Supporting Information for "Embracing Data Incompleteness for Better Earthquake Forecasting"

L. Mizrahi[1], S. Nandan[1], and S. Wiemer[1]

[1]Swiss 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},

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

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

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

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

$$a' := \alpha' + \rho \cdot \gamma, \tag{S3}$$

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

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 $\{e_1, \ldots, e_n\}$, 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.

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

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

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

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 $\lambda_i = \lambda(t_i)$,

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

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

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

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

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

and

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

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

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

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. The 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,

$$d' := d \cdot e^{\Delta m \cdot \gamma} \tag{S13}$$

ensures that

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

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

$$k'_0 \coloneqq k_0 \cdot e^{\Delta m \cdot \gamma \cdot \rho}. \tag{S15}$$

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

$$\mu' = \mu \cdot e^{-\beta \cdot \Delta m}. \tag{S16}$$

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 × 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 $100'001^{st}$ 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

$$N_{fc} = \frac{N_{train} \cdot T_{test}}{T_{train} \cdot N_{bins}}, \tag{S17}$$

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.

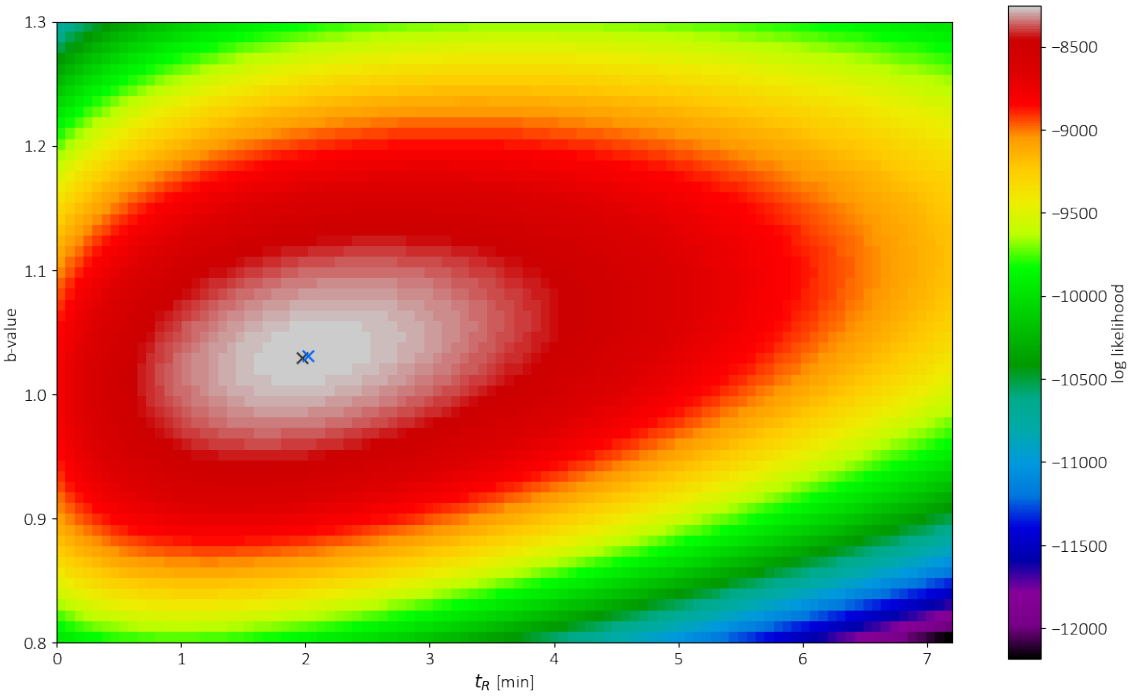

Figure S1: Log likelihood of observing the test data for different values of t_R and b-value, when current rate is known. Black cross indicates true values used in simulation, blue cross indicates maximum likelihood estimators obtained.

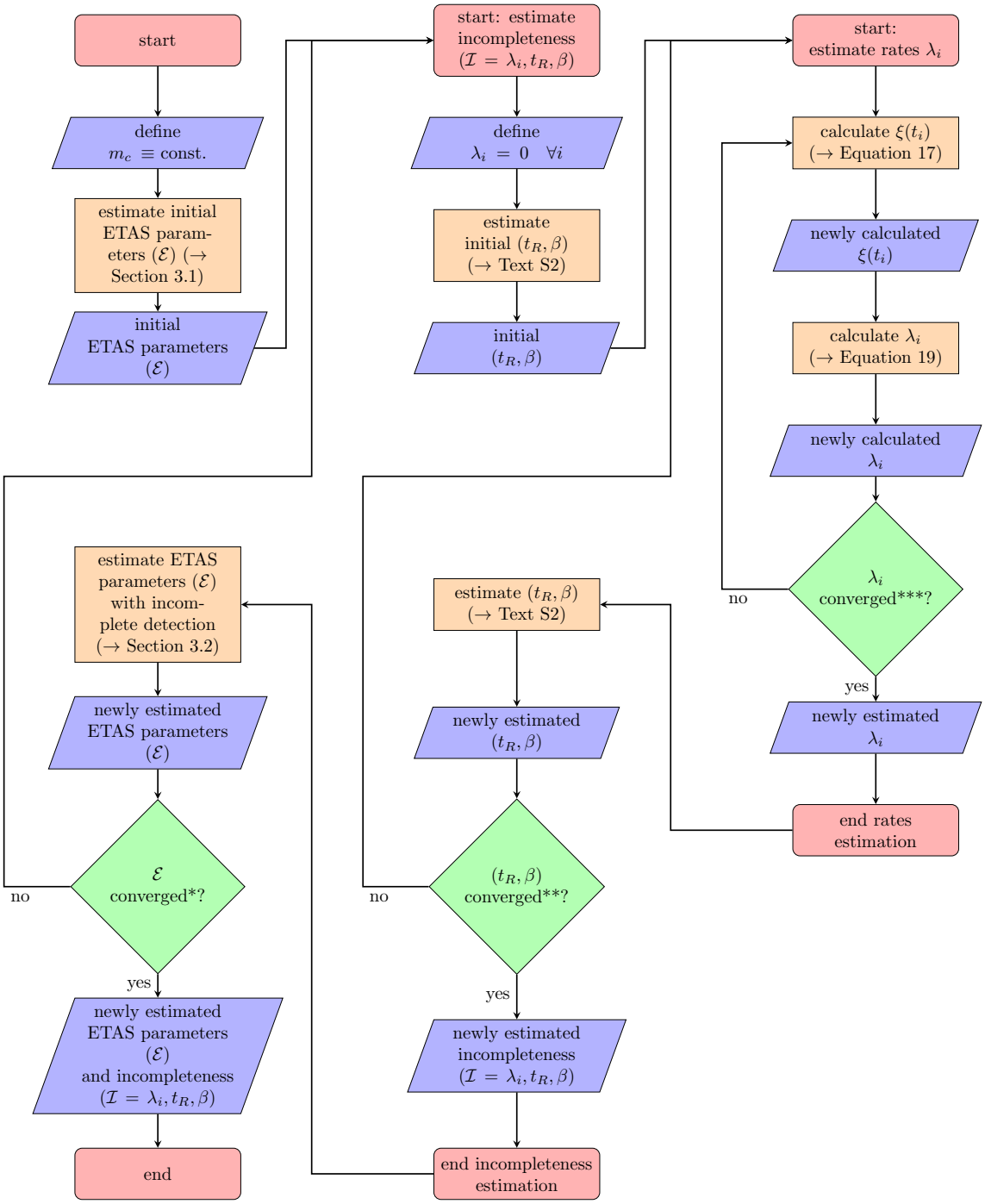

Figure S2: Flow diagram of PETAI inversion. Caption on next page.

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 $^*\theta = 10^{-3}$, $^{**}\theta = 10^{-12}$, $^{***}\theta = 1$.

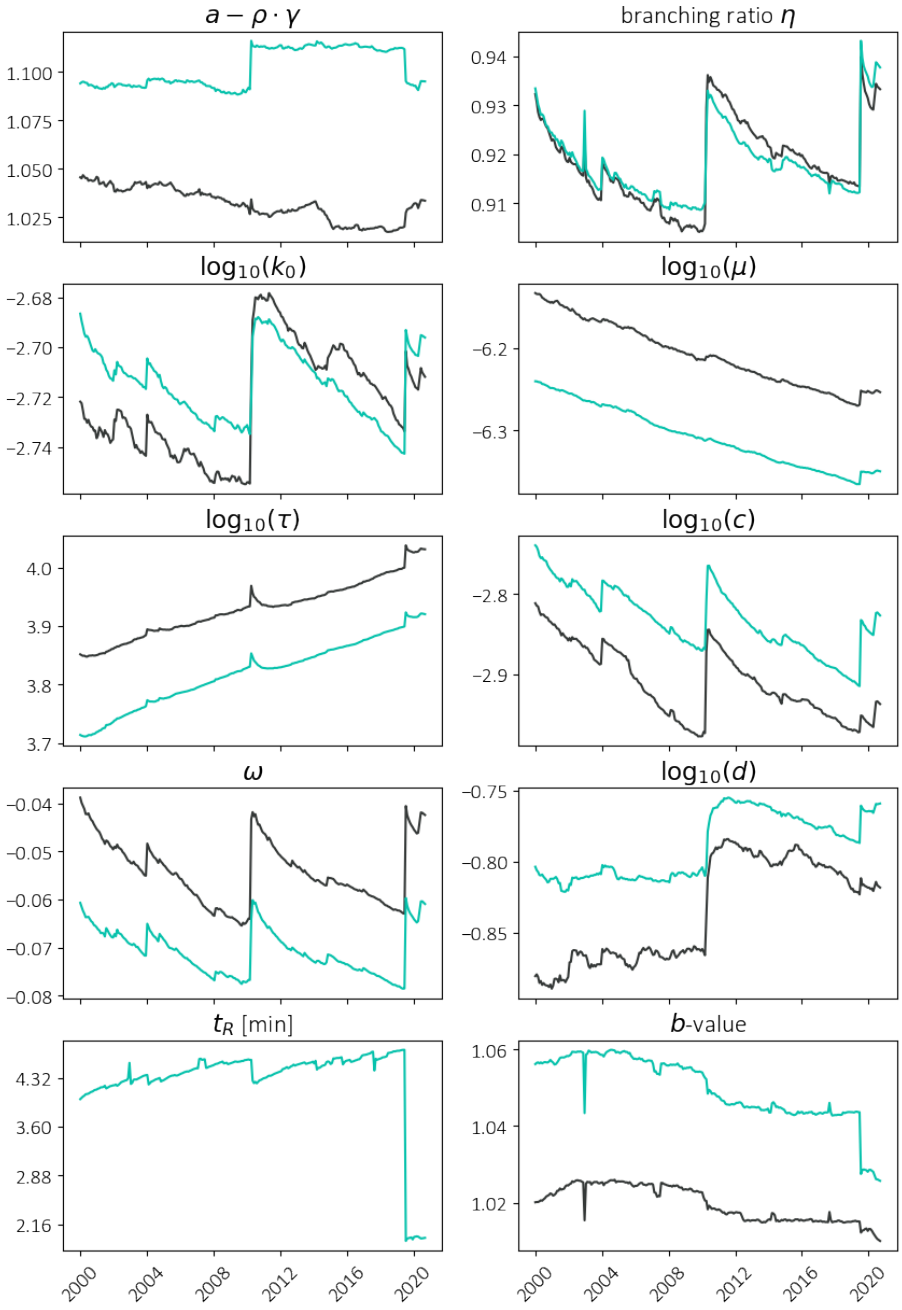

Figure S3: Evolution of ETAS and PETAI parameter estimates with increasing training catalog, when using standard inversion (black lines) and when using PETAI inversion (turquoise lines). The evolution for t_R is only given for PETAI inversion because it does not exist in standard ETAS.