User Manual „ETAS Estimation“

In this manual, it is described how to

- prepare required datasets (catalog, polygon)
- specify the ini file (input settings),
- run an ETAS model estimation,
- understand and interpret outputs
- check plausibility of model results

Citations:

- The exemplary earthquake event set for California has been downloaded from the Southern California Earthquake Data Center (https://scedc.caltech.edu/data/alt-2011-dd-hauksson-yang-shearer.html, last accessed on October 25, 2021).
- Christian Grimm implemented the code in Matlab R2019a.
- The standard functionality of the model goes back to the implementation of Jalilian (2019) in the *R* package *ETAS:*

  Jalilian, A. (2019). ETAS: An R package for fitting the space-time ETAS model to earthquake data, J. Stat. Softw. 88, no. 1, 1–39, doi: 10.18637/jss.v088.c01.

- Additional features such as anisotropic spatial kernels and short-term aftershock incompleteness refer to the model theory described in detail in the following papers:

  C. Grimm, M. Käser, S. Hainzl, M. Pagani, and H. Küchenhoff (2021). Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model, Bull. Seismol. Soc. Am. XX, 1–20, doi: 10.1785/0120210097

  Hainzl, S. (2021). ETAS-Approach Accounting for Short-Term Incompleteness of Earthquake 753 Catalogs. Bulletin of the Seismological Society of America. doi: 10.1785/0120210146.

  C. Grimm, S. Hainzl, M. Käser, and H. Küchenhoff (2022). Solving three major biases of the ETAS model to improve forecasts of the 2019 Ridgecrest sequence., doi: https://doi.org/10.21203/rs.3.rs-1128731/v1 (accepted)

Prepare Required Datasets

An ETAS model estimation run requires as input data

- An earthquake catalog
- A spatial polygon
- An ini file with estimation settings

In the following, the structure of these input types is described in detail.

Earthquake Catalog:

The earthquake catalog has to be a table with the following columns:

Column	Data Type	Description
id	scalar	Unique event IDs

date	datetime	Event times in datetime format
lon	scalar	Longitude of earthquake
lat	scalar	Latitude of earthquake
mag	scalar	Magnitude of earthquake
depth	scalar	Depth of earthquake (positive values)

Example:

1 id	2 date	3 lon	4 lat	5 mag	6 depth
1	1981-01-01 01:4...	-118.8129	33.7305	2.2700	7.0370
2	1981-01-01 04:1...	-115.9670	33.2532	2.2600	6.4080
3	1981-01-01 05:2...	-117.3044	34.1806	2.3700	6.2520
4	1981-01-01 05:3...	-117.3029	34.1792	1.6000	6.3420
5	1981-01-01 08:2...	-117.1632	34.0072	1.8800	15.8750
6	1981-01-01 10:5...	-117.3054	34.0953	1.4000	16.2900
7	1981-01-01 15:0...	-118.5618	35.3143	2.1600	4.9200
8	1981-01-01 17:4...	-116.7636	33.5038	1.5000	4.3730

<u>Spatial Polygon:</u>

The spatial polygon is a .mat file with two numerical arrays

- *polygonTarget (Nx2):* Target polygon that defines target space window, for which ETAS parameters should be estimated
- *polygonComple (Mx2):* Complementary polygon that defines a larger space window. Events that are inside of *polygonComple*, but outside of *polygonTarget*, can trigger events inside of *polygonTarget,* but are not estimated themselves.

In both arrays, the **first column** denotes the **longitudes** and the **second column** denotes the **latitudes** of the polygon boundary points.

In terms of the space window, *polygonTarget* needs to be a subset of *polygonComple*. It is recommended, that *polygonComple* goes about one geographical degree beyond *polygonTarget*.

Example:

polygonTarget — 7x2 double

	1	2
1	-114.9000	32.6000
2	-116.1000	34.8000
3	-117.9000	36.5000
4	-120.1000	34.3000
5	-117.8000	33.3000
6	-117.2000	32.6000
7	-114.9000	32.6000
8		
9		
10		

polygonComple — 5x2 double

	1	2
1	-121.1000	31.6000
2	-113.9000	31.6000
3	-113.9000	37.5000
4	-121.1000	37.5000
5	-121.1000	31.6000
6		
7		
8		
9		
10		

Specify Ini File

<u>Ini File with Estimation Settings:</u>

For each model run, you need to create (or modify) an ini file with all relevant model estimation settings. The ini file must be a .mat file including four structures:

- **ModelSettings:** General settings for model type, initial parameters
- **TargetSettings:** Paths to catalog and polygons, definition of time, magnitude and depth window

- **TimeSettings:** Definition of the temporal constraint of direct aftershock triggering
- **SpaceSettings:** Settings of the spatial kernel and for the anisotropic rupture estimation

In the following, the fields of each structure are described in detail.

ModelSettings:

The fields “modelType” and “spaceModel” are the main settings of the model and have significant impact on estimation results.

Field	Data Type	Available Inputs	Description
modelType	String	‘ETAS’ ‘ETAS-Incomplete’ (default)	General type of ETAS model: ‘ETAS-Incomplete’ is the preferred default, as this model accounts for short-term incompleteness.
spaceModel	String	‘aniso’ (default) ‘iso’	Type of spatial kernel: ‘aniso’ is the preferred default, as this model accounts for anisotropic spatial kernels.

The initial parameter values should be edited only

- to speed up the code, if parameter estimates of a similar model are known
- for advanced, scientific use of the model (such as sensitivity studies)

Field	Data Type	Recommended Inputs	Unit	Description
mu_ini	Scalar	1		Should be fixed
A_ini	Scalar	0.01		Should be smaller than 0.1
alpha_ini	Scalar	2		Should be in the range of 1.5 to 2.3
c_ini	Scalar	0.01	days	Should be below 0.1
p_ini	Scalar	1		Should be in the range of 0.8 to 1.2
D_ini	Scalar	0.1	Km^2	Should be below 10
gamma_ini	Scalar	2		Should be below 2.5
q_ini	Scalar	2		Should be below 5
Tb_ini	Scalar	60	sec	Should be below 30 minutes
b_ini	Scalar	1		Should be around 1

TargetSettings:

Column	Data Type	Recommended Inputs	Unit	Description
pathCatalog	String			Full path to the earthquake catalog input file (.mat file; see above)
pathPolygon	String			Full path to the spatial polygon input file (.mat file; see above) to define target and complementary space window
tStart	Datetime			Start date for target time window (e.g. Jan 1, 1982, 00:00:00)
tEnd	Datetime			End date for target time window (e.g. Jan 1, 2020, 00:00:00)
tIni	Datetime	Recommended 3-12 months before tStart		Start date of complementary time window. Events after tIni, but before tStart, can trigger into the target time window [tStart, tEnd], but are not estimated themselves
Mc	Scalar	Should be >= completeness magnitude of catalog	Mw	Cut-off magnitude to apply to the catalog. All events below Mc are removed from the event set.
Mmax	Scalar	Should be >= largest magnitude of catalog	Mw	Maximum possible magnitude in the modeled seismic region. This input is only

				used as an upper magnitude boundary to estimate the branching ratio.
maxDepth	String	Positive numbers	km	Cut-off depth to apply to the catalog. All deeper events are removed from the event set.

SpaceSettings:

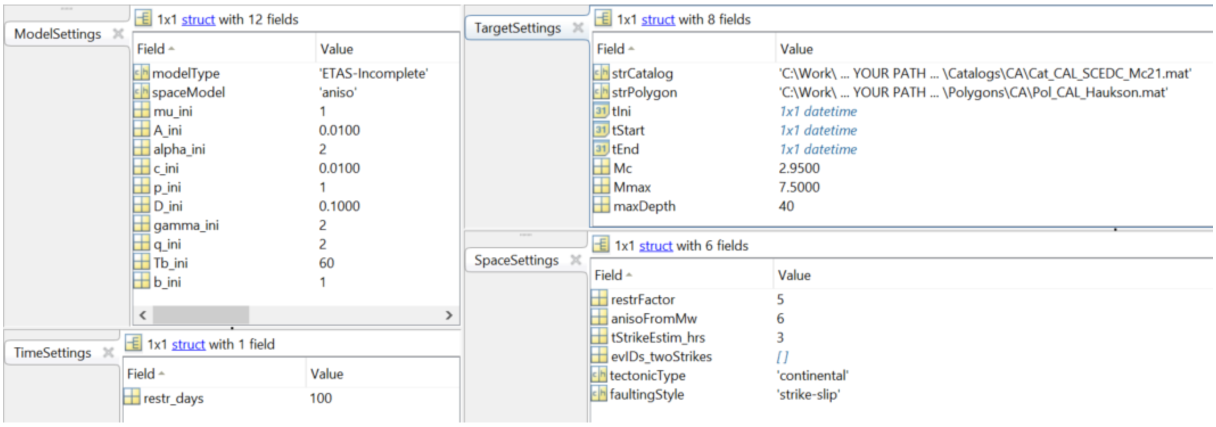
All fields but “restrFactor” are only relevant if ModelSettings.spaceModel = ‘aniso’.

Column	Data Type	Recommended / Available Inputs	Unit	Description
restrFactor	Scalar	Range of 1 to 5		Restriction of spatial trigger kernel to “restrFactor * ruptureLength”
anisoFromMw	Scalar	around 6.0	Mw	All events above this magnitude threshold are modelled anisotropically
tStrikeEstim_hrs	Scalar	e.g. 3	hours	Time window of aftershocks that are considered to fit the segment’s orientation (i.e. strike and epicenter location) of an anisotropic rupture.
evIDs_twoStrikes	Empty or Scalar	[] *Event ID*		Empty is default. If an event ID is submitted, this event is duplicated (each version has event weight 0.5) and modeled with two different strikes. This functionality was specifically designed for the 2019 Ridgecrest M6.4 event.
tectonicType	String	‘continental’ ‘subduction’		Tectonic zone types that determine the rupture length scaling law
faultingStyle	String	‘reverse’ ‘normal’ ‘strike-slip’ ‘all’		Faulting style that determine the rupture length scaling law

TimeSettings:

Note that the field “restrDays” has a strong effect on computation time. The ETAS model only estimates aftershock trigger rates between events that are no more than “restr_days” apart in time. Therefore, a larger value leads to a higher volume of possible trigger associations.

Column	Data Type	Recommended / Available Inputs	Unit	Description
Restr_days	Scalar	Range from 100 to 365	days	Temporal restriction of direct aftershock triggering. Events with larger time difference cannot be associated as aftershock pair.



Call ETAS Estimation Function

An ETAS model estimation run is called via the function “estimate_etasModel.m”. This function requires four inputs:

- **pathIniFile:** Full path to an .mat ini file containing the structures “ModelSettings”, “TargetSettings”, “SpaceSettings”, “TimeSettings” (see description above)
- **pathResultsFile:** Full path to directory and filename (with .mat ending) where ETAS estimation results should be saved. If directory does not yet exist, it is created.
- **makeLog:** Boolean value; If true, an estimation log file is saved under the results file path (recommended)
- **saveDiagnosticPlots:** Boolean value; If true, some diagnostic plots (e.g. aftershock productivity, Omori law, spatial kernels) are saved under the results file path (recommended)

Understanding and interpreting outputs

The ETAS model program saves a .mat result file at the desired *pathResultsFile*. This .mat file contains the following variables:

Variable	Data Type	Description
ClustersETAS	Table	Earthquake clusters in the original catalog, identified by the ETAS model. Contains list of event IDs, cluster size, mainshock id (mId), mainshock date (mDate), mainshock magnitude (mMag), largest foreshock or aftershock magnitude (aMag) and further information regarding the cluster’s branching tree
EstimSettings	Structure	Contains the ini file settings structures (see details above)
IterationLog	Structure	Evolution of loglikelihood function values, gradients, AIC values, parameter estimates and integrated ETAS rates over the course of iterations. Rows stand for iterations. Columns in fields “paramETAS” and “loglik_grad” represent the 10 parameters.
ModelFuncs	Structure	Internal ETAS functions representing the ETAS model definition by Grimm et al, 2021). Functions are needed for plotting and simulations. Do not manually edit.

ModelSummary	Structure	Summarizes diagnostic values. Number of events observed and number of events estimated should match closely. Branching ratio should be smaller than 1, unless an unstable sequence is estimated.
paramETAS	Numerical Vector	Resulting parameter estimates. Note: Parameter D, Tb and b are converted to internal ETAS model units, different to the units of the initial parameters submitted in the ini file settings. **To compare with ini setting parameters, convert as follows:** • D: units degree^2 instead of km^2 (convert by factor 111^2) • Tb: units days instead of seconds (convert by factor 24*60*60) • Beta: Gutenberg-Richter parameter to the exponential base instead of base 10 (convert to b-value by factor 1/log(10))
paramETAS_descrip	Table	Lists resulting parameter names and units in comparison with initial parameters (submitted by ini file) and converting factors

Check Plausibility of Model Results

The following quick checks help to diagnose whether the ETAS model results are robust:

- **Number of target events:** The estimated number of target events (“integrated rate”) should closely match the number of target events in the original catalog (see “ModelSummary” in .mat results file or display at the end of iteration)
    - **If not close:** Check whether IterationLog.integratedRate shows that at an intermediate iteration, the integrated rate closely matched the number of target events in the original catalog. If so, check in IterationLog.paramETAS whether parameter estimates substantially changed since that iteration. If no relevant different in parameter values shows up, the “miss-match” of of number of target events can be neglected.
- **Gradients:** Gradients should be close to zero at the end of the iterations.
- **Branching Ratio:** The branching ratio should be smaller than 1 (see “ModelSummary” or display at the end of iteration), unless a known, unstable sequence is estimated (e.g. Ridgecrest M6.4 sequence)
- **Diagnosis Plots:** Plots of aftershock productivity, Omori law etc should look reasonable (e.g. compare aftershock productivity with Bath’s law)