

Embracing Data Incompleteness for Better Earthquake Forecasting

Leila Mizrahi¹, Shyam Nandan², and Stefan Wiemer²

¹Swiss Seismological Service

²ETH

November 22, 2022

Abstract

We propose two new methods to calibrate the parameters of the epidemic-type aftershock sequence (ETAS) model based on expectation maximization (EM) while accounting for temporal variation of catalog completeness. The first method allows for model calibration on earthquake catalogs with long history, featuring temporal variation of the magnitude of completeness, m_c . This extended calibration technique is beneficial for long-term probabilistic seismic hazard assessment (PSHA), which is often based on a mixture of instrumental and historical catalogs. The second method jointly estimates ETAS parameters and high-frequency detection incompleteness to address the potential biases in parameter calibration due to short-term aftershock incompleteness. For this, we generalize the concept of completeness magnitude and consider a rate- and magnitude-dependent detection probability – embracing incompleteness instead of avoiding it. Using synthetic tests, we show that both methods can accurately invert the parameters of simulated catalogs. We then use them to estimate ETAS parameters for California using the earthquake catalog since 1932. To explore how the newly gained information from the second method affects earthquakes' predictability, we conduct pseudo-prospective forecasting experiments for California. Our proposed model significantly outperforms the base ETAS model, and we find that the ability to include small earthquakes for simulation of future scenarios is the main driver of the improvement. Our results point towards a preference of earthquakes to trigger similarly sized aftershocks, which has potentially major implications for our understanding of earthquake interaction mechanisms and for the future of seismicity forecasting.

Embracing Data Incompleteness for Better Earthquake Forecasting

L. Mizrahi^{1*}, S. Nandan¹, and S. Wiemer¹

¹Swiss Seismological Service, ETH Zurich

Key Points:

- Two methods are proposed to invert ETAS parameters when catalog completeness varies with time.
- We find that the ability to include small events in ETAS simulations leads to superior forecasts.
- Our results suggest that earthquakes may tend to trigger aftershocks of similar size.

*Sonneggstrasse 5, 8092 Zurich, Switzerland

Corresponding author: Leila Mizrahi, leila.mizrahi@sed.ethz.ch

Abstract

We propose two new methods to calibrate the parameters of the epidemic-type aftershock sequence (ETAS) model based on expectation maximization (EM) while accounting for temporal variation of catalog completeness. The first method allows for model calibration on earthquake catalogs with long history, featuring temporal variation of the magnitude of completeness, m_c . This extended calibration technique is beneficial for long-term probabilistic seismic hazard assessment (PSHA), which is often based on a mixture of instrumental and historical catalogs. The second method jointly estimates ETAS parameters and high-frequency detection incompleteness to address the potential biases in parameter calibration due to short-term aftershock incompleteness. For this, we generalize the concept of completeness magnitude and consider a rate- and magnitude-dependent detection probability – embracing incompleteness instead of avoiding it. Using synthetic tests, we show that both methods can accurately invert the parameters of simulated catalogs. We then use them to estimate ETAS parameters for California using the earthquake catalog since 1932. To explore how the newly gained information from the second method affects earthquakes’ predictability, we conduct pseudo-prospective forecasting experiments for California. Our proposed model significantly outperforms the base ETAS model, and we find that the ability to include small earthquakes for simulation of future scenarios is the main driver of the improvement. Our results point towards a preference of earthquakes to trigger similarly sized aftershocks, which has potentially major implications for our understanding of earthquake interaction mechanisms and for the future of seismicity forecasting.

Plain Language Summary

Some earthquakes, especially smaller ones, are not detected by the instruments installed to detect them. Our capability to detect earthquakes varies with time, on one hand because more and better instruments are being deployed over the years, leading to long-term changes of detection capability. On the other hand, earthquakes are more difficult to be detected during times when several earthquakes are shaking the ground simultaneously. This manifests itself in short-term changes of detection capability. Incomplete detection can lead to different kinds of bias in seismicity models used for earthquake forecasting. We propose two methods which allow us to calibrate these models while accounting for long-term (first model) and short-term (second model) changes in detection capability, which in turn allows us to use a larger and more representative fraction of the available data. We test both methods on synthetic data and then apply them to the Californian earthquake catalog. Finally, we test how the second model can help us become better at forecasting. We find that being able to include small earthquakes in our simulations leads to superior forecasts. Also, our results suggest that earthquakes may preferentially trigger aftershocks of similar size, which can have major implications for earthquake forecasting.

Keywords

data incompleteness, model inversion, ETAS, earthquake forecasting

1 Introduction

Earthquakes trigger aftershocks, which can trigger their own aftershocks, and so on. Because of this, they cluster in space and time. The clustering behavior of earthquakes is relevant in a variety of contexts, and in particular it is relevant for seismicity forecasting. Epidemic-type aftershock sequence (ETAS) models (see Ogata, 1998; Veen and Schoenberg, 2008; Nandan et al., 2017) intrinsically account for the spatio-temporal

59 clustering of earthquakes, and they have been shown to be among the best-performing
60 earthquake forecasting models available today (Nandan et al., 2019c).

61 1.1 Variation of m_c

62 A fundamental requirement for parameter estimation of the ETAS model is the com-
63 pleteness of the training catalog above a magnitude threshold m_c , the magnitude of com-
64 pleteness. If the catalog were incomplete, aftershocks of undetected events could be wrongly
65 interpreted as aftershocks of events in the catalog, which would distort our view on the
66 triggering behaviour of earthquakes. Numerous approaches to estimate m_c have been
67 proposed (Wiemer and Wyss, 2000; Cao and Gao, 2002; Woessner and Wiemer, 2005;
68 Amorese, 2007; Rydelek and Sacks, 1989; see Mignan and Woessner (2012) for an overview).
69 While in many cases, a global value of m_c is established for the entire catalog, it is well
70 known that m_c can vary with space. Several estimation techniques for spatial variations
71 of m_c exist (see e.g. Wiemer and Wyss (2000), or Mignan et al. (2011) with a Bayesian
72 approach). In addition to varying in space, m_c can vary with time for reasons such as
73 gradual improvement of the seismic network, software upgrades, and so on (Woessner
74 and Wiemer, 2005; Amato and Mele, 2008; Nanjo et al., 2010; Hutton et al., 2010; Mignan
75 and Chouliaras, 2014). Schorlemmer and Woessner (2008) proposed a probabilistic method
76 to calculate a space-time dependent m_c based on station data. One of the few assump-
77 tions they made is that data comes from a period of homogeneous recording, which ex-
78 plicitly excludes periods of aftershock clustering.

79 This raises another important cause of variations in time of m_c : Short-term after-
80 shock incompleteness (STAI). Because earthquakes strongly cluster in time, seismic net-
81 works can only capture a subset of events during periods of high activity (Kagan, 2004).
82 Hainzl (2016b) and Hainzl (2016a) modeled STAI based on the short-term rate of earth-
83 quakes, bringing into relation true and apparent triggering laws. Stallone and Falcone
84 (2020) have proposed a method to stochastically replenish catalogs suffering from STAI,
85 to be used for better operational earthquake forecasting and hazard assessment, albeit
86 without addressing effectiveness of the method in this regard. van der Elst (2021) recently
87 described a robust b -value estimator which can be applied during incomplete aftershock
88 sequences and provides an instantaneous estimate of time-varying m_c . With m_c usually
89 being assumed constant in the ETAS parameter inversion, the crucial requirement of com-
90 pleteness of the training catalog is not fulfilled during large aftershock sequences, which
91 has been shown to cause a substantial bias of the estimated parameters (Seif et al., 2017).
92 Approaching this issue, Omi et al. (2014) described a method to estimate parameters
93 of the ETAS model from incompletely observed aftershock sequences, by statistically mod-
94 elling detection deficiency. On a similar note, Zhuang et al. (2017) showed that estimat-
95 ing ETAS parameters using a replenished catalog is more stable with respect to cutoff
96 magnitude.

97 1.2 The importance of earthquakes below m_c

98 Besides the incompleteness-induced bias of estimated ETAS parameters, a second,
99 distinct biasing effect of incompleteness on ETAS models is the forced omission of after-
100ershocks triggered by unobserved events in the forecast. Although small earthquakes
101 trigger fewer aftershocks than large ones do, Marsan (2005), as well as Helmstetter et
102 al. (2005) found that small earthquakes, being more numerous, are as important as large
103 ones for earthquake triggering. Helmstetter et al. (2006) thereafter proposed a forecast-
104 ing model based on ETAS with a focus on the inclusion of aftershocks of small earth-
105 quakes in their forecast. They modeled a time-varying $m_c(t)$ during aftershock sequences
106 and inflated the triggering contribution of events above $m_c(t)$ to account for aftershocks
107 triggered by those below $m_c(t)$. Helmstetter et al. (2007) used the analogous technique
108 in the time-independent context and the resulting model was shown to outperform all
109 other time-independent models participating in the prospective CSEP forecasting exper-

110 iment in California (Strader et al., 2017). The success of their high-resolution forecast
 111 can potentially be attributed to the inclusion of small earthquakes when forecasting large
 112 ones. On the other hand, Nandan et al. (2019a) found that earthquakes tend to pref-
 113 erentially trigger similarly sized aftershocks, which would reduce the contribution of small
 114 events in triggering larger ones. Besides their potential to cumulatively contribute to af-
 115 tershock triggering, the large number of earthquakes below m_c can help highlighting the
 116 underlying fault structure.

117 The idea that earthquakes below the threshold of (complete) detection are relevant
 118 for our understanding of earthquakes' clustering behavior was thoroughly discussed by
 119 Sornette and Werner (2005a). They pointed out the important distinction between min-
 120 imum triggering magnitude m_0 and magnitude of completeness, giving constraints for
 121 m_0 . More generally, Sornette and Werner (2005b) found that assuming m_0 to be equal
 122 to m_c leads to a biased view of the ongoing triggering. They predicted that the branch-
 123 ing ratio will be drastically underestimated, and the fraction of non-triggered events will
 124 be overestimated, if m_0 is much smaller than m_c .

125 1.3 Method 1: ETAS calibration with time-varying m_c

126 Although short- and long-term variation of m_c and the biasing effects on ETAS pa-
 127 rameter estimates caused by data incompleteness are known and discussed (Hainzl (2016b);
 128 Seif et al., 2017; Zhuang et al. (2017)), nearly all applications of the ETAS model assume
 129 for simplicity a global magnitude of completeness for the entirety of the training period.
 130 In order to be complete for the entire training period, the modeller is often forced to use
 131 very conservative estimates of m_c , as a result completely ignoring abundant and high-
 132 quality data from more complete periods. Alternatively, to benefit from the abundance
 133 of smaller magnitude earthquakes in more complete periods in model training, the du-
 134 ration of the training period is often restricted. Parameters estimated in this way can
 135 however be dominated by one or two sequences and may not represent long-term behav-
 136 ior, making the use of ETAS models for long term probabilistic seismic hazard assess-
 137 ment (PSHA) non-ideal. This problem is further exacerbated due to a substantial re-
 138 duction in the size of catalog eligible for ETAS model calibration due to the assumption
 139 of conservatively constant m_c . Instead, the modellers rely on smoothed seismicity ap-
 140 proaches based on declustered catalogs (see e.g. Gerstenberger et al., 2020; Petersen et
 141 al., 2018; Wiemer et al., 2009), which has been found to be a problematic approach due
 142 to the biasing effects of declustering on the size distribution of mainshocks, and thus on
 143 the estimated seismic hazard (Mizrahi et al., 2021). In this regard, Llenos and Michael
 144 (2020) proposed an approach to calculate regionally optimized background earthquake
 145 rates from ETAS to be used for the U.S. Geological Survey National Seismic Hazard Model
 146 (NSHM), stressing the need for methods to address catalog heterogeneities such as time-
 147 dependent incompleteness.

148 **Goal 1:** In this article, we propose a modified ETAS parameter inversion method
 149 for training catalogs with time-varying completeness magnitude $m_c(t)$. This simultane-
 150 ously allows the the inclusion of historical data in the parameter inversion, as well as the
 151 inclusion of small magnitude events, which make up a large fraction of data and can en-
 152 able the ability to more clearly illuminate faults. ETAS models can hence be trained on
 153 a more representative and informative set of data, which in some areas facilitates a more
 154 correct approach to PSHA.

155 1.4 Method 2: an approach to generalize the concept of m_c

156 The clustering behavior of earthquakes, which is at the heart of the success of ETAS
 157 models, is also the cause of STAI, an important source of data incompleteness and hence
 158 a source of bias in ETAS. With the second method proposed in this article, we want to
 159 utilize the knowledge about clustering derived using the ETAS model to quantitatively

160 estimate the level of completeness of a catalog at any given time, and then use this knowl-
 161 edge to minimize the incompleteness-induced bias in the ETAS model.

We approach this issue by generalizing the notion of m_c , moving from a binary completeness space (complete versus incomplete) to a continuous-valued completeness space by means of a magnitude-dependent detection probability – embracing incompleteness instead of avoiding it. Deriving inspiration from Hainzl (2016b) and Hainzl (2016a), we model the probability of an earthquake to be detected, $f(m, \lambda(t))$, depending on its magnitude m and on the current local earthquake rate $\lambda(t)$. Besides a description of detection probability in terms of magnitude and current rate, any such model intrinsically requires an ancillary model to describe the temporal evolution of rates. Conveniently, the ETAS model provides a simple way of calculating such time-dependent event rates as

$$\lambda(t) = \mu + \sum_{i:t_i < t} \iint_R g(m_i, t - t_i, x - x_i, y - y_i) dx dy. \quad (1)$$

That is, the sum of background rate μ and the rate of all aftershocks of previous events e_i , throughout the region R . Note that for simplicity, by integrating over space this formulation of λ is spatially invariant. Here, m_i , t_i and (x_i, y_i) denote the magnitude, time, and epicentre of event e_i . The aftershock triggering rate $g(m, \Delta t, \Delta x, \Delta y)$ describes the rate of aftershocks triggered by an event of magnitude m , at a time delay of Δt and a spatial distance $(\Delta x, \Delta y)$ from the triggering event. We here use the definition

$$g(m, \Delta t, \Delta x, \Delta y) = \frac{k_0 \cdot e^{a(m-m_{ref})}}{e^{\frac{-(\Delta t+c)^{1+\omega}}{\tau}} \cdot ((\Delta x^2 + \Delta y^2) + d \cdot e^{\gamma(m-m_{ref})})^{1+\rho}}, \quad (2)$$

162 as in Nandan et al. (2017).

163 The time-dependent event rate $\lambda(t)$, and hence the temporal evolution of detec-
 164 tion probability, depends on the estimated ETAS parameters ($\mathcal{E} = (\mu, k_0, a, c, \omega, \tau, d, \gamma, \rho)$)
 165 which describe $g(m, \Delta t, \Delta x, \Delta y)$. Forming a loop of subordination, the ETAS param-
 166 eter estimates in turn depend on the temporal evolution of detection probability, as knowl-
 167 edge about catalog incompleteness lets us correct for this incompleteness during param-
 168 eter inversion.

169 **Goal 2:** In this article, we describe a method to estimate ETAS parameters (\mathcal{E})
 170 when time-varying probabilistic detection incompleteness of the training catalog is given,
 171 and vice-versa a method to estimate high-frequency probabilistic detection incompleteness
 172 (\mathcal{I}) when ETAS parameters are known. An overarching algorithm leverages this cir-
 173 cular dependency and jointly estimates both \mathcal{I} and \mathcal{E} in a self-consistent manner.

174 The first method described in this article allows $m_c(t)$ as an input to the ETAS pa-
 175 rameter calibration, which makes it powerful in a long-term context. This second method
 176 is complementary to the first one, addressing the additional challenge of estimating short-
 177 term variations of completeness.

178 1.5 Paper outline

179 The remainder of the paper is structured as follows. Section 2 describes the earth-
 180 quake catalog that was used in this analysis. The modified ETAS parameter inversion
 181 methods are presented in Section 3.1 for time-varying m_c , and in Section 3.2 for time-
 182 varying probabilistic detection incompleteness. Sections 3.3 and 3.4 describe the formu-
 183 lation of probabilistic detection incompleteness and the algorithm for joint estimation
 184 of ETAS parameters and detection probability. Section 4 presents synthetic tests for both
 185 methods, and Section 5 presents applications of both methods to the Californian data.
 186 Section 6 describes pseudo-prospective forecasting experiments used to assess the im-

187 pact of the newly acquired information on the forecastability of earthquakes in Califor-
 188 nia. Finally, in Section 7, we present our conclusions.

189 2 Data

190 In this article, we use the ANSS Comprehensive Earthquake Catalog (ComCat) pro-
 191 vided by the U.S. Geological Survey. We adopt the preferred magnitudes as defined in
 192 ComCat, and use as study region the collection area around the state of California as
 193 proposed in the RELM testing center (Schorlemmer and Gerstenberger, 2007). We con-
 194 sider events of magnitude $M \geq 0.0$, with magnitudes rounded into bins of size $\Delta M =$
 195 0.1 . For the major part of the study, the time frame used is January 1, 1970 until De-
 196 cember 31, 2019. For the analysis of long-term variations in m_c , we extend the time frame
 197 to start on January 1, 1932, when instrumentation was introduced to the Californian seis-
 198 mic network (Felzer, 2007).

199 Whenever ETAS parameters are inverted, we use the first fifteen years of data to
 200 serve as auxiliary data. Earthquakes in the auxiliary catalog may act as triggering earth-
 201 quakes in the ETAS model, but not as aftershocks. Thus, the start of the primary cat-
 202 alog is either January 1985, or January 1947.

203 To estimate a constant magnitude of completeness of the catalog, we use the method
 204 described by Mizrahi et al. (2021) with an acceptance threshold value of $p = 0.1$, which
 205 yields $m_c = 3.1$ for the time period between 1970 and 2019. This method is adapted
 206 from Clauset et al. (2009) and jointly estimates m_c and the b -value of the Gutenberg-
 207 Richter law (Gutenberg and Richter, 1944) describing earthquake size distribution. It
 208 compares the Kolmogorov-Smirnov (KS) distance between the observed cumulative dis-
 209 tribution function (CDF) and the fitted GR law to KS distances obtained for magnitude
 210 samples simulated from said GR law. A value of m_c is accepted if at least a fraction of
 211 $p = 0.1$ of KS distances is larger than the observed one.

212 3 Model

213 3.1 ETAS parameter inversion for time-varying m_c

Consider an earthquake catalog

$$C = \{e_i = (m_i, t_i, x_i, y_i), i \in \{1, \dots, n\}\} \quad (3)$$

214 consisting of events e_i of magnitudes m_i which occur at times t_i and locations (x_i, y_i) .
 215 Furthermore, consider a time-varying magnitude of completeness $m_c(t)$ defined for all
 216 t_i . We say that the catalog is complete if $m_i \geq m_c(t_i) \forall i$.

217 To calibrate the ETAS model, the nine parameters to be optimized are the back-
 218 ground rate μ and the parameters $k_0, a, c, \omega, \tau, d, \gamma, \rho$ which parameterize the aftershock
 219 triggering rate $g(m, t, x, y)$ given in Equation 2. We build on the expectation maximiza-
 220 tion (EM) algorithm to estimate the ETAS parameters (Veen and Schoenberg, 2008).
 221 In this algorithm, the expected number of background events \hat{n} and the expected num-
 222 ber of directly triggered aftershocks \hat{l}_i of each event e_i are estimated in the expectation
 223 step (E step), along with the probabilities p_{ij} that event e_j was triggered by event e_i ,
 224 and the probability p_j^{ind} that event e_j is independent. Following the E step, the nine pa-
 225 rameters are optimized to maximize the complete data log likelihood in the maximiza-
 226 tion step (M step). E and M step are repeated until convergence of the parameters. The
 227 usual formulation of the EM algorithm defines

$$p_{ij} = \frac{g_{ij}}{\mu + \sum_{k:t_k < t_j} g_{kj}}, \quad (4)$$

$$p_j^{ind} = \frac{\mu}{\mu + \sum_{k:t_k < t_j} g_{kj}}, \quad (5)$$

228 with $g_{kj} = g(m_k, t_j - t_k, x_j - x_k, y_j - y_k)$ being the aftershock triggering rate of e_k at
 229 location and time of event e_j . For a given target event e_j , Equations (4-5) define p_{ij} to
 230 be proportional to the aftershock occurrence rate g_{ij} , and p_j^{ind} to be proportional to the
 231 background rate μ . As an event must be either independent or triggered by a previous
 232 event, the normalization factor $\Lambda_j := \mu + \sum_{k:t_k < t_j} g_{kj}$ in the denominator of (4-5) stip-
 233 ulates that $p_j^{ind} + \sum_{k:t_k < t_j} p_{kj} = 1$. This relies on the assumption that all potential
 234 triggering earthquakes of e_j were observed, that is, all events prior to t_j above the refer-
 235 ence magnitude (minimum considered magnitude), m_{ref} were observed. To fulfill this
 236 requirement, most applications of the method define m_{ref} to be equal to the constant
 237 value of m_c .

238 For the case of time-varying $m_c(t)$, we define $m_{ref} := \min_i \{m_c(t_i)\}$, the minimum
 239 $m_c(t_i)$ for times t_i of events in the complete catalog. This implies that for the times when
 240 $m_c(t) > m_{ref}$ the requirement of complete recording of all potential triggers may be
 241 violated. Events whose magnitudes fall between m_{ref} and $m_c(t)$ are not part of the com-
 242 plete catalog and are considered to be unobserved (even though they may have been de-
 243 tected by the network). Hence, the normalization factor Λ_j (the denominator of Equa-
 244 tions 4-5) needs to be adapted to account for the possibility that e_j was triggered by an
 245 unobserved event.

Consider

$$\xi(t) = \frac{\int_{m_{ref}}^{m_c(t)} f_{GR}(m) \cdot G(m) dm}{\int_{m_c(t)}^{\infty} f_{GR}(m) \cdot G(m) dm}, \quad (6)$$

the ratio between the expected number of events triggered by unobserved events and the
 expected number of events triggered by observed events at time t . Here, $f_{GR} = \beta \cdot e^{-\beta \cdot (m - m_{ref})}$
 is the probability density function of magnitudes according to the GR law, and $G(m) =$
 $\int_0^{\infty} \iint_R g(m, t, x, y) dx dy dt$ is the total number of expected aftershocks larger than m_{ref}
 of an event of magnitude m . Note that in the calculation of $G(m)$ we make the simpli-
 fying assumption that the considered region R extends infinitely in all directions, allow-
 ing a facilitated, asymptotically unbiased estimation of ETAS parameters (Schoenberg,
 2013). Analogously,

$$\zeta(t) = \frac{\int_{m_{ref}}^{m_c(t)} f_{GR}(m) dm}{\int_{m_c(t)}^{\infty} f_{GR}(m) dm} \quad (7)$$

is the ratio between the expected number of unobserved events and the expected num-
 ber of observed event at time t . If $\beta > a - \rho\gamma$, both $\xi(t)$ and $\zeta(t)$ are well-defined and
 we have that

$$\xi(t) = e^{-(a - \beta - \rho\gamma) \cdot \Delta m} - 1, \quad (8)$$

$$\zeta(t) = e^{\beta \cdot \Delta m} - 1, \quad (9)$$

where $\Delta m = m_c(t) - m_{ref}$. The condition that β is larger than the productivity ex-
 ponent $\alpha = a - \rho\gamma$ (see Text S1) is generally fulfilled in naturally observed catalogs
 (Helmstetter, 2003). If this were not the case, earthquake triggering would be dominated
 by large events and one would need to introduce a maximum possible magnitude for both
 denominators to be finite. The normalization factor Λ_j consists of the sum of background
 rate and aftershock rates of all events which happened prior to e_j . In the case of time-
 varying m_c , besides the possibilities of being a background event or being triggered by
 an observed event, the event e_j can also be triggered by an unobserved event. We thus

generalize Λ_j by adding to the rate of aftershocks g_{kj} of each observed triggering event e_k the expected rate of aftershocks of unobserved triggering events at that time, $g_{kj} \cdot \xi(t_k)$. This yields $\Lambda_j = \mu + \sum_{k:t_k < t_j} g_{kj} \cdot (1 + \xi(t_k))$ and thus the generalized definition of p_{ij} and p_j^{ind} is given by

$$p_{ij} = \frac{g_{ij}}{\mu + \sum_{k:t_k < t_j} g_{kj} \cdot (1 + \xi(t_k))}, \quad (10)$$

$$p_j^{ind} = \frac{\mu}{\mu + \sum_{k:t_k < t_j} g_{kj} \cdot (1 + \xi(t_k))}. \quad (11)$$

Note that the probability p_{uj} that event e_j was triggered by an unobserved event is given such that $p_j^{ind} + p_{uj} + \sum_{k:t_k < t_j} p_{kj} = 1$. In the above equations, the special case of $m_c(t) \equiv m_{ref}$ is accounted for when $\xi(t) \equiv 0$. In this special case, \hat{n} and \hat{l}_i are obtained by summing independence probabilities ($\hat{n} = \sum_j p_j^{ind}$) and triggering probabilities ($\hat{l}_i = \sum_j p_{ij}$), respectively. In the generalized case however, \hat{n} and \hat{l}_i are the estimated number of background events and aftershocks above m_{ref} , which includes unobserved events. Similarly to inflating the triggering power, we hence inflate the observed event numbers to account for unobserved events. Whenever an event is observed at time t_j , we expect that $\zeta(t_j)$ events occurred under similar circumstances (with same independence and triggering probabilities), but were not observed. This yields

$$\hat{n} = \sum_j p_j^{ind} \cdot (1 + \zeta(t_j)), \quad (12)$$

$$\hat{l}_i = \sum_j p_{ij} \cdot (1 + \zeta(t_j)). \quad (13)$$

246 With these adapted definitions of p_{ij} , p_j^{ind} , \hat{n} and \hat{l}_i (Equations 10 - 13), ETAS param-
247 eters can be inverted using the procedure described by Veen and Schoenberg (2008).

248 3.2 ETAS parameter inversion for time-varying probabilistic detection

To overcome the binary view of completeness which forces us to disregard earthquakes which were detected but happen to fall between m_{ref} and $m_c(t)$, we can take the generalization of the EM algorithm for ETAS parameter inversion one step further by introducing a time and magnitude-dependent probability of detection,

$$f: \mathbb{R}_{\geq m_{ref}} \times \mathbb{R} \longrightarrow [0, 1] \\ (m, t) \mapsto p.$$

To be able to account for such a probabilistic concept of catalog completeness in the ETAS inversion algorithm, one needs to generalize $\xi(t)$ and $\zeta(t)$ (equations 6 and 7). In contrast to before, the magnitude of an event does not determine whether or not the event has been detected. We therefore adapt the bounds of integration in numerator and denominator such that all events above magnitude m_{ref} are considered. To obtain the expected number of earthquakes triggered by observed and unobserved events, the integrands are multiplied by the probability of the triggering events to be observed, $f(m, t)$, or unobserved, $(1 - f(m, t))$, respectively. The generalized formulations of $\xi(t)$ and $\zeta(t)$ then read

$$\xi(t) = \frac{\int_{m_{ref}}^{\infty} (1 - f(m, t)) \cdot f_{GR}(m) \cdot G(m) dm}{\int_{m_{ref}}^{\infty} f(m, t) \cdot f_{GR}(m) \cdot G(m) dm}, \quad (14)$$

and

$$\zeta(t) = \frac{\int_{m_{ref}}^{\infty} (1 - f(m, t)) \cdot f_{GR}(m) dm}{\int_{m_{ref}}^{\infty} f(m, t) \cdot f_{GR}(m) dm}. \quad (15)$$

249 For compatible choices of $f(m, t)$, $f_{GR}(m)$, $G(m)$, we find that $\xi(t)$ and $\zeta(t)$ are well-
 250 defined. Consider for instance the special case of binary detection, where $f(m, t)$ is de-
 251 fined via the Heaviside step function H as $f_{bin}(m, t) = H(m - m_c(t))$, which is equal
 252 to 1 if $m \geq m_c(t)$ and 0 otherwise. This is the case discussed in the previous section,
 253 for which we have well-definedness if $\beta > a - \rho\gamma$.

254 The reference magnitude m_{ref} is a model constant. Smaller values of m_{ref} allow
 255 the modeller to use a larger fraction of the observed catalog, which can be especially use-
 256 ful in regions with less seismic activity.

257 Note that both generalizations of the ETAS inversion algorithm (for time-varying
 258 completeness or for time-varying probabilistic detection) can without further modifica-
 259 tion be applied when m_c or detection probability vary with space. The formulation is
 260 based on the assumption that the behaviour of observed events is locally representative
 261 (in space and/or time) of the behaviour of unobserved events.

262 3.3 Rate-dependent probabilistic detection incompleteness

263 In this section we present our approach to define $f(m, t)$, where the temporal com-
 264 ponent is purely driven by the current rate of events $\lambda(t)$. Note that this means we only
 265 capture changes in detection due to changes in short-term circumstances, and neglect
 266 long-term changes due to network updates. We make the following simplifying assump-
 267 tions.

- 268 • Any earthquake will obstruct the entire seismic network from detecting smaller
 269 earthquakes for a duration of t_R (recovery time of the network).
- 270 • Magnitudes of events which are simultaneously blocking the network are distributed
 271 according to the time-invariant Gutenberg-Richter law which also describes the
 272 magnitude distribution of the full catalog (Gutenberg and Richter, 1944).

273 De Arcangelis et al. (2018) found that short-term aftershock incompleteness can be well
 274 explained in terms of overlapping seismic records, while instrumental coverage of an area
 275 plays a subsidiary role. Nevertheless, assuming t_R to be independent of the magnitude
 276 of the event, and independent of the spatial distance between the event and the locations
 277 of interest, is certainly a major simplification which could be refined in subsequent stud-
 278 ies.

The probability $f(m, t)$ of an earthquake to be detected is then given by the prob-
 ability of it being the largest of all the earthquakes which are currently blocking the net-
 work.

$$f(m, t) = \left(1 - e^{-\beta \cdot (m - m_{ref})}\right)^{t_R \cdot \lambda(t)}. \quad (16)$$

279 Here, $t_R \cdot \lambda(t)$ represents the expected number of events blocking the network at time
 280 t and $1 - e^{-\beta \cdot (m - m_{ref})}$ is the probability of any given earthquake's magnitude falling
 281 between m_{ref} and m , where $\beta = b \cdot \ln 10$ is the exponent in the GR law with basis e .
 282 Thus, $f(m, t)$ is the probability that in the set of $t_R \cdot \lambda(t)$ events currently blocking the
 283 network, all of them have a magnitude of less than m , which is the condition for an event
 284 of magnitude m to be detected. Because the time-dependence of $f(m, t)$ is solely con-
 285 trolled by the time-dependence of λ , we here use the terms $f(m, t)$ and $f(m, \lambda)$ inter-
 286 changeably.

Setting $\kappa := -\frac{a - \rho\gamma}{\beta}$, we obtain

$$\xi(t) = \frac{1}{(\kappa + 1) \cdot \text{B}(\kappa + 1, t_R \cdot \lambda(t) + 1)} - 1, \quad (17)$$

$$\zeta(t) = t_R \cdot \lambda(t), \quad (18)$$

287 so long as $\beta > a - \rho\gamma$, where B is the Beta function. A positive background rate $\mu >$
 288 0 ensures $\lambda(t) > 0 \quad \forall t$. Expressions analogous to (17) and (18) hold when alternative
 289 exponents are chosen instead of $t_R \cdot \lambda$ in the definition of $f(m, \lambda)$ (Equation 16).

290 The network recovery time t_R and the current event rate $\lambda(t)$ at the times t_i of all
 291 earthquakes e_i need to be estimated from the data.

292 **3.4 Estimating probabilistic epidemic-type aftershock incompleteness** 293 **(PETAI)**

294 **3.4.1 Estimation of (t_R, β) when λ_i are known**

295 Text S2 and Figure S1 describe how t_R and β can be jointly estimated using a max-
 296 imum likelihood approach for the case when current event rates $\lambda_i = \lambda(t_i)$ are known.
 297 In reality, λ_i have to be estimated themselves.

298 **3.4.2 Estimation of λ_i when ETAS parameters and (t_R, β) are known**

They depend on one hand on the ETAS parameters (see Equation 1). On the other
 hand, the sum of aftershocks of previous earthquakes in the definition of $\lambda(t)$ (Equation
 1) does not account for aftershocks of events that were not detected. As in the ETAS
 parameter inversion, to account for aftershocks of undetected events in the calculation
 of $\lambda(t)$, we inflate the triggering power of each event e_i by a factor of $1 + \xi(t_i)$ and de-
 fine

$$\lambda(t) = \mu + \sum_{i:t_i < t} (1 + \xi(t_i)) \cdot \iint_R g(m_i, t - t_i, x - x_i, y - y_i) dx dy. \quad (19)$$

299 **3.4.3 Estimation of λ_i and (t_R, β) when ETAS parameters are known**

300 $\xi(t)$ however requires knowledge of (t_R, β) . This implies that even when ETAS pa-
 301 rameters are fixed, an additional, lower-level circular dependency dictates the relation-
 302 ship between $(\lambda_i)_{i=1, \dots, n}$ and (t_R, β) .

303 To fully estimate the high-frequency probabilistic detection incompleteness, given
 304 fixed ETAS parameters, we recursively re-estimate $(\lambda_i)_{i=1, \dots, n}$ and (t_R, β) , until (t_R, β)
 305 converges, starting with an informed or random initial guess for (t_R, β) .

306 **3.5 PETAI inversion algorithm**

307 The overarching joint inversion of ETAS parameters (\mathcal{E}) and high-frequency de-
 308 tection incompleteness ($\mathcal{I} = (\lambda_i, t_R, \beta)$) starts with estimating ETAS parameters in the
 309 usual way, i.e. using the algorithm described in Section 3.1, with a time-independent com-
 310 pleteness magnitude $m_c (= m_{ref})$ above which all events are detected. It then recursively
 311 re-estimates (\mathcal{I}) and (\mathcal{E}) until convergence of the ETAS parameters. The result is a prob-
 312 abilistic, epidemic-type aftershock incompleteness (PETAI) model. With all its compo-
 313 nents having now been described, Figure S2 shows the flow diagram of the PETAI inver-
 314 sion algorithm. A simplified illustration of the inversion algorithm is shown in Figure
 315 1. Starting with the initial ETAS parameters obtained assuming constant m_c , event rates
 316 can be calculated at each point in time as the sum of aftershock rates of all previous events,
 317 plus the background rate. Given these event rates, the detection probability function is
 318 calibrated, which then provides insight into the temporal evolution of catalog (in-)completeness.
 319 ETAS parameters can then be re-estimated, now also using data below m_c , by account-
 320 ing for the estimated incompleteness. With this new set of ETAS parameters, event rates
 321 can be re-calculated, upon which detection probability is re-calibrated, and so on, un-
 322 til all convergence criteria are satisfied.

4 Synthetic tests

4.1 Long-term variation of m_c

To test the ETAS parameter inversion for time-varying m_c , we generate a complete synthetic catalog using ETAS and then artificially impose a given $m_c(t)$ on the catalog. Assuming $m_c(t)$ to be known, we use the method described in Section 3.1 to infer the parameters used in the simulation.

We estimate $m_c(t)$ based on the Californian catalog described in Section 2 with a time horizon from 1932 to 2019. Fixing the b -value we had estimated for the main catalog (1970 - 2019, $M \geq 3.1$, $b = 1.01$), we estimate m_c for successive 10 year periods starting in 1932. The last period then comprises only 8 years of data. Estimation of m_c is analogous to the main catalog, using the method of Mizrahi et al. (2021) with an acceptance threshold of $p = 0.1$, but keeping $b = 1.01$ fixed. This yields

$$m_c(t) = \begin{cases} 4.3 & \text{for } t \text{ between 1932 and 1941,} \\ 3.9 & \text{for } t \text{ between 1942 and 1951,} \\ 4.3 & \text{for } t \text{ between 1952 and 1961,} \\ 3.4 & \text{for } t \text{ between 1962 and 1971,} \\ 3.1 & \text{for } t \text{ between 1972 and 1981,} \\ 3.3 & \text{for } t \text{ between 1982 and 1991,} \\ 2.4 & \text{for } t \text{ between 1992 and 2001,} \\ 2.8 & \text{for } t \text{ between 2002 and 2011,} \\ 3.6 & \text{for } t \text{ between 2012 and 2019.} \end{cases} \quad (20)$$

The large increase in m_c for the years 2012 to 2019 is due to the Ridgecrest events in 2019. Although the period affected by aftershock incompleteness only makes up a small fraction of the 8 year period, our method with an acceptance threshold of $p = 0.1$ yields a conservative estimate of m_c . To avoid such an effect, one could use shorter than 10 year periods, or use different methods to estimate time-varying m_c . Note that our method to invert ETAS parameters for time-varying m_c (Section 3.1) accepts $m_c(t)$ as an input and works independently of how this $m_c(t)$ was obtained. We here want to keep the focus on the parameter inversion and thus choose the described approach to estimate $m_c(t)$ due to its simplicity.

To mimic a realistic scenario, we simulate the synthetic catalog using parameters obtained after applying ETAS parameter inversion for time-varying m_c on the California data, with two manual corrections.

The first correction is done because it has been shown that certain assumptions in the ETAS model such as a spatially isotropic aftershock distribution or a temporally stationary background rate, as well as data incompleteness can lead to biased estimations of the productivity exponent (Hainzl et al., 2008; Hainzl et al., 2013; Seif et al., 2017). This bias can lead to a lack of clustering when catalogs are simulated. We thus use an artificially increased productivity exponent for our simulation as described in Text S1.

Secondly, we reduce the background rate μ . In this way, the size of the simulated catalog is reduced such that inversion requires a reasonable amount of computational power, even for large regions and time horizons. The final parameters used for the simulation of the catalog can be found in Table 1, first column.

A catalog of events of magnitude $M \geq 2.4 = m_{ref}$ is simulated as described in Text S3 for the time period of January 1832 to December 2019 in a square of 40° lat \times 40° long. Because of missing long-term aftershocks in the beginning of the simulated catalog, we allocate a burn period of 100 years in the beginning of the simulated period and

are left with a catalog from 1932 to 2019. The starting year of our synthetic catalog coincides with the introduction of instrumentation in California (Felzer, 2007). This allows us to impose the $m_c(t)$ history observed in California on the synthetic catalog by discarding all events e_i for which $m_i < m_c(t_i)$.

We apply the ETAS inversion for time-varying m_c with the here-obtained $m_c(t)$ (equation 20) to the synthetic catalog. The first two columns of Table 1 show the ETAS parameters used in the simulation of the synthetic catalog, and inverted from the synthetic catalog. The parameters estimated from the synthetic catalog lie reasonably close to the truth, confirming the correctness of the method.

4.2 PETAI

To test the PETAI inversion algorithm, a synthetic catalog is created as follows. We use the parameters obtained after applying the PETAI inversion algorithm to the California data (1970 to 2019) with $m_{ref} = 2.5$.

The value of m_{ref} is chosen to achieve a balance between the amount of data available for the inversion and the computational power required to process such an amount of data. For the same reasons as before, we modify the parameters to obtain a corrected productivity exponent as described in Text S1. Although this time the starting parameters were estimated using PETAI, which already accounts for short-term aftershock incompleteness, the effects of the assumptions of isotropy and stationary background rate are not addressed, leaving a still non-negligible negative bias.

We reduce the background rate μ for the same reasons as before. The final parameters used for the simulation of the catalog can be found in Table 1, third column.

Using these parameters, we simulate as described in Text S3, a synthetic catalog that resembles the Californian catalog, for the period between 1850 and 2020 in a square of $40^\circ \text{ lat} \times 40^\circ \text{ long}$. As in the previous case, because of missing long-term aftershocks in the beginning of the simulated catalog, we discard the first 100 years of data and are left with a catalog from 1950 to 2020. Based on this catalog and given the ETAS parameters used for simulation, we calculate the current event rate at the time of each event in the catalog. As the current event rate is to a large extent driven by aftershock rates of earlier events, we expect overestimation of detection probabilities, as well as overestimation of independence probabilities, during the beginning of the time period (Wang et al., 2010; Schoenberg et al., 2010; Nandan et al., 2019a). For this reason, we allocate another 20 years of burn period, leaving us with a catalog starting in 1970.

Using the detection probability function given by Equation (16) and parameters estimated from the Californian catalog (Table 1, Column 3), we calculate for each event its probability of being detected, and according to this probability we randomly decide for each event whether it has been detected or not. The subset of all events that were detected is then our test catalog. While we use β as estimated in California, we adapt the value of t_R so that the fraction of undetected events in the catalog corresponds to the fraction of estimated undetected events inferred for California. This estimated number of undetected events is obtained by summing $\zeta(t_i)$, the expected number of unobserved events per observed event, which is estimated as a component of the PETAI inversion, over all occurrence times t_i of events in the primary catalog. Applied to California, we arrive at 5041.74 undetected events, which make up 6.25% of the inferred total. The same fraction of undetected events is obtained for the synthetic catalog when a value of 5 minutes ($10^{-2.46}$ days) is used for t_R .

401

4.2.1 Inverted number of undetected events

402

403

404

405

406

407

408

409

410

411

412

Figure 2 (a) shows the series of events of the synthetic test catalog over the primary time period in blue, with the undetected synthetic events marked in black. The number of undetected events is 1282, which makes up 6.25% of the original synthetic catalog. Figure 2 (b) shows cumulative number of undetected synthetic events over time in black, compared to the cumulative inferred number of undetected events in blue. Overall, it is estimated 1068.88 events were undetected. While this underestimates the true number of 1282 undetected events, the major part of events can be reconstructed, with accurate timing. As a comparison, we fitted a GR law to the test catalog and calculated the difference in event numbers between the extrapolated fit and the test catalog. The expected number of undetected events obtained this way is 583.13, which is less than half of the true number of undetected events.

413

414

415

416

417

418

419

420

421

422

423

424

425

Figure 2 (c) shows inferred versus actual number of undetected events for different assumed detection efficiencies. Starting from the same synthetic catalog, different values for t_R yield different detection probabilities. In addition to the t_R value of 5 minutes, which yields a realistic fraction of thrown out events and is marked with a star, we test t_R values of 1, 1.97, 10, 30, 60, 180, 360, and 720 minutes (crosses), leading us to throw out between 2.88% and 39.23% of events. Note that that value of 1.97 minutes for t_R corresponds to the value inferred from the Californian catalog. The estimated number of undetected events is reasonably accurate when the percentage of undetected events is small. For large portions of undetected events above 30%, the number of undetected events is increasingly underestimated, such that only 4521.06 undetected events instead of 7365 are estimated for a t_R of 720 minutes. Such high values of t_R , which correspond to very high fractions of undetected $M \geq 2.5$ events in California, are probably unrealistic.

426

4.2.2 Inverted parameters

427

428

429

430

431

Table 1 shows the ETAS parameters and (t_R, β) that were used in the simulation of the synthetic catalog (Column 3). The parameters inverted from the synthetic test catalog are shown in Column 4. With the exception of t_R and τ , the inverted parameters correspond well to the parameters used in the simulation. A possible explanation for the underestimation of τ is the finiteness of the catalog time window.

432

433

434

435

436

The hypothesis that a finite time horizon causes underestimation of τ is further supported by the observation that in the case of the long-term synthetic catalog since 1932, τ is only slightly underestimated, and in fact estimates for τ show a modest increase with increasing duration of catalog (see Figure S3). Such a bias is expected also when parameters are inverted with the usual method.

437

438

439

Besides τ , the recovery time t_R of the network is clearly underestimated, which might explain the overly conservative estimate of undetected events. Despite this underestimation, we are able to infer a major part of undetected events.

440

5 Application to California

441

5.1 Inverted Parameters

442

443

444

445

446

447

Table 2 contains ETAS parameters, β , and, if applicable, t_R estimates obtained when applying different inversion algorithms to Californian data. Additionally, the resulting values for productivity exponent $\alpha = a - \rho\gamma$ and branching ratio η (see Equation S1) are provided. The first column shows the results of applying the usual inversion method as described in Section 3.1 with a constant completeness magnitude of $m_c \equiv 3.1$ to the main catalog (1970 to 2019). The second column shows the parameters inverted when

448 time-varying completeness (Equation 20) is accounted for and thus historical data from
 449 1932 to 2019 can be used with a reference magnitude of $m_{ref} = 2.4$.

450 The results of applying PETAI inversion to the main catalog (1970 to 2019) with
 451 a reference magnitude of $m_{ref} = 2.5$ is given in Column four. Note that the estima-
 452 tion of β is independent of the ETAS parameter estimates for the first two applications,
 453 but not so in the case of PETAI inversion.

454 For comments on the inverted parameters, see Text S4, and for comments on com-
 455 putational time required for the two described inversion methods, see Text S5.

456 5.2 Incompleteness Insights Through PETAI

457 In addition to a new set of estimated ETAS parameters, applying the PETAI in-
 458 version to the Californian catalog produces further interesting outputs. Similarly to the
 459 case of the synthetic catalog, Figure 3 (a) shows the estimated cumulative number of un-
 460 detected events over time. As expected, the increase is predominantly step-wise, caused
 461 by short, incomplete periods during aftershock sequences, and long, complete periods
 462 in-between. While the total expected number of undetected events is at 5041.74, the ex-
 463 trapolated number obtained from a fitted GR law is only 88.91. Although the true num-
 464 ber of undetected events can never be known, the synthetic test suggests that the PETAI
 465 result is more reliable, and thus the GR law extrapolation would be a severe underest-
 466 mation of the true number of undetected events.

467 The magnitude-dependent detection probability evolution is illustrated in Figure
 468 3 (b). Magnitudes with detection probabilities of up to 99.9%, 99%, 90% and 50% are
 469 shown in light blue, dark blue, black, and yellow, respectively. The white area corresponds
 470 to detection probabilities higher than 99.9%. In around 84% of event times t_i , events of
 471 magnitude $M \geq 4$ are expected to be detected with a probability of 99.9% or more. Sim-
 472 ilarly, in 82% of event times t_i , $M \geq 3$ events are expected to be detected with a prob-
 473 ability of 99% or more. Spikes of incompleteness during large sequences lead to detec-
 474 tion probabilities of less than 50% for smaller events, in the most extreme case for events
 475 of magnitude $M \leq 3.47$.

476 As expected, periods of elevated incompleteness coincide with the periods of rapid
 477 increase in undetected events shown in (a). The last step in (a), which corresponds to
 478 the 2019 Ridgecrest sequence, is extraordinarily large compared to all previous steps. This
 479 is most likely explained by the fact that the sequence was better recorded than compa-
 480 rable sequences in previous years. When detection the capability of the seismic network
 481 improves, this would lead to a shorter recovery time t_R . Because we have assumed t_R
 482 to be stationary for simplicity, a larger number of recorded events will lead to a smaller
 483 estimated detection probability, which in turn leads to larger numbers of expected un-
 484 detected events. In future versions of the model, to avoid such artifacts, it would be ad-
 485 visable to combine the possibility of including long-term changes in completeness (as in
 486 the model described in Section 3.1) with rate-dependent aftershock incompleteness by
 487 means of a non-stationary t_R .

488 Figure 3 (c) - (g) shows excerpts of Figure 3 (b) for the 1989 M6.9 Loma Prieta,
 489 the 1992 M6.1 Joshua Tree and 7.3 Landers, the 1994 M6.7 Northridge, the 1999 M7.1
 490 Hector Mine, and the 2019 M6.4 and 7.1 Ridgecrest events. The x -axes are logarithmic
 491 with reference point one day prior to the (first) mainshock, and range from 2 hours be-
 492 fore until 30 days after that mainshock (60 days for (g)). x -axis tick labels represent time
 493 since the (first) mainshock.

494 Figure 3 (h) visualizes the range of observed states of detection efficiency during
 495 the 2019 Ridgecrest sequence. Current rate of events $\lambda(t_i)$ at all occurrence times t_i of
 496 events in the catalog are estimated as part of the PETAI inversion algorithm, and each

497 λ_i corresponds to a certain shape of the detection probability function $f(m, \lambda(t_i))$. Fig-
 498 ure 3 (h) shows the detection probability as a function of magnitude when the rate λ_i
 499 is fixed, for the selection of $\lambda_i = \lambda(t_i)$ at times t_i that that are highlighted in Figure
 500 3 (g). Note that time differences in Figure 3 (h) refer to the M7.1 mainshock. Prior to
 501 the large events, detection is almost perfect for all magnitudes. After the M6.4 event,
 502 detection is weakened and recovers with time, until the M7.1 mainshock, when it is again
 503 weakened. Around 15 minutes after the earthquake, events of magnitude below 3.0 still
 504 have almost no chance to be detected, with M3.5 events having roughly a 50% chance
 505 to be detected. After three hours, detection has already clearly improved, although M2.5
 506 events are still almost surely not detected. After six days, the detection probability func-
 507 tion only slightly differs from the detection probability before the main events happened.

508 6 Pseudo-prospective forecasting experiments

509 To better understand if and how the PETAI model can improve earthquake fore-
 510 casts, we conduct a pseudo-prospective forecasting experiment, designed to answer sev-
 511 eral questions: Does the PETAI model outperform the current state of the art? If so,
 512 what is the role of the newly estimated ETAS parameters in this improvement? Simi-
 513 larly, what is the role of newly included small earthquakes in this improvement? How
 514 do the models perform for different target magnitude thresholds?

515 6.1 Competing models

516 We compare four models.

- 517 1. The base ETAS model assumes perfect detection above a constant $m_c = 3.1$ and
 518 represents the current state of the art.
- 519 2. PETAI, the alternative model, has two modifications to the base model. Firstly,
 520 it uses improved ETAS parameter estimates that were obtained in the PETAI in-
 521 version with a reference magnitude m_{ref} of 2.5. Secondly, magnitude $M \geq 2.5$
 522 earthquakes are allowed to trigger and be triggered. For this, the events in the train-
 523 ing catalog, which act as triggering earthquakes in the simulation, have their trig-
 524 gering capability inflated by $1 + \xi(t)$, as estimated in the PETAI inversion.

525 Two intermediate models are assessed to dissect the effect of the two modifications.

- 526 3. `par_only` uses improved parameter estimates, but only models $M \geq 3.1$ events
 527 assuming perfect detection there (i.e. $\xi(t) \equiv 0$). In this case, the parameters ob-
 528 tained for the PETAI model have to be transformed to be compatible with a refer-
 529 ence magnitude of $m_{ref} = 3.1$ as described in Text S6.
- 530 4. Vice-versa, `trig_only` models $M \geq 2.5$ events using the inverted $\xi(t)$ for inflated
 531 triggering, but does not use the improved ETAS parameter estimates. In this case,
 532 the parameters obtained for the base model have to be transformed to be com-
 533 patible with a reference magnitude of $m_{ref} = 2.5$ as described in Text S6.

534 6.2 Experiment setup

535 For a testing period length of 30 days, we define a family of training and testing
 536 periods such that the testing periods are consecutive and non-overlapping. Each train-
 537 ing period ends with the starting date of its corresponding testing period. The starting
 538 date of the first testing period is January 1st, 2000.

539 For each testing period, all competing models are trained based on the correspond-
 540 ing training data. Figure S3 shows the parameter evolution with increasing training pe-
 541 riod obtained with standard ETAS and PETAI inversion. Then, forecasts are issued with

each model through simulation of 100'000 possible continuations of the training catalog. This is done by simulating Type I earthquakes (the cascade of aftershocks of earthquakes in the training catalog) and Type II earthquakes (simulated background earthquakes and their cascade of aftershocks) similarly to how it is described by Nandan et al. (2019b). The algorithm used is described in detail in Text S3.

The performance of each model is evaluated by calculating the log-likelihood of the testing data given the forecast. Two competing models can be compared by calculating the information gain (IG) of the alternative model M_{alt} over the null model M_0 , which is simply the difference in log-likelihood of observing the testing data. The mean information gain (MIG) is calculated as the mean over all testing periods.

As a benchmark, we additionally calculate the IG of the null model versus the flat model, which gives a spatially and temporally invariant forecast.

For details on the flat model, on the calculation of the log-likelihood and on the conditions under which one model is considered superior over another, see Text S7 and Nandan et al. (2019b).

6.3 Results

Figure 4 shows the results of the pseudo-prospective forecasting experiments. From top to bottom, the panels correspond to increasing target magnitude thresholds. The left column of panels shows the cumulative information gain of the three alternative models compared to the null model, which is indicated as a black horizontal line. The number of events for all testing periods combined, and the total information gain of the null model versus the flat model, are indicated as text. The column of panels in the middle shows the mean information gain matrix when each model is compared to each other model. Green cells indicate positive information gain, while pink cells correspond to negative information gain. The significance level of potential outperformance is shown in the right column of panels in a matrices similar to the middle panels. Cells highlighted in green indicate a p -value of less than 0.05 and thus, significant outperformance. For p -values between 0.05 and 1, cell color ranges from dark grey to white.

For a target magnitude threshold of $m_t = 3.1$, PETAI as well as trig_only significantly outperform the other two models with p -values of virtually 0 and a mean information gain of 0.97 and 0.94, respectively. Note that this improvement is over a very strong null model, which has a total information gain of 49'246 (i.e. a MIG of 202.66) over the flat model. PETAI has a slightly positive but not statistically significant information gain compared to trig_only. On the other hand, par_only does not outperform the null model. This suggests that the main driver of the improvement of the forecast is the inclusion of small events between $M_{2.5}$ and $M_{3.1}$ in the simulations, which is possible due to the estimated $\xi(t_i)$ obtained in the PETAI inversion.

With increasing values for m_t , the mean information gain values generally decrease, and almost no model significantly outperforms any other model. Occasionally, par_only is outperformed by the base model or by trig_only. Both of these observations suggest that taking into account information about smaller earthquakes mainly helps forecasting the like. ETAS parameters calibrated using large amounts of small earthquakes will reflect more precisely the behaviour of small earthquakes and are unnecessary, if not disadvantageous, when describing large ones. Moreover, simulating aftershocks of small earthquakes is the key ingredient for improved forecasting of similarly-sized events. Although theoretically, small earthquakes can trigger large ones, and their relative numerosity implies significant contribution the overall triggering (Marsan, 2005; Helmstetter et al., 2005; Sornette and Werner, 2005a), we find that the beneficial effect vanishes when forecasting large events. Helmstetter et al. (2006) compared the probability gain of their time-dependent model versus their similar but time-independent model and found that prob-

ability gain decreases with an increasing target magnitude threshold. They speculated that this observation may be due to a smaller sample size when target magnitude threshold increases. In our case, this effect is observed at considerably large sample sizes of 3601, 1111, 307, and 85 events for $m_t = 3.5, 4.0, 4.5,$ and 5.0 . Another possible explanation for this effect is provided by the findings of Nandan et al. (2019a) and Nandan et al. (2021), that earthquakes tend to preferentially trigger aftershocks of similar size and that accounting for this preference in the ETAS model produces superior forecasts. Their results explain the improved forecast of small events when small events are used for simulation, as well as the vanishing of this improvement when the magnitude difference between newly included events and target events becomes large. This could furthermore serve as an alternative explanation of the results of Helmstetter et al. (2006).

7 Conclusion

We propose a modified algorithm for the inversion of ETAS parameters when m_c varies with time, and an algorithm for the joint inversion of ETAS parameters and probabilistic, epidemic-type aftershock incompleteness. We test both methods on synthetic catalogs, concluding that they are able to accurately invert the parameters used for simulation of the synthetics. The given formulations are rather general and can equally be applied to spatial or spatiotemporal variations of m_c , as well as to any suitable definition of a detection probability function.

Two potential use cases are the estimation of ETAS parameters based on the Californian catalog since 1932 with long-term fluctuations of m_c between 4.3 and 2.4, and the estimation of ETAS parameters and short-term aftershock incompleteness based on the incomplete Californian catalog of events above $M2.5$. The latter is further used to test the forecasting power of small earthquakes. Results of numerous pseudo-prospective forecasting experiments suggest that

- Information about small earthquakes significantly and substantially improves forecasts of similar-sized events.
- Main driver of this improvement is the simulation of aftershocks of small events, made possible thanks to high-frequency estimates of incompleteness.
- Information about small earthquakes does not significantly affect the performance of large event forecasts.

This supports previous findings of Nandan et al. (2019a) and Nandan et al. (2021), that earthquakes preferentially trigger aftershocks of similar size.

Our results have potentially major implications for the future of earthquake forecasting. Thanks to the here-presented algorithms, ETAS models may be calibrated for regions with low seismicity where the usual inversion algorithms would fail due to missing data.

Furthermore, the newly gained insights from forecasting experiments guide us in the search of the next generation earthquake forecasting models. Besides other discussed topics such as anisotropy, temporally or spatially non-stationary background rate (Hainzl et al., 2008; Hainzl et al., 2013; Nandan et al., 2020), the importance of a magnitude-dependent distribution of aftershock magnitudes is emphasized.

Acknowledgments

The data used for this analysis is available through the website <https://earthquake.usgs.gov/earthquakes/search/> (U.S. Geological Survey, Earthquake Hazards Program, 2017). The authors wish to thank Sebastian Hainzl, Andrew Michael and Andrea Llenos for insightful discussions and helpful feedback on earlier versions of

639 this article. This work has received funding from the Eidgenössische Technische Hochschule
 640 (ETH) research grant for project number 2018-FE-213, “Enabling dynamic earthquake
 641 risk assessment (DynaRisk)” and from the European Union’s Horizon 2020 research and
 642 innovation program under Grant Agreement Number 821115, real-time earthquake risk
 643 reduction for a resilient Europe (RISE).

644 References

- 645 Amato, A., & Mele, F. (2008). Performance of the ingv national seismic network from
 646 1997 to 2007. *Annals of Geophysics*.
- 647 Amorese, D. (2007). Applying a change-point detection method on frequency-magnitude
 648 distributions. *Bulletin of the Seismological Society of America*, *97*(5), 1742–1749.
- 649 Cao, A., & Gao, S. S. (2002). Temporal variation of seismic b-values beneath northeast-
 650 ern japan island arc. *Geophysical research letters*, *29*(9), 48–1.
- 651 Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empir-
 652 ical data. *SIAM review*, *51*(4), 661–703.
- 653 De Arcangelis, L., Godano, C., & Lippiello, E. (2018). The overlap of aftershock coda
 654 waves and short-term postseismic forecasting. *Journal of Geophysical Research:
 655 Solid Earth*, *123*(7), 5661–5674.
- 656 Felzer, K. R. (2007). Appendix i: Calculating california seismicity rates. *US Geol. Surv.
 657 Open-File Rept. 2007-1437I*.
- 658 Gerstenberger, M. C., Marzocchi, W., Allen, T., Pagani, M., Adams, J., Danciu, L., Field,
 659 E. H., Fujiwara, H., Luco, N., Ma, K.-F., et al. (2020). Probabilistic seismic haz-
 660 ard analysis at regional and national scales: State of the art and future challenges.
 661 *Reviews of Geophysics*, *58*(2), e2019RG000653.
- 662 Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin
 663 of the Seismological Society of America*, *34*(4), 185–188.
- 664 Hainzl, S. (2016a). Apparent triggering function of aftershocks resulting from rate-dependent
 665 incompleteness of earthquake catalogs. *Journal of Geophysical Research: Solid
 666 Earth*, *121*(9), 6499–6509.
- 667 Hainzl, S. (2016b). Rate-dependent incompleteness of earthquake catalogs. *Seismolog-
 668 ical Research Letters*, *87*(2A), 337–344.
- 669 Hainzl, S., Christophersen, A., & Enescu, B. (2008). Impact of earthquake rupture ex-
 670 tensions on parameter estimations of point-process models. *Bulletin of the Seis-
 671 mological Society of America*, *98*(4), 2066–2072.
- 672 Hainzl, S., Zakharova, O., & Marsan, D. (2013). Impact of aseismic transients on the es-
 673 timation of aftershock productivity parameters. *Bulletin of the Seismological So-
 674 ciety of America*, *103*(3), 1723–1732.
- 675 Helmstetter, A. (2003). Is earthquake triggering driven by small earthquakes? *Physical
 676 review letters*, *91*(5), 058501.
- 677 Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2006). Comparison of short-term and
 678 time-independent earthquake forecast models for southern california. *Bulletin
 679 of the Seismological Society of America*, *96*(1), 90–106.
- 680 Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2007). High-resolution time-independent
 681 grid-based forecast for $m \geq 5$ earthquakes in california. *Seismological Research
 682 Letters*, *78*(1), 78–86.
- 683 Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2005). Importance of small earthquakes
 684 for stress transfers and earthquake triggering. *Journal of Geophysical Research:
 685 Solid Earth*, *110*(B5).
- 686 Hutton, K., Woessner, J., & Hauksson, E. (2010). Earthquake monitoring in southern
 687 california for seventy-seven years (1932–2008). *Bulletin of the Seismological So-
 688 ciety of America*, *100*(2), 423–446.
- 689 Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earth-
 690 quake source. *Bulletin of the Seismological Society of America*, *94*(4), 1207–1228.

- 691 Llenos, A. L., & Michael, A. J. (2020). Regionally optimized background earthquake rates
692 from etas (robere) for probabilistic seismic hazard assessment. *Bulletin of the*
693 *Seismological Society of America*, *110*(3), 1172–1190.
- 694 Marsan, D. (2005). The role of small earthquakes in redistributing crustal elastic stress.
695 *Geophysical Journal International*, *163*(1), 141–151.
- 696 Mignan, A., Werner, M., Wiemer, S., Chen, C.-C., & Wu, Y.-M. (2011). Bayesian esti-
697 mation of the spatially varying completeness magnitude of earthquake catalogs.
698 *Bulletin of the Seismological Society of America*, *101*(3), 1371–1385.
- 699 Mignan, A., & Chouliaras, G. (2014). Fifty years of seismic network performance in greece
700 (1964–2013): Spatiotemporal evolution of the completeness magnitude. *Seismo-*
701 *logical Research Letters*, *85*(3), 657–667.
- 702 Mignan, A., & Woessner, J. (2012). Estimating the magnitude of completeness for earth-
703 quake catalogs. *Community Online Resource for Statistical Seismicity Analy-*
704 *sis*, 1–45.
- 705 Mizrahi, L., Nandan, S., & Wiemer, S. (2021). The Effect of Declustering on the Size Dis-
706 tribution of Mainshocks. *Seismological Research Letters*. [https://doi.org/10.](https://doi.org/10.1785/0220200231)
707 [1785/0220200231](https://doi.org/10.1785/0220200231)
- 708 Nandan, S., Ouillon, G., & Sornette, D. (2019a). Magnitude of earthquakes controls the
709 size distribution of their triggered events. *Journal of Geophysical Research: Solid*
710 *Earth*, *124*(3), 2762–2780.
- 711 Nandan, S., Ouillon, G., & Sornette, D. (2021). Triggering of large earthquakes is driven
712 by their twins. *arXiv preprint arXiv:2104.04592*.
- 713 Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019b). Forecasting the full dis-
714 tribution of earthquake numbers is fair, robust, and better. *Seismological Re-*
715 *search Letters*, *90*(4), 1650–1659.
- 716 Nandan, S., Ouillon, G., Sornette, D., & Wiemer, S. (2019c). Forecasting the rates of fu-
717 ture aftershocks of all generations is essential to develop better earthquake fore-
718 cast models. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8404–8425.
- 719 Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017). Objective estimation of spa-
720 tially variable parameters of epidemic type aftershock sequence model: Appli-
721 cation to california. *Journal of Geophysical Research: Solid Earth*, *122*(7), 5118–
722 5143.
- 723 Nandan, S., Ram, S. K., Ouillon, G., & Sornette, D. (2020). Is the earth crust operat-
724 ing at a critical point? *arXiv preprint arXiv:2012.06013*.
- 725 Nanjo, K., Ishibe, T., Tsuruoka, H., Schorlemmer, D., Ishigaki, Y., & Hirata, N. (2010).
726 Analysis of the completeness magnitude and seismic network coverage of japan.
727 *Bulletin of the Seismological Society of America*, *100*(6), 3261–3268.
- 728 Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals*
729 *of the Institute of Statistical Mathematics*, *50*(2), 379–402.
- 730 Omi, T., Ogata, Y., Hirata, Y., & Aihara, K. (2014). Estimating the etas model from
731 an early aftershock sequence. *Geophysical Research Letters*, *41*(3), 850–857.
- 732 Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Rukstales, K. S., Mc-
733 Namara, D. E., Williams, R. A., Shumway, A. M., Powers, P. M., Earle, P. S.,
734 et al. (2018). 2018 one-year seismic hazard forecast for the central and eastern
735 united states from induced and natural earthquakes. *Seismological Research Let-*
736 *ters*, *89*(3), 1049–1061.
- 737 Rydelek, P. A., & Sacks, I. S. (1989). Testing the completeness of earthquake catalogues
738 and the hypothesis of self-similarity. *Nature*, *337*(6204), 251–253.
- 739 Schoenberg, F. P. (2013). Facilitated estimation of etas. *Bulletin of the Seismological So-*
740 *ciety of America*, *103*(1), 601–605.
- 741 Schoenberg, F. P., Chu, A., & Veen, A. (2010). On the relationship between lower mag-
742 nitude thresholds and bias in epidemic-type aftershock sequence parameter esti-
743 mates. *Journal of Geophysical Research: Solid Earth*, *115*(B4).
- 744 Schorlemmer, D., & Gerstenberger, M. (2007). Relm testing center. *Seismological Re-*
745 *search Letters*, *78*(1), 30–36.

- 746 Schorlemmer, D., & Woessner, J. (2008). Probability of detecting an earthquake. *Bulletin of the Seismological Society of America*, *98*(5), 2103–2117.
- 747
- 748 Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating etas: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449–469.
- 749
- 750
- 751 Sornette, D., & Werner, M. J. (2005a). Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, bath’s law, and observed aftershock sequences. *Journal of Geophysical Research: Solid Earth*, *110*(B8).
- 752
- 753
- 754
- 755 Sornette, D., & Werner, M. J. (2005b). Apparent clustering and apparent background earthquakes biased by undetected seismicity. *Journal of Geophysical Research: Solid Earth*, *110*(B9).
- 756
- 757
- 758 Stallone, A., & Falcone, G. (2020). Missing earthquake data reconstruction in the space-time-magnitude domain.
- 759
- 760 Strader, A., Schneider, M., & Schorlemmer, D. (2017). Prospective and retrospective evaluation of five-year earthquake forecast models for california. *Geophysical Journal International*, *211*(1), 239–251.
- 761
- 762
- 763 U.S. Geological Survey, Earthquake Hazards Program. (2017). Advanced national seismic system (anss) comprehensive catalog of earthquake events and products: Various. <https://doi.org/10.5066/F7MS3QZH>
- 764
- 765
- 766 van der Elst, N. J. (2021). B-positive: A robust estimator of aftershock magnitude distribution in transiently incomplete catalogs. *Journal of Geophysical Research: Solid Earth*, e2020JB021027.
- 767
- 768
- 769 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.
- 770
- 771
- 772 Wang, Q., Jackson, D. D., & Zhuang, J. (2010). Missing links in earthquake clustering models. *Geophysical Research Letters*, *37*(21).
- 773
- 774 Wiemer, S., Giardini, D., Fäh, D., Deichmann, N., & Sellami, S. (2009). Probabilistic seismic hazard assessment of switzerland: Best estimates and uncertainties. *Journal of Seismology*, *13*(4), 449.
- 775
- 776
- 777 Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan. *Bulletin of the Seismological Society of America*, *90*(4), 859–869.
- 778
- 779
- 780 Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, *95*(2), 684–698.
- 781
- 782
- 783 Zhuang, J., Ogata, Y., & Wang, T. (2017). Data completeness of the kumamoto earthquake sequence in the jma catalog and its influence on the estimation of the etas parameters. *Earth, Planets and Space*, *69*(1), 36.
- 784
- 785

8 Tables

Table 1: ETAS and PETAI parameters used and inferred in synthetic tests.

parameter	$m_c(t), m_{ref} = 2.4$		$f(m, t), m_{ref} = 2.5$	
	simulation	estimated	simulation	estimated
$\log_{10}(\mu)$	-7.20	-7.07	-8.00	-7.99
$\log_{10}(k_0)$	-3.26	-3.25	-3.24	-3.23
a	2.90	2.86	2.82	2.95
$\log_{10}(c)$	-2.52	-2.49	-2.85	-2.72
ω	-0.02	-0.02	-0.06	-0.11
$\log_{10}(\tau)$	3.57	3.58	3.92	3.62
$\log_{10}(d)$	-0.86	-0.81	-0.76	-0.78
γ	1.35	1.36	1.22	1.31
ρ	0.67	0.71	0.67	0.75
$\log_{10}(t_R)$	n/a	n/a	-2.46	-3.44
β	n/a	n/a	2.37	2.34

Table 2: ETAS and PETAI parameters inferred for California. First column shows parameters when constant m_c of 3.1 is assumed. Second and third column show parameters when time-varying m_c is accounted for, and fourth and fifth column show parameters when PETAI inversion is applied. Note that the originally derived parameters are given in Columns 1, 2, and 4. Columns 3 and 5 show the parameters of Columns 2 and 4, transformed (as described in Text S6) to a reference magnitude of 3.1 to allow comparison with Column 1. Productivity exponent $\alpha = a - \rho\gamma$ and branching ratio η are not directly inverted but inferred from the inverted parameters.

parameter	$m_c \equiv \text{const.}$	$m_c(t)$		$f(m, t)$	
m_{ref}	3.1	2.4	3.1	2.5	3.1
$\log_{10}(\mu)$	-6.86	-5.97	-6.68	-6.35	-6.97
$\log_{10}(k_0)$	-2.53	-2.63	-2.36	-2.70	-2.49
a	1.74	1.86	1.86	1.92	1.92
$\log_{10}(c)$	-2.97	-2.52	-2.52	-2.85	-2.85
ω	-0.05	-0.02	-0.02	-0.06	-0.06
$\log_{10}(\tau)$	4.03	3.57	3.57	3.92	3.92
$\log_{10}(d)$	-0.51	-0.86	-0.45	-0.76	-0.45
γ	1.19	1.35	1.35	1.22	1.22
ρ	0.60	0.67	0.67	0.67	0.67
$\log_{10}(t_R)$	n/a	n/a	n/a	-2.86	-2.86
β	2.33	2.32	2.32	2.37	2.37
$a - \rho\gamma$	1.03	0.95	0.95	1.09	1.09
η	0.94	0.95	0.95	0.93	0.93

9 Figures

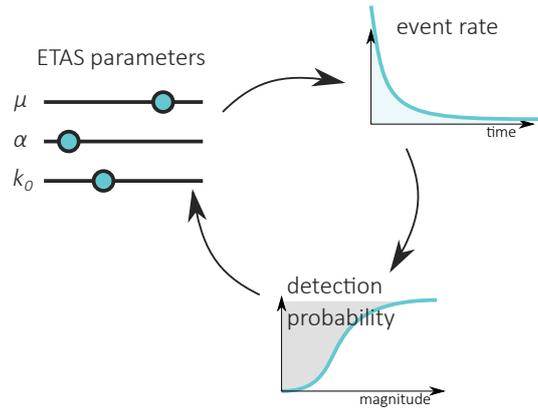


Figure 1: Simplified schematic illustration of PETAI inversion.

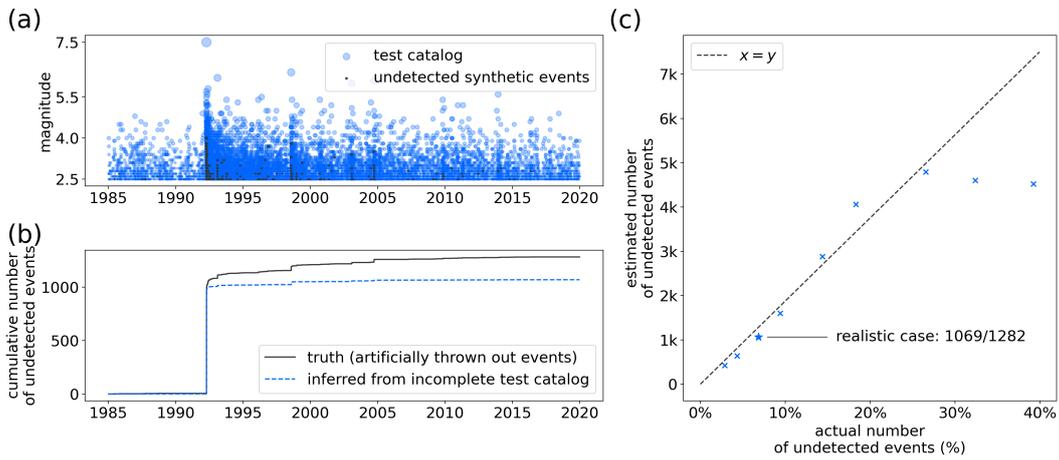


Figure 2: (a) Earthquake magnitudes over time for the test catalog (blue). Events marked in black were simulated, but declared as undetected. (b) Cumulative number of unobserved events over time. Black line marks the truth, blue line is inferred from the test catalog using PETAI. (c) Estimated number of undetected events versus actually thrown out events, for different assumed detection efficiencies. Star marks the realistic test catalog used in (a) and (b), crosses represent the same original catalog with different values for t_R used to sample out undetected events.

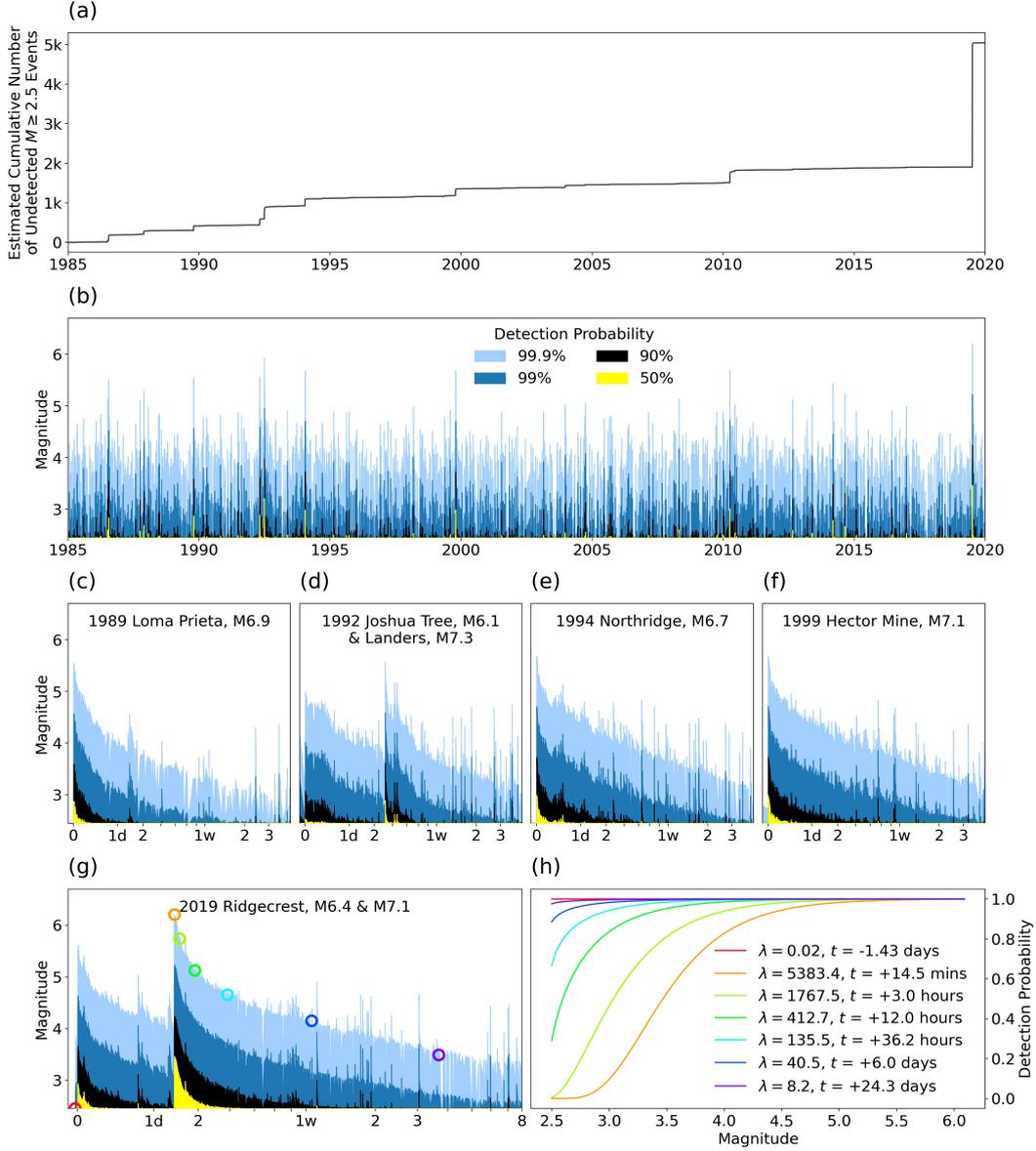


Figure 3: Aftershock Incompleteness in California. (a) Estimated cumulative number of undetected events over time. (b) Evolution of magnitude-dependent detection probability. Yellow indicates a detection probability of 50% or less. Black, dark blue, and light blue indicate detection probabilities of up to 90%, 99%, and 99.9%, respectively. White area represents detection probabilities higher than 99.9%. (c)-(g) Excerpts of (b) for selected large events. x -axes are logarithmic with reference point one day prior to (first) mainshock, and range from 2 hours before to 30 days after that mainshock (60 days for (g)). x -axis tick labels represent time since the (first) mainshock. Colored circles in (g) represent selected times t_i and corresponding magnitude of 99.9% detection. (h) Detection probability function $f(m, \lambda = \lambda(t_i))$ snapshots for the times that are highlighted in (g). Time deltas are given with respect to the M7.1 mainshock.

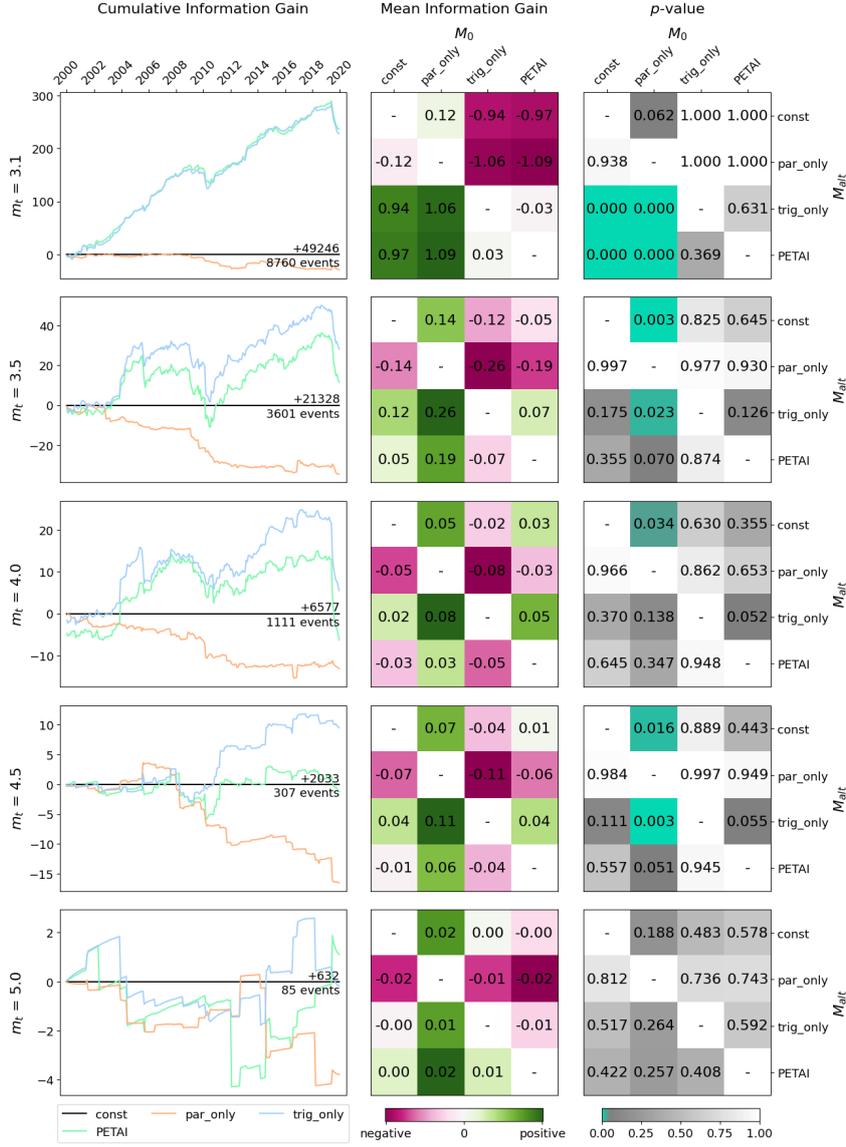


Figure 4: From left to right: Cumulative information gain for the alternative models versus the null model, mean information gain matrix, and corresponding p -value matrix comparing all competing models. Matrix entries represent the test of superiority of M_{alt} (y -axis) versus M_0 (x -axis). From top to bottom: target magnitude thresholds m_t of 3.1, 3.5, 4.0, 4.5, and 5.0. Indicated as text in the left panels is the cumulative information gain of the null model versus the flat model, and the number of events in all testing periods combined. Note the different y -axes for the left panels. Also note that the color scheme for the middle panels is different between threshold magnitudes m_t and normalized with respect to the maximum absolute mean information gain for that m_t . Color coding for the panels on the right is such that p -values of 0.05 and below are green, and transition from grey to white between 0.05 and 1.

Supporting Information for “Embracing Data Incompleteness for Better Earthquake Forecasting”

L. Mizrahi¹, S. Nandan¹, and S. Wiemer¹

¹Swiss Seismological Service, ETH Zurich

Contents of this file

- **Text S1** Branching ratio (η) and corrected productivity exponent (α)
- **Text S2** Estimation of β and t_R , given ETAS parameters and current rates
- **Text S3** Catalog simulation
- **Text S4** Comments on parameters inverted from the Californian catalog
- **Text S5** Comments on computational time
- **Text S6** Parameter transformation for reference magnitude changes
- **Text S7** Forecast evaluation

- **Figure S1** Log likelihood of observing the test data for different values of t_R and b -value, when current rate is known
- **Figure S2** Flow diagram of PETAI inversion
- **Figure S3** Evolution of ETAS and PETAI parameter estimates with increasing training catalog

Text S1 Branching ratio (η) and corrected productivity exponent (α)

Branching ratio (η)

The branching ratio η is defined as the expected number of direct aftershocks (larger than m_{ref}) of any earthquake larger than m_{ref} ,

$$\eta = \int_{m_{ref}}^{\infty} f_{GR}(m) \cdot G(m) dm, \quad (\text{S1})$$

where $f_{GR} = \beta \cdot e^{-\beta \cdot (m - m_{ref})}$ is the probability density function of magnitudes according to the GR law, and $G(m) = \int_0^{\infty} \iint_R g(m, t, x, y) dx dy dt$ is the total number of expected aftershocks of an event of magnitude m . We make the simplifying assumption that the considered region R extends infinitely in all directions, allowing a facilitated, asymptotically unbiased estimation of ETAS parameters (Schoenberg, 2013). It follows easily that

$$\eta = \frac{\beta \cdot k_0 \cdot \pi \cdot d^{-\rho} \cdot \tau^{-\omega} \cdot e^{c/\tau} \cdot \Gamma(-\omega, c/\tau)}{\rho \cdot (\beta - (a - \rho\gamma))}, \quad (\text{S2})$$

if $\beta > a - \rho \cdot \gamma$, where $\Gamma(s, x) = \int_x^{\infty} t^{s-1} e^{-t} dt$ is the upper incomplete gamma function.

Corrected productivity exponent (α)

Due to the magnitude dependency of the spatial triggering kernel, our parameter a does not strictly correspond to a as described in Veen and Schoenberg (2008). For comparability, we define the productivity exponent $\alpha := a - \rho \cdot \gamma$. We fix $\alpha' = 2.0$ and from this derive new values for a and k_0 , keeping the branching ratio η constant. In particular, we define

$$a' := \alpha' + \rho \cdot \gamma, \quad (\text{S3})$$

$$k_0' := k_0 \cdot \frac{\beta - (a' - \rho \cdot \gamma)}{\beta - (a - \rho \cdot \gamma)}. \quad (\text{S4})$$

It can be easily shown that in this way, the branching ratio η remains the same as long as $\beta - (a - \rho \cdot \gamma) < 0$. Note that the condition $\beta < \alpha$ for Equation S2, where α is the productivity exponent, is generally fulfilled in naturally observed catalogs.

Text S2 Estimation of β and t_R , given ETAS parameters and current rates

In the case when the true ETAS parameters, as well as the current event rates $\lambda(t_i)$ for all events e_i in the primary catalog $\{e_1, \dots, e_n\}$, are known, the GR-law exponent β and the network recovery time t_R can be estimated by optimizing the log-likelihood \mathcal{LL} of observing the catalog at hand.

$$\mathcal{LL} = \sum_{i=1}^n (\ln(\nu_i + 1) - \ln N) \quad (\text{S5})$$

$$+ \sum_{i=1}^n \left(\nu_i \cdot \ln(1 - e^{-\beta \cdot (m_i - m_{ref})}) \right) \quad (\text{S6})$$

$$+ \sum_{i=1}^n (\ln \beta - \beta \cdot (m_i - m_{ref})), \quad (\text{S7})$$

where $N = \sum_{i=1}^n (\nu_i + 1)$, and $\nu_i = t_R \cdot \lambda(t_i)$ is the expected number of events blocking the network at time t_i . The expression for \mathcal{LL} given above is valid in general for alternative exponents ν_i in the definition of detection probability (Equation 16 in the article). \mathcal{LL} is derived from the likelihood \mathcal{L}_i of an event of magnitude m_i to occur and to be observed during a current event rate of $\lambda_i = \lambda(t_i)$,

$$\mathcal{L}_i = f_{emp}(\lambda_i) \cdot f_{GR}(m_i) \cdot f_{det}(m_i, \lambda_i), \quad (\text{S8})$$

where $f_{GR}(m)$ is the probability density function of magnitudes given by the GR law, $f_{det}(m, \lambda)$ is the detection probability as defined in Equation 16 in the article, and

$$f_{emp}(\lambda) = \begin{cases} \frac{t_R \cdot \lambda + 1}{\sum_i (t_R \cdot \lambda_i + 1)}, & \text{if } \lambda \in \{\lambda_1, \dots, \lambda_n\} \\ 0, & \text{otherwise} \end{cases} \quad (\text{S9})$$

is the empirical density function of event rates. $f_{emp}(\lambda)$ is defined such that

$$\sum_{i=1}^n f_{emp}(\lambda_i) = 1 \quad (\text{S10})$$

and

$$f_{emp}(\lambda_i) \propto \frac{1}{\int_{m_{ref}}^{\infty} f_{GR}(m) \cdot f_{det}(m, \lambda_i) dm} = \lambda_i \cdot t_R + 1, \quad \forall i = 1, \dots, n. \quad (\text{S11})$$

Without the latter condition (Equation S11), we would wrongly assume that the values $\lambda(t_i)$ were uniformly drawn from the true distribution of event rates. However, in our sample of λ_i , large values of λ are underrepresented, because during times t when $\lambda(t)$ is high, events are less likely to be detected, and those times and their corresponding rates are thus less likely to be part of our sample. Defining $f_{emp}(\lambda_i)$ to be inversely proportional to the fraction of events that are observed when the current rate is λ_i corrects for this underrepresentation. This yields

$$\mathcal{L}_i = \frac{\nu_i + 1}{\sum_j (\nu_j + 1)} \cdot \beta \cdot e^{-\beta \cdot (m_i - m_{ref})} \cdot \left(1 - e^{-\beta \cdot (m_i - m_{ref})}\right)^{\nu_i}, \quad (\text{S12})$$

which explains the term for \mathcal{LL} given above. Figure S1 shows the log likelihood of the synthetic test catalog for different values of t_R and β when λ_i are known. The crosses highlight that the resulting estimators match the ground truth parameters used in the simulation of the catalog.

Text S3 Catalog simulation

The following algorithm is used to simulate the continuation of a training catalog.

Note that the synthetic catalogs referred to in Sections 4.1 and 4.2 in the article are not continuations of a training catalog, hence generation 0 (defined in the following) consists only of background events. The locations of these background events are uniformly distributed in the study region. Also in the case of synthetic catalog simulation, the “testing period”, which is referred to below, is the period for which one wishes to simulate a catalog. Where different models are mentioned, the base model is used for synthetic catalog simulation.

1. Background events are simulated for the testing period.
 - Number of background events is drawn from a Poisson distribution with mean as given by the ETAS background rate.
 - Occurrence times are drawn from a uniform distribution within the testing period.
 - Locations are drawn from the locations of events in the training catalog, weighted by their probability of being background events. The locations are then randomly displaced by a distance drawn from a normal distribution with mean 0 and standard deviation of 0.1° .
 - Magnitudes are drawn from a GR law with exponent β as estimated in the PETAI inversion (for PETAI and trig_only). For the base model and par_only, we use the β estimate obtained when using the formula proposed by Tinti and Mulargia (1987) for binned magnitude values, using magnitudes $M \geq 3.1$ in the training catalog.
2. The training catalog together with the simulated background events make up generation 0. $i_{gen} := 0$.
3. Expected number of aftershocks is calculated for all events of generation 0. In the case of the PETAI and the trig_only model, the average number of aftershocks triggered by any event e_i in the training catalog is inflated by $1 + \xi(t_i)$.
4. Actual number of aftershocks of each event is randomly drawn from a Poisson distribution with mean as calculated in the previous step.
5. Aftershocks of the current generation i_{gen} are simulated.
 - Aftershock time distance to its parent event is randomly generated according to the estimated ETAS time kernel. If aftershock time falls out of the testing period, this aftershock is discarded.
 - Aftershock spatial distance to its parent event is randomly generated according to the estimated isotropic ETAS spatial kernel. If aftershock location falls out of the considered polygon, this aftershock is discarded.
 - Aftershock magnitude is generated according to the GR law with exponent β (same as for the background events).
6. The newly generated aftershocks now make up the next generation $i_{gen} + 1$.
7. We move on to the next generation. $i_{gen} := i_{gen} + 1$
8. Expected number of aftershocks is calculated for all events of generation i_{gen} . Continue with step 4.

The algorithm terminates when no aftershocks fall into the testing period anymore, which is expected to happen in a finite amount of time if the branching ratio $\eta < 1$.

Text S4 Comments on parameters inverted from the Californian catalog

Table 2 in the article contains ETAS parameters, β , and, if applicable, t_R estimates obtained when applying different inversion algorithms to Californian data. Additionally, the resulting values for productivity exponent $\alpha = a - \rho\gamma$ and branching ratio η (see Equation S2) are provided. The first column shows the results of applying the usual inversion method as described in Section 3.1 in the article, with a constant completeness magnitude of $m_c \equiv 3.1$ to the main catalog (1970 to 2019). The second column shows the parameters inverted when time-varying completeness (Equation 20 in the article) is accounted for and thus historical data from 1932 to 2019 can be used with a reference magnitude of $m_{ref} = 2.4$.

The results of applying PETAI inversion to the main catalog (1970 to 2019) with a reference magnitude of $m_{ref} = 2.5$ is given in Column four. Note that the estimation of β is independent of the ETAS parameter estimates for the first two applications, but not so in the case of PETAI inversion.

To allow a better comparison between parameters inverted using different methods when m_{ref} varies, the third and fifth columns show the parameters of Columns two and four, respectively, after having been translated to a value of $m_{ref} = 3.1$ as described in Section Text S6.

Interestingly, the estimate of τ clearly decreases with an increase of the time horizon of the catalog, although usually in this case one would expect an increase of τ . The less pronounced decrease of τ in case of PETAI inversion speaks against the possibility that the decrease is caused by inclusion of lower magnitude earthquakes revealing previously unseen earthquake interactions. This indicates that the lower value of τ may actually better reflect the long-term behavior of earthquake interaction.

Another counter-intuitive observation is the increase of c for both new inversion techniques, in particular for PETAI inversion. The parameter c has been interpreted to reflect aftershock incompleteness (Kagan, 2004; Lolli and Gasperini, 2006; Hainzl, 2016) and would thus be expected to decrease when this effect is accounted for by the model (Seif et al., 2017). The observed higher value of c even after accounting for STAI thus requires a different interpretation of c . Narteau et al. (2009) have brought the parameter in relation with differential stress and the intensity of stress re-distribution. Another possible interpretation provided by Lippiello et al. (2007) is based on the dynamical scaling hypothesis in which time differences relate to magnitude differences. The dependence of c of the cutoff magnitude as proposed by Shcherbakov et al. (2004) can qualitatively explain our observations: The value inverted for c is highest in the case of $m_{ref} = 2.4$, and lowest for $m_{ref} = 3.1$. Note that such a dependency is not accounted for in our model, and thus the values in Columns two and three, and four and five of Table 2 in the article, respectively, are identical. Overall, one should be careful to not over-interpret this estimate of c . After all, c is overestimated in the PETAI synthetic test and hence an observed increase in c might be a consequence of complex interdependencies of all parameters involved.

While the branching ratio η does not substantially vary with the different inversion methods, we observe a slightly increased productivity exponent for the PETAI inversion. This is expected given the results of Seif et al. (2017), with the extent of the observed increase being in line with their estimated extent of underestimation for the productivity exponent.

The value of β shows an increase from 2.33 to 2.37, which translates to a b -value increase from 1.01 to 1.03, when STAI is accounted for in the PETAI inversion. This is expected due to the underestimated number of small events caused by STAI.

Text S5 Comments on computational time

There are two aspects to consider when discussing the computational time of the here presented parameter inversion techniques. On one hand, the increased complexity of the algorithms plays an important role. In particular, the PETAI inversion comprises multiple loops of ETAS and incompleteness estimation. Although convergence was usually reached after 4 iterations, this still implies a minimum factor of 4 in terms of computation time which is only required for ETAS inversion, on top of which comes the time needed for the estimation of detection parameters and event rates. The second factor, which contributes even more to an increase of computation time, is the increased size of the catalog which is available to be used. For our application to Californian data, the number of events used in the PETAI inversion increases by a factor of 3.78 because the minimum considered magnitude is reduced from 3.1 to 2.5. This leads the number of pairs of potentially related events to increase from 7.3 million to 47.1 million. While this causes a substantial increase in run time, educated initial guesses for ETAS parameter inversions can substantially reduce run time.

In contrast to the PETAI inversion, the run time of the ETAS parameter inversion with time-varying m_c is barely affected by model complexity. During synthetic experiments, we found run time to be comparable to the run time of usual ETAS inversion when the number pairs of potentially related events was similar.

Text S6 Parameter transformation for reference magnitude changes

With the exception of μ , k_0 and d , all parameters are m_{ref} -agnostic, and the three exceptions can easily be adjusted to another reference magnitude as follows. Denote by Δm the difference between new and original reference magnitude, $\Delta m = m'_{ref} - m_{ref}$. Then,

$$d' := d \cdot e^{\Delta m \cdot \gamma} \quad (\text{S13})$$

ensures that

$$d \cdot e^{\gamma \cdot (m - m_{ref})} = d' \cdot e^{\gamma \cdot (m - m'_{ref})}. \quad (\text{S14})$$

Stipulating that the branching ratio η remains unchanged, it follows that

$$k'_0 := k_0 \cdot e^{\Delta m \cdot \gamma \cdot \rho}. \quad (\text{S15})$$

The adaptation of the background rate μ follows trivially from the GR law,

$$\mu' = \mu \cdot e^{-\beta \cdot \Delta m}. \quad (\text{S16})$$

When comparing the parameter estimates obtained when assuming $m_{ref} = m_c \equiv 3.6$, we transform them to refer to $m'_{ref} = 3.3$ before calculating their distance to the generating parameters. Also, for μ , d , k_0 , c and τ , we apply the \log_{10} before calculating differences, so that the compared values are in the same order of magnitude for all parameters.

Text S7 Forecast evaluation

The performance of each model is evaluated by calculating the log-likelihood of the testing data given the forecast. Specifically, we calculate the log-likelihood of N_i earthquakes to occur in each

bin b_i of a spatial grid of 0.1° latitude \times 0.1° longitude. Here, N_i is the number of earthquakes that actually occurred during the testing period in spatial bin b_i .

The log-likelihood for b_i is calculated based on the smoothed estimate of the probability of N_i earthquakes to occur in b_i , where the probability estimate is based on the 100'000 simulations of the model in question. For smoothing we use Gaussian kernels with adaptive bandwidth as described by Nandan et al. (2019), with a fixed value of $\Omega = 3.0$. To avoid arbitrary likelihood values due to extrapolation, we define a water-level likelihood for event counts larger than the maximum simulated event count in the respective bin. This waterlevel probability is defined as a uniform value of $100'001^{-1}/n_{extr}$, where n_{extr} is the number of event counts larger than the maximum observed and smaller than a generously high maximum possible event count. Symbolically, this suggests that all other possible event counts could have been simulated in the 100'001st simulation. Inevitably, the probabilities for non-extrapolated event counts are proportionally reduced such that the probabilities of all possible event counts add up to 1.

The total log-likelihood of the testing data is then given by the sum of log-likelihoods over all bins b_i .

Two competing models can be compared by calculating the information gain (IG) of the alternative model M_{alt} over the null model M_0 , which is simply the difference in log-likelihood of observing the testing data. The mean information gain (MIG) is calculated as the mean over all testing periods. We accept the superiority of M_{alt} over M_0 when we reject the null hypothesis that M_{alt} does not outperform M_0 . To decide whether to reject the null hypothesis, we perform a one-sided t-test on the set of IGs for all testing periods, and we reject the null hypothesis when a p -value of less than 0.05 is observed.

The flat model forecasts the same number of events in all spatial bins. This number of events forecasted, N_{fc} , is given by

$$N_{fc} = \frac{N_{train} \cdot T_{test}}{T_{train} \cdot N_{bins}}, \quad (\text{S17})$$

where n_{train} is the number of events observed in the training period, T_{train} is the length of the training period in days, $T_{test} = 30$ is the testing period length, and N_{bins} is the total number of spatial bins. The log-likelihood for the flat model is calculated assuming a Poisson distribution of event numbers with mean N_{fc} in each spatial bin.

References

- Hainzl, S. (2016). Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs. *Journal of Geophysical Research: Solid Earth*, *121*(9), 6499–6509.
- Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of the Seismological Society of America*, *94*(4), 1207–1228.
- Lippiello, E., Bottiglieri, M., Godano, C., & de Arcangelis, L. (2007). Dynamical scaling and generalized omori law. *Geophysical Research Letters*, *34*(23).
- Lolli, B., & Gasperini, P. (2006). Comparing different models of aftershock rate decay: The role of catalog incompleteness in the first times after main shock. *Tectonophysics*, *423*(1-4), 43–59.
- 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.
- Narteau, C., Byrdina, S., Shebalin, P., & Schorlemmer, D. (2009). Common dependence on stress for the two fundamental laws of statistical seismology. *Nature*, *462*(7273), 642–645.
- Schoenberg, F. P. (2013). Facilitated estimation of etas. *Bulletin of the Seismological Society of America*, *103*(1), 601–605.
- Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating etas: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449–469.
- Shcherbakov, R., Turcotte, D. L., & Rundle, J. B. (2004). A generalized omori’s law for earthquake aftershock decay. *Geophysical research letters*, *31*(11).
- Tinti, S., & Mulargia, F. (1987). Confidence intervals of b values for grouped magnitudes. *Bulletin of the Seismological Society of America*, *77*(6), 2125–2134.
- 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.

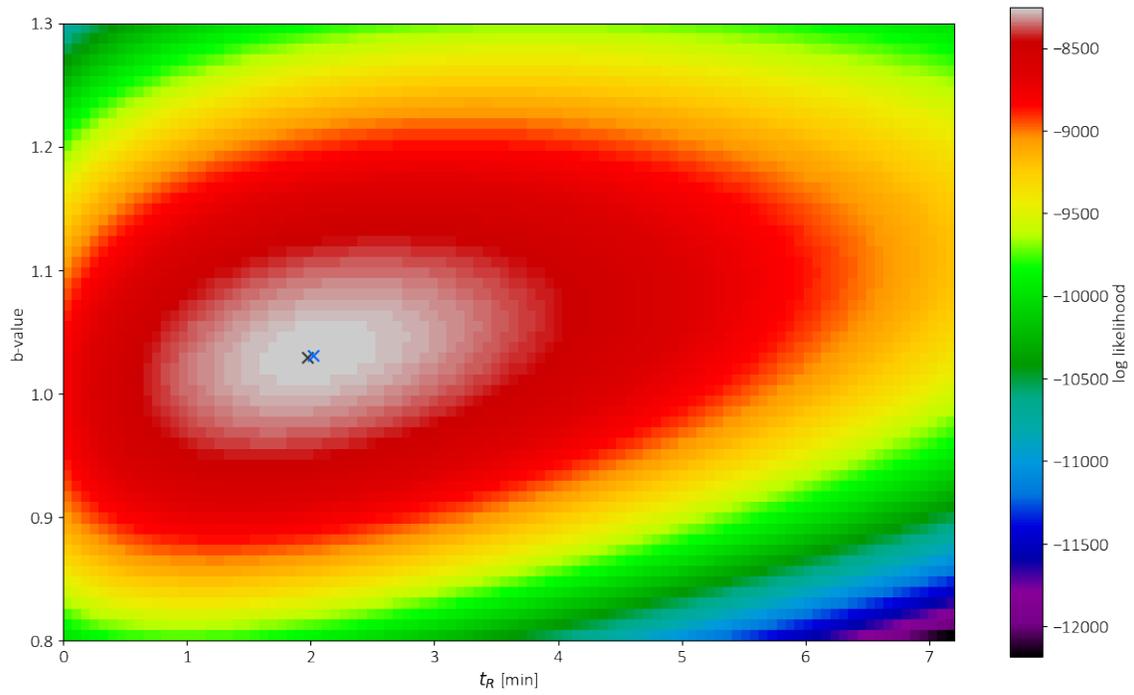


Figure S1: Log likelihood of observing the test data for different values of t_R and b -value, when current rate is known. Black cross indicates true values used in simulation, blue cross indicates maximum likelihood estimators obtained.

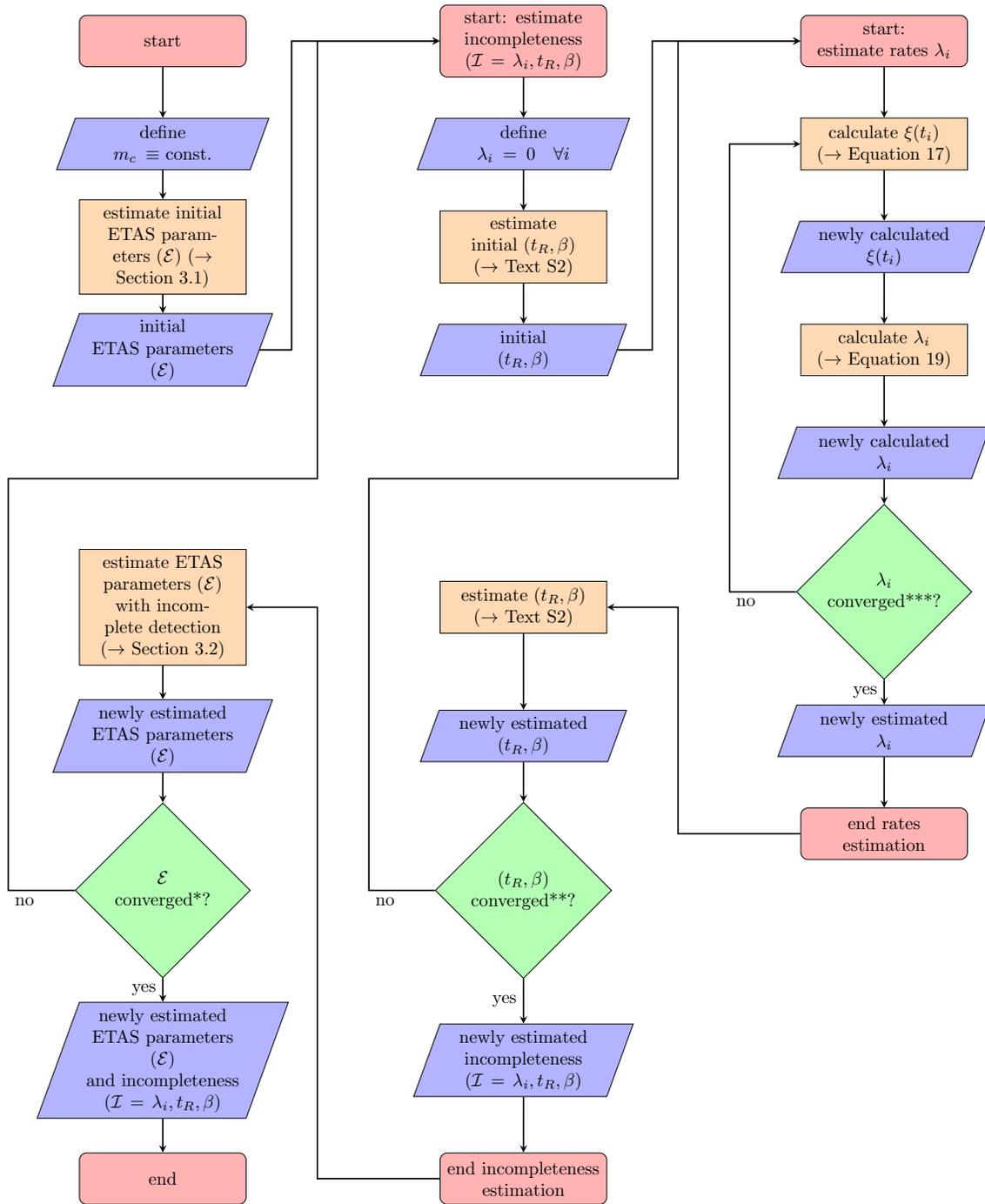


Figure S2: Flow diagram of PETAI inversion. Caption on next page.

Figure S2: (Previous page.) Flow diagram of PETAI inversion. Main algorithm starts at top left and ends at bottom left. The middle column describes the estimation of incompleteness ($\mathcal{I} = \lambda_i, t_R, \beta$) when ETAS parameters (\mathcal{E}) are given. Note that the estimation of $(\lambda_i)_{i=1, \dots, n}$ when ETAS parameters and (t_R, β) are fixed requires yet another loop to obtain self-consistency, as updating λ_i (step Λ) leads to changes in the inflation factor $1 + \xi(t_i)$, which forces one to update $(\lambda_i)_{i=1, \dots, n}$. This sub-sub-algorithm is visualized in the right column of the flow diagram. Process boxes are linked to corresponding methods and equations described in this article.

*, **, ***: Convergence is reached when the estimated values of the k^{th} iteration, \hat{a}_k , lie very close to the estimated values of the previous iteration, that is, if $\sum_{a \in A} |\hat{a}_k - \hat{a}_{k-1}| \leq \theta$. Here, A is the set of values that are tested for convergence, $*A = \mathcal{E}$, $**A = \{t_R, \beta\}$, $***A = \{\lambda_i, i = 1, \dots, n\}$. For convergence threshold θ we use $*\theta = 10^{-3}$, $**\theta = 10^{-12}$, $***\theta = 1$.

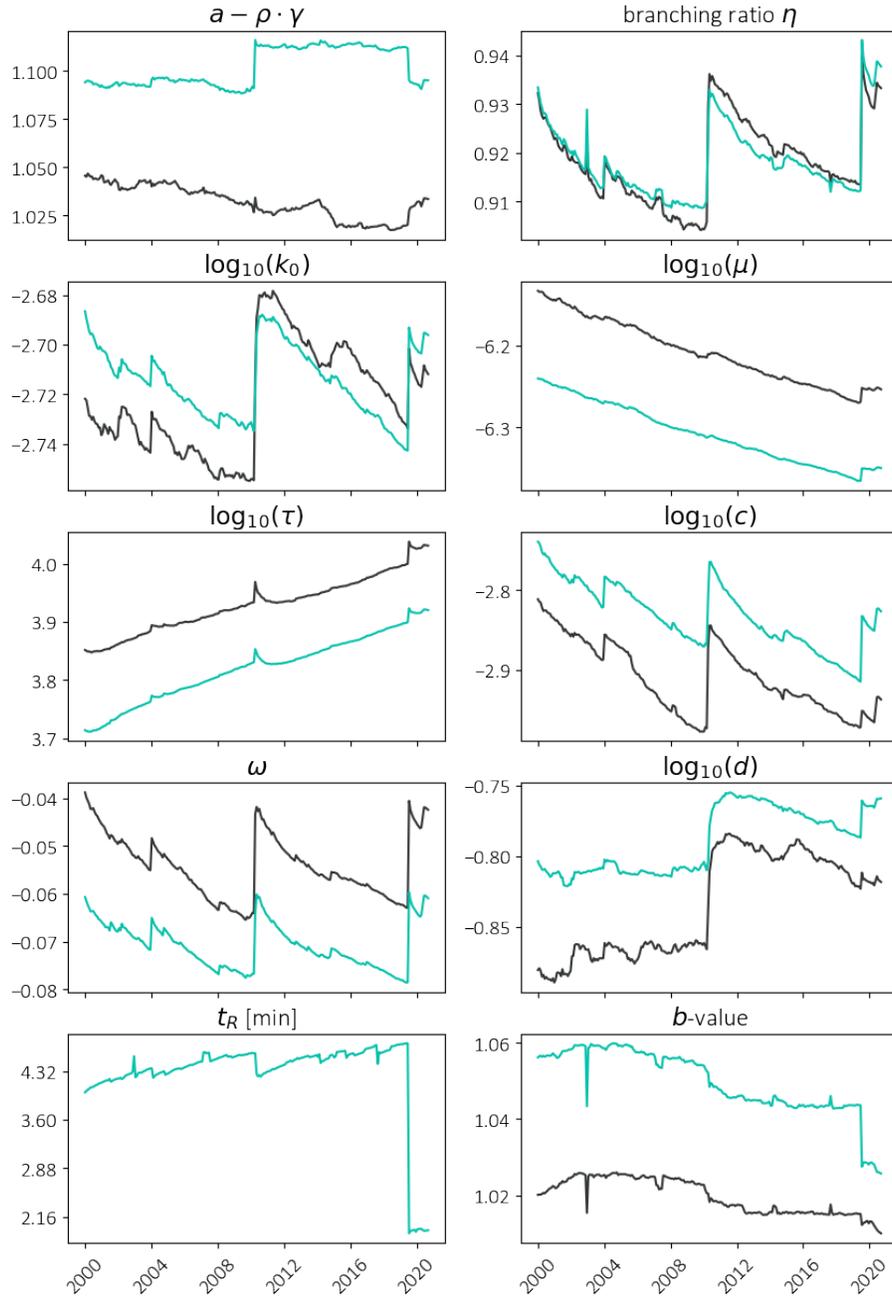


Figure S3: Evolution of ETAS and PETAI parameter estimates with increasing training catalog, when using standard inversion (black lines) and when using PETAI inversion (turquoise lines). The evolution for t_R is only given for PETAI inversion because it does not exist in standard ETAS.