How ETAS Can Leverage Completeness of Modern Seismic Networks Without Renouncing Historical Data

EGU General Assembly 2020: Sharing Geoscience Online

Leila Mizrahi[1]

Shyam Nandan[1]

Stefan Wiemer[1]

Contact: leila.mizrahi@sed.ethz.ch

[1] Swiss Seismological Service, ETH Zurich, Switzerland

Schweizerischer Erdbebendienst
Service Sismologique Suisse
Servizio Sismico Svizzero
Swiss Seismological Service

ETH zürich

Abstract

The Epidemic-Type Aftershock Sequence (ETAS) model is often used to describe the spatio-temporal distribution of earthquakes. A fundamental requirement for parameter estimation of the ETAS model is the completeness of the catalog above a magnitude threshold m_c. m_c is known to vary with time for reasons such as gradual improvement of the seismic network, short term aftershock incompleteness and so on. For simplicity, nearly all applications of the ETAS model assume a global magnitude of completeness for the entirety of the training period. However, in order to be complete for the entire training period, the modeller is often forced to use very conservative estimates of m_c, as a result completely ignoring abundant and high-quality data from the recent periods, which falls below the assumed m_c. Alternatively, to benefit from the abundance of smaller magnitude earthquakes from the recent period in model training, the duration of the training period is often restricted. However, parameters estimated in this way may be dominated by one or two sequences and may not represent long term behaviour.

We developed an alternative formulation of ETAS parameter inversion using expectation maximization, which accounts for a temporally variable magnitude of completeness and the triggering power of unobserved events. Results of a synthetic test suggest that the parameter bias introduced by successive application of simulation and inversion decreases substantially with an increasing fraction of data used in the inversion.

To test the adequacy of such a technique, we evaluate its forecasting power on the Southern California catalog, compared to the constant completeness magnitude ETAS base model. In Southern California, m_c since 1971 is estimated to be around 3.3, and a general decreasing trend in the temporal evolution of m_c can be observed. Both models are trained on the primary catalog with identical time horizon. While the reference model is solely based on information about earthquakes of magnitude 3.3 and above, our alternative represents completeness magnitude as a monotonically decreasing step-function, starting at 3.3 and assuming values down to 1.9 in more recent times.

To compare the two models, we issue forecasts by repeated probabilistic simulation of earthquake interaction scenarios, and evaluate those forecasts by assessing the likelihood of the actual occurrences under each of the alternatives. As a measure to quantify the difference in performance between the two models, we calculate the mean information gain due to model extension for different spatial resolutions, different temporal forecasting horizons, and different target magnitude ranges.

Preliminary results of the Southern California pseudo-prospective forecasting experiment suggest that the forecasting power of such a model increases significantly with the amount of data available, indicating substantial importance of the method for the future of probabilistic seismic hazard assessment.

Epidemic-Type Aftershock Sequence (ETAS) Model

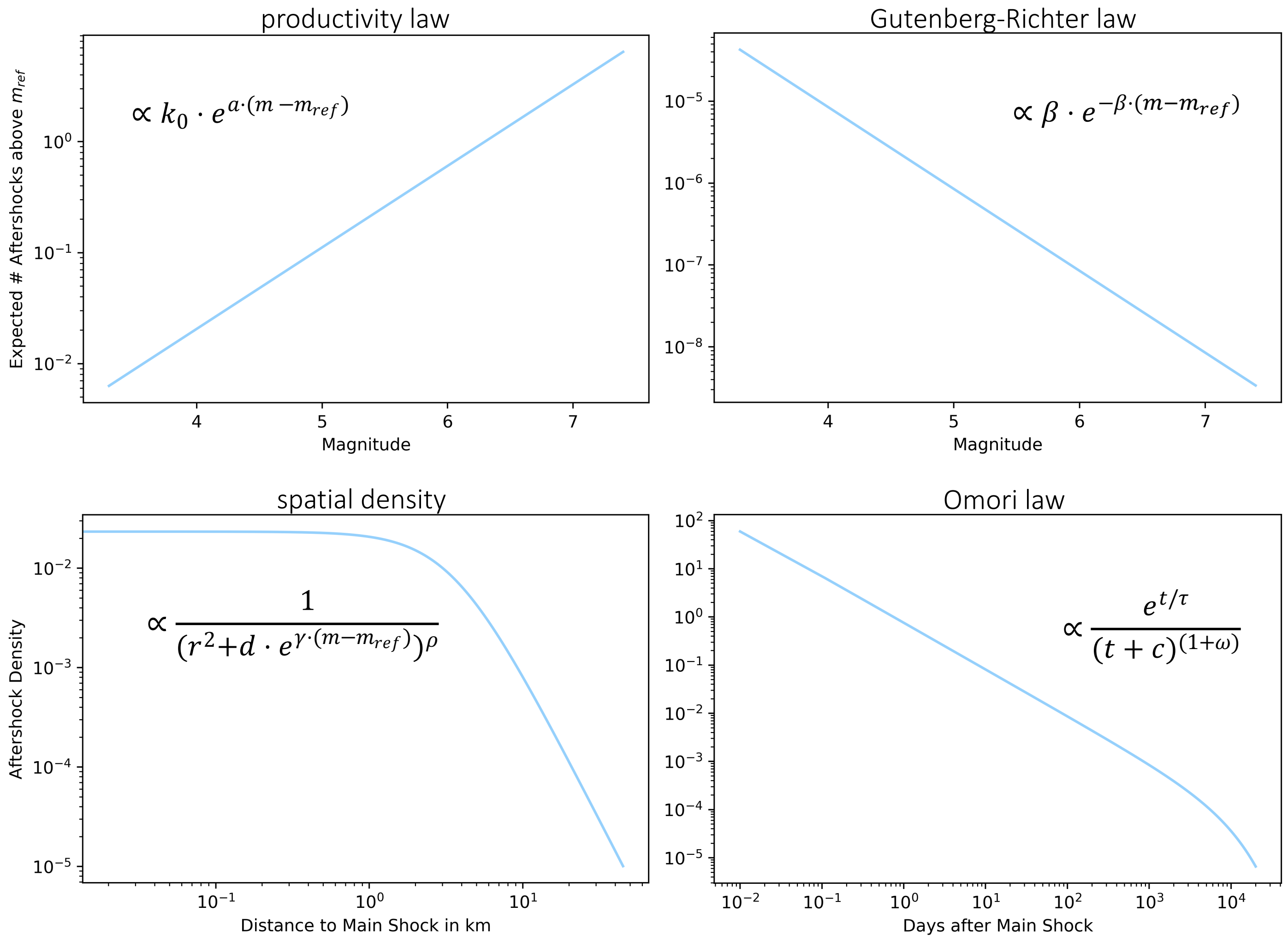
ETAS distinguishes **background events** and **triggered events**

All aftershocks can recursively trigger own aftershocks

Aftershock triggering is based on four **empirical principles**

To issue a forecast, one needs to:

- Estimate parameters describing these empirical laws
- Simulate thousands of possible scenarios

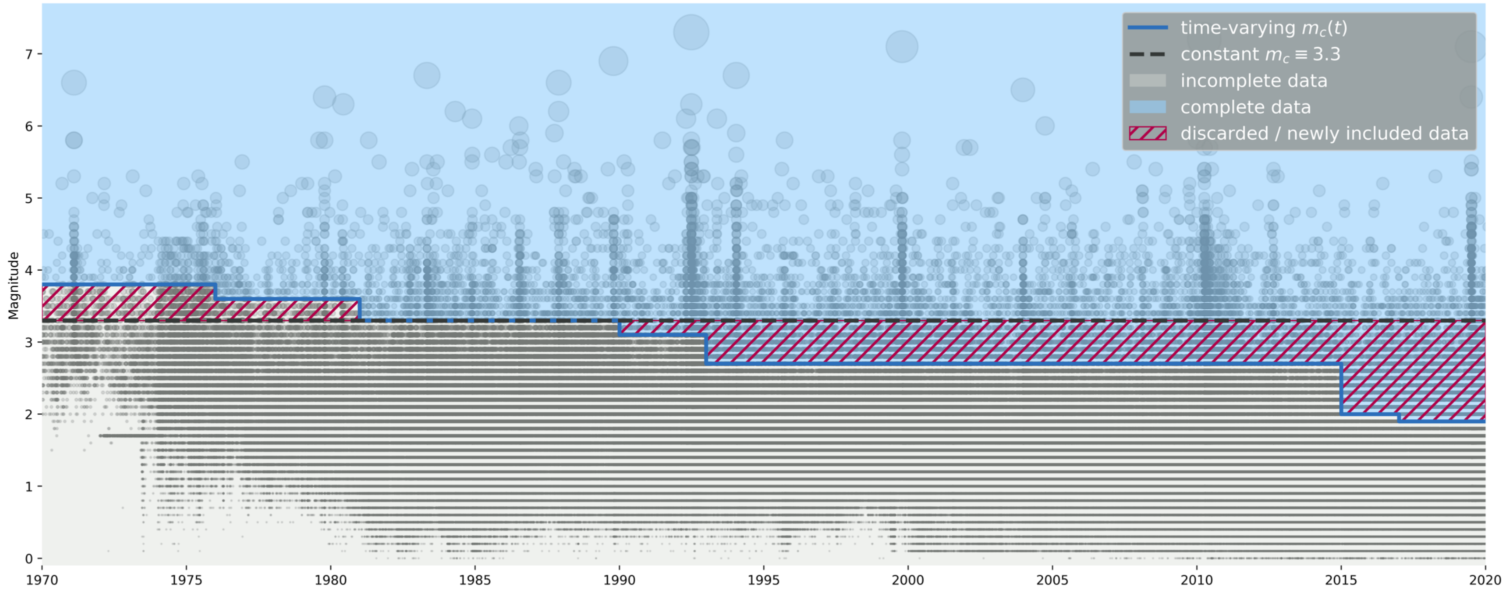


Simplifying assumptions:

- Only events above a **reference magnitude** $\boldsymbol{m_{ref}}$ are assumed to trigger or be triggered
- Because for parameter calibration, training catalog is required to be **complete above** $\boldsymbol{m_{ref}}$, one usually assumes $m_{ref} = m_c$

m_c: minimum magnitude of completeness of the catalog

Allowing Time-Varying Magnitude of Completeness



Earthquakes $M \geq 0.0$ in Southern California since 1970
Blue: complete data. Red: Discarded / newly included data when assuming time-varying instead of constant m_c

 # events above time-varying m_c: 40'800
 # events above constant m_c: 11'400

- Usually in ETAS, completeness magnitude *(m_c)* is assumed to be constant in time
- In reality, m_c changes with time, due to improvements of seismic network, short term aftershock incompleteness, and so on
- Time-varying completeness allows to use abundant and high-quality data from the recent periods without losing the long-term time horizon of the training catalog

Synthetic Experiment: Setup

How does the **fraction Φ of data** above completeness which is available for inversion affect the **accuracy of parameter estimates?**

Test data: ETAS-simulated catalog

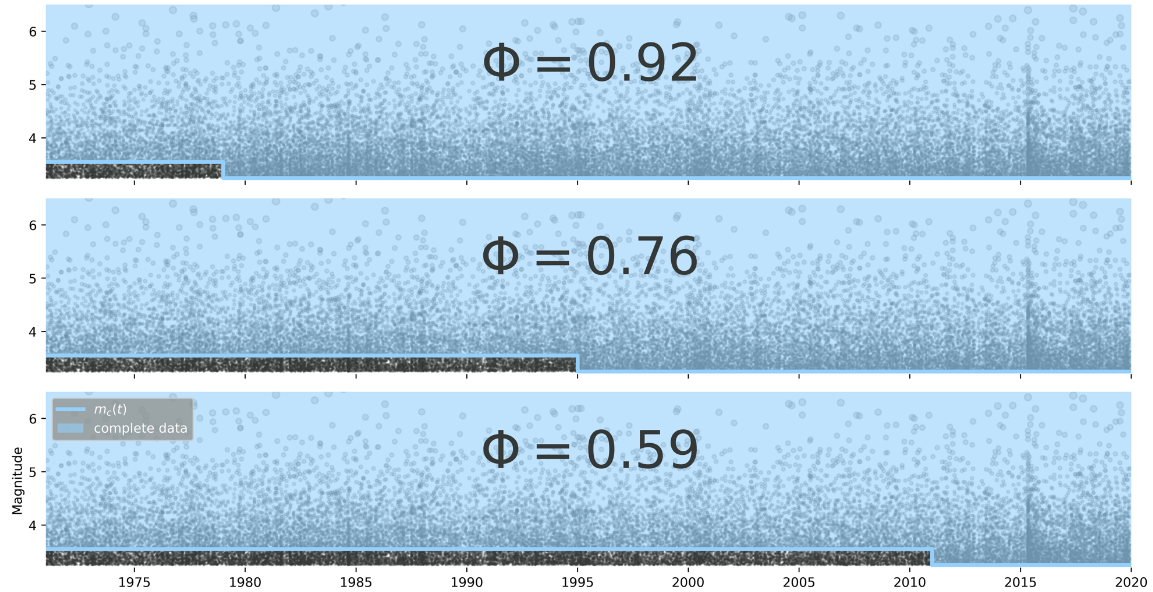
- m_{ref} = 3.3
- Latitude: -140 to -100
- Longitude: 20 to 60
- Time: 1971 to 2020

Parameter	Value	
μ	$10^{-7.5}$	*background rate*
k_0	$10^{-2.6}$	*productivity parameters*
a	1.6	
c	$10^{-2.95}$	*time kernel parameters*
ω	-0.02	
τ	4.0	
d	$10^{-0.45}$	*spatial kernel parameters*
γ	1.01	
ρ	0.5	
b	1.0	*GR law exponent*

ETAS parameters used for catalog simulation

Family of time-varying completeness magnitude step functions to control completeness fraction:

$$m_c^{t_0}(t) = \begin{cases} 3.6, & t < t_0 \\ 3.3, & t \geq t_0 \end{cases}$$



Different $m_c(t)$ step functions applied to synthetic catalog
Blue: complete data.
Φ: fraction of data above m_{ref} which falls above $m_c(t)$

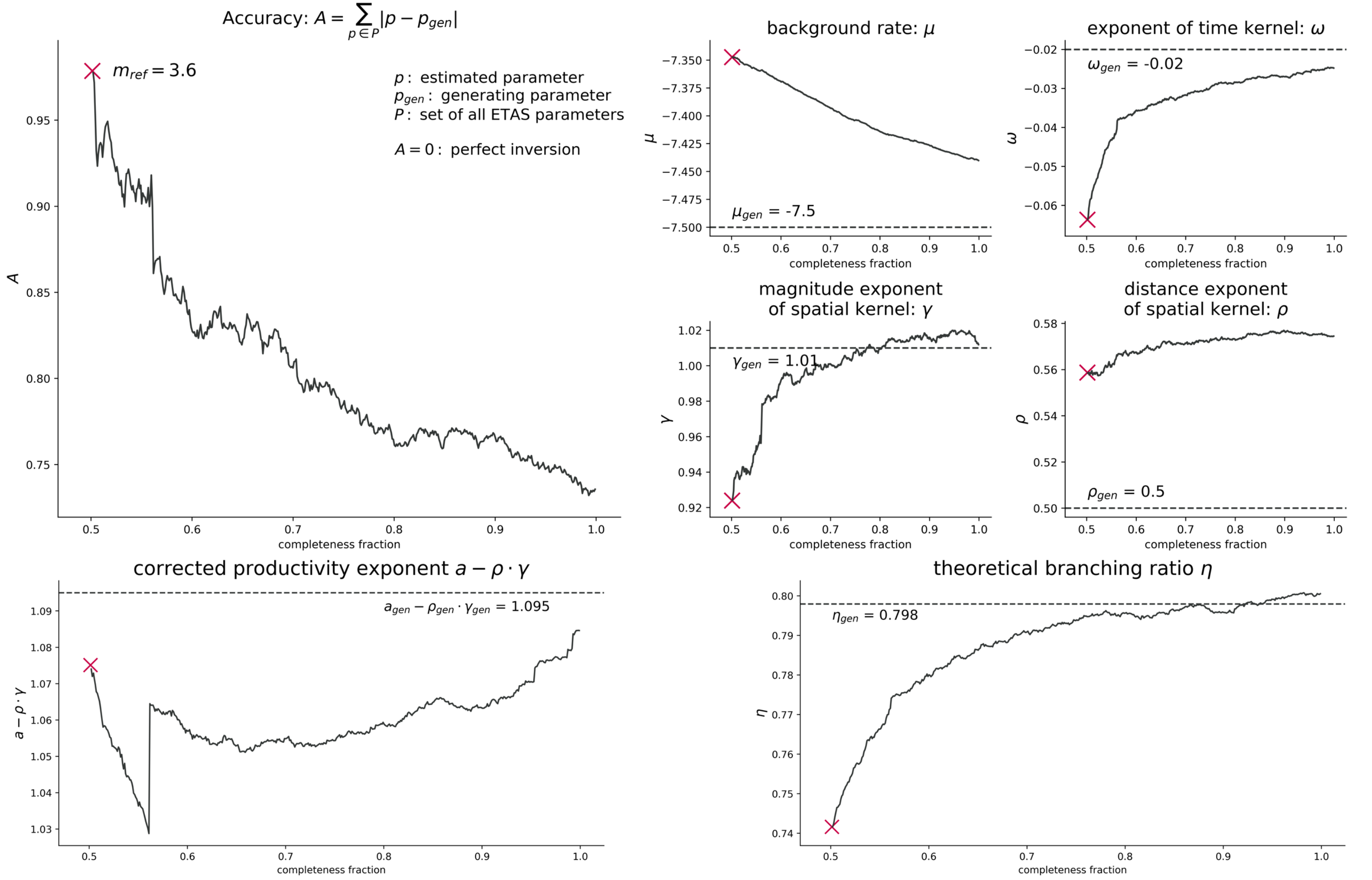
Synthetic Experiment: Results

We reformulated the commonly used ETAS inversion using expectation maximization (Veen and Schoenberg, 2008), to account for

- time-varying magnitude of completeness
- triggering power of unobserved events

Overall accuracy improves with increasing *Φ*.

While individual parameters may show no or minor improvement, more meaningful quantities such as the branching ratio η improve substantially



Evolution of selected quantities and ETAS parameters estimated with increasing Φ. Dashed line marks generating parameters' equivalent. For m_{ref} = 3.6, μ, k_0 and d are corrected to match m_{ref} = 3.3. For the cumulative absolute difference to generating parameters (top left) we compare log_{10} of μ, d, k0, c and τ, instead of the values themselves.

Reference: Veen, A., & Schoenberg, F. P. (2008). Estimation of space–time branching process models in seismology using an em–type algorithm. *Journal of the American Statistical Association*, *103*(482), 614-624.

Pseudo-Prospective Forecasting Experiment (Ongoing)

In Southern California: Does including low-magnitude earthquake data yield **better forecasts?**

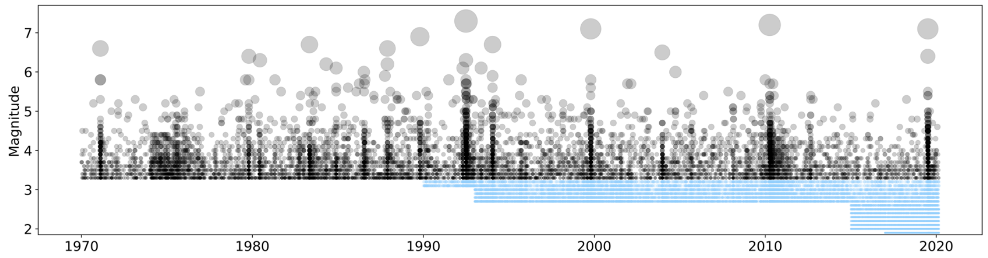
Two competing models:

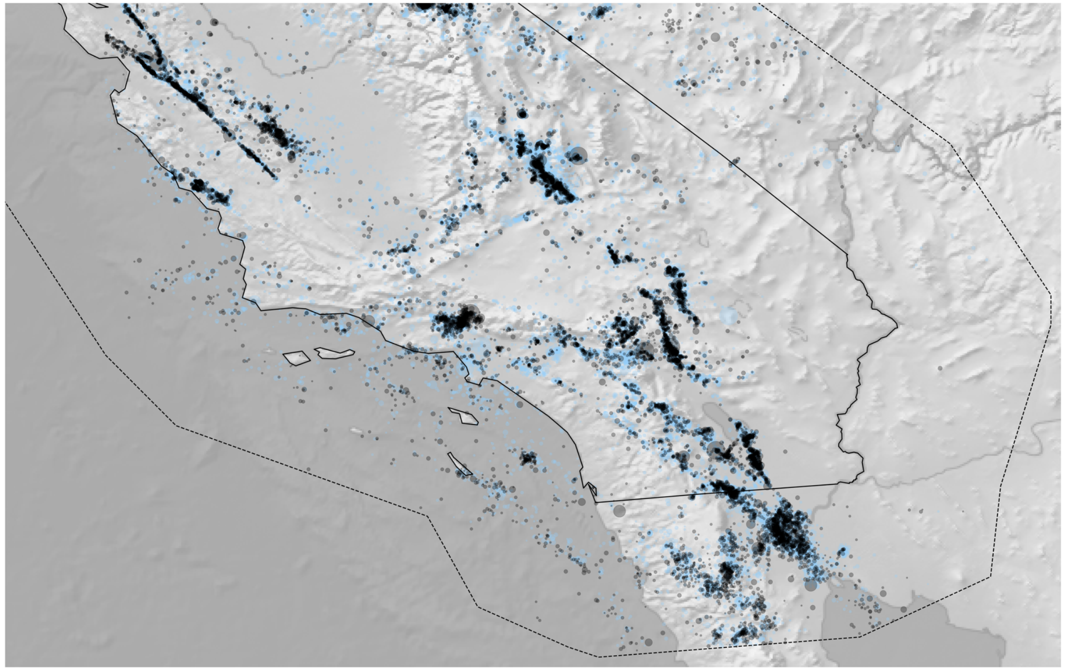
Model 1: constant $m_c = 3.3$

- Parameters inverted with standard method
- Simulation of aftershocks of earthquakes above M3.3

Model 2: time-varying $m_c(t)$

- Parameters inverted with reformulated method
- Simulation of aftershocks of earthquakes above $m_c(t)$





Data used to train (and test) the two models. Earthquakes marked in black are used for both models, those marked in blue are only used in Model 2. Dashed line marks the area in consideration.

Pseudo-Prospective Forecasting Experiment (Ongoing)

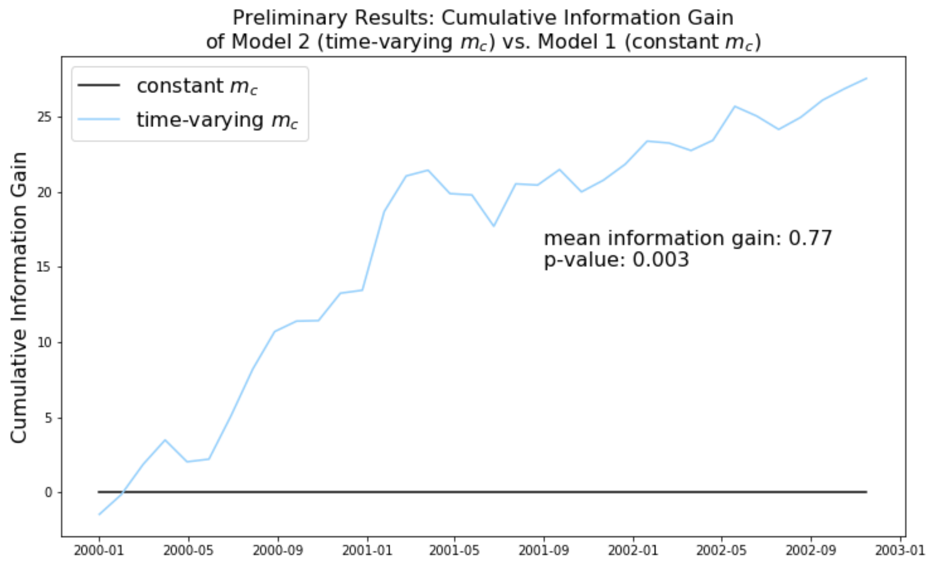
Pseudo-Prospective Forecasting Setup:

- Non-overlapping 30 day forecast testing periods, starting on January 1st, 2000, ending in January 2020
- Spatial resolution: 0.1° lat x 0.1° long (≈ 10km x 10km)
- Magnitude threshold for target events: 3.3

Here showing results for the first 36 testing periods (until November 2002)

Preliminary Results:

Including low-magnitude data when training the ETAS model **significantly improves** its forecasting accuracy.



Cumulative information gain of Model 2 (time-varying m_c) over Model 1 (constant m_c). Information gain is defined as difference in log-likelihood of observed truth in each model (as proposed by Nandan et al., 2019).

Reference: Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019). Forecasting the Full Distribution of Earthquake Numbers Is Fair, Robust, and Better. *Seismological Research Letters*, *90*(4), 1650-1659.