# What Should PSH Maps Do and How Well Do They Do It?

Seth Stein[1], Edward M. Brooks[1], Bruce D. Spencer[2], Kris Vanneste[3], Thierry Camelbeeck[3], Bart Vleminckx[3], Leah Salditch[1], Antonella Peresan[4]

[1]Earth & Planetary Sciences & Institute for Policy Research, Northwestern University, Evanston IL USA

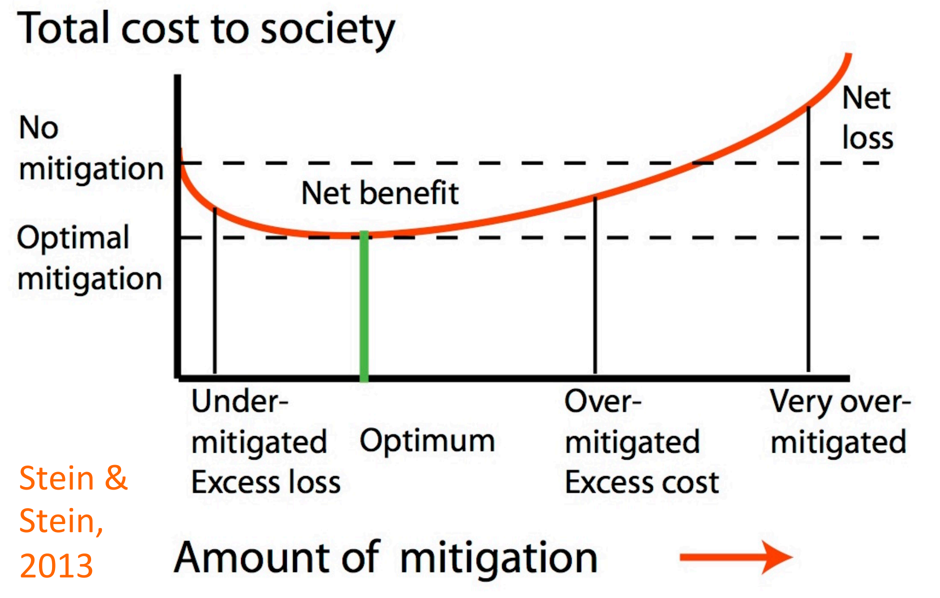
[2]Statistics & Institute for Policy Research, Northwestern University

[3]Royal Observatory of Belgium, Brussels, Belgium

[4]National Institute of Oceanography and Experimental Geophysics, Udine, Italy

# View: PSH maps are forecasts of future shaking and should be evaluated as forecasts

## Challenge: Although PSHA has been widely used for 40+ years, we and others are only starting to study how well it does what it's supposed to do



Stein & Stein, 2013

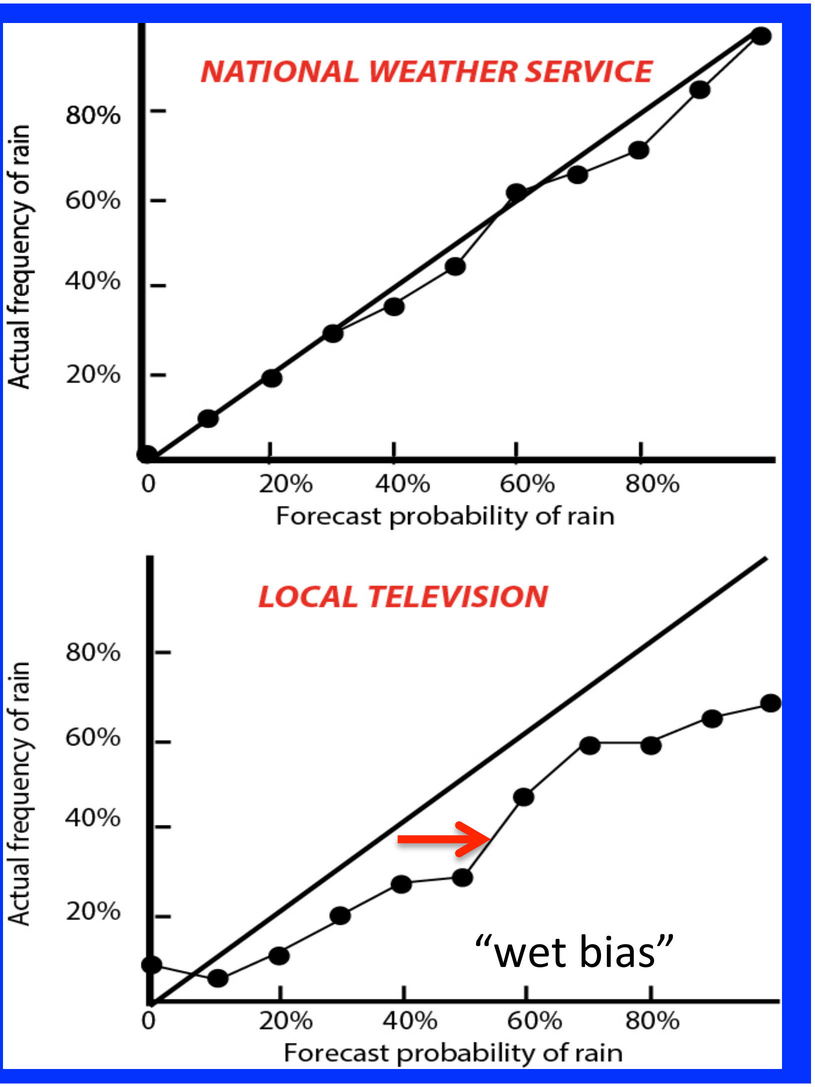
Scientific goal: Evaluate PSHA performance to better understand and improve maps

Policy goal: Assess how good maps should be to better help engineers and policy makers develop cost-effective hazard mitigation policies given other societal needs

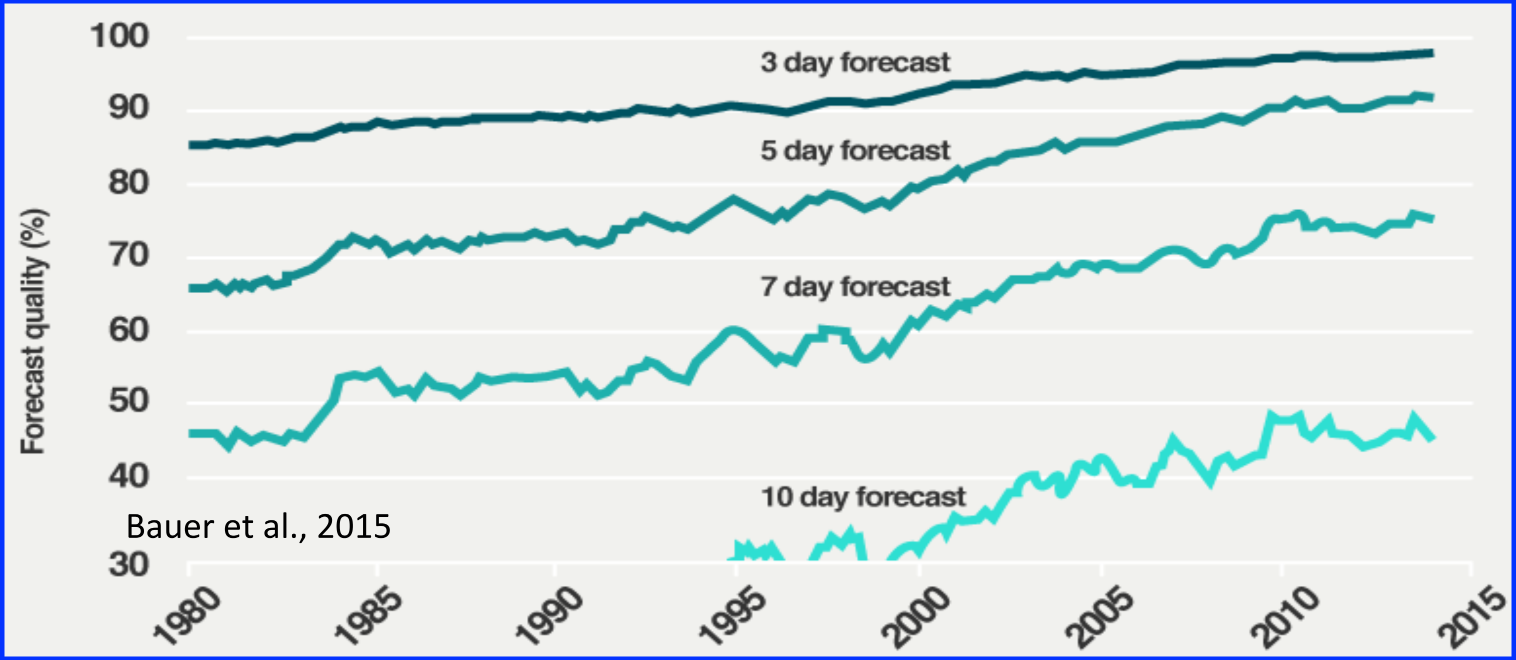
# Meteorology shows assessment is crucial for improvement

Weather forecasts are evaluated to assess how well their predictions matched what occurred:

"it is difficult to establish goals for any project designed to enhance forecasting performance without an unambiguous definition of what constitutes a good forecast." (Murphy, 1993)



# Assessing weather forecast performance has helped improve forecasts



Bauer et al., 2015

Current 5-day forecast is as good as 3-day forecast 20 years ago.
Useful (60%) forecast window has increased by a day every 10 years

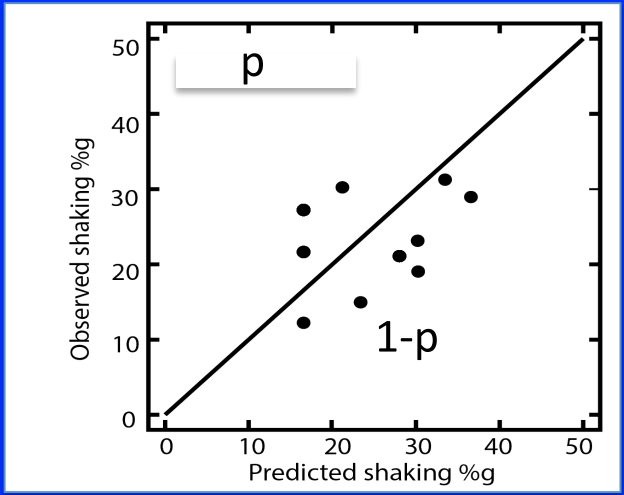
## How well do probabilistic seismic hazard maps forecast future shaking?

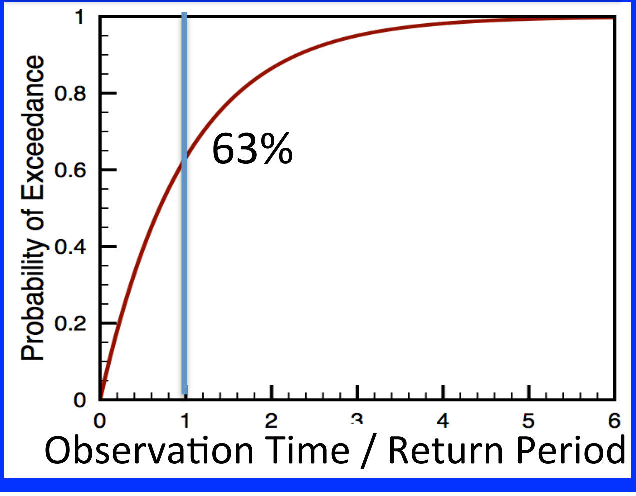
As for any forecast, this question involves verification and validation.

Verification involves assessing how well the algorithm used to produce maps implements the conceptual PSHA model (*"have we built the map right?"*).

Validation asks how well a map forecasts the shaking that actually occurs (*"have we built the right map?"*).

# Expected PSH map performance





At a point on the map, the probability *p* that during *t* years of observations shaking will exceed a value on a map with a τ-year return period is assumed to be

$$p = 1 - exp(-t/\tau)$$

The fraction of sites at which observed shaking exceeds the mapped value should behave the same way.

Shaking predicted by a map with a τ-year return period should be exceeded at 1% of the sites in $t = \tau/100$ years, 10% in $t = \tau/10$ years, and 63% in $t = \tau$ years.

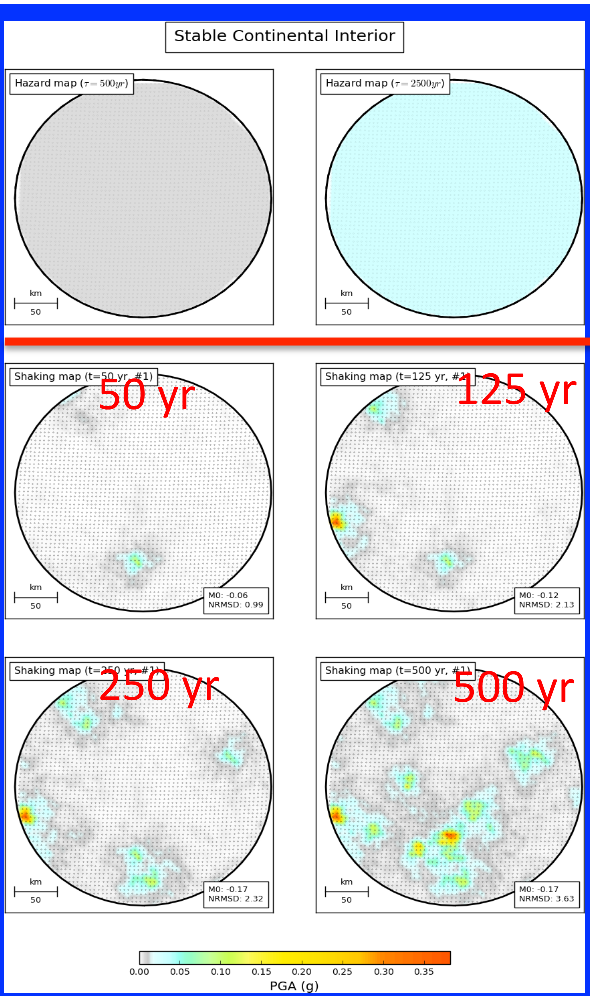
## Verification: how should ideal PSH maps work?

We simulate the shaking history of an area with assumed distribution of earthquakes, frequency-magnitude relation, temporal occurrence model, and ground-motion prediction equation.

We compare the “observed” shaking at many sites over time to that predicted by a hazard map generated for the same set of parameters.

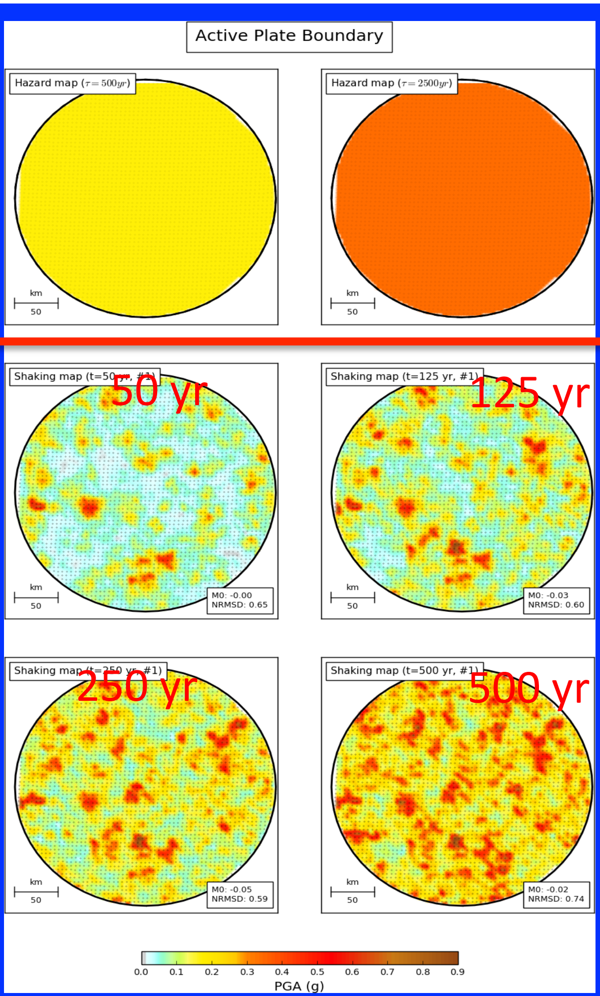
Ideal case gives insight into real map performance.

## Hazard maps for 500 & 2500 year return periods



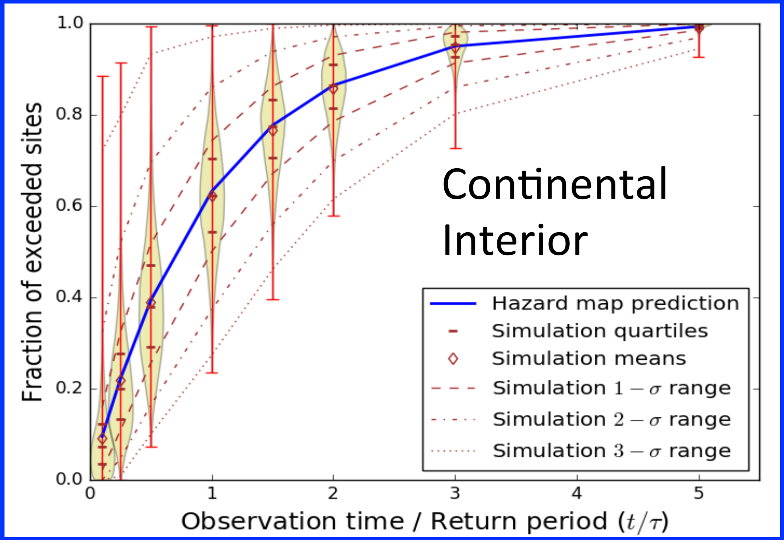
Simulated maximum shaking

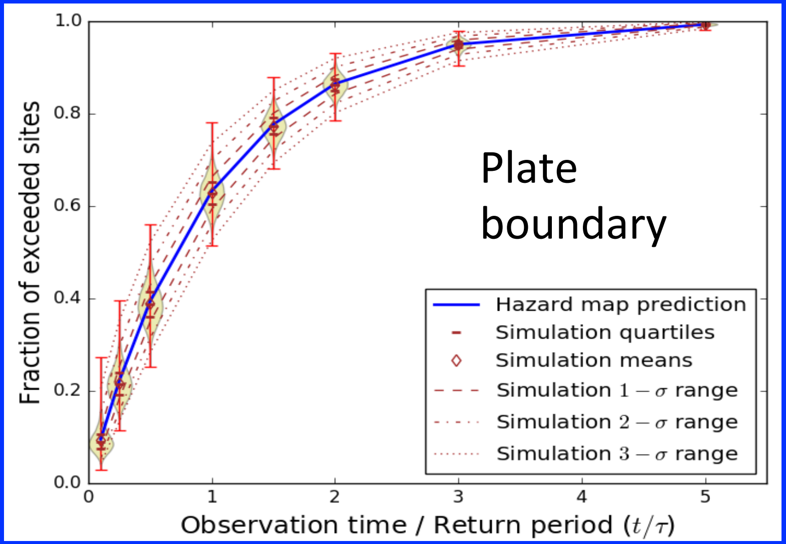
Some sites above map, some below. Shaking in big quakes greater than in map, as expected



Vanneste et al., 2017

# Results (10,000 runs)

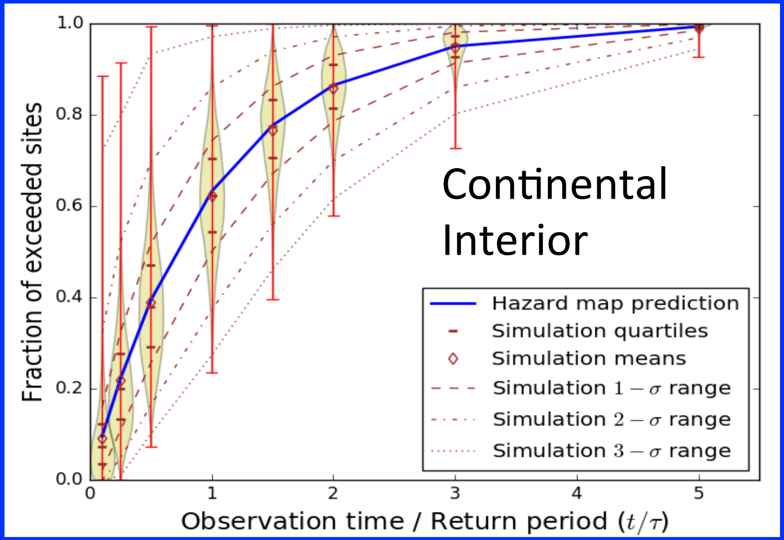


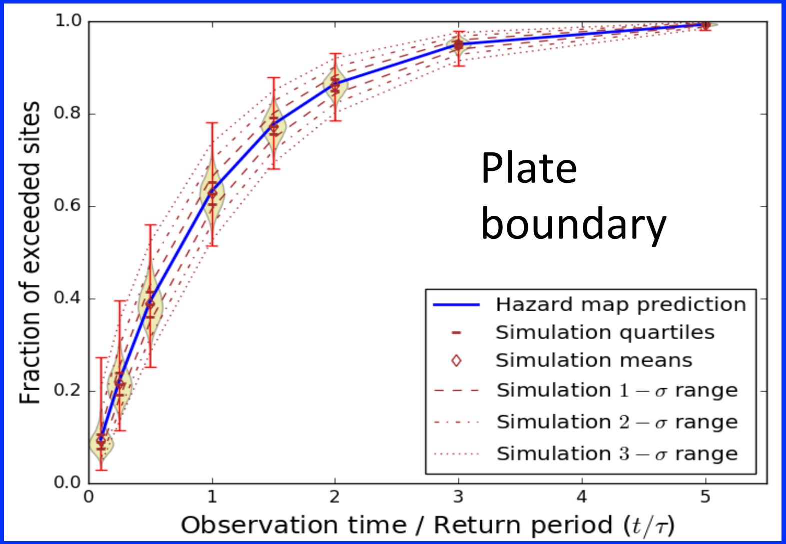


- Mean of ensemble consistent with map prediction
- Scatter decreases for longer simulations (increasing t/τ), because as observation time increases, the largest earthquakes and resulting shaking are increasingly likely to have occurred.
- For the same reason, scatter is much less for the more active plate boundary

Vanneste et al., 2017

# Results (10,000 runs)

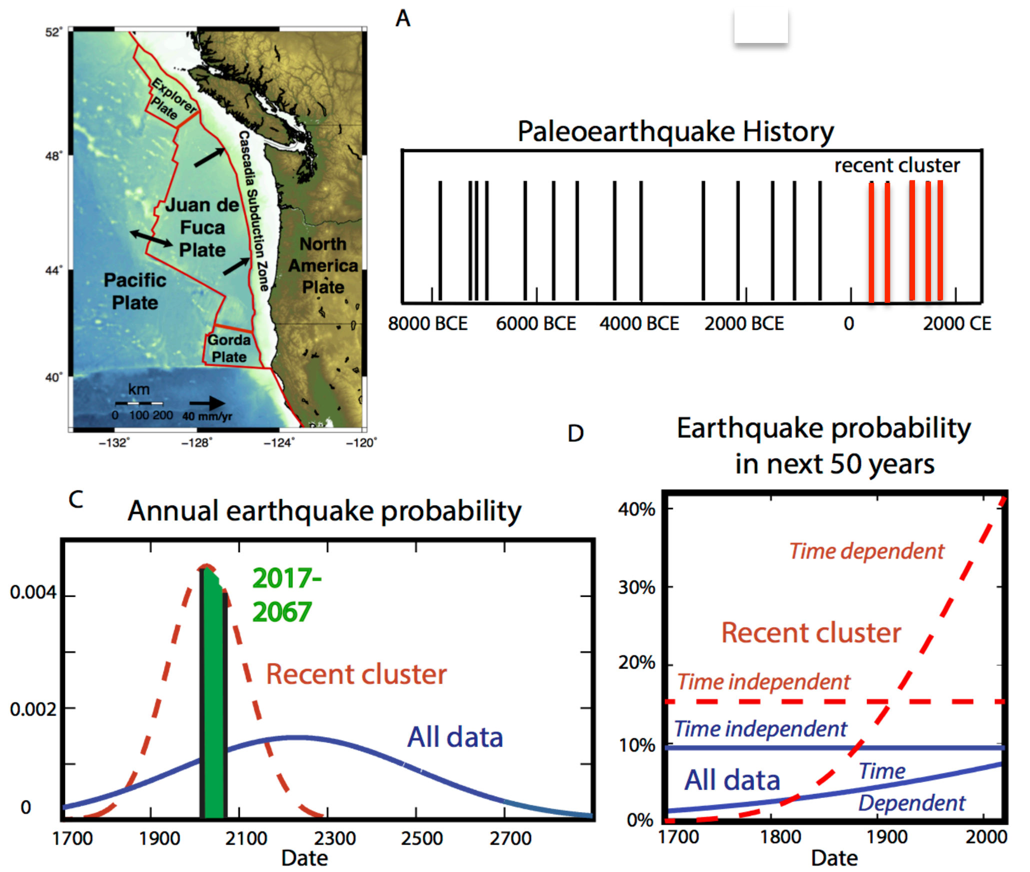




- An earthquake history can yield a fractional exceedance significantly higher or lower than that predicted while still being consistent with the hazard map.
- A real map involves assumptions about more complicated source geometries and recurrence rates, which are unlikely to be exactly correct, so scatter likely to be greater & bias is possible.

Vanneste et al., 2017

# Choice of parameters: Cascadia paleoearthquake record



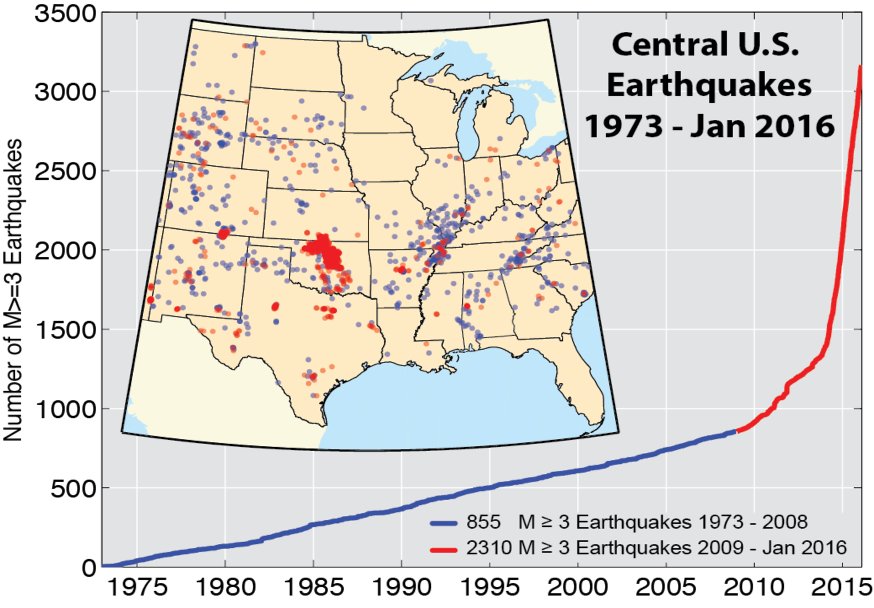
Recent cluster (1500 yr) gives mean recurrence μ = 326 yr and σ = 88 yr

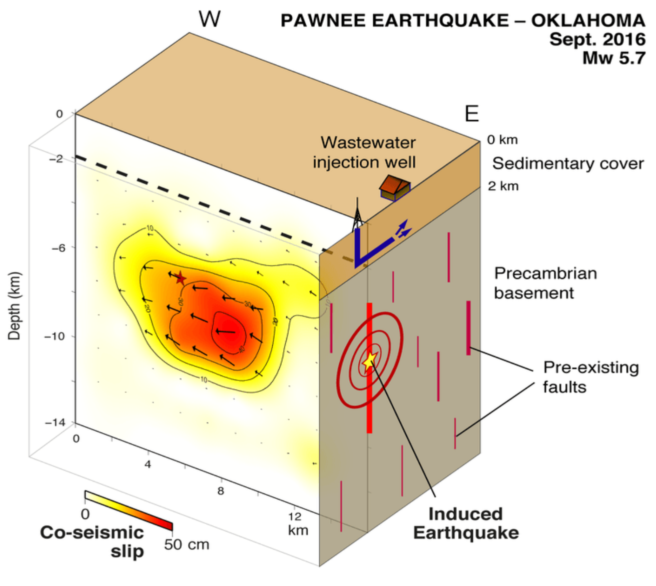
Past 10,000 yr gives larger mean recurrence μ = 530 yr and σ = 271 yr

Stein et al., 2017

Recent cluster predicts 6x higher probability of quake in next 50 yr

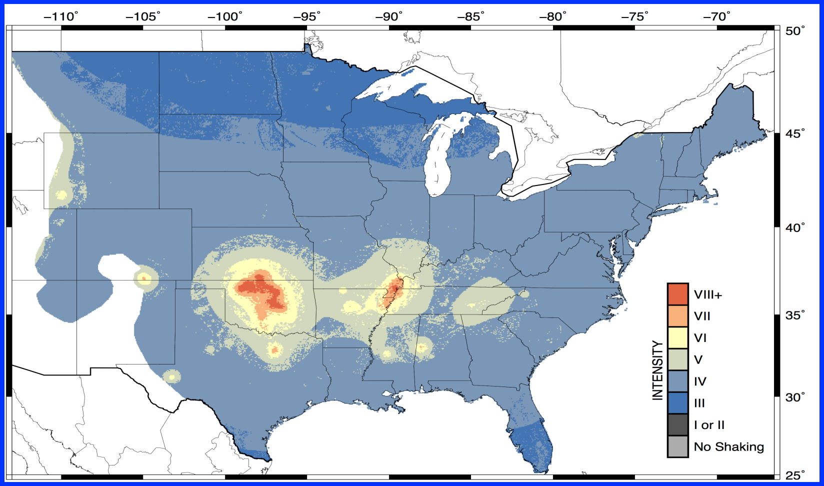
# Validation study #1: earthquakes in Central U.S. increasing due to wastewater injection from oil & gas production





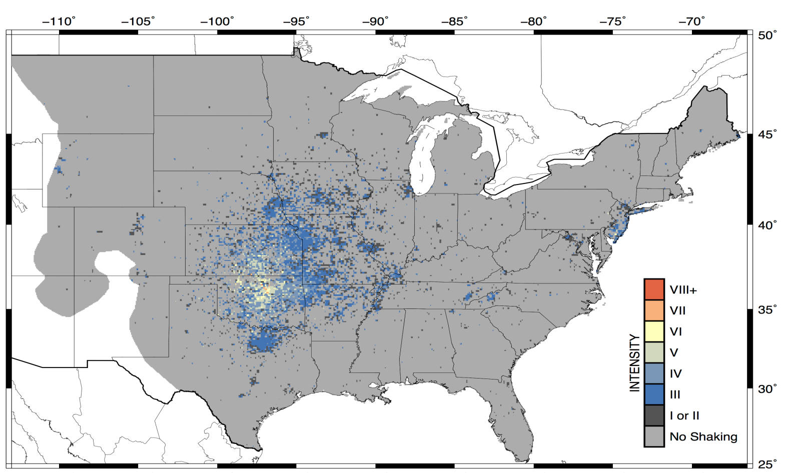
USGS has started making one-year hazard maps to better understand the hazard resulting from human activities. These are ideal for validation study due to short time scales and good shaking data.

## 2016 One-Year Seismic Hazard Map



Describes both induced and natural shaking expected in 2016.
1% probability of exceedance for a one-year window.

## Observed Shaking in 2016: Did You Feel It?



Record of maximum felt shaking reported to DYFI.
Grey regions had no reports, but may have experienced shaking.

Brooks et al., 2017

# Comparing Observed and Predicted Shaking

**Fractional Exceedance metric M0**

$M0 = |f-p|$

p = Predicted fraction of site exceedances relative to map

f = Fraction that actually exceeded

Implicit in definition of PSHA map

Binary (above/below), neglects magnitude of misfit

**Squared Misfit metric M1**

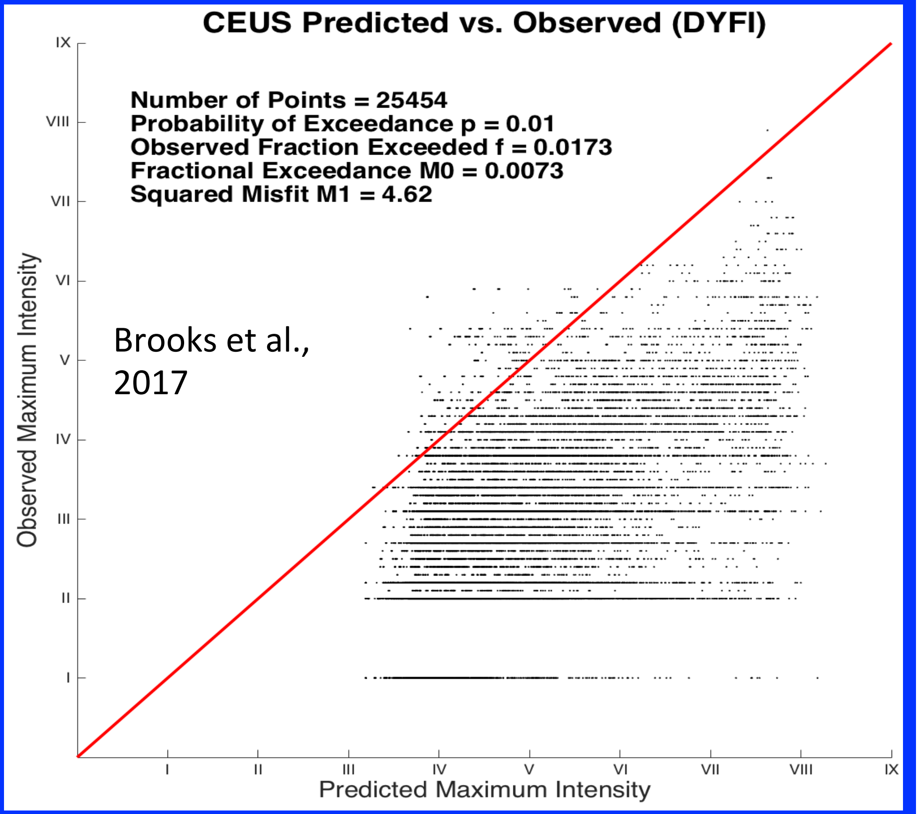
$M1 = \sum(x_i-s_i)^2/N$

$x_i$ = Maximum observed shaking at site i

$s_i$ = Predicted shaking at site i

Like visual comparison; how similar are observed and predicted spatial distributions?

Not what PSHA seeks



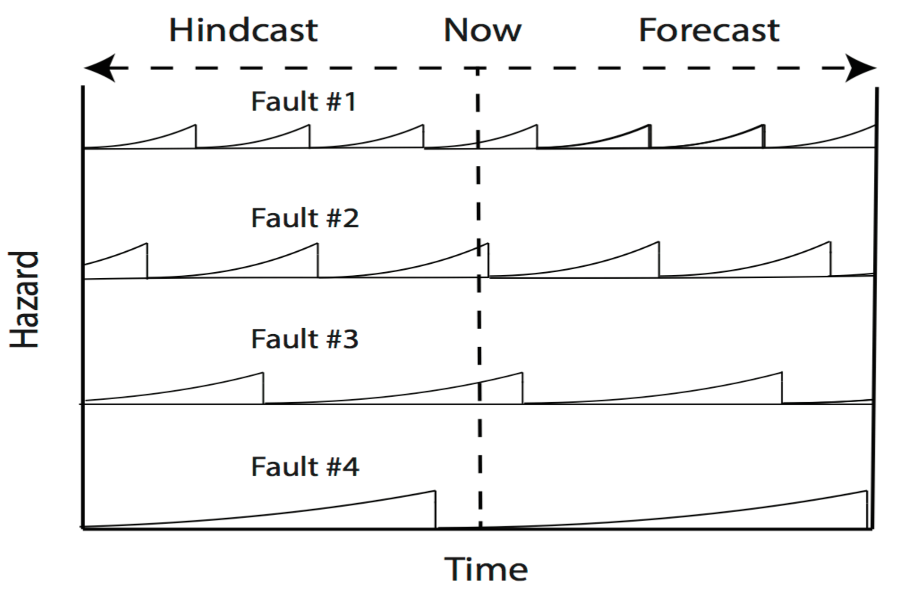
Brooks et al., 2017

Map did very well by both metrics and thus can be valuable for policy making

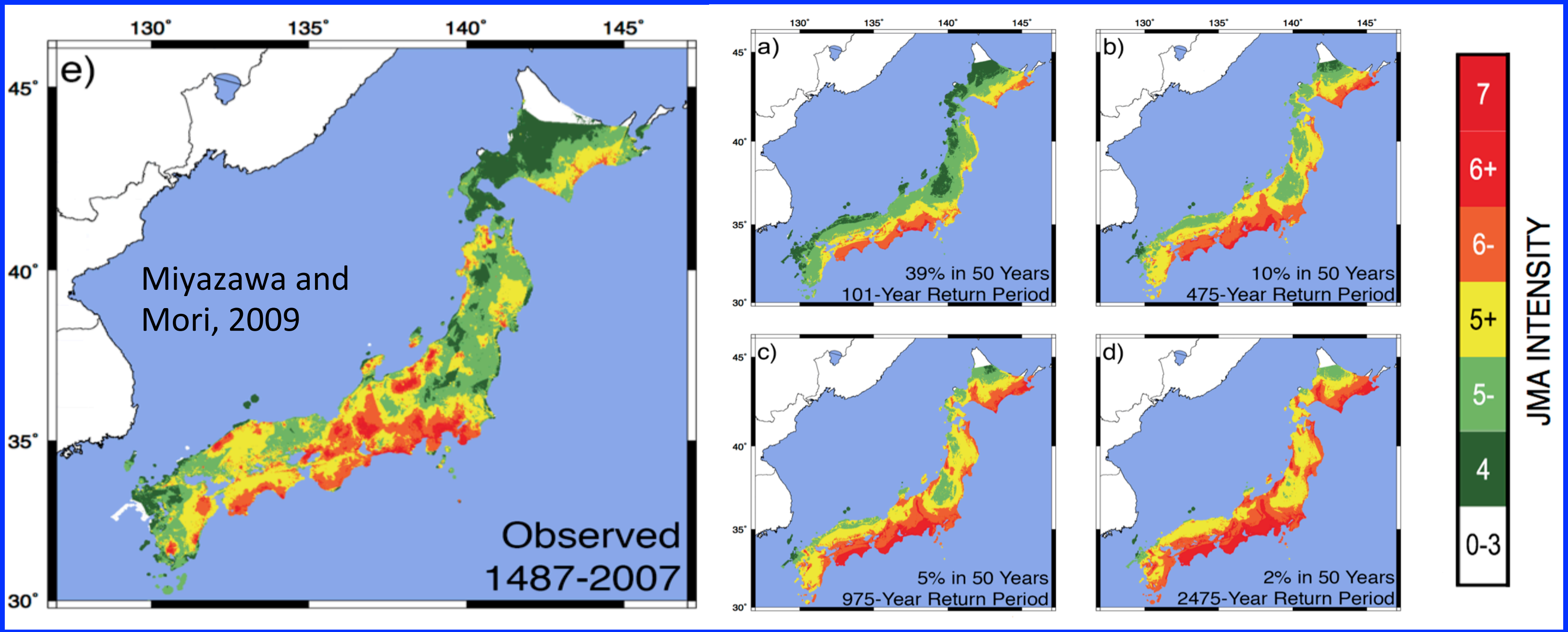
In areas of natural seismicity, the short time since hazard maps began to be made is a challenge for assessing how well they work.

Hindcasts offer long records, but are not true tests, as they compare maps to data that were available when the map was made.

Still, because shaking data were not used in map, we get useful insight to help identify strengths & weaknesses and hence improve maps.



Brooks et al., 2016

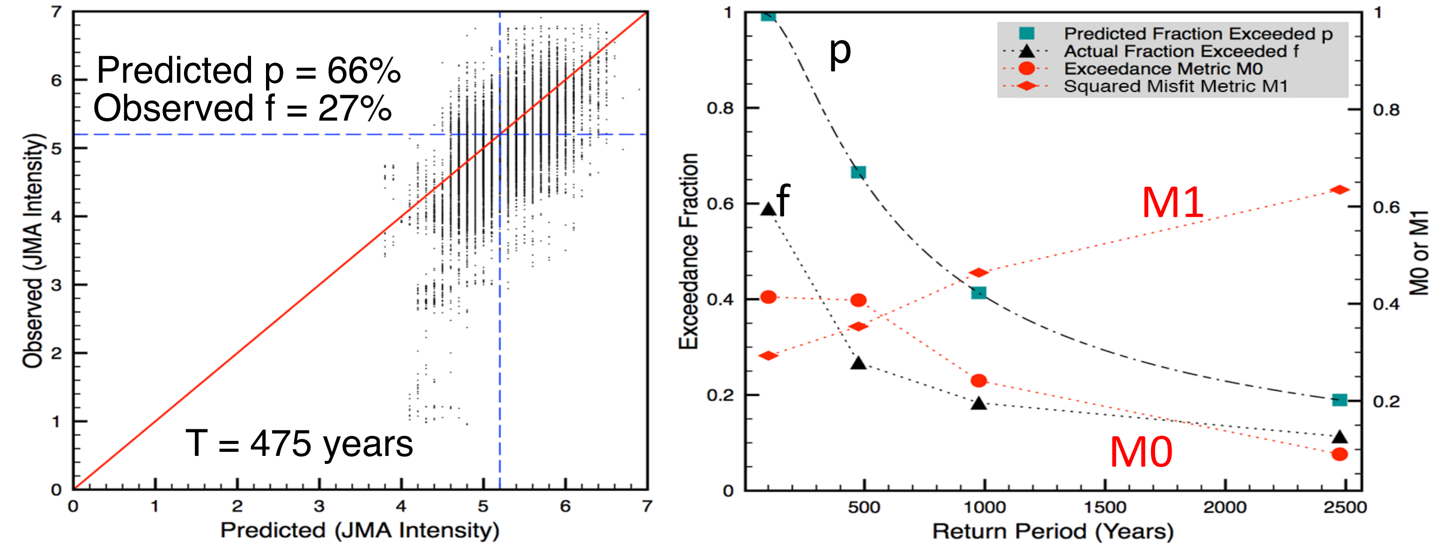


Validation study #2: compare 510-year shaking record to 2008 Japanese National Hazard (JNH) maps using both exceedance (M0) and squared misfit (M1) metrics

Brooks et al., 2016

# Compare Predicted and Observed Shaking

- Maps generally overpredict, so f<p
- Predicted and observed exceedance fractions & difference decrease with return period , so exceedance metric (M0) does
- Squared misfit metric (M1) is higher for long return periods, since not enough time has passed to observe many large events expected in 975- and 2475-year maps

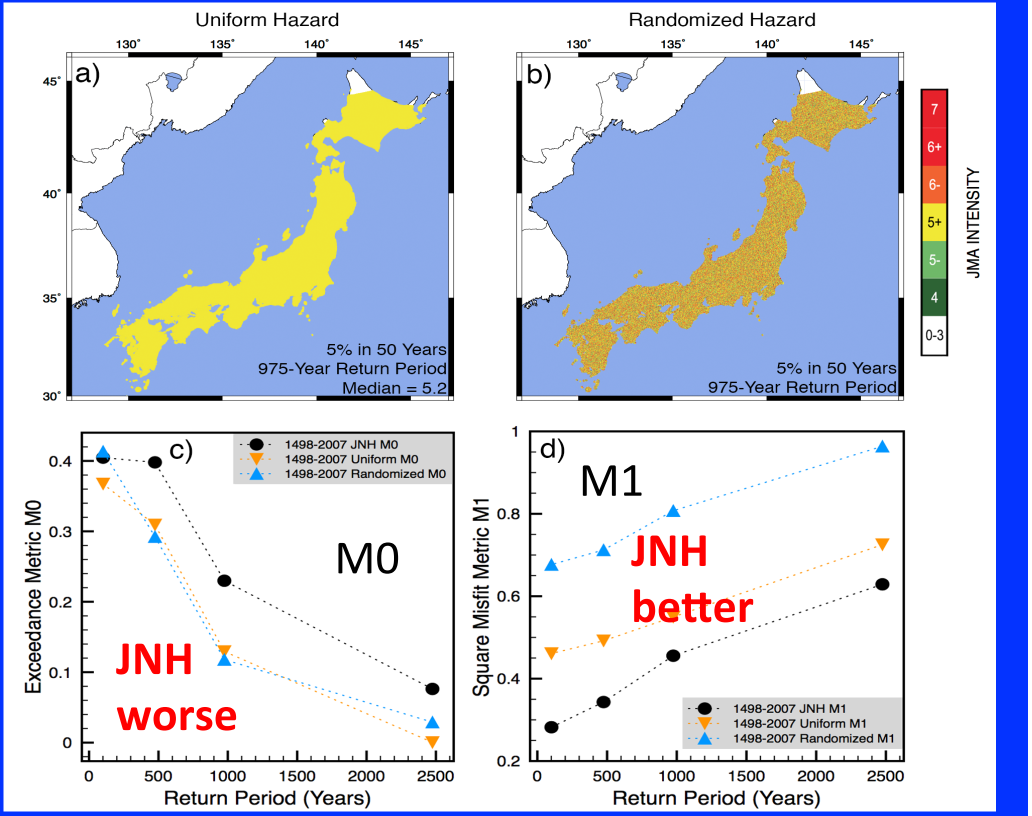


Brooks et al., 2016

# Could uniform & random maps be better?

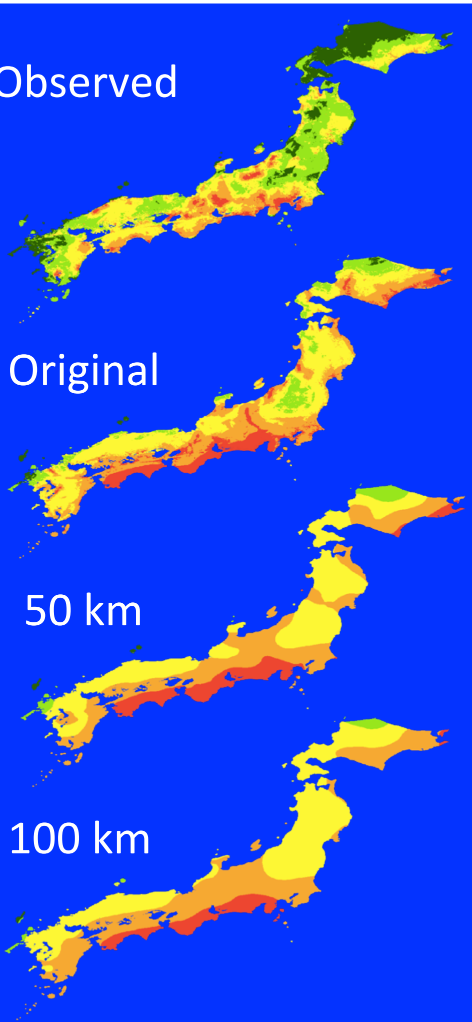
Geller (2011) argued that "all of Japan is at risk from earthquakes, and the present state of seismological science does not allow us to reliably differentiate the risk level in particular geographic areas," so a map showing uniform hazard would be preferable.

Test: By exceedance metric M0 uniform and randomized maps do better, but by misfit metric M1 detailed JNH maps perform better.



Brooks et al., 2016

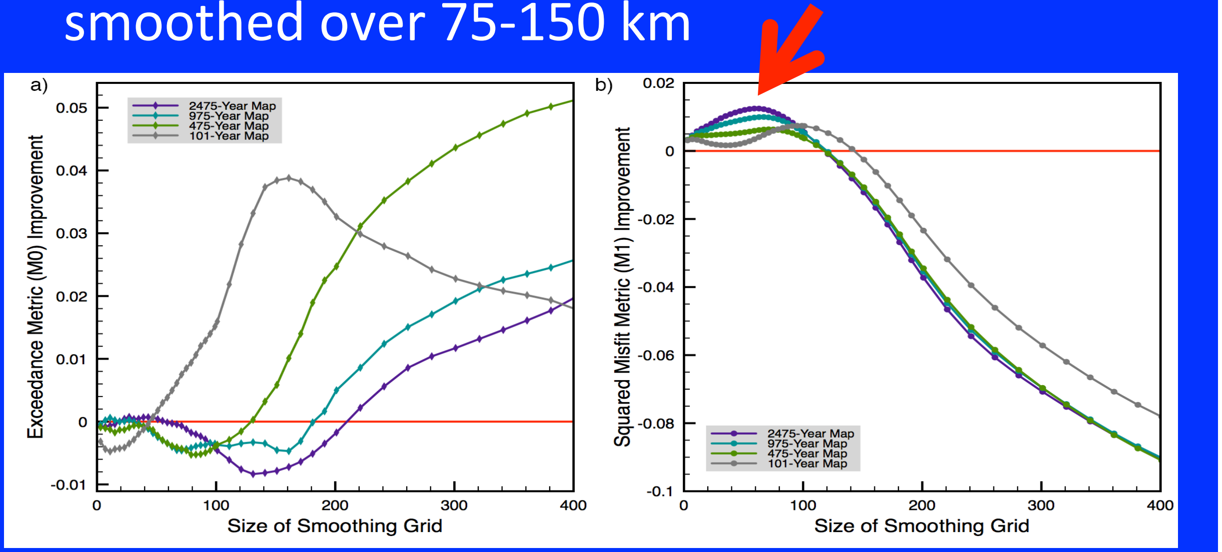
# Smoothed Maps



Brooks et al., 2017

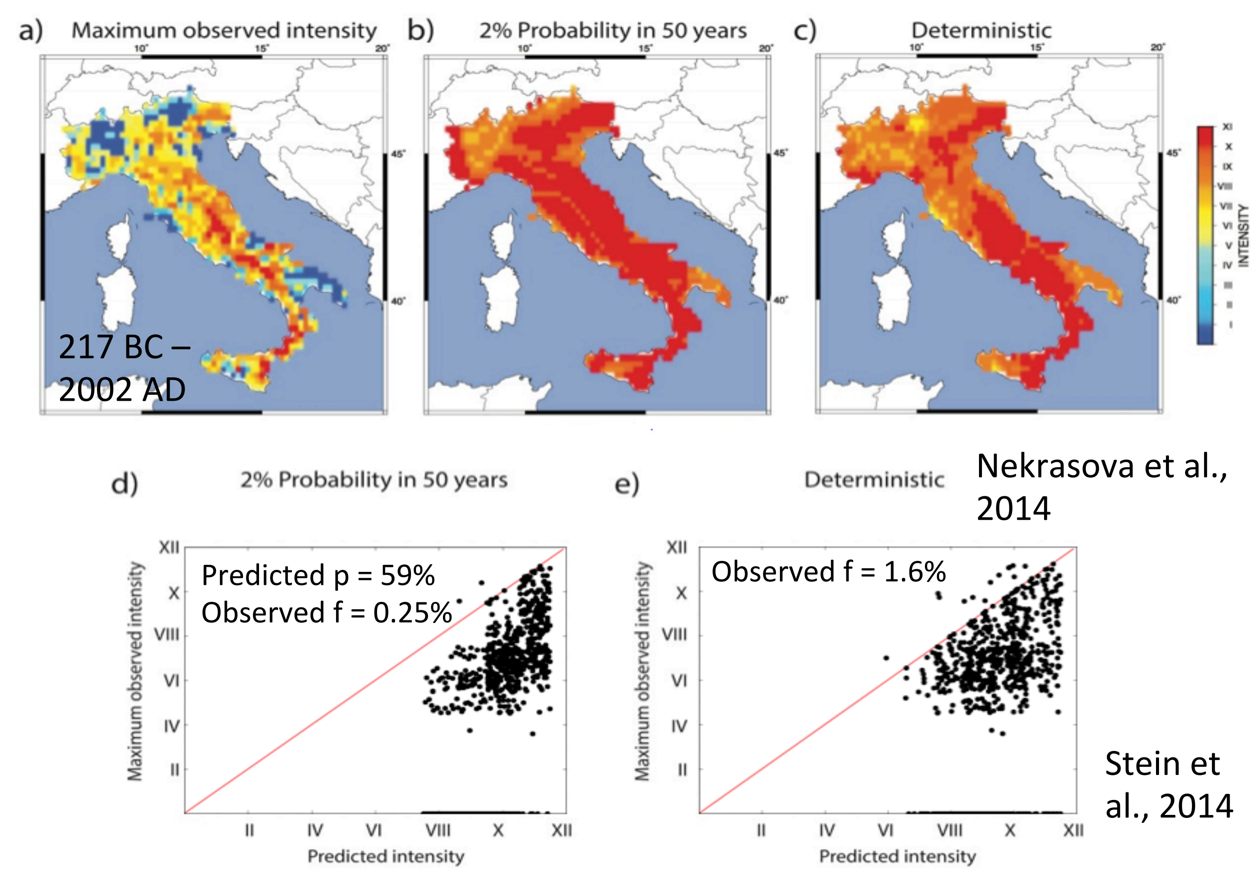
How detailed should a map be?

- By exceedance metric (M0) map improves by smoothing over larger areas
- By squared misfit (M1), map performs best if smoothed over 75-150 km



Intermediate level of detail may be best

# Validation study #3:



217 BC – 2002 AD

Nekrasova et al., 2014

Predicted p = 59%
Observed f = 0.25%

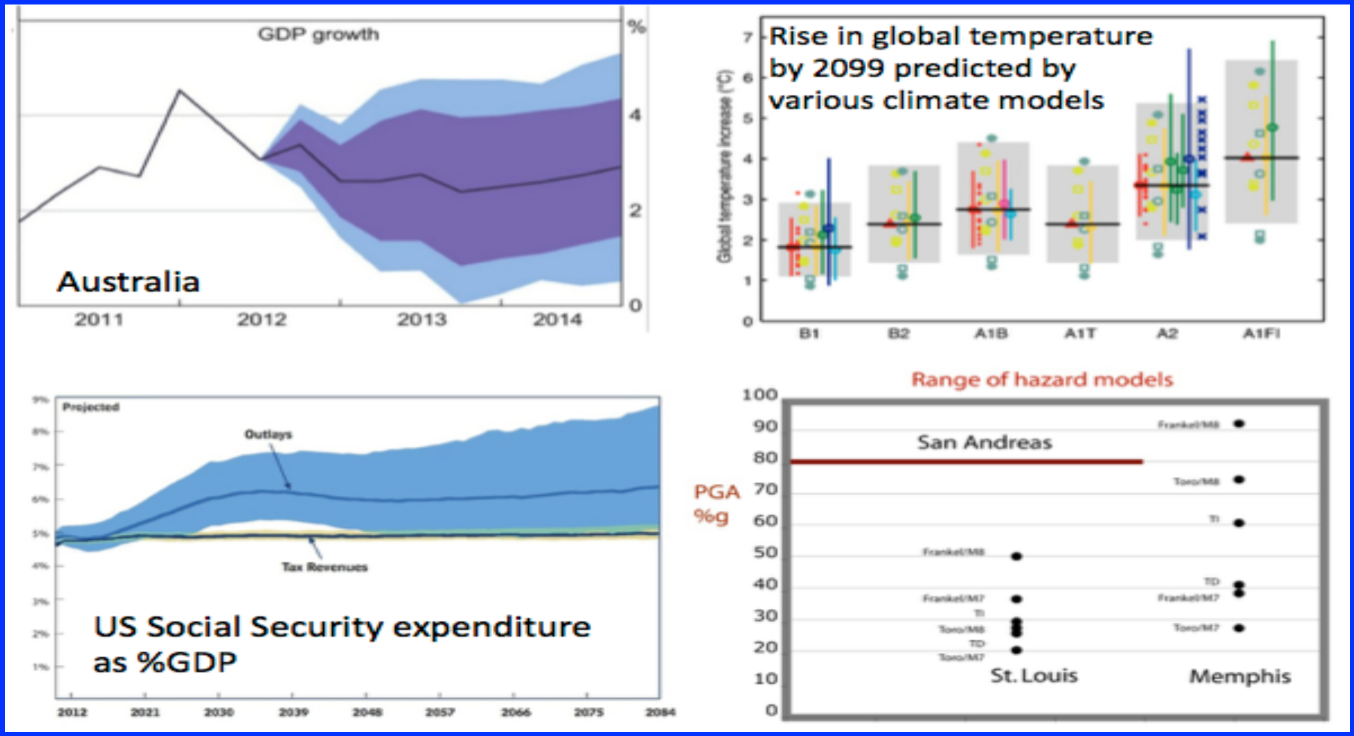
Observed f = 1.6%

Stein et al., 2014

Both probabilistic & deterministic maps for Italy significantly overestimate shaking reported over ~2200 years, a time span comparable to map return period

Misfit indicates problems in the data, maps, or both.

# Probabilistic forecasts are increasingly used in applications including meteorology, finance, and demography because showing the probability of different outcomes is useful in policy making and gives an estimate of forecast uncertainty.



Stein
et al., 2015

# Hard to tell if probabilistic forecasts are bad.

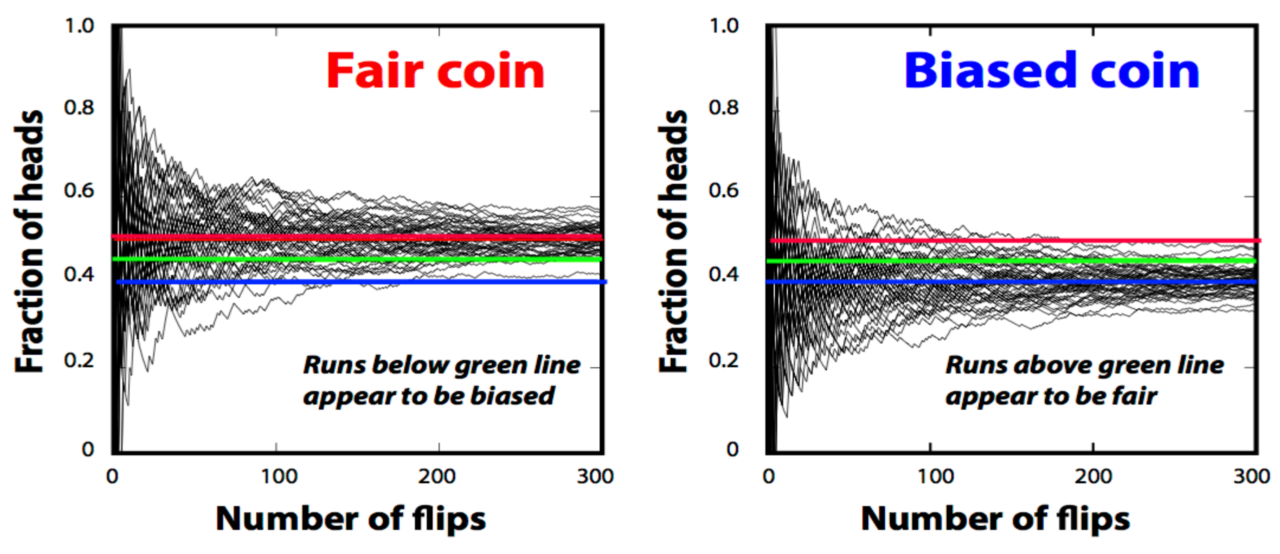
Because they allow low-probability extreme events,
such an event does not demonstrate a weakness in the model.

When spring 2012 was the wettest on record in Britain despite being forecast as dry, the Meteorological Office admitted that its forecast was "not helpful" but likened it to the guide to a horse race - "any of the outcomes could occur, but some are more likely than others."

Met Office three-month forecast was 'not helpful'

Assessing whether a probabilistic forecast did poorly/well because of bias or bad/good luck is like trying to tell if a coin is fair - equally likely to come up heads or tails if flipped - or biased.

The mean of a suite of runs converges on the expected value with smaller standard deviation as the run gets longer.



However, for a single run, even if very long, it can be hard to distinguish fair from biased.

# Summary

PSHA appears internally consistent: the most likely exceedance distribution is that predicted by a map with correct parameters.

Shaking in big earthquakes often greater than in map, as expected

A single earthquake history can yield fractional exceedance significantly higher or lower than predicted

Because reality gives only one earthquake history, it is hard to assess whether a bad (or good) fit between a map and actual shaking arises by chance or reflects a map biased by poor parameter choices.

Assessing and exploring map performance should improve maps for both scientific and policy purposes