Beyond Validation

Uncovering the True Value of Catastrophe Models

Dr. Tom Philp

CEO, Maximum Information

Oasis Insight Bermuda

June 2024

MAXIMUM INFORMATION

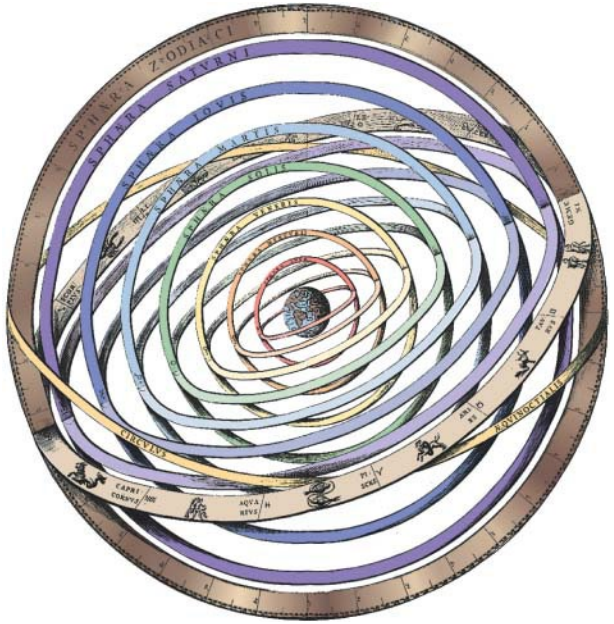
Contents

Geocentrism vs Heliocentrism

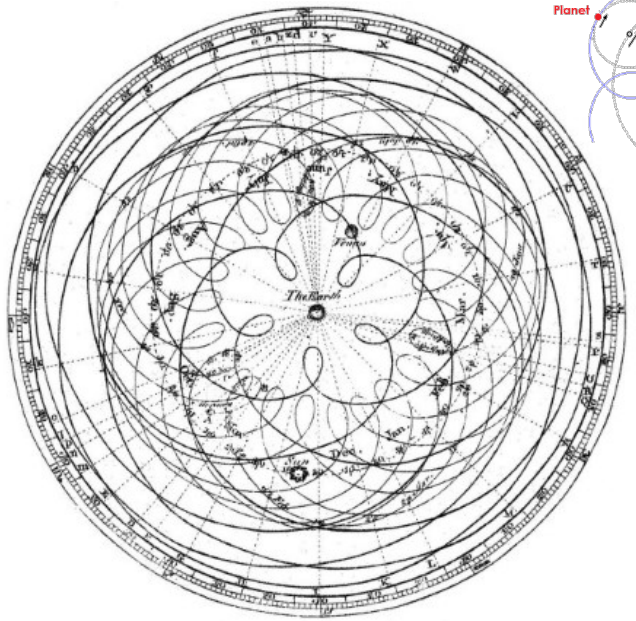
Or: How I learned to stop Validating and Love Evaluation

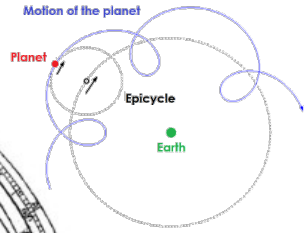
The Geocentric (Ptolemaic) Universe Models

Eudoxus (380 B.C.) → Aristotle (350 B.C.) → Ptolemy (150 A.D.)



- Comets must exist within the sphere of earth because they don't move in perfect circles
- Everything beyond the earth & comets are the Heavens and are thus perfect and thus move in perfect circles



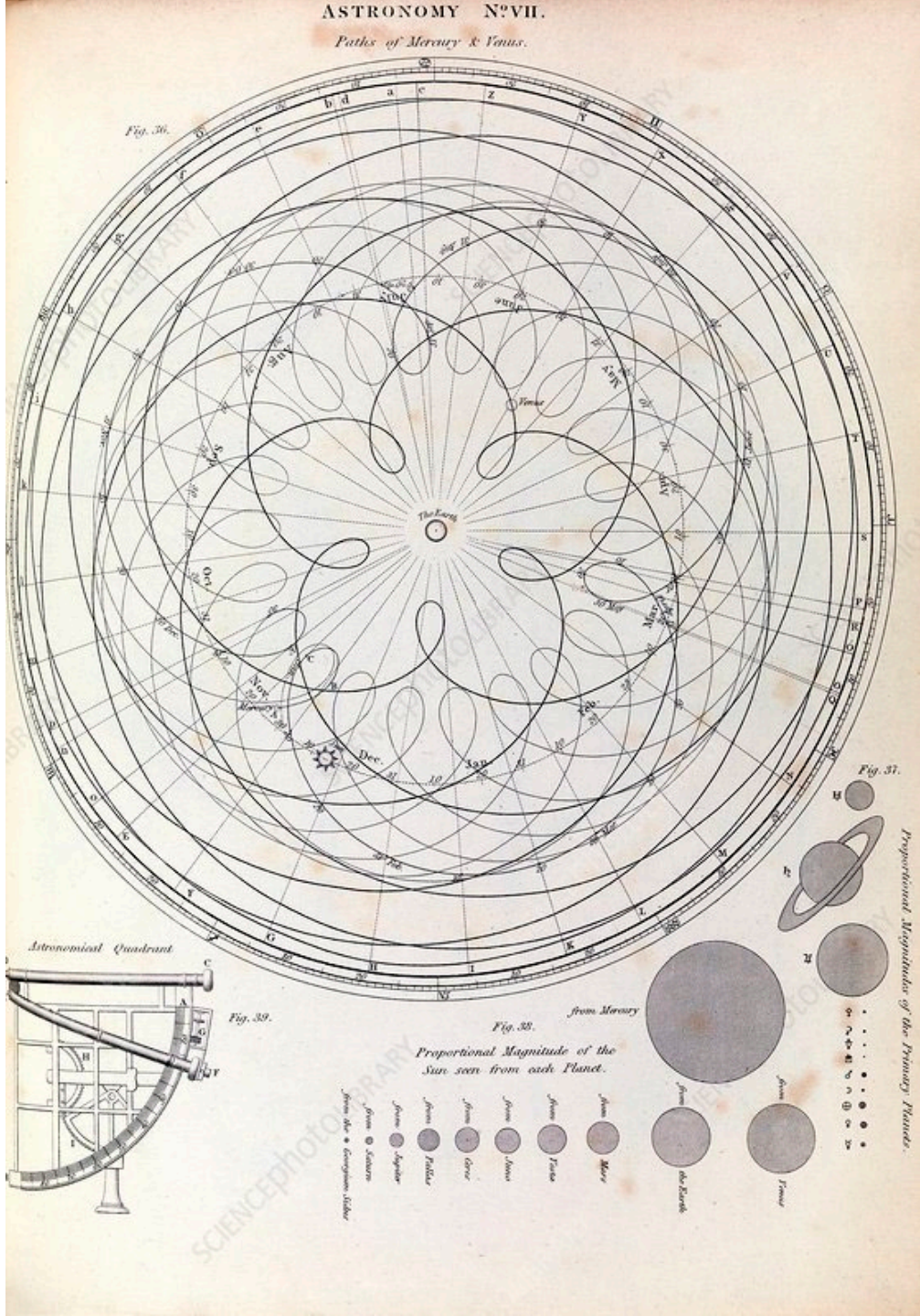


Validating the Geocentric

- Geocentric models were regularly challenged & updated because new, unpredicted observations needed new explanations.

- By the time of Ptolemy in 150 AD, there were 93 tweaks (AKA MODEL ADJUSTMENT FACTORS) to the Celestial Sphere models, for predicting locations of the Sun, Moon, and 5 known planets.

- By creatively introducing complex retrograde motions & varying planetary speeds, Ptolemy's model was largely considered accurate and predictive until...



The Heliocentric (Copernicus') Model

- Copernicus placed sun at centre (1543).
- More predictive accuracy immediately.
- Sensible assumptions – e.g. uniform speeds of planets
- Didn't throw the baby out with the bathwater – retained some Ptolemaic thinking .

Evaluation (not Validation) of Quantitative Models

← TIME100 CLIMATE

Naomi Oreskes

Professor, Harvard University

Kayana Szymczak

BY NAOMI ORESKES
NOVEMBER 16, 2023 7:35 AM EST

"The Ptolemaic system of astronomy is a historical example of a model that was empirically adequate but based on a wrong conceptualization."

Evaluation (Not Validation) of Quantitative Models

Naomi Oreskes*

Gallatin School of Individualized Study, New York University, New York, New York

"Calling a model validated does not make it valid."

Validation vs Evaluation

Validation

- How well does this model line up with experience?
- Foundation of Known Truth: I (can) have a good handle on what truth is.

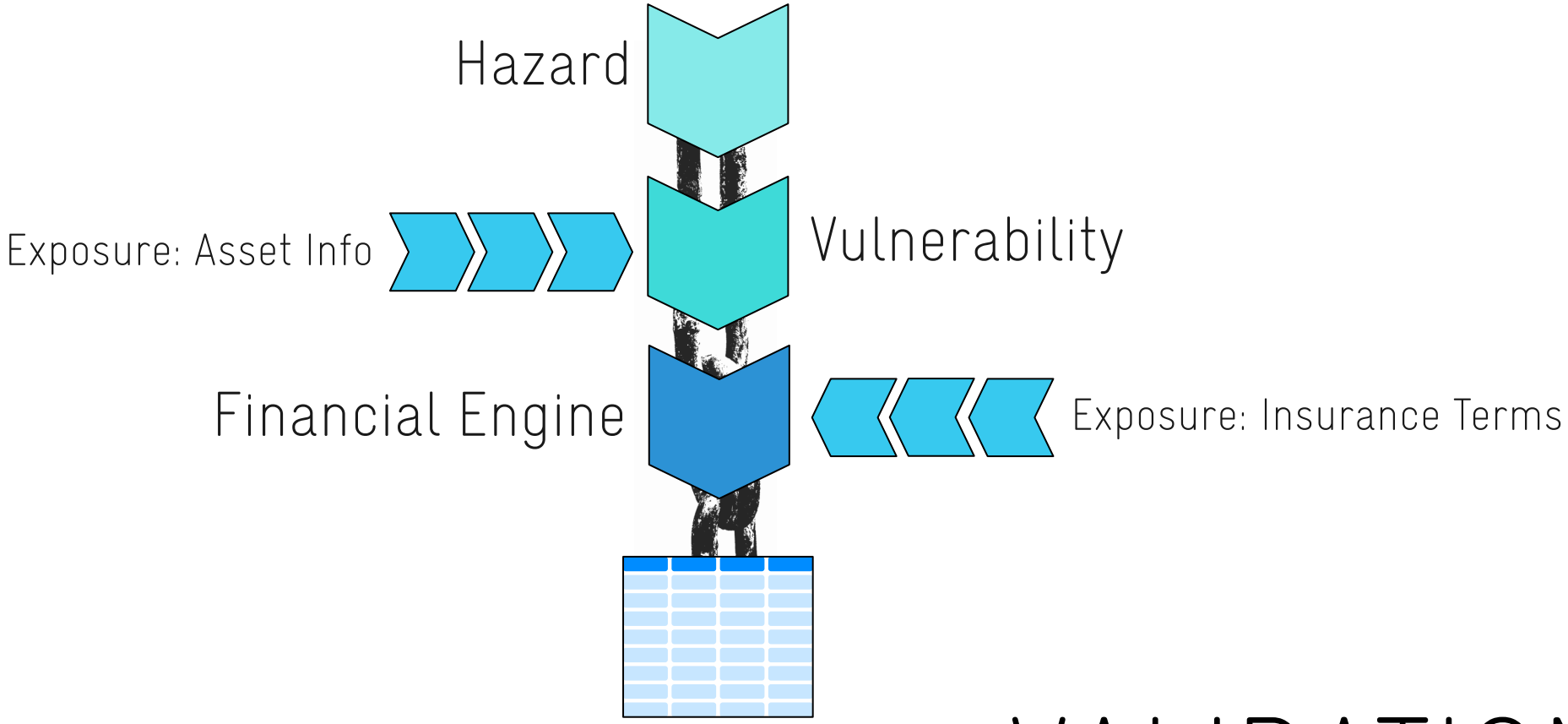
Evaluation

- How well should this model line up with experience?
- Foundation of Value: Why am I using is model in the first place and who/what decision is it useful for?

Cat Modelling Chains

Are we currently chained to a loss-table focused worldview?

The Cat Modelling Chain



→ VALIDATION

What does the Cat Modelling Validation Chain Enforce?

- A Quasi-Ptolemaic Worldview:
  1. An inability to see the independent value of the many components of the chain.
  2. A faith-based loss-centric ideology that validating model losses is the key to reliable model implementation

Breaking the Chains

An Upcoming Paradigm Shift in Cat Model use, driven by:

1) Operational Multi-Model Platforms

2) Evaluation, not Validation

Breaking the Chains:
For Underwriters & Portfolio Managers

- Evaluation Question: Does the event set produce reasonable spatial relationships across the model region (not necessarily tied to history)?

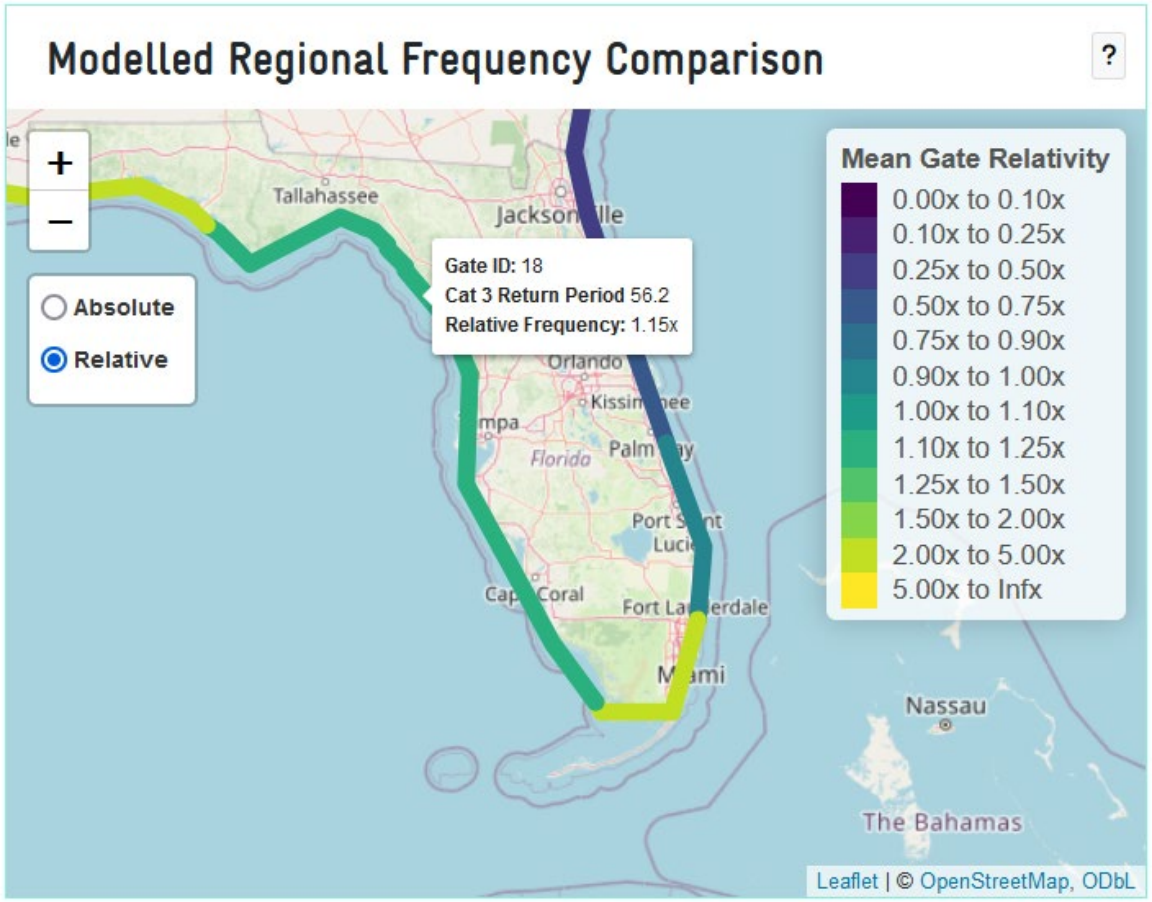
- Real-world UW/Portfolio Question: What does the model tell me about risks/opportunities based on mismatch between raw history & stochastic set?

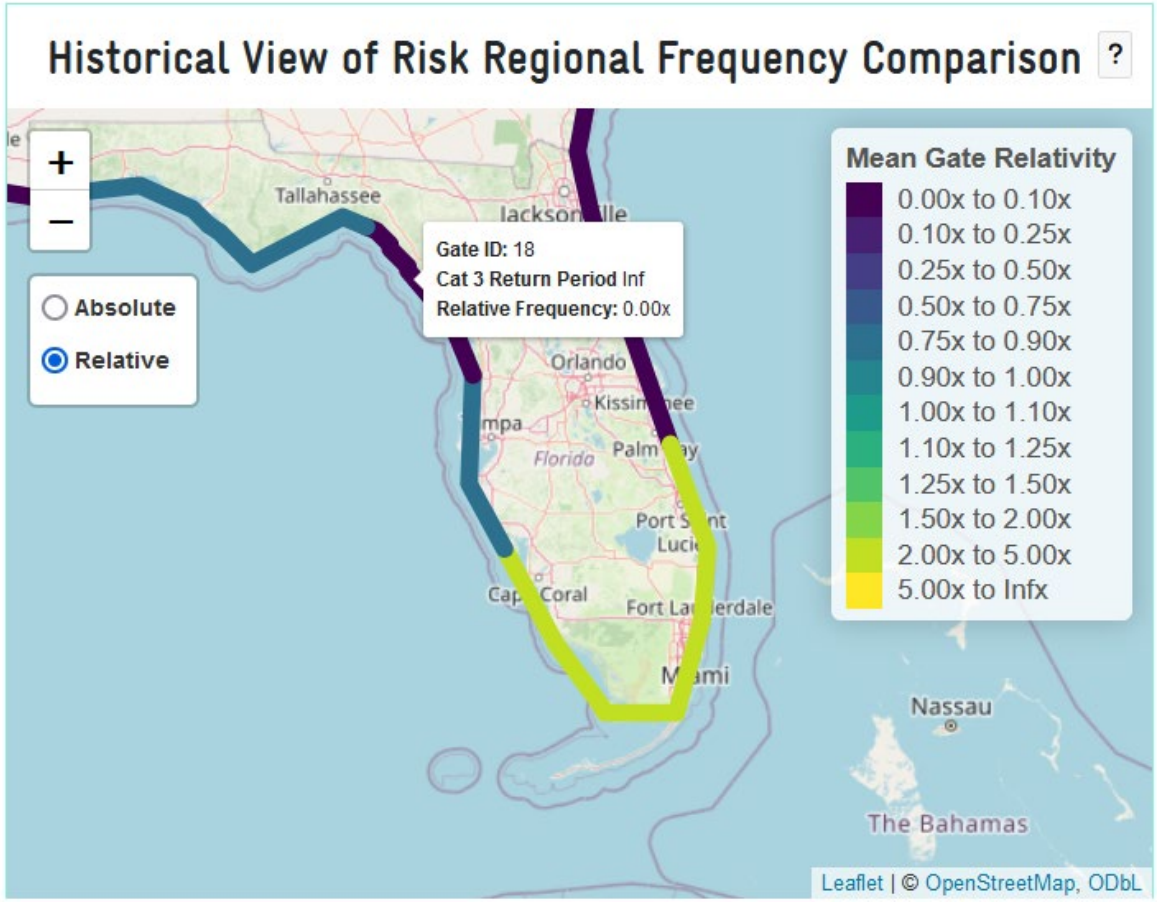
Stochastic Set Relativities

- KEY POINT (1): Stochastic hazard sets contain huge amount of very important info that is often obscured by a focus on loss output.

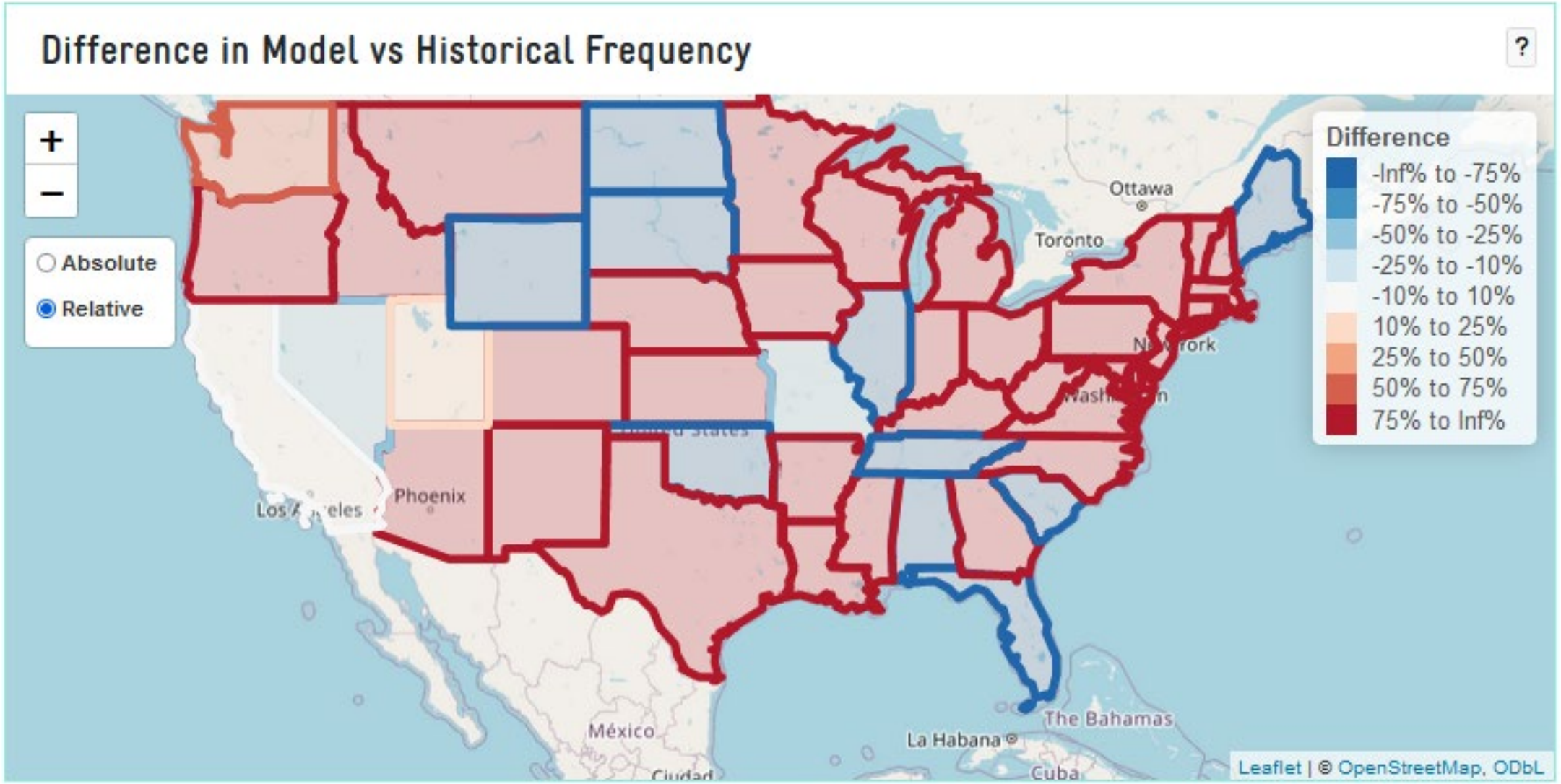
- KEY POINT (2): Pure "Validation" against historical observations is nonsensical; the models were built to move away from history because the observational record is sparse

UWs: Cat 3+ FL Hurricane (200km gates)





Portfolio Managers: Mag 7.5+ US Earthquake



Breaking the Chains:
For Exposure Managers

Evaluation Question: What are the full ranges of potential losses from exposures in my portfolio?

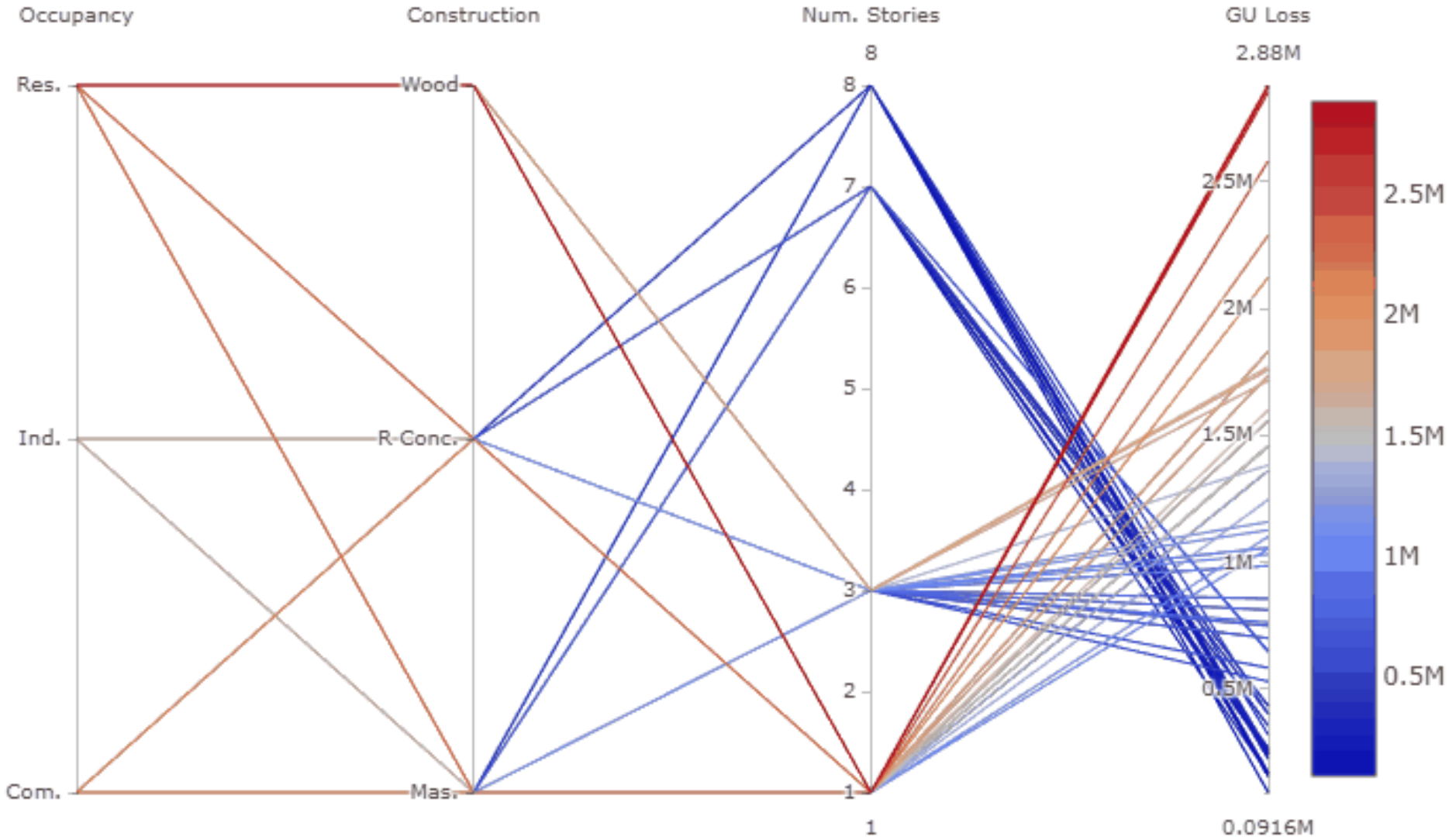
Real-world Exposure Management Question: What combinations of modifiers lead to outsized losses that could foreseeably cause us to be seen as a sort thumb?

Sore Thumb Vulnerability Detection

- KEY POINT 1: Vulnerability data is sparse at best; can we ever hope to validate at the peril-region scales necessary?

- KEY POINT 2: What key aspects of model output can keep us up at night?

What combinations lead to sore thumbs?



Breaking the Chains:
For Claims Managers

Evaluation Question: Do hazard event footprints capture important damage & loss driving features?

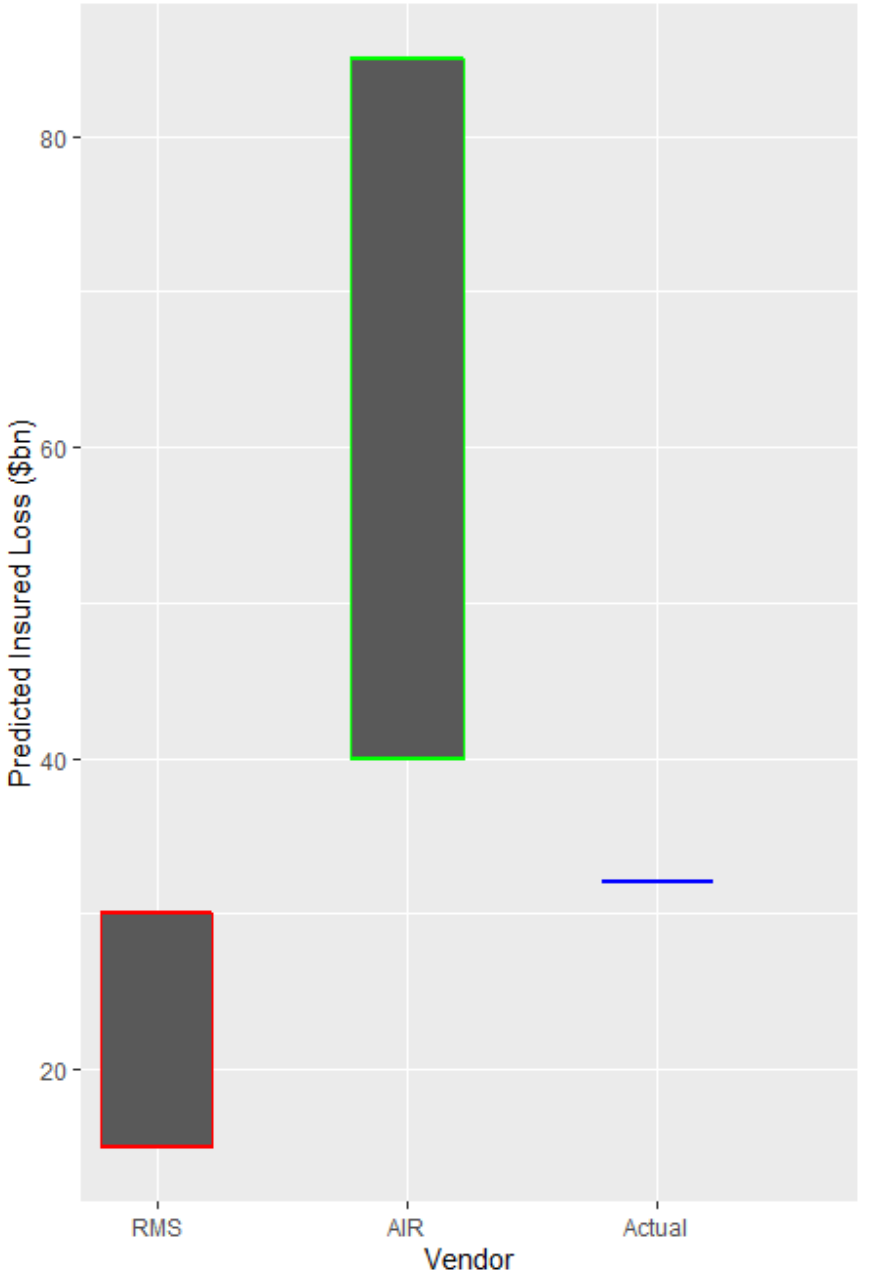
Real-World Claims Manager Question: Can I expect that model events are going to give me a fair representation of on the ground claims realities as and when a single event occurs? If so, what types of claims are we likely to see coming in?

Live Event Response: Capturing Loss

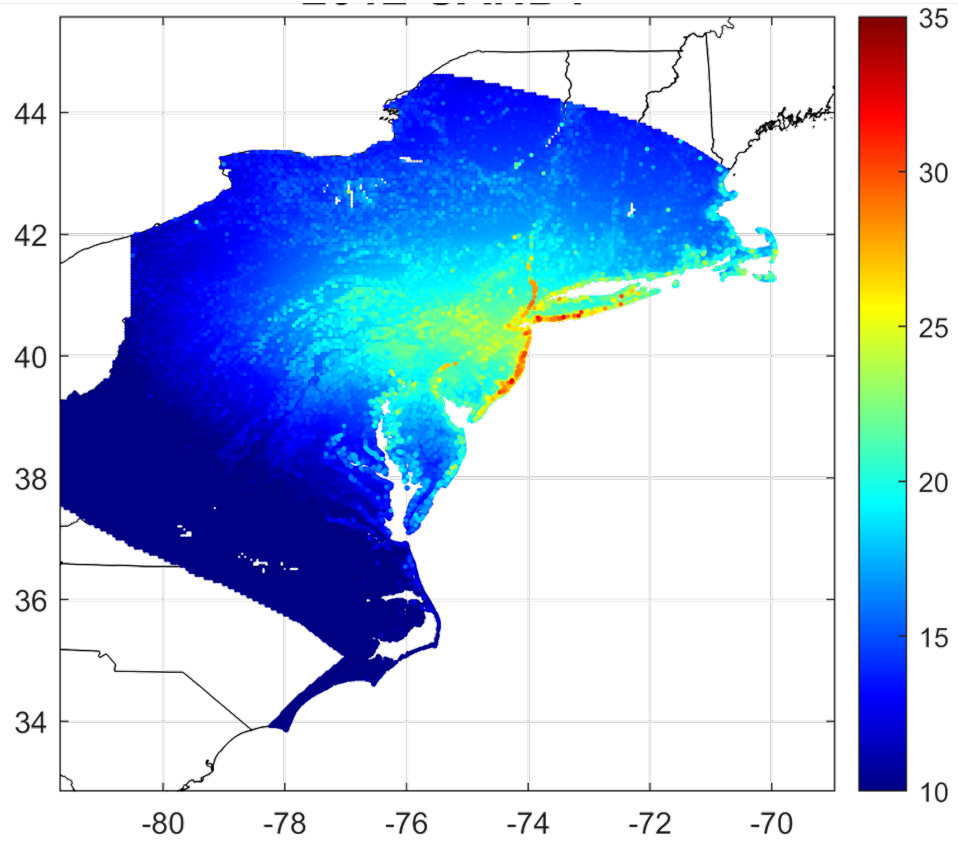
- KEY POINT 1: Hazard modelling is (largely) prohibitively computationally expensive to do very accurately at site-specific spatial scales. Some level of aggregation is usually needed to stabilize damage/loss estimates.

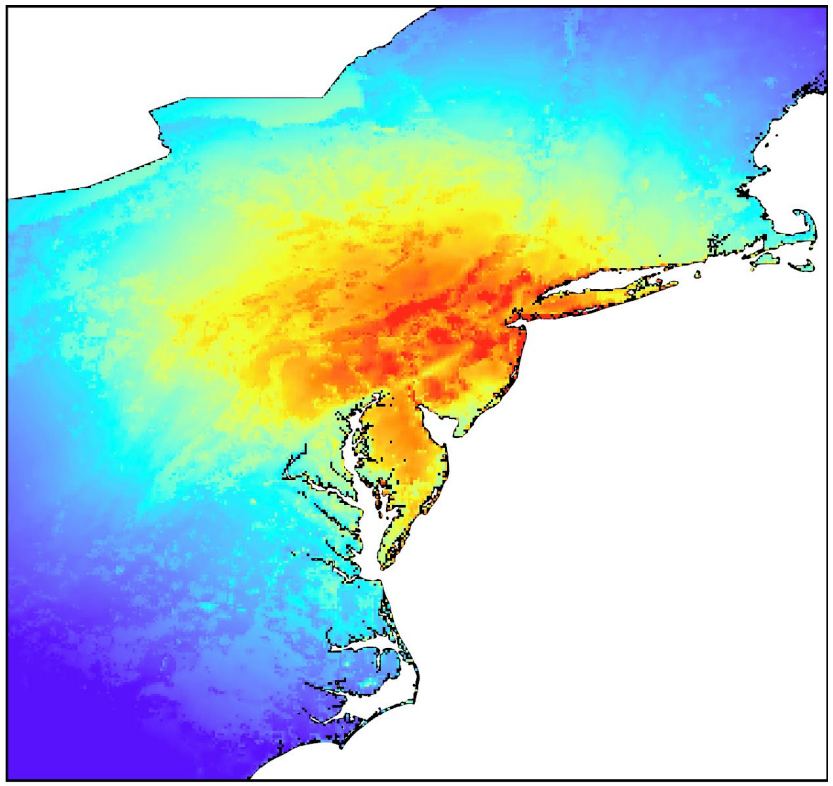
- KEY POINT 2: Even when aggregated, key assumptions may be missing. Are we likely to have missed any? Do I need multiple models?

Hurricane Maria Loss, 2017



Footprint Analysis: Superstorm Sandy 2012

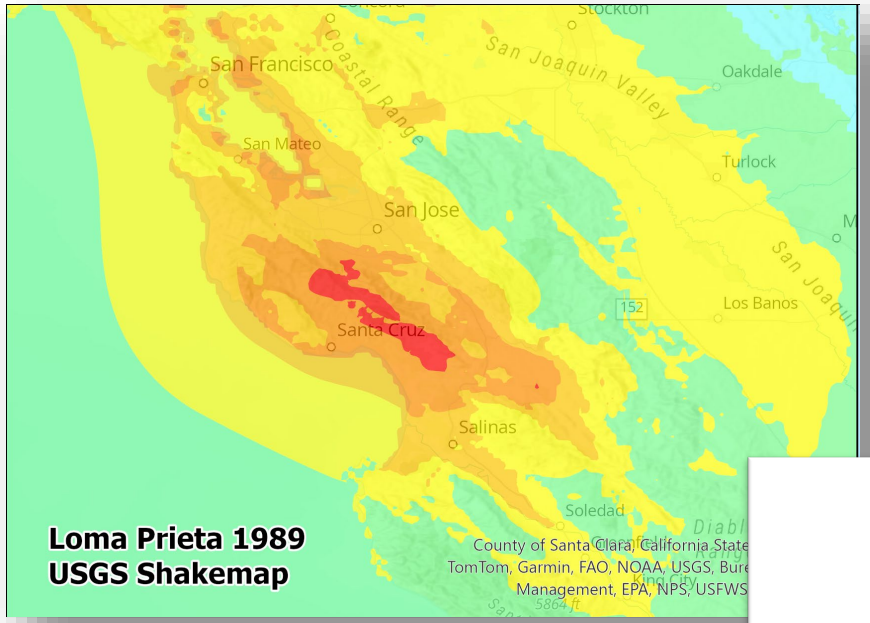


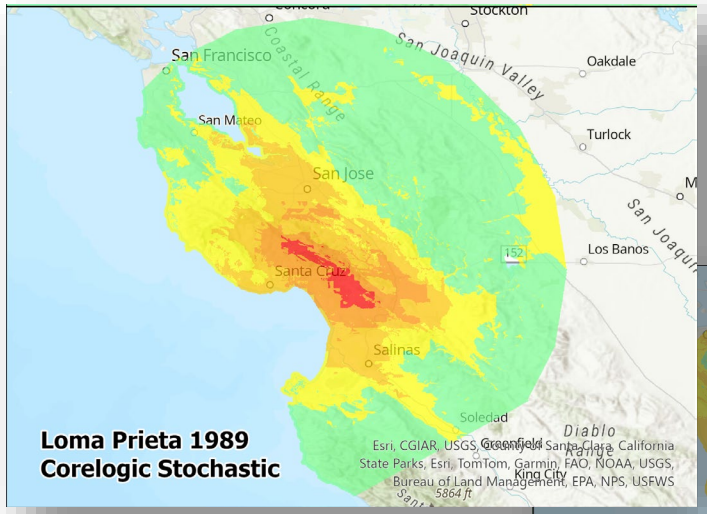


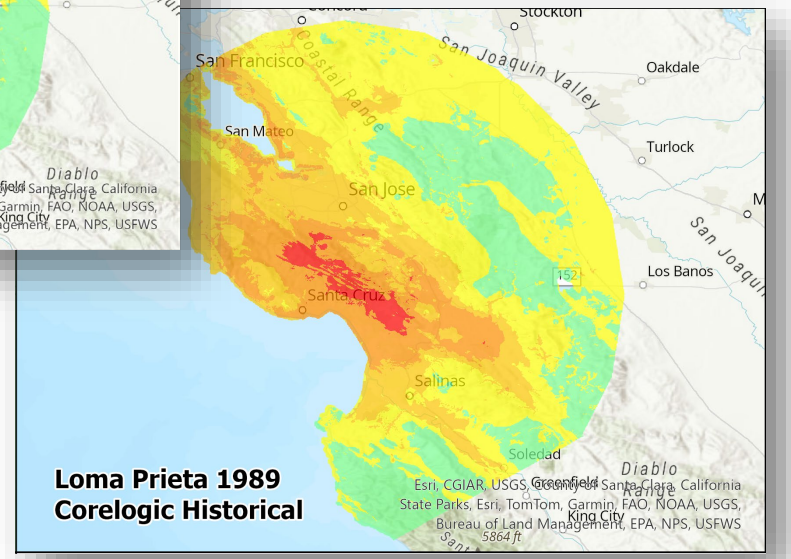
ARA

NCAR

Footprint Analysis: Loma Prieta EQ 1989







PSA at 1.0 s

- Not Felt (<.17 %g)
- Weak (.17 - 1.4 %g)
- Light (1.4 - 3.9 %g)
- Moderate (3.9 - 9.2 %g)
- Strong (9.2 - 18 %g)
- Very Strong (18 - 34 %g)
- Severe (34 - 65 %g)
- Violent (65 - 124 %g)
- Extreme (>124 %g)

USGS

CoreLogic®

Conclusions

- Validation is a necessary early step in model assessment but…

- At *best*, validation only narrows our cat risk worldview to a finite set of (potentially unsensible) metrics that may miss the real value of cat models. At *worst*, it may incorrectly tie us into the use of a model framework that is not actually predictive, but has been highly attuned and thus appears to be.

- "Modelers and policymakers must continue to work toward finding effective ways to evaluate and judge the quality of their models, and to develop appropriate terminology to communicate these judgments." Oreskes, 1998.

Thank You

tom@maxinfo.io

MAXIMUM INFORMATION