

COMPUTER SCIENCE MONOGRAPHS

*A Publication
of
The Institute of Statistical Mathematics*

**STATISTICAL ANALYSIS OF SEISMICITY - UPDATED VERSION
(SASeis2006)**

by

Y. Ogata

April 2006

Computer Science Monographs

Editors

Junji NANAOKO(*in chief*)

Satoshi ITO

Nobuhisa KASHIWAGI

Yumi TAKIZAWA

The series of Computer Science Monographs publishes the statistical software and the articles on computer applications originating at The Institute of Statistical Mathematics. All communications relating to this publication should be addressed to the Editorial Office, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan.

The electronic version of this issue is downloadable at
<http://www.ism.ac.jp/>

Statistical Analysis of Seismicity - updated version (SASeis2006)

Ogata, Yoshihiko

The Institute of Statistical Mathematics

Introduction	2
Programs AFTPOI and RAFTPOI	4
1. The Poisson process of the Omori-Utsu law and residual Poisson process	4
1.1 <i>The MLE computation</i>	4
1.2 <i>The residual point process</i>	5
2. Data [work.etas].....	5
3. Control file [aftpoi.open].....	6
4. Remarks.....	7
4.1 <i>Initial values for parameters in control file [aftpoi.open]</i>	7
4.2 <i>Abnormal termination</i>	7
4.3 <i>Standard errors</i>	7
5. Graphs	7
Overview of the program [eptren]	9
Programs ETAS and RETAS	12
The ETAS model and the residual point process.....	12
1.1 <i>The MLE computation</i>	12
1.2 <i>The residual point process of the ETAS model</i>	13
2. Programs and data	14
2.1 <i>Exact and approximate algorithms</i>	14
2.2 <i>Precursory and target intervals</i>	15
2.3 <i>Data file [work.etas]</i>	15
2.4 <i>Control file [etas.open]</i>	16
3. Graphs	17
4. Remarks.....	18
4.1 <i>Application to aftershock sequences</i>	18
4.2 <i>Abnormal termination</i>	19
4.3 <i>Test data set</i>	19
Program ETASIM	20
Control file [etasim.open]	20
Overview of the program [pgraph]	21

Introduction

This package consists of stand-alone programs, AFTPOI, EPTREN, ETAS, ETASIM, PGRAPH, RAFTPOI and RETAS. The source files are written in FORTRAN77, and can be compiled by GNU FORTRAN “g77” on any platforms where GNU FORTRAN is available. These programs only output numerical files for graphics. We also provide modules for plotting figures in postscript format in the same directory, written in R, a widely-used free statistical programming language. All the names are corresponding to the names of the FORTRAN sources and the output files.

Some test data files for these programs are included. They are either simulated data or actual earthquake (aftershocks) data. It is recommended that you first try these data files to understand how each program works.

All programs are downloadable in zip format from the "*Computer Science Monographs*" page in WWW pages of the Institute of Statistical Mathematics. They can be also retrieved from

<http://bemlar.ism.ac.jp/www2/SASeisUpCollection/SASeis2006/> or

<http://www.ism.ac.jp/~ogata/Statsei4/download.html>

The latter is the same as the one distributed at the 4th Statistical Seismology Workshop held at Hayama Campus of the Graduate University of Advanced Studies, Japan, during 9-13 January 2006.

AFTPOI: Maximum likelihood estimates (MLEs) of parameters in the Omori-Utsu (modified Omori) formula for the decay of occurrence rate of aftershocks with time, which is formulated as a non-stationary Poisson process.

EPTREN: MLEs of parameters in a non-stationary Poisson process with a rate function being an exponential polynomial or an exponential Fourier series. The optimal order of the polynomial and series can be determined by the minimized AIC value.

ETAS: MLEs of parameters in the ETAS model for general seismicity and aftershock sequences.

ETASIM: Simulation of earthquake data set based on the ETAS model.

PGRAPH: To display some elementary statistical features of dataset of a point process or its residual point process.

RAFTPOI: Occurrence times of earthquakes are transformed using the Omori-Utsu formula with

the MLEs by AFTPOI. This is called as the residual data and can be used for the diagnostic analysis or the detection of anomalies in the aftershock sequence.

RETAS: Occurrence times of earthquakes are transformed using the ETAS model with the MLEs. This is called as the residual data and can be used for the diagnostic analysis or the detection of anomalies in the aftershock sequence.

Some technical notes are as follows:

1. If the programs crash with an error message "cannot find cygwin1.dll", copy cygwin1.dll to c:\windows\system32.
2. To install updated version of R for windows, MacOS or some other UNIX based systems, please go to "<http://www.r-project.org/>".
3. To install most recent version of cygwin, please go to "<http://www.cygwin.com/>".

Acknowledgements

I am grateful Jancang Zhuang for his advices on R commands, and also for carefully reading the manuals for SASeis2006.

References.

- Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of American Statistical Association*, Application, Vol. 83, No. 401, pp. 9-27
- Ogata, Y. and Katsura, K. (1985) Fortran programs for Statistical Analysis of Series of Events (SASE) consisting of the programs EPTREN, LINLIN, SIMBVP, LINSIM and PGRAPH included in *Time Series and Control Program Package*, **TIMSAC-84** (joint with H. Akaike, T. Ozaki, M. Ishiguro, G. Kitagawa, Y. Tamura, E. Arahata, K. Katsura and R. Tamura), Computer Science Monograph, No. 22/23, The Institute of Statistical Mathematics, Tokyo, Japan.
- Utsu, T. and Ogata, Y. (1997) Computer program package: Statistical Analysis of point processes for Seismicity, **SASeis**, *IASPEI Software Library* for personal computers, the International Association of Seismology and Physics of Earth's Interior in collaboration with the American Seismological Society, Vol. 6, pp. 13-94.

Programs AFTPOI and RAFTPOI

Estimation of parameter values in the Omori-Utsu Law for the decay rate of aftershocks, which is also called “the modified Omori Law” by Tokuji Utsu.

1. The Poisson process of the Omori-Utsu law and residual Poisson process

1.1 The MLE computation

This FORTRAN program computes the maximum likelihood estimates (MLEs) of the parameter values in the Omori-Utsu formula (the modified Omori formula; Omori, 1894; Utsu, 1961) which represent the decay law of aftershock activity in time (see also Utsu *et al.*, 1995). In these equations, $f(t)$ represents the rate of aftershock occurrence at time t , where t is the time measured from the origin time of the main shock. μ , K , c and p are non-negative constants. μ represents constant-rate background seismicity which may be included in the aftershock data.

$$f(t) = \mu + K / (t + c)^p$$

The FORTRAN program was written by Ogata (1983), where the negative log-likelihood function is minimized by the Davidon-Fletcher-Powell algorithm. Namely, starting from a given set of initial guess of the parameters that is given in the control file [**aftpoi.open**], the program [**aftpoi.f**] repeats calculations of the function values and its gradients at each step of parameter vector. At each cycle of iteration, the linearly searched step (**lambda**), negative log-likelihood value (**-LL**), and two estimates of square sum of gradients are shown. At each linear search of the likelihood computation, **-LL** and the 4 parameter values are shown. The value **-LL** decreases and tends to a final value. one of the square sum of gradients becomes nearly zero and the iteration is terminated. The program [**aftpoi**] always call the dataset named [**work.etas**] with the format as described below.

In the present case, the following results are displayed, where $AIC = -2 LL + 2 \times$ (number of searched variables), and the number = 4 in this case. The calculated record of the program [**aftpoi**] is stored by the name [**aftpoi.print**] in the directory of [**Calculations**] for your initial check of the program.

1.2 The residual point process

The FORTRAN program [**raftpoi.f**] computes the following output for displaying the goodness-of-fit of the Omori-Utsu model to the data. The cumulative number of earthquakes at time t since t_0 is given by the integration of $f(t)$ in Func6 with respect to the time t ,

$$F(t) = \mu (t - t_0) + K \{ c^{1-p} - (t - t_i + c)^{1-p} \} / (p - 1)$$

where the summation of i is taken for all data event. The output of the program is given in the file [**work.res**] which includes the column of $\{F(t_i), i = 1, 2, \dots, N\}$. This file is used for displaying the cumulative curve and magnitude v.s. transformed time $F(t_i)$. For the users who have the free graphic software **R**, I set a module [**r.raftpoi**] to display them, in addition to the module [**r.seis**] for displaying the cumulative curve and magnitude v.s. ordinary time t_i . If the observed rate of occurrence is compared with the calculated one from the model, period of decreased or increased seismic activity (relative quiescence or activation) can be recognized. The calculated record of the program [**raftpoi**] is stored by the name [**raftpoi.print**] in the directory of [**Calculations**] for you to check the program. The program [**raftpoi**] always call the dataset named [**work.res**] with the format as described below.

2. Data [**work.etas**]

The data file for this program is a sequential file with a 9 columns-format of the sequential number, longitude, latitude, magnitude, time from the mainshock in days, depth, year, month, and day, as is given by the prototype data file [**work.etas**]. The first row of the data is the title. The second row corresponds to the main shock (not used in this program, but is used in other programs such as [**etas.f**]). The time (usually in days) is measured from the main shock ($t = 0$). If aftershock in an early stage of the sequence (e.g., from $t = 0$ to 0.01 day) are not included because of incomplete observation, it is better to set T_{start} at the beginning of the aftershock data (e.g., $T_{start} = 0.01$). $0 < T_{start} < t_1 < \dots < t_i < t_{i+1} < \dots < t_n < T_{end}$ in 5th column of [**work.etas**]. Magnitudes in 4th column are used only in selecting the aftershocks by giving a threshold magnitude.

3. Control file [aftpoi.open]

The file [aftpoi.open] provides the input-variables such as the restricting range of data and initial values of parameters.

The first row of [aftpoi.open] consists of two numbers; the first number is always set 6 for the selection of the subroutine [func6] for the Omori-Utsu model, and then for the second number, any dummy number can be set.

The second row consists of three numbers; zts , zte and $tarst$ in the above illustrated variables when the file is applied to the program [aftpoi.f]. When the file is applied to the program [raftpoi.f] for the diagnostic outputs, the last number can be $zrend$ ($\geq zte$) that is the last time of the period of available data. The example shows the first 6.5days of aftershock period since the mainshock due to [work.etas], where the target interval is [$tarst$, zte] days.

The third row consists of two numbers; threshold magnitude M_{th} and reference magnitude M_z . Here I have taken the magnitude of the mainshock as the reference magnitude, but you may take $M_z = M_{th}$ for general seismicity. Note that the MLE of the parameter K differs depending on the reference parameter.

The fourth row provides the initial estimates of the five parameters μ , K , c , α and p when this is applied the program [aftpoi.f], but α ($=0.0$) is dummy parameter to make same format as [etas.open]. If you set $\mu = 0.0$ in the initial value, this variable remains 0.0 throughout the calculations. Also, if you set $p = 1.0$, this variable remains 1.0 (the original Omori formula) throughout the calculations. When this control file is used for the program [retas.f] to make the diagnostic outputs, these values in this row have to be the MLEs estimated by the program [aftpoi.f].

The fifth row consists of three numbers; zts , zte and $tarst$ only used when the file is applied to the program [raftpoi.f] for the diagnostic outputs, the last number must be zte that is the end of the target period.

The sixth row is the same as the fourth row, only used for the program [raftpoi.f], and the seventh row of the value of the negative of the maximum likelihood corresponding to the MLEs, used just for the user' note to compare goodness-of-fit with other cases or models.

4. Remarks

4.1 Initial values for parameters in control file [*aftpoi.open*]

The program works for initial values given fairly arbitrary. A suggested set of the initial values could be set in the control file [**aftpoi.open**]. Type an appropriate value if you wish to change it. Unreasonable initial values may cause an overflow error. If **-LL** value and gradients do not seem to converge, we recommend you to use the last parameter estimates as the initial values in the control file [**etas.open**] and restart the execution.

4.2 Abnormal termination

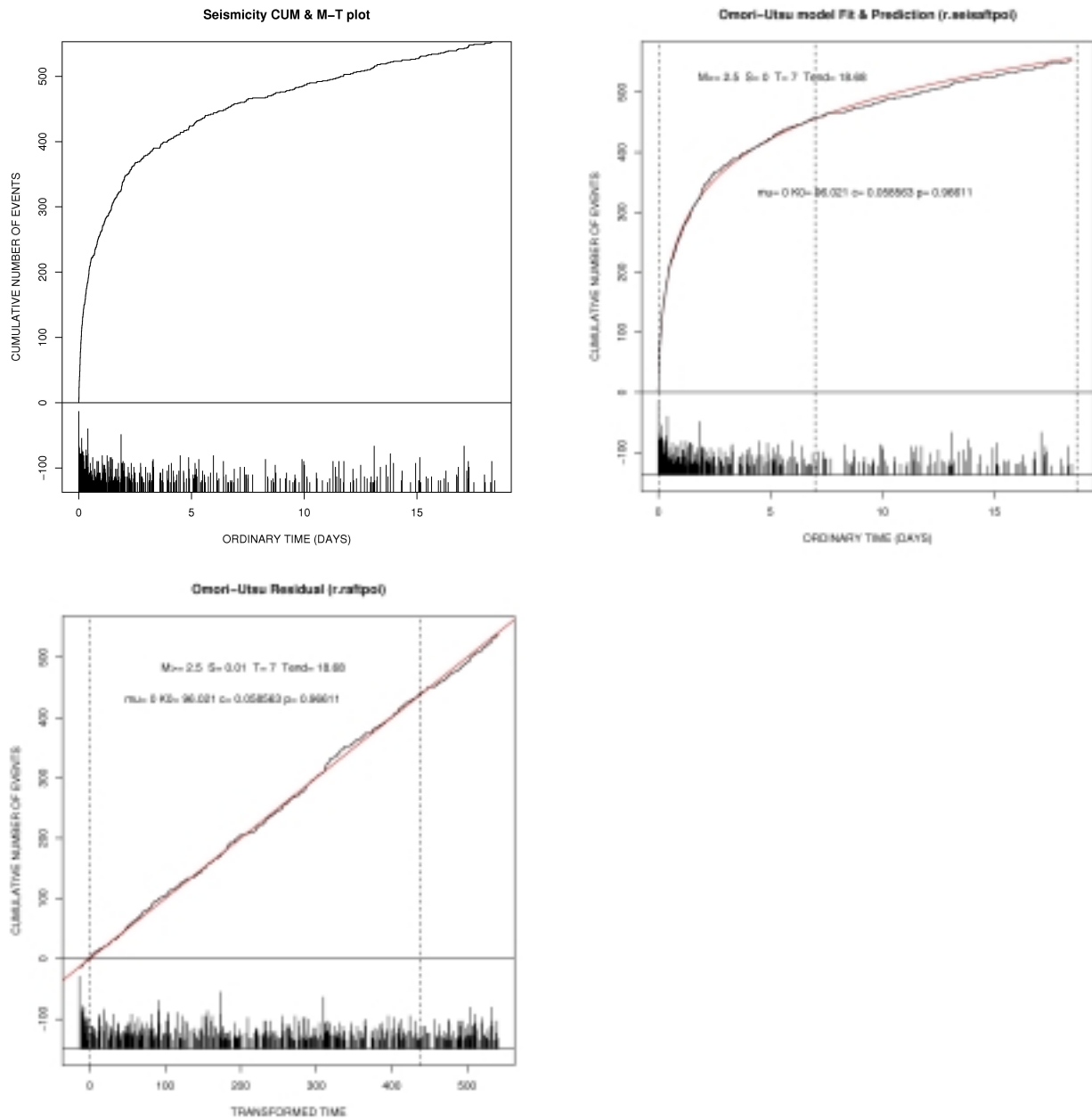
If the data are very different from the proposed model, an overflow error may occur or the solution may not converge within the fixed number of iterations ($30N$, N is the number of parameters). The results in this case are often incorrect and not recommended to adopt as the estimates.

4.3 Standard errors

The standard errors (SD) for estimates are computed by taking the square root of the corresponding diagonal element of the Inverse Fisher matrix.

5. Graphs

The R language modules [**r.seis**], [**r.seispoi**] and [**r.raftpoi**] are used to plot the graphs of cumulative number and magnitude of earthquakes against the ordinary time and transformed time, respectively: These modules show the files [**work.etas**] and [**work.res**]. The last graph includes printed number with certain time values for vertical dotted lines showing target interval and MLEs in control file [**aftpoi.open**]. The outputs of the graphs in postscript format, named as **seis.ps**, **seispoi.ps**, and **raftpoi.ps** are given in the same directory.



The R language modules [**r.raftpoi**] and [**r.seisftpoi**] shows the theoretical curve of cumulative frequency (red color) which should be a straight line by definition of the transformed time (if the model is correct), and the cumulative frequency curves of the data (red color).

The graphical free software R is also installed in the other directory based on the Microsoft Windows XP (R Gui). To command in the R console, write

```
> source("r.*")
```

where * correspond above R-module, then the graphical window should appear and draw

appropriate figures.

The FORTRAN programs are originally designed and programmed (1983) and reprogrammed (December 2005) by Yosihiko Ogata, Institute of Statistical Mathematics, Tokyo, Japan. The R-module was designed and programmed by Yosihiko Ogata (2003).

References

- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, AC-19, 716-723.
- Ogata, Y., Estimation of parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure, *J. Phys. Earth*, **31**, 115-24, 1983.
- Omori, F., On the after-shocks of earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo*, **7**, 111-200, 1894.
- Utsu, T., A statistical study on the occurrence of aftershocks, *Geophys. Mag.*, **30**, 521-605, 1961.
- Utsu, T., Y. Ogata, and R. S. Matsu'ura, The centenary of the Omori formula for a decay law of after-shock activity, *J. Phys. Earth*, **43**, 1-33, 1995.

Overview of the program [eptren]

This program computes the maximum likelihood estimates (MLEs) of the coefficients A_1, A_2, \dots, A_n in an exponential polynomial

$$f(t) = \exp\{A_1 + A_2 t + A_3 t^2 + \dots\} \quad (1)$$

or $A_1, A_2, B_2, \dots, A_n, B_n$ in a Poisson process model with an intensity taking the form of an exponential Fourier series

$$f(t) = \exp\{A_1 + A_2 \cos(2\pi t/P) + B_2 \sin(2\pi t/P) + A_3 \cos(4\pi t/P) + B_3 \sin(4\pi t/P) + \dots\} \quad (2)$$

which represents the time varying rate of occurrence (intensity function) of earthquakes in a region. These two models belong to the family of non-stationary Poisson processes. The optimal order n can be determined by minimize the value of the Akaike Information Criterion (AIC) (cf., Akaike, 1974).

The maximum number of coefficients is limited to 21 due to the dimension set in the FORTRAN

source. The coefficients A_1, A_2, \dots, A_{21} can be determined for equation (1) and $A_1, A_2, B_2, \dots, A_{10}, B_{10}$ can be determined for equation (2) at the most.

Since a non-stationary Poisson process is assumed in the data, we need careful interpretation of the results when the model is applied to data containing earthquake clusters such as aftershocks, foreshocks, and earthquake swarms.

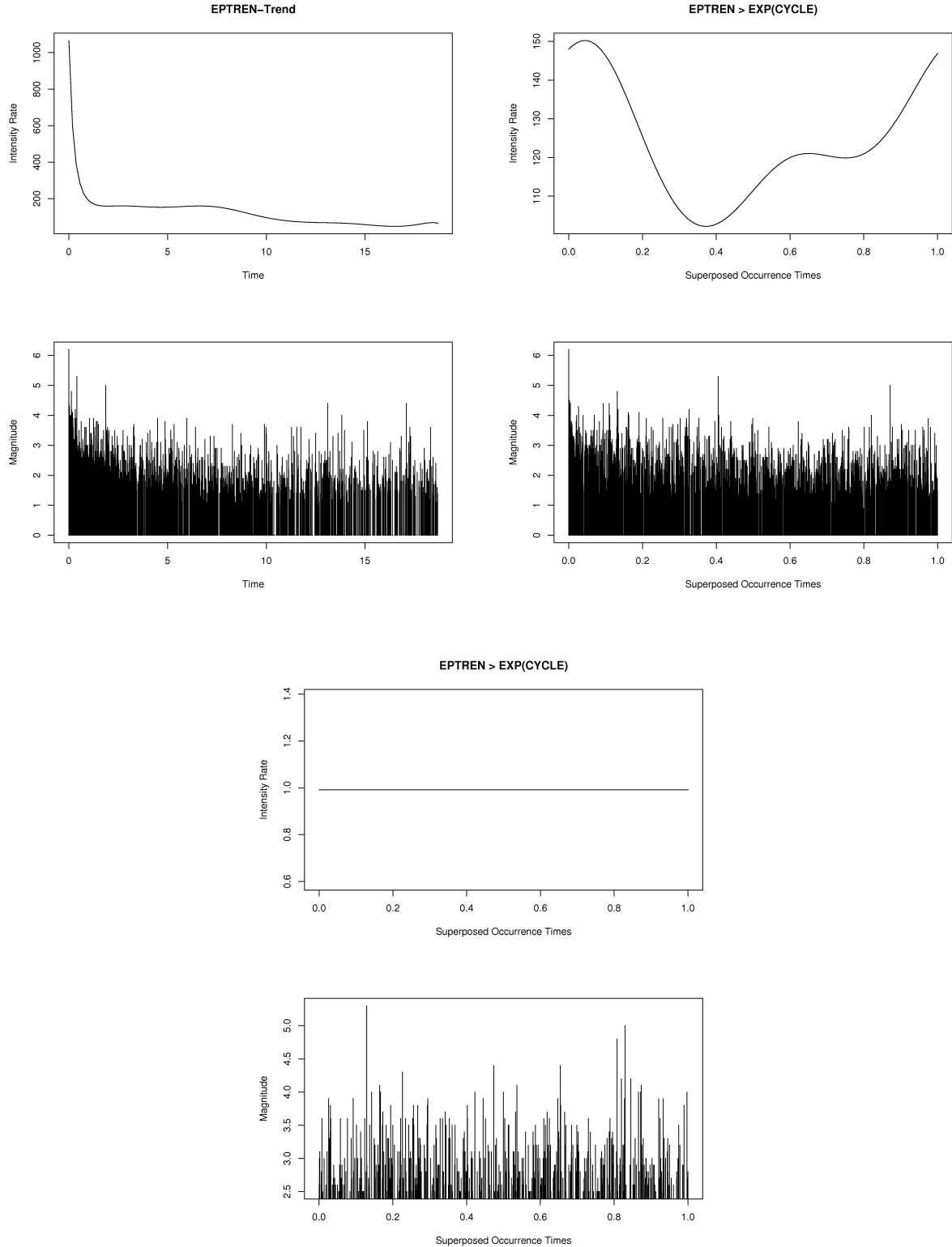
R-modules are also attached in order display the outputs of numerical data for plotting graphs of the estimated functions.

Structure of the program is the following.

```
[eptren]
|-----[inputs]
|-----[reduc1]
|-----[reduc2]
|-----[davidn]-----[funct]
|           |-----[hesian]
|           |-----[linear]----[funct]
|
|-----[fincal]
|-----[output]---[printr]----[trenfn]
|-----[cyclfn]
```

1. The name of the input data file can be either [**work.etas**] or [**work.res**] with their format as describe before. The choice of either data is controlled by the input control file. The range of x-coordinate is limited to the non-negative valued variable. When dataset [**work.etas**] includes remarkable clusters, the interpretation of the trend and cycles must be carefully made.
2. The input control file [**eptren.open**]selects either trend or cycle fitting and specify necessary variables.

3. The output files of [eptren] are [out.eptren1] and [out.eptren2]. The R-module [r.eptrend] or [r.epcycle] is used to plot figures in both console and postscripts [eptrend.ps] and [epcycle.ps].



4. Calculated records of the program [**eptren**] are stored files [**eptren??.print**] in the directory of [**Calculations**], where each “?” is either 0, 1 or 2, which is given in [**eptren.open**].

This program was originally designed (January 1985) and revised (December 2005) by Yosihiko Ogata, and programmed and also reprogrammed by Koichi Katsura. The R-module was designed and programmed by Jiancang Zhuang and Yosihiko Ogata (December 2005).

References

- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **AC-19**, 716-723.
- Lewis, P.A.W. and G.S. Shedler (1976). Statistical analysis of non-stationary series in a database system, *IBM. J. Res. Develop.*, **20**, 465-481.
- Macleane, C.J. (1974). Estimation and testing of an exponential polynomial rate function within the non-stationary Poisson process, *Biometrika*, **61**, 81-86.

Programs ETAS and RETAS

Estimation of parameter values in the ETAS model and diagnostic analysis of seismic activities

The ETAS model and the residual point process

1.1 The MLE computation

The FORTRAN program [**etas.f**] computes the maximum likelihood estimates (MLEs) of five parameters of the Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1988) calculated for a set of data on the occurrence times and magnitudes of earthquakes from an earthquake hypocenter catalog. An example of data [**work.etas**] is given in the same directory. The executive file of the present FORTRAN source code is compiled by GNU-FORTRAN on any platforms that GNU FORTRAN is available. It should also be able to be compiled on any system where GNU FORTRAN is available.

The ETAS model is a point-process model representing the activity of earthquakes of magnitude M_z and larger occurring in a certain region during a certain interval of time. The total number of such earthquakes is denoted by N . The seismic activity includes primary activity of constant occurrence rate μ in time (Poisson process). Each earthquake (including aftershock of another earthquake) is followed by its aftershock activity, though only aftershocks of magnitude M_z and larger are included in the data. The aftershock activity is represented by the Omori-Utsu formula (modified Omori formula; see section 1 of Program AFT) in the time domain. The rate of aftershock occurrence at time t following the i th earthquake (time: t_i , magnitude: M_i) is given by

$$n_i(t) = K \exp[-(M_i - M_z) / (t - t_i + c)^p], \quad \text{for } t > t_i \quad (1)$$

where K , μ , c , and p are constants, which are common to all aftershock sequences in the region. The rate of occurrence of the whole earthquake series at time t becomes

$$\lambda(t) = \mu + \sum_i n_i(t). \quad (2)$$

The summation is done for all i satisfying $t_i < t$. Five parameters μ , K , c , μ , and p represent characteristics of seismic activity of the region. The dataset should include the occurrence times t_i of earthquakes associate with M_i in the period $[T_{start}, T_{end}]$ such as the prototype data file [**work.etas**] given in this directory as an example. The calculated record of the program [**etas**] is stored by the name [**etas.print**] in the directory of [**Calculations**] for your initial check of the program.

1.2 The residual point process of the ETAS model

The FORTRAN program [**retas.f**] computes the following output to display the goodness-of-fit of the ETAS model to the data. The cumulative number of earthquakes at time t since t_0 is given by the integration of (2) with respect to the time t ,

$$\Lambda(t) = \mu(t - t_0) + K \sum_i \exp[-(M_i - M_z) / (t - t_i + c)^p] / (p - 1), \quad (3)$$

where the summation of i is taken for all data event. The output of the program is given in the file [**work.res**] which includes the column of $\{\Lambda(t_i), i = 1, 2, \dots, N\}$. This file is used to display the cumulative curve and magnitude versus transformed time $\Lambda(t_i)$. The calculated record of the program [**retas**] is stored by the name [**retas.print**] in the directory of [**Calculations**] for your initial check of the program.

For the users who have the free graphic software **R**, see Section 3 (Graphs) below. I have written a R-module [**r.retas**] to display them, in addition to the module [**r.seis**] for displaying the cumulative curve and magnitude v.s. ordinary time t_i . If the observed rate of occurrence is compared with the calculated one from the model, period of decreased or increased seismic activity (relative quiescence or activation) can be recognized. [**r.seisetas**] See section 3 (Graphs) below.

2. Programs and data

2.1 Exact and approximate algorithms

A FORTRAN program [**etas.f**] consists of two (exact and approximated) versions of the calculation algorithm for the maximization of likelihood. The approximated one (Ogata et al., 1993) to reduce the processing time for relatively large data sets (number of quakes N larger than several hundreds). The processing time is proportional to N^2 for the exact version but proportional to N for the approximation version. The approximated version uses the subroutine [**func9**] which runs at one of the five levels 1, 2, 4, 8, and 16. The higher level means faster processing but lower accuracy. At levels 1 or 2, the results are very close to the exact solution by the exact version which uses the subroutine [**func4**]. The subroutines and the levels are selected by the control file [**etas.open**], in addition to other input variables such as the range of data components that restrict data and initial values of parameters.

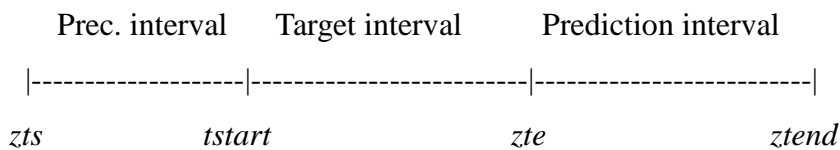
The maximum number of earthquakes can be processed by ETAS and RETAS is 17,777. The data file may contain as many as 17,777 quakes based on the set dimensions of the ETAS program.

Starting from a given set of initial guess of the parameters that is given in the control file [**etas.open**], the program [**etas.f**] repeats calculations of the function values and its gradients at each step of parameter vector. At each cycle of iteration, the linearly searched step (**lambda**), negative log-likelihood value (**-LL**), and two estimates of square sum of gradients are shown. At each linear search of the likelihood computation, **-LL** and the 5 parameter values are shown. The value **-LL** decreases and tends to a final value. With the convergence of **-LL**, one of the sum of squared gradients becomes nearly zero and the iteration is terminated. The time of the termination is

displayed. In the present case, the following results are displayed, where $AIC = -2 \mathbf{LL} + 2 \times 5$. If $-\mathbf{LL}$ value and gradients do not seem to converge, we recommend you to use the last parameter estimates as the initial values in the control file [**etas.open**] and restart the execution.

2.2 Precursory and target intervals

The period for which the ETAS model parameters are computed is called target interval. The seismicity in this period may be affected by earthquakes which occurred before this period due to the long-lived nature of aftershock activity. To consider this effect, a time interval precursory to the target interval (called precursory interval) is chosen and aftershock activities following earthquakes in this period are taken into computation. When the ETAS parameters are estimated for two or more successive periods (usually divided at turning points of seismicity), it is better to set a precursory period before each target period (see Ogata, 1992, for instance).



2.3 Data file [*work.etas*]

The data file for ETAS should have a sequential file with a 9 columns-format of the sequential number, longitude, latitude, magnitude, time from the mainshock in days, depth, year, month, and day, as is given by the prototype data file [**work.etas**]. The time is usually measured in days. Time data t_i in the 5th column should be given at least to the fourth place of decimals (in double precision), since the differences between them are important. The first row of the data is the title. Any copy of [**work.etas**] is saved by the name [***.etas**] where * is an alphabetical word.

2.4 Control file [*etas.open*]

Furthermore, the file [**etas.open**] controls the data reading such as [**work.etas**] and other input variables such as the range of data components that restrict data and initial values of parameters, $ztstart$, $ztend$, which represent the beginning and the end of the interval for which data are prepared, respectively. Any copy of [**etas.open**] is saved by the name [**etas.***] where * is an alphabetical word, corresponding to the data [***.etas**].

The first row consists of two numbers; either 4 or 9 for the selection of the subroutine [**func4**] or [**func9**], respectively, and then one of 1, 2, 4, 8, or 16 when [**func9**] is used; when [**func4**] is used any number should be set.

The second row consists of three numbers; zts , zte and $tarst$ in the above illustrated variables when the file is applied to the program [**etas.f**]. When the file is applied to the program [**retas.f**] for the diagnostic outputs, the last number can be $ztend$ ($\geq zte$) that is the last time of the period of available data. The example shows the first 6.5days of aftershock period since the mainshock due to [**work.etas**], where the target interval is [0.03, 6.5]days and the data of first 0.03days period is only used as the history of the ETAS model.

The third row consists of two numbers; threshold magnitude M_{th} and reference magnitude M_z . Here I have taken the magnitude of the mainshock as the reference magnitude, but you may take $M_z = M_{th}$ for general seismicity. Note that the MLE of the parameter K differs depending on the reference parameter.

The fourth row provides the initial estimates of the five parameters μ , K , c , and p when this is applied the program [**etas.f**], and this is the last row when we use for the program [**etas.f**]. If you set $\mu = 0.0$ in the initial value, this variable remains 0.0 throughout the calculations. Also, if you set $p = 1.0$ (the original Omori formula), this variable remains 1.0 throughout the calculations. When this control file is used for the program [**retas.f**] to make the diagnostic outputs, these values in this row have to be the MLEs estimated by the program [**etas.f**].

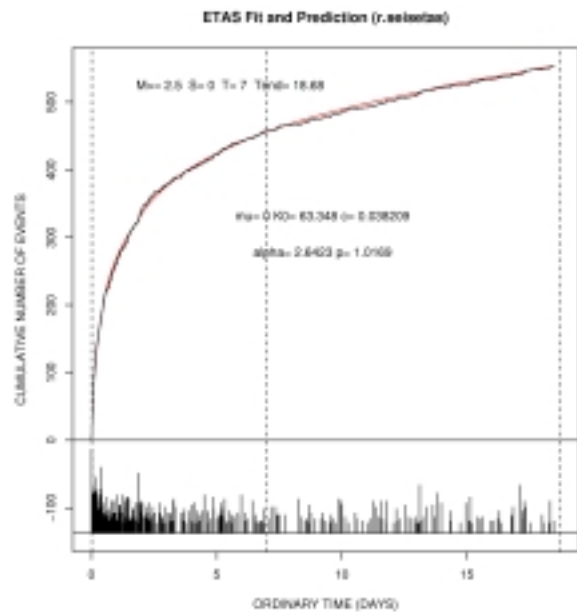
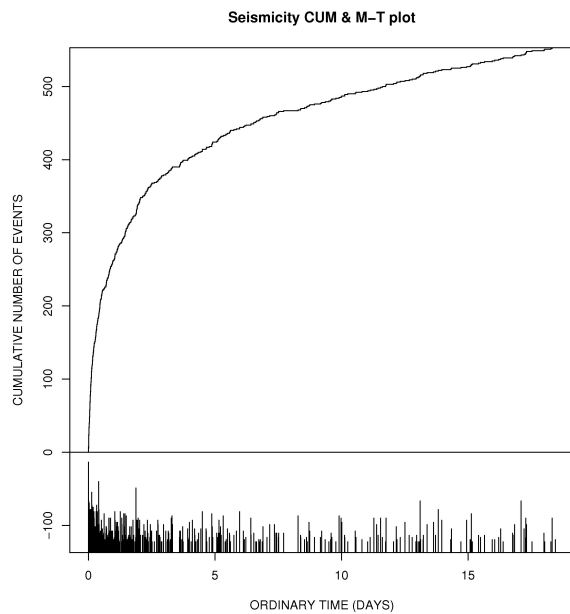
The fifth row consists of three numbers; zts , zte and $tarst$. This row is only used when the file is applied to the program [**retas.f**] for the diagnostic outputs, the last number must be zte that is the end of the target period.

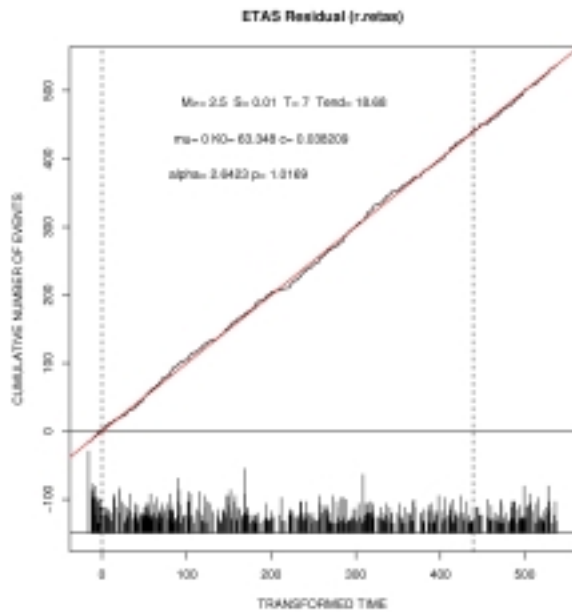
The sixth row is the same as the fourth row, only used for the program [**retas.f**], and the seventh

row of the value of the negative of the maximum likelihood value corresponding to the MLEs, used just for the user' note to compare with other cases or models.

3. Graphs

The R language modules [**r.seis**], [**r.seisetas**] and [**r.retas**] are used to plot the graphs of cumulative number and magnitude of earthquakes against the ordinary time and transformed time, respectively. The transformation is made in such a way that the theoretical cumulative frequency curve becomes a straight line. These use the files [**work.etas**] and [**work.res**], respectively. The last graph includes printed number with certain time values for vertical dotted lines showing target interval and MLEs in control file [**etas.open**]. The outputs of the graphs in postscript format, named as **seis.ps**, **seisetas.ps**, and **retas.ps** are given in the same directory.





The R language modules [**r.retas**] and [**r.seisetas**] shows the theoretical curve of cumulative frequency (red color) which is a straight line by definition of the transformed time, and the cumulative frequency curves of the data (red color).

The graphical free software R is also installed in the other directory based on the Microsoft Windows XP (RGui). To command in the R console, write

```
> source('r.*')
```

where * correspond above R-module, then the graphical window should appear and draw appropriate figures.

4. Remarks

4.1 Application to aftershock sequences

When the ETAS model is applied to an aftershock sequence, the main shock should always be included in the data. Since the primary seismicity (background seismicity) μ is usually absent in the aftershock sequence data, it may be preferable to use a model with $\mu = 0$. To constrain μ to be 0, set 0.0 as the initial value for m in the control file [**etas.open**].

4.2 Abnormal termination

If the data are very different from the proposed model, an overflow error may occur or the solution may not converge within the fixed number of iterations ($30N$, where N is the number of parameters). The results in this case are often incorrect and not recommended to adopt as the estimates.

4.3 Test data set

The test data made from the JMA earthquake catalogs named as [**work.etas**] is attached, which is same as [**miyagiN.etas**]. This is the aftershock data of 26th July 2003 earthquake of M6.2 at the northern Miyagi-Ken, northern Japan. There is another dataset [**fukuokaW.etas**], which is the aftershock data of 16th August 2005 earthquake of M7.2 at the offshore of western Fukuoka-Ken, Kyushu, Japan.

The FORTRAN programs are originally designed and programmed (1985) by Yosihiko Ogata, Institute of Statistical Mathematics, Tokyo, Japan. The R-module was designed and programmed by Yosihiko Ogata (2003).

References

- Ogata, Y., Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statist. Assoc.*, **83**, No.401, 9-27, 1988.
- Ogata, Y., Statistical model for standard seismicity and detection of anomalies by residual analysis for point process, *Tectonophysics*, **169**, 1-16, 1989.
- Ogata, Y., Detection of precursory relative quiescence before great earthquakes through a statistical model, *J. Geophys. Res.*, **97**, 19845-19871, 1992.
- Ogata, Y., R.S. Matsu'ura, and K. Katsura, Fast likelihood computation of Epidemic Type Aftershock Sequence model, *Geophys. Res. Lett.*, **20**, 2143-2146, 1993.
- Utsu, T., Y. Ogata, and R. S. Matsu'ura, The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth*, **43**, 1-33, 1995.

Program ETASIM

These programs produce simulated data files for given sets of parameters in the point process model used in ETAS. See Ogata (1981) or Ogata (1998) for theoretical basis. It is noted that the intensity defined by a combination of parameter values should be non-negative and well-defined. Due to some combinations of parameter values, the simulated data can be explosive. The simulated events are given in the file [**work.etasim**] in the same format as [**work.etas**].

Control file [**etasim.open**].

The parameters of the ETAS model and other necessary parameters should be given in the control file [**etasim.open**]. There are two options; either simulating magnitude by Gutenberg-Richter's Law ($ic = 0$) or using magnitudes from [**work.etas**] ($ic =$ any number except 0, say 1). For the first option, you have to provide b -value of G-R law and number of events to be simulated. The second option simulates the same number of events that are not less than threshold magnitude in the data [**work.etas**], and simulation starts after a precursory period depending on the same history of events in [**work.etas**] in the period. Therefore the occurrence times of simulated events in [**work.etasim**] is the same as those in [**work.etas**] during the precursory period.

The first row consists of two numbers, the first number is ic and the second number is b -value. If ic is not 0 second number can be any integer (dummy number).

The second row consists of two numbers, the first number is $tstart$ where $[0, tstart]$ is precursory period. The second number is the number of the simulated events if $ic = 0$, otherwise dummy number.

The third row consists of two numbers, the first number is cutoff (threshold) magnitude of the simulated data, and the second number is the reference magnitude to calculate the intensity rate of the ETAS model. This should be the same reference magnitude given in [**etas.open**]. For example, this could be the magnitude of the mainshock in case of aftershock simulation.

The fourth row consists of five numbers of the ETAS parameters, which are μ , K , c , α and p in the order from the first to the fifth number.

Calculated record of the program [**etasim**] is stored by the name [**etasim.print**] in the directory of [**Calculations**] for your initial check of the program.

The FORTRAN programs are originally designed and programmed (1985) by Yosihiko Ogata, Institute of Statistical Mathematics, Tokyo, Japan. The R-module was designed and programmed by Yosihiko Ogata (2003).

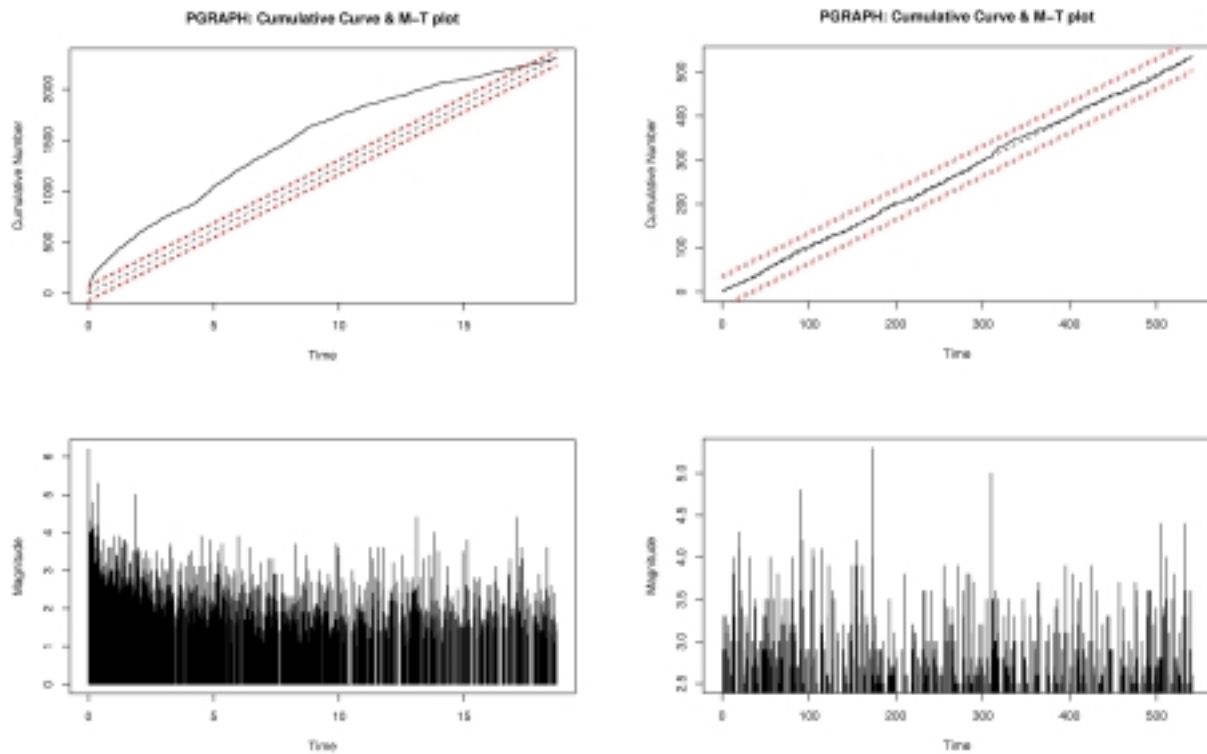
References

- Ogata, Y, On Lewis' simulation method for point processes, *IEEE Information Theory*, IT-27, 23-31, 1981.
- Ogata, Y. (1998) Space-time point-process models for earthquake occurrences, *Ann. Inst. Statist. Math.*, **50**, 379-402.

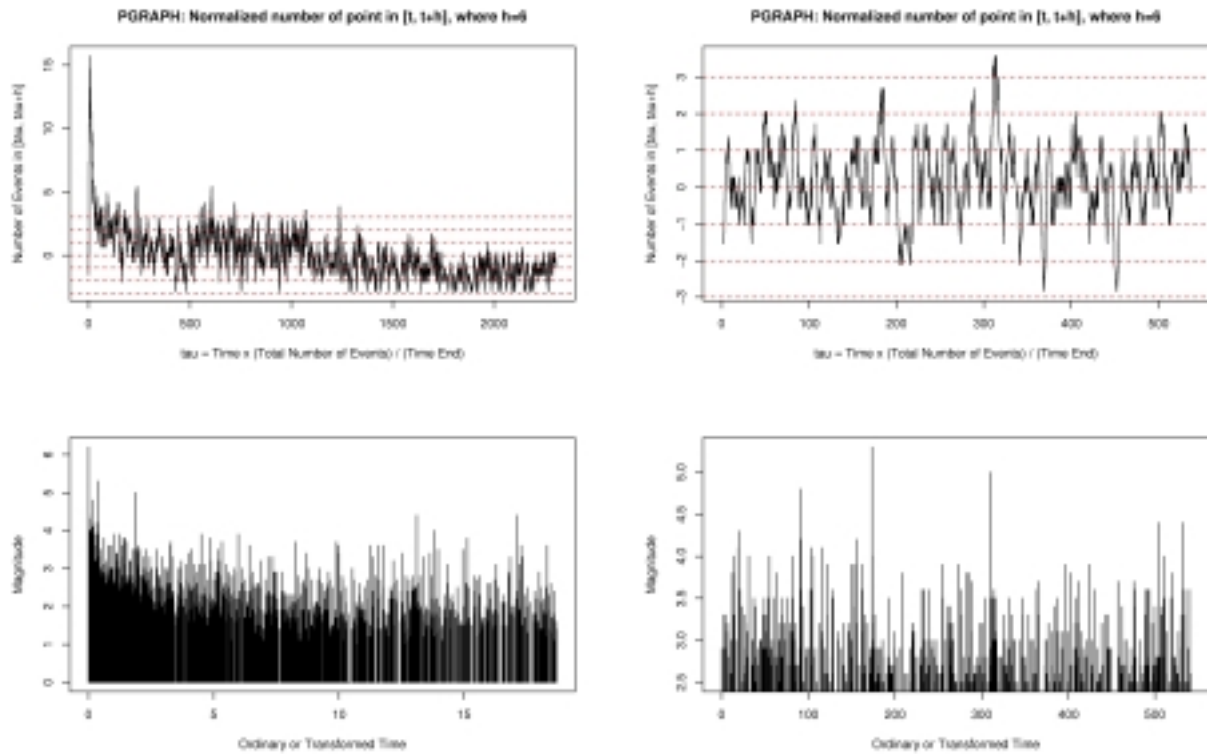
Overview of the program [pgraph]

This provides the following several graphical outputs for the point process data set.

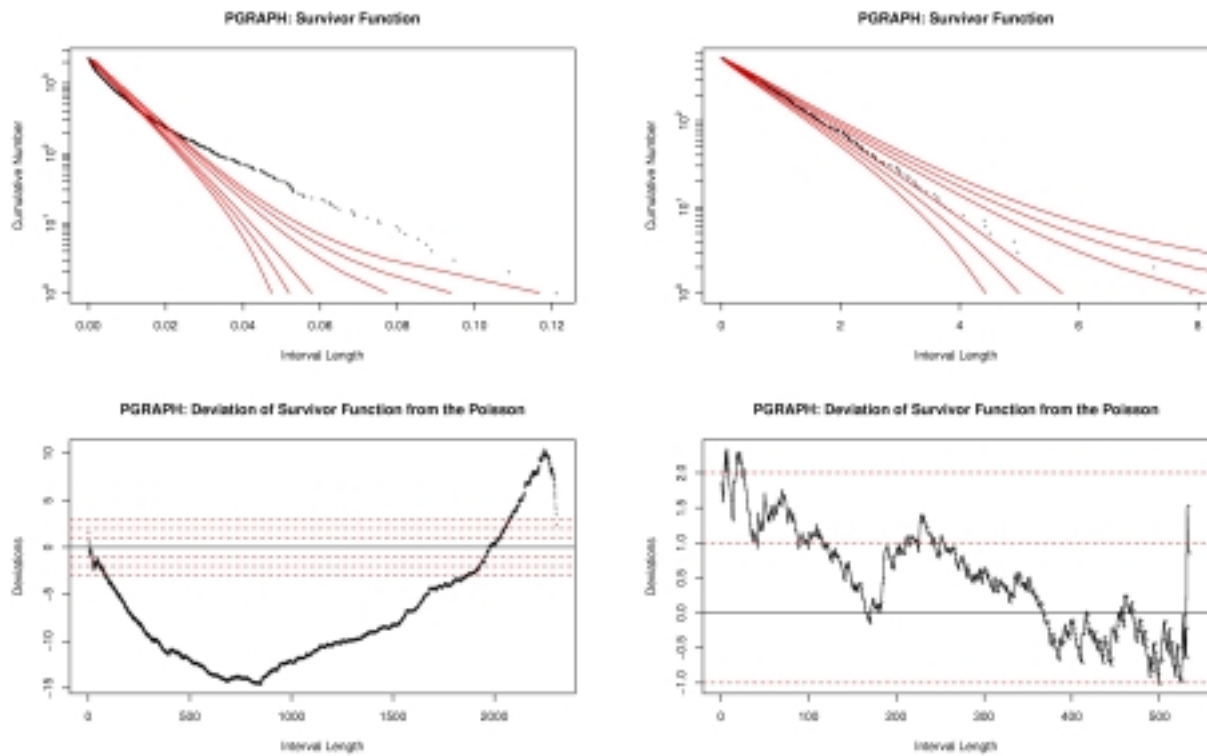
1. Cumulative numbers of events versus time, and positions of spikes (subroutine cumlat), and output file is [**out.pgCumMT**]. The R-module [**r.pgCumMT**] produces these figures in both console and postscript [**pgCumMT.ps**].



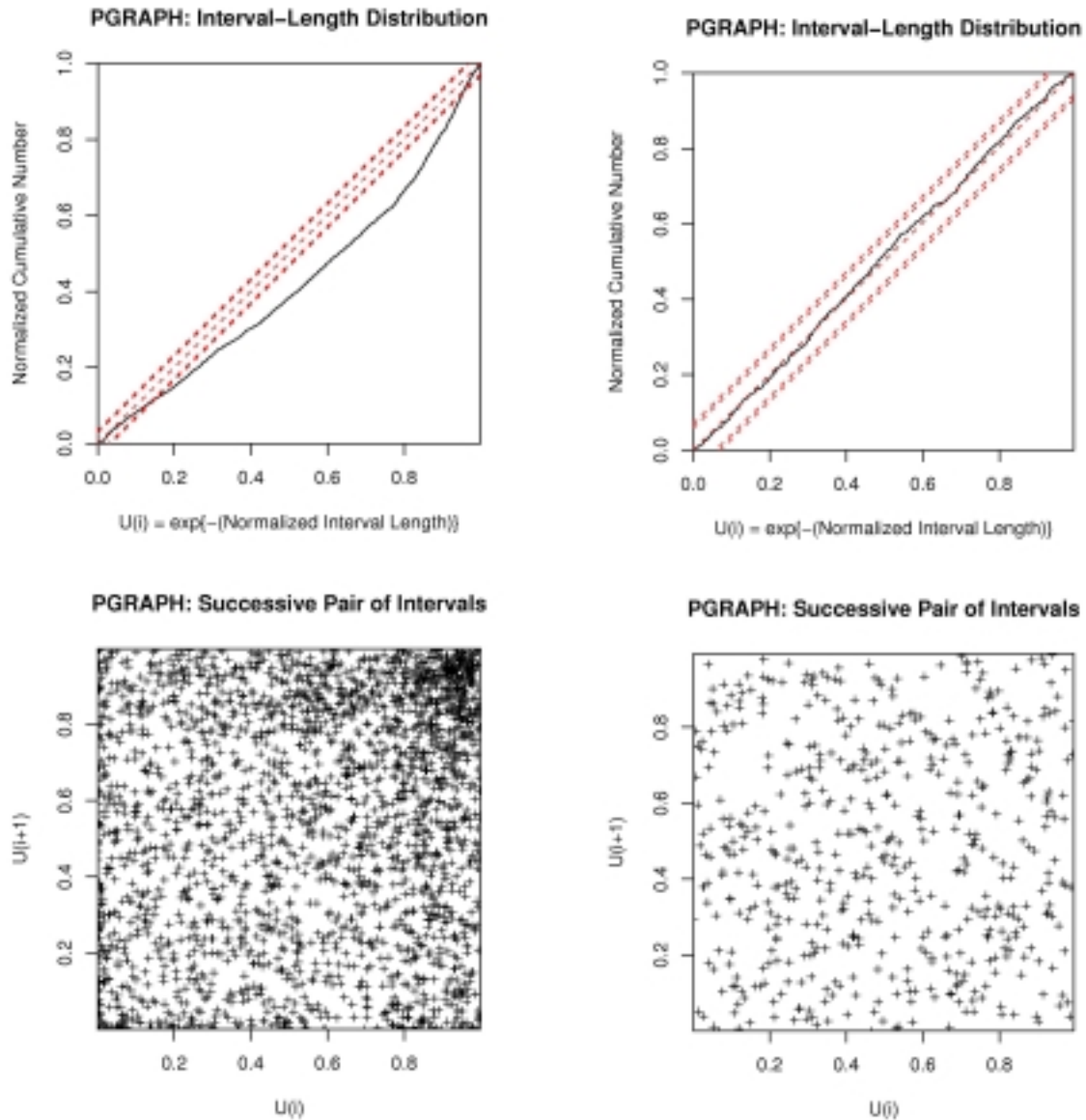
2. Series of counted events in the moving interval with a fixed length. This is a kind of moving average. Dotted lines indicate $i \times$ (standard error), $i = 1, 2, 3$, assuming the stationary poisson process (subroutine count1 or count2), and output file is [out.pgPTnum]. The R-module [r.pgPTnum] produces these figures in both console and postscript [pgPTnum.ps].



3. The log survivor curve with $i \times$ (standard error), $i = 1, 2, 3$, assuming the stationary poisson process, and the similar graph in which (x, y) plots are rotated and shifted in such a way that the standard error lines and expectation lines are parallel (subroutine `surviv`), and output file is `[out.pgSurviv]` and `[out.pgSurDev]`. The R-module `[r.pgSurviv]` produces these figures in both console and postscript `[pgSurviv.ps]`.

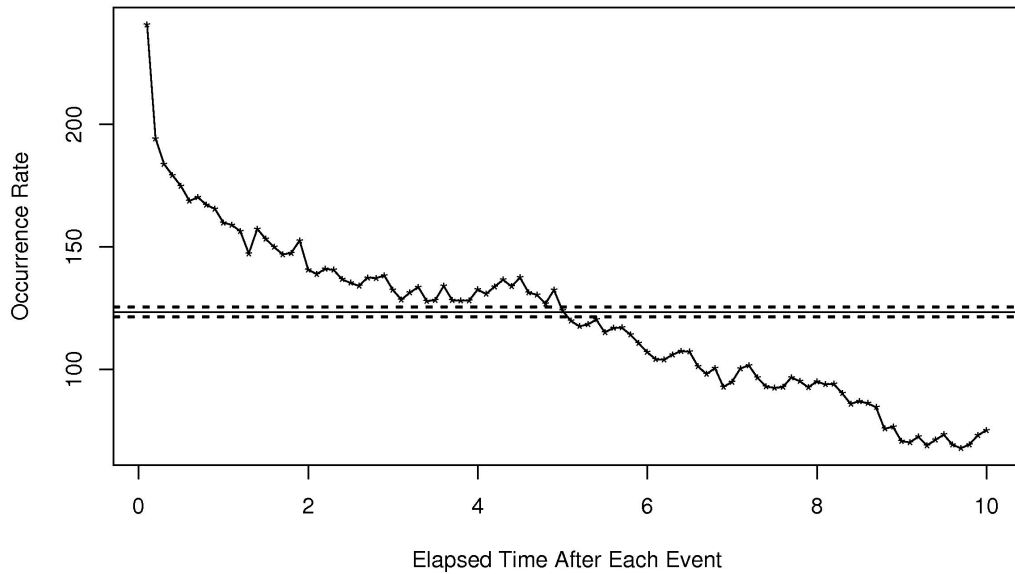


4. Empirical distribution of $u(i) = \exp\{-\mu x(i)\}$ where μ is mean occurrence rate and $x(i)$ is the i -th interval length between consecutive points, and lines of .95 and .99 significance bands of the two-sided Kolmogorov-Smirnov test assuming the uniform distribution are given. The related graph of $(u(i), u(i+1))$ plots are also carried out (subroutine `unifrm`), and output file is `[out.pgInter1]` and `[out.pgInter2]`. The R-module `[r.pgInterP]` produces these figures in both console and postscript `[pgInterP.ps]`.

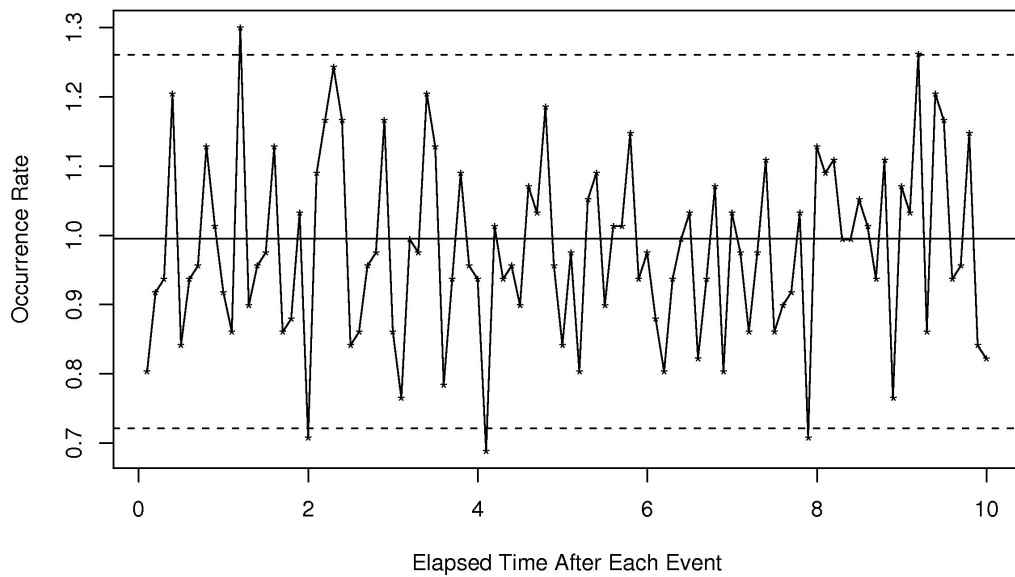


5. Estimated intensity $m_f(t)$ under the palm probability. This is related to the covariance density $c(t)$ by the relation of $c(t) = \mu^* \{m_f(t) - \mu\}$, where m is the mean intensity of the point process. The 0.95 and 0.99 error bands are provided assuming the stationary Poisson process (subroutine `palmxy`, `palmpr`), and output file is [**out.pgPalm**]. To make figure R-module [**r.pgPalm**] is used and also produces postscript output [**pgPalm.ps**].

PGRAH: Palm Intensity

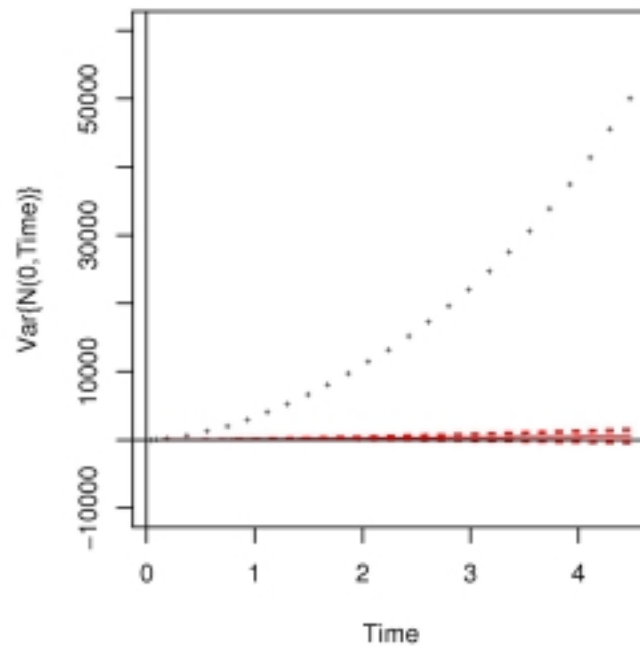


PGRAH: Palm Intensity

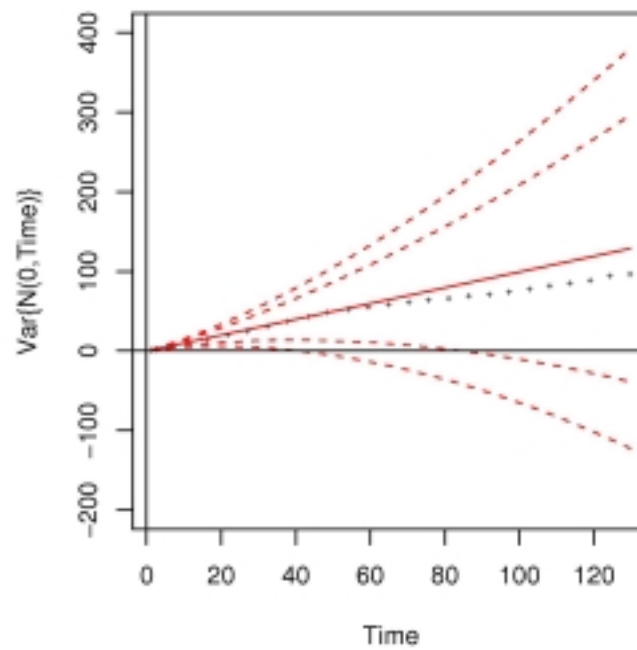


6. Estimation of variance time curve with the 0.95 and 0.99 error lines assuming the stationary Poisson process (subroutine `vtcxyp`, `vtcprt`), and output file is [`out.pgVTC`]. The R-module [`r.pgVTC`] produces these figures in both console and postscript [`pgVTC.ps`].

PGRAPH: Variance-Time Curve



PGRAPH: Variance-Time Curve



7. Subroutine structure of the FORTRAN program [**pgraph.f**]

[pgraph]

```
|-----[input]
|-----[count1]----[shimiz]
|-----[surviv]-----[errbr2]-----[plsinv]
|           |-----[unifrm]----[unitsq]
|
|-----[palmpr]
|-----[vtc]
|-----[vtcpert]
```

8. The name of the analyzing data and its format is either [**work.etas**] or [**work.res**] and their format. The choice of either data is controlled by the following control file. The range of x-coordinate is limited to the non-negative valued variable.

9. Control file is [**pgraph.open**] in which necessary variables to read are given and explained.

10. Calculated record of program [**pgraph**] is stored by the name [**pgraph.print**] and [**pgraphRes.print**] corresponding to the dataset [**work.etas**] and [**work.res**], respectively, in the directory of [**Calculations**] for you to check of the execution program.

This program was originally designed (January 1985) and revised (December 2005) by Yoshihiko Ogata, and programmed and also reprogrammed by Koichi Katsura. The R-module was designed and programmed by Jiancang Zhuang and Ogata (December 2005).

References

Cox, D.R. and Lewis, P.A.W. (1966). *The Statistical Analysis of Series of Events*, Methuen & Co. Ltd, London.

The Institute of Statistical Mathematics

**4-6-7 Minami-Azabu, Minato-ku
Tokyo 106-8569, Japan**