazard curves $f^{(l)}$ sampled at $x_0 = 0.29$, from which we induce the EED: $P(\Phi;\mathcal{E}) = \mathrm{Beta}[1.0,10.7]$





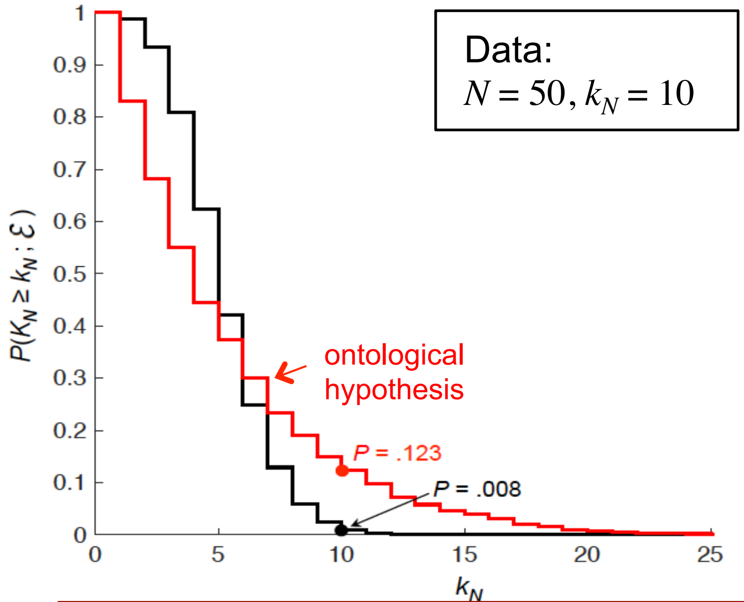
Test 1 of the Single-Location, Time-Independent Model

- Preselect $x_0 = .29$ as the exceedance threshold and $\alpha = .05$ as the significance level
- Observe maximum intensities at a single location in each of N disjunct intervals T_n, assigning $e_n = 1$ if $x_n > x_0$ and $e_n = 0$ if $x_n \le x_0$
- Compute the exceedance score k_N by summing the binary sequence $\{e_n : n = 1, 2, \dots, N\}$
- Test distribution conditional on Φ is binomial

$$P(K_N \geq k_N \mid \Phi) = \sum_{n=k_N}^{N} \binom{N}{n} \Phi^n (1-\Phi)^{N-n}$$

- Ontological test distribution is a binomial mixture

$$P(K_N \geq k_N; \mathcal{E}) = \int_0^1 P(K_N \geq k_N \mid \Phi)\, dP(\Phi\,; \mathcal{E})$$

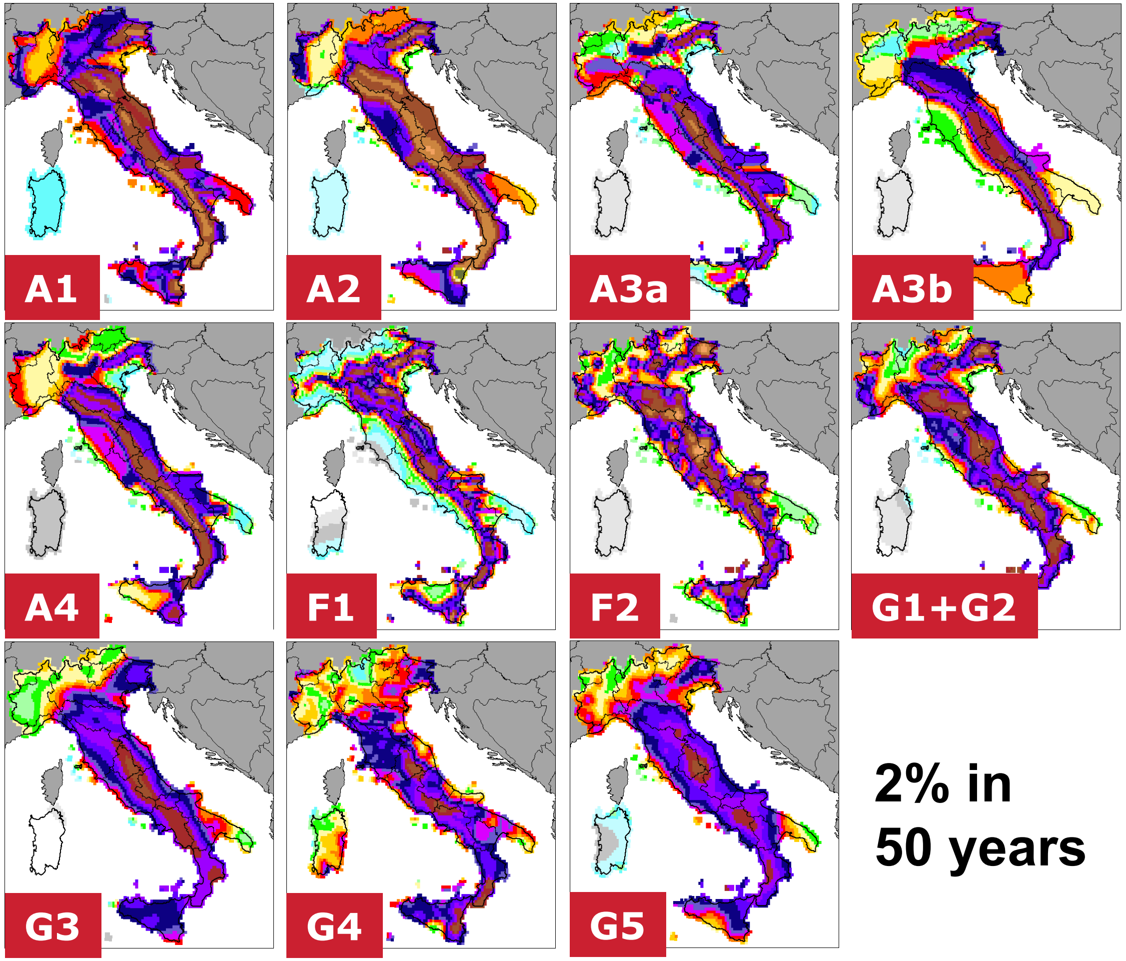


The mean-hazard model fails ($P = .008$) the test, but the complete probabilistic model does not ($P = .123$)

- **Probability is a frequency** of events in a ***experimental concept*** (aleatory variability)
- This frequency is unknown and this uncertainty is the ***epistemic uncertainty***
- ***All models are wrong,*** but some can be ***“right”*** in a defined experimental concept; the experimental concept can be related to the ***“usefulness”*** of the model

An application: The new seismic hazard model for Italy

(thanks to Carlo Meletti and all the other Italian colleagues!)



2% in 50 years

Legenda

< 0.025
0.025-0.050
0.050-0.075
0.075-0.100
0.100-0.125
0.125-0.150
0.150-0.175
0.175-0.200
0.200-0.225
0.225-0.250
0.250-0.275
0.275-0.300
0.300-0.350
0.350-0.400
0.400-0.450
0.450-0.500
0.500-0.600
0.600-0.700
0.700-0.800
0.800-0.900
0.900-1.000
1.000-1.250
1.250-1.500
1.500-1.750
1.750-2.000

$\{F_m(x), \pi_m\}$ Set of earthquake rate models/ GMPEs/hazard models, and their "weight"

A – Seismotectonic zonations

F – Faults-based

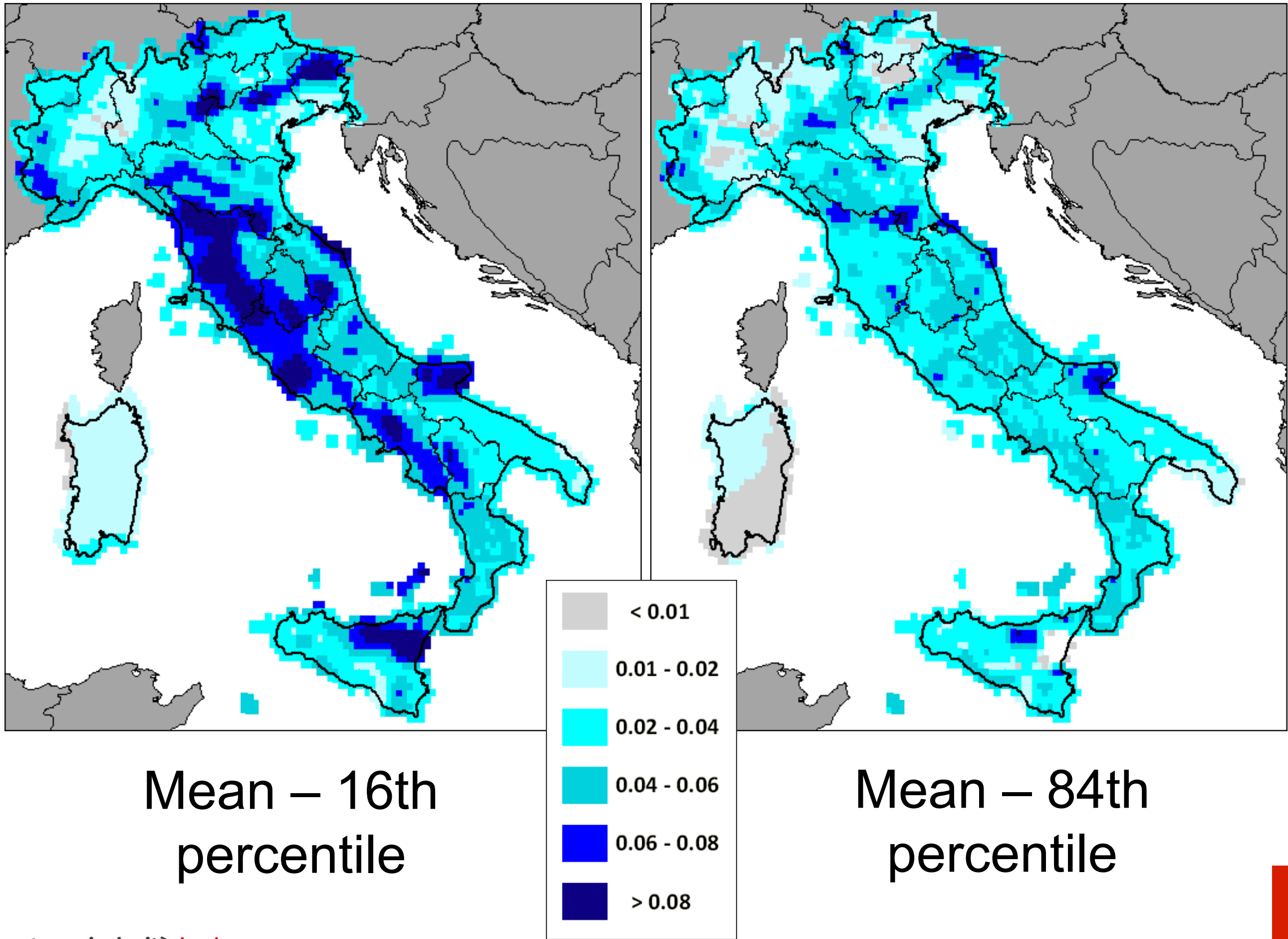
G – gridded seismicity

G3 and **G4** – based on deformation data

PRELIMINARY RESULTS

The case of earthquake rate models

- All models explore their **own epistemic uncertainty**
- All considered models have been **tested for consistency** with past data (pseudo-validation through **CSEP-type** tests).
- All models have been set up **independently** from the others
- All models are **weighted** according to **three independent procedures**:
  - the scoring through retrospective testing;
  - the evaluation of the reliability of models through an experts' elicitation session;
  - the correlation among outcomes



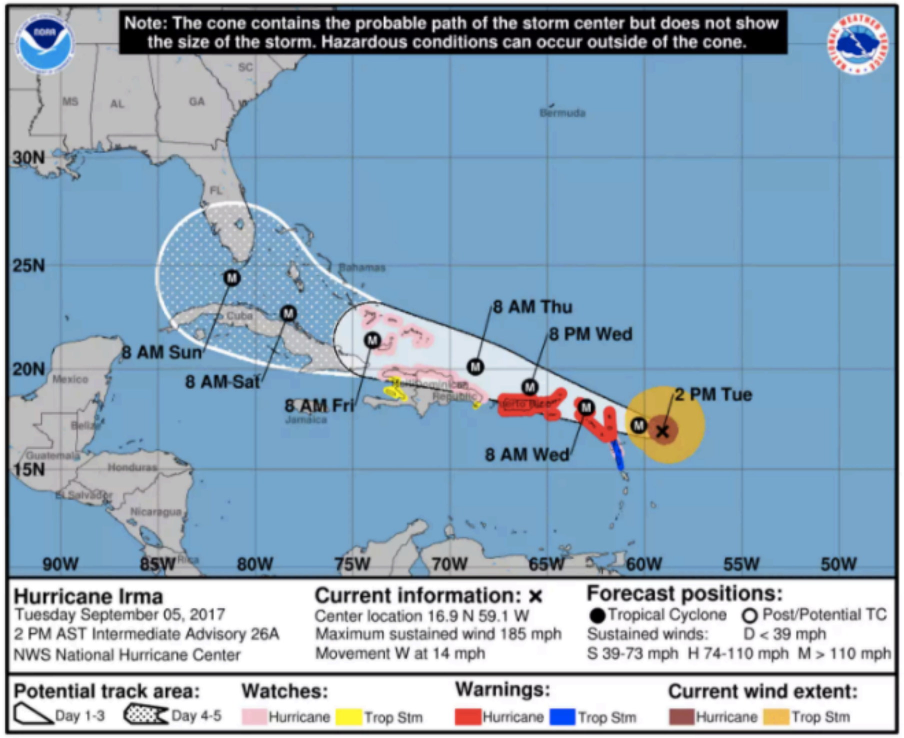
Mean – 16th percentile

Mean – 84th percentile

A spatial representation of **epistemic uncertainty** among models (10% in 50 years)

PRELIMINARY RESULTS

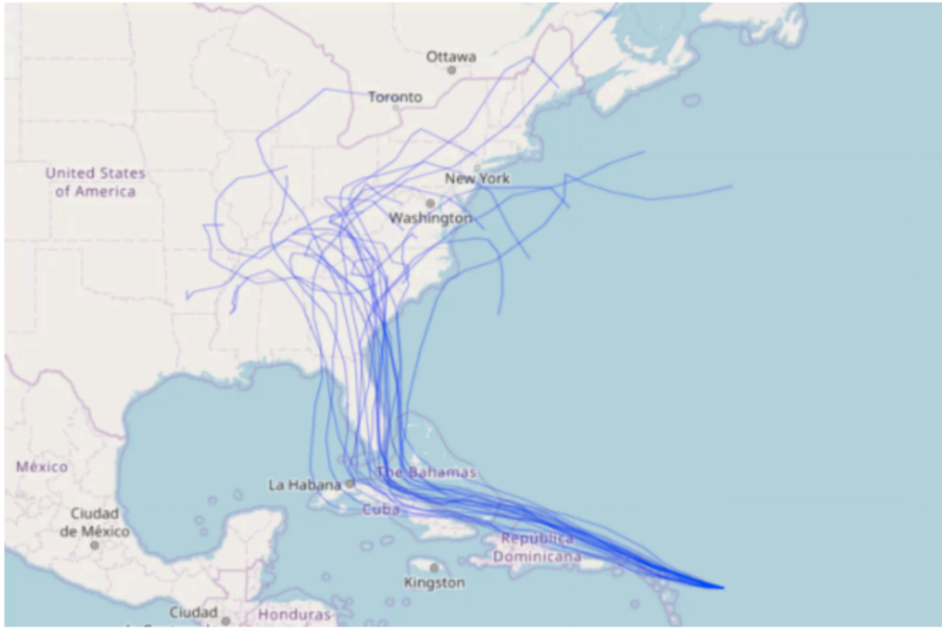
Hurricane Irma (Sept. 6, 2017)



The cone of uncertainty for Hurricane Irma as of 2 p.m. Tuesday. (National Hurricane Center)

The cone of uncertainty
(aleatory variability of one single forecast)

Spaghetti models (ensemble modeling)

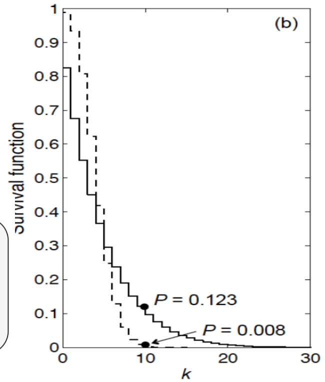


Spaghetti model plot from the GFS model run Tuesday morning for Hurricane Irma. (StormVistaWxModels.com)

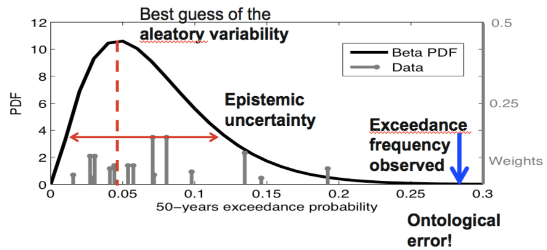
Points to **take home**

A **full** description of the hazard is essential to keep hazard analysis into a **scientific domain** (it allows a proper validation)

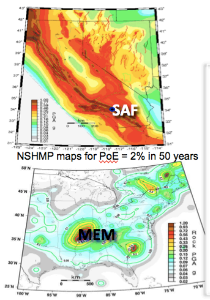
The growing ability of scientists to make **accurate predictions** about **natural phenomena** provides convincing evidence that **we really are gaining in our understanding of how the world works**.

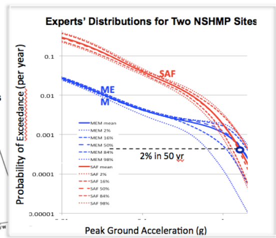


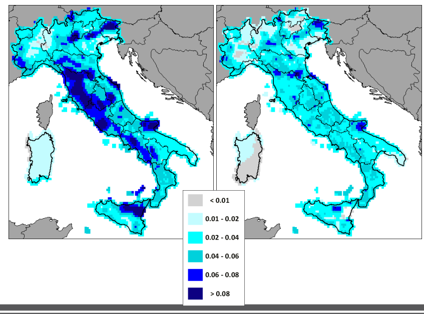
A **full** description of the hazard requires that probability has to be described by a **distribution** (which describes **aleatory variability and epistemic uncertainty**) instead of one single value

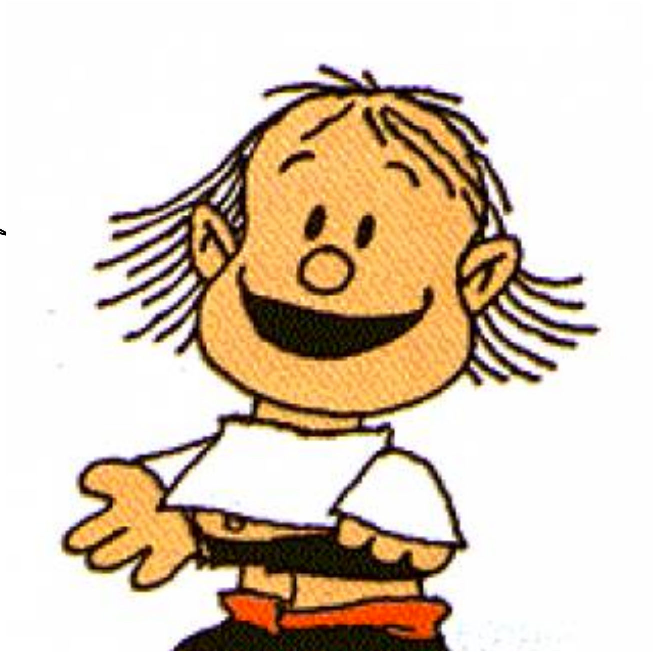


A **full** description of the hazard is important for **interpreting** correctly the **hazard outcomes**, and for **decision makers** (***mean*** **hazard is not the hazard)**









Thanks!

warner.marzocchi@ingv.it