Some reasoning on the improvement of the ETAS modeling at the occurrence of the 2016 Central Italy seismic sequence

Anna Maria Lombardi

Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

annamaria.lombardi@ingv.it

Abstract

This study presents an application of the ETAS model to the first 20 days of the 2016 Central Italy sequence. Despite of the provisional nature of data, the model is able to describe the occurrence rate, but for the first hours after the mainshock occurrence. A sensitivity analysis of the model to two uncertainty sources, the model parameters and the occurrence history, shows that the second has a main role in controlling the performance of the ETAS model. Previous results, together with the clear inability of ETAS to forecast the occurrence of a sequence before its starting time, give important suggestions about possible improvements.
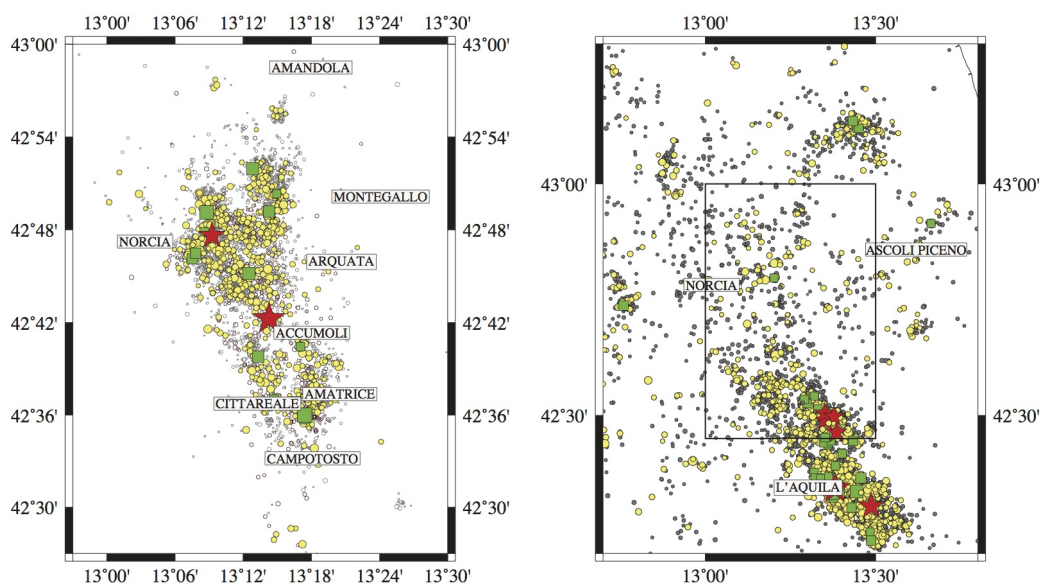
I. Introduction

The use of time-dependent models for practical purposes has been introduced recently in Italy to forecast seismicity (Marzocchi et al. 2012; 2014). Time variations of the seismic rate are more evident over the short-term (days to weeks), and in particular after a large earthquake. The modeling of these time variations, and its use for practical purposes, led to the development of operational earthquake forecasting (OEF), a set of procedures for gathering and disseminating information about the time dependence of seismic hazards (Marzocchi et al., 2014).

Case studies may serve as a good basis to develop OEF efficient procedures, able both to forecast the spatio-temporal evolution of sequences and to improve the knowledge about the main events. This paper is a case study of the 2016 Central Italy sequence, still on-going, by mean of the Epidemic Type Aftershock Sequences (ETAS) model (Ogata, 1998; Lombardi, 2015; 2016a; 2016b).

Table 1: *ETAS parameters*

	INGV Bulletin (1478 events) $ML \geq 2.5$	CSIv1.1 (391 events) $ML \geq 3.0$	CPTI15 (85 events) $Mw \geq 5.0$
μ	2.0e-02	2.7e-03	3.9e-04
	(1.9, 2.2) e-02	(7.3, 3.2) e-03	(3.6, 4.2) e-04
k	4.7e-02	4.8e-02	1.8e-02
	(4.4, 5.2) e-02	(4.4, 6.1) e-02	(1.1, 2.5) e-02
p	1.11e+00	1.10e+00	1.0e+00
	(1.10, 1.14) e+00	(1.09, 1.13) e+00	(0.9, 1.1) e+00
c	1.2e-02	1.4e-02	9.0e-02
	(1.1, 1.7) e-02	(1.1, 2.3) e-02	(0.2, 1.0) e-01
α	1.2e+00	1.1e+00	8.0e-01
	(1.1, 1.3) e+00	(0.8, 1.2) e+00	(0.4, 1.2) e+00
d	7.0e-01	1.2e+00	3.7e+00
	(6.0, 7.0) e-01	(1.1, 1.5) e+00	(2.3, 5.7) e+00
q	1.57e+00	1.55e+00	2.2e+00
	(1.55, 1.61) e+00	(1.52, 1.62) e+00	(2.1, 3.0) e+00
γ	3.6e-01	~ 0	1.0e+00
	(3.1, 4.0)e-01		(0.7, 1.5) e+00
b	1.1	0.90	0.9
	(1.05, 1.15)	(0.85, 0.95)	(0.8, 1.0)
$LogL$	-8263	-2852	-1406
	(-8266, -8263)	(-2854, -2852)	(-1408, -1406)

Figure 1: *Map of the seismicity occurred in the area interested by the 2016 Central Italy sequence. Gray and yellow circles mark the events with magnitudes ML < 2.5 and 2.5 ≤ ML < 4.0, respectively. Green square and red stars marks the events with magnitude 4.0 ≤ ML < 5.0 and ML ≥ 5.0, respectively. The left panel show the first 20 days of the 2016 sequence (August 24 - September 15) in the area struck by the sequence [12.95-13.5E, 42.45-43.0N]. In the right panel the seismicity occurred before the 2016 sequence (April 16 2005 - August 24 2016) in the larger area [12.7-13.8E, 42.0-43.5N], used to estimate the ETAS model, is shown (the square identifies the area interested by the 2016 sequence). The strongest events in the southern part refer to the 2009 L'Aquila sequence.*

It has two main goals: to check the performance of the model on very provisional data and to provide some points of interest for a possible improvement of the model.

II. Estimating the ETAS Model

On August 24, 2016, at 01:36 UTC, a magnitude ML 6.0 earthquake hit part of the Central Italy area, close to the village of Amatrice and Accumuli, causing a number of human losses and significant economic damages.

The evolution of the seismic sequence, after the ML 6.0 event, is modeled in the present study through the ETAS model. The seismicity data used here are downloaded from the site of the official INGV bulletin (www.iside.rm.ingv.it), from April 16 2005, day marking the start-up of a new seismic network, up to September 15 2016. I estimate the ETAS model on seismicity occurred in the region

[12.7-13.8E, 42.0-43.5N] (in the following called estimating region), including the whole 2009 L' Aquila sequence (see Figure 1). This region includes the area [12.95-13.5E, 42.45-43.0N], at the present interested by the 2016 Central Italy sequence (in the following called testing region).

I estimate the ETAS model by using the data from April 16 2005 to August 24 2016, occurred in the estimating region, with magnitude above ML2.5. More than 60% of events belongs to the 2009 L'Aquila sequence. The choice of the threshold magnitude is due to the incompleteness of data below ML2.5, soon after the occurrence of the strongest event of the 2009 L'Aquila sequence (Lombardi, 2016a).

The version of the ETAS model used here has been implemented in SEDAv1.0 (Statistical Earthquake Data Analysis), a statistical software freely provided via the Zenodo open access platform
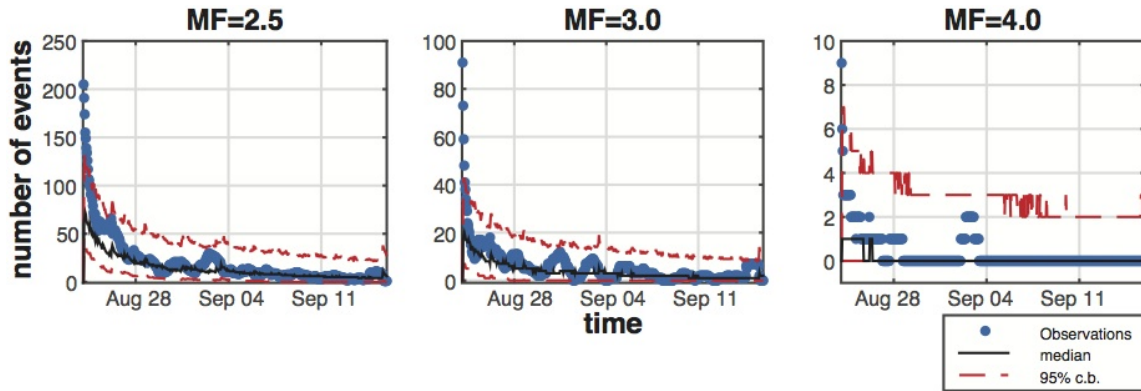
Figure 2: *Comparison between the observed and the expected number of events for the threshold forecast magnitude M_F= 2.5, 3.0 and 4.0 for the first 20 days of the sequence (Aug 24 -Sep 15)*

(https://zenodo.org/record/55277; Lombardi, 2016a; Lombardi, 2016b).

For the present study, I use a background grid covering the testing area and with a step of 0.01^o. By applying the SEDAv1.0 estimating tool (based on the maximum likelihood criterion), I found the parameters reported in Table 1, by mean of 1000 runs (Lombardi, 2015).

The numbers in the brackets are the 95% confidence bounds of the parameters values founded in all runs. They have a double meaning: to represent the epistemic uncertainty of the model and to summarize a possible correlation of model parameters, causing a possible multimodality of log-likelihood function.

III. Measuring the performance of the ETAS model

I make two types of analysis, separately. In each of them I measure the sensitivity of the ETAS model to its two uncertainty sources: the occurrence history and the parameters (including the background spatial distribution).

In the first analysis I check the sensitivity of the ETAS model to the occurrence history and I mimic real-time forecast calculations, by using the better ETAS parameters, listed in Table 1, estimated from the Bulletin data. Specifically, I compare the observed number of events with what expected by the ETAS model, in the testing area (Number of events test, Lombardi,

2016a). Specifically, I simulate 10000 ETAS time histories, for N_D overlapping interval times of one day $\{D_i, i = 1, ..., N_D\}$, updated each hour, starting from the time occurrence of the mainshock (August 24 2016, 01:36:32, ML6.0) up to September 15 2016. The simulations are done by the tool implemented in SEDAv1.0 and based on the thinning method (Ogata, 1998; Lombardi 2016b). For each D_i, I import the actual history occurred up to the starting time of D_i and I simulate the events inside. This test allows to both quantify the agreement between model and observations and to measure the sensitivity of the model to the history H_t. In this first step the parameters sensitivity is not investigated.

In Figure 2 I show the comparison between the expected and the observed number of events with magnitude above the threshold forecast magnitude M_F=2.5, 3.0 and 4.0. Specifically, I show the median expected number of events (black line) together with the 95% confidence bounds (red lines), to quantify the uncertainty of predictions. The confidence interval defined by the ETAS model contains the observed number of events with $ML \geq 2.5$, except in the first 8 hours of the sequence. Anyway, the observations are above the median forecasts uninterruptedly in the first three days, showing a systematic underestimation. After this period, the observations (blue points) and the median forecasts (black line) overlap. These
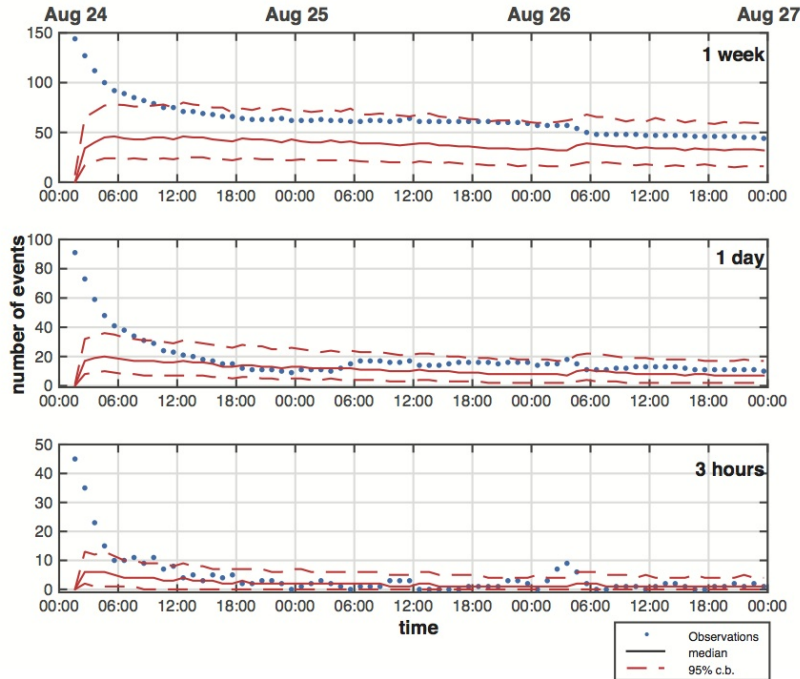
Figure 3: *Comparison between the observed and the expected number of events for M_F= 3.0 and for interval times of 1 week, 1 day and 3 hours for the first 3 days of the sequence (Aug 24-27)*

results improve for larger values of M_F, for which the period of underestimation is smaller (Figure 2).

Previous results do not depend on the length of the forecast interval D_i. None significant difference is found for interval times D_i of 3 hours, 1 day and 1 week, in the first 3 days of the sequence (Figure 3). The only difference is that the observed number of events is steadily above the median, for intervals D_i of 1 week.

Previous calculations are a measure of the agreement between the actual history and what expected by the model: this last is fixed, whereas the history changes among simulations (Figures 2 and 3). The second methodology is a retrospective forecast analysis devoted to check the sensitivity of the ETAS model to the input parameters (including the background spatial distribution). Specifically, I compute the expected number of events N_F for each of 1000 models obtained by applying the SEDAv1.0 estimating tool (see previous section),

including the actual time history up to the end of the forecast interval. The calculations are done for interval times of 1 day, updated each hour, and for M_F=2.5, integrating the conditional intensity of the ETAS model (Lombardi, 2015). As Figure 4 shows, the parameter sensitivity of the model is negligible. The better approximation of observations in the first hours of the sequence, respect to previous analysis (see Figure 2), is due to inclusion of the whole actual history in calculations.

By keeping in mind the provisional nature of the earthquakes data, previous analyses show that the ETAS model is able to describe the sequence, except at the beginning of the sequence. Clearly, the ETAS model does not have any predictive ability before the main event. The probability to have an event above ML5.5 in the next day, computed at 00:00 of Aug 24, is 10^{-5} in the whole area, mainly given by the background rate. Improving the forecast capability before the occurrence of a sequence, of the ETAS or similar stochastic models, is a
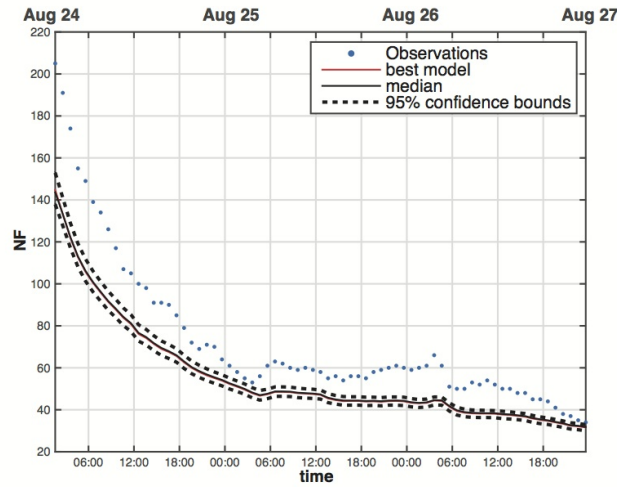
Figure 4: *Parameters sensitivity analysis for the ETAS model. The plot compares observations and forecasts for M_F= 2.5 and for the first 3 days of the sequence (Aug 24-27)*

challenge to which the seismological statistical community must respond.

IV. IMPROVING THE MODEL: THE INTEGRATION OF DIFFERENT DATA

The improving of the forecast capability of the ETAS model is a very difficult process that cannot surely be resolved in the present study. This section is devoted to discuss some very preliminary results, derived by the comparison of different types of earthquakes catalogs. The inefficiency of the ETAS model to predict imminent sequences, following the occurrence of strong events, mainly derives from the poor modeling of the background. Firstly, the poissonian model neglects possible long-term temporal variations of the seismic rate. Second, the temporal window covered by instrumental catalogs should be too small to infer the actual spatial distribution of background.

Regarding the 2016 sequence, I compare the informations coming from three different catalogs: the official INGV bulletin, used before, the instrumental CSIv1.1 (Castello et al. 2007; http://csi.rm.ingv.it/) and the historical CPTI15 catalogs (Rovida et al., 2016; http://emidius.mi.ingv.it/CPTI15-

DBMI15/). The CSIv1.1 is available for the period 1981-2002, but I use the data from July 1 1984, above ML3.0, doing to network changes during the early 1980s. The CPTI15 collects the strong earthquakes occurred in Italy from 1000 to 2014, but I consider the events above Mw5.0 after 1600. Figure 5 is a comparison of the background estimation for the three catalogs. Table 1 reports the ETAS parameters for all of them. Both the INGV Bulletin and the CSIv1.1 catalogues identify the areas around Norcia and Campotosto cities as the most hazardous regions. Anyway, the size and the bounds of these areas are defined by the sequences included in the data (the 1997-1998 Colfiorito sequence, at north-west, for the CSIv1.1, and the 2009 L'Aquila sequence, at south-east, for the INGV Bulletin). The small size of the CPTI15 catalogues does not allow to reach detailed results. It seems to confirm what obtained from the other two catalogues, but assign a larger relative seismic potential to the area around Accumuli, struck by the 2016 earthquake.

In summary, this analysis shows that the instrumental catalogs might cover a too small temporal window to infer a reliable spatial distribution of background, that might result driven by the occurrence of sequences.
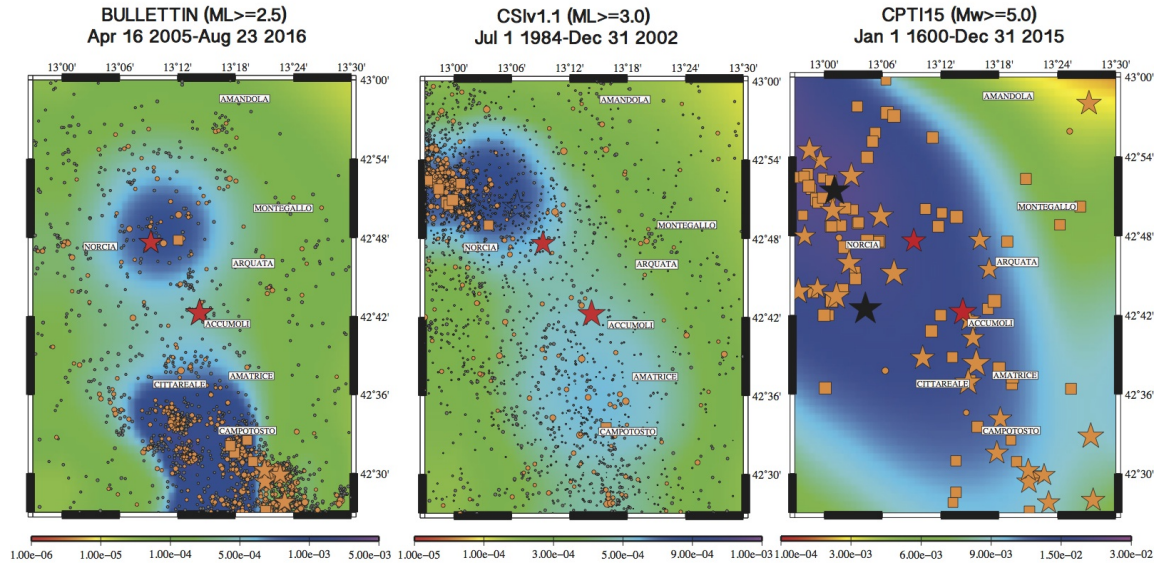
Figure 5: *Comparison between the background spatial probability distributions obtained from three catalogs: the instrumental official INGV Bulletin, the CSIv1.1 and the historical CPTI15 catalogue. Gray circles mark the events below magnitude 2.5; orange circles, squares and stars mark the events with magnitude M < 4.0, 4.0 ≤ M < 5.5 and 5.5 ≤ M < 6.5, respectively. Black stars marks the events with M ≥ 6.5. The red stars mark the ML 6.0 and ML5.4 events of the 2016 Central Italy sequence, respectively, both occurred at August 24. The magnitude scales are ML for the INGV bulletin and CSIv1.1 and Mw for CPTI15. The right panel shows the events above Mw4.0, for consistency with the other panels, but the background rate is estimated on events above Mw5.0.*

V. Conclusions and perspectives

The preliminary results shown here highlight some important issues:

- The ETAS model is able to describe the first part of the 2016 Central Italy sequence, also with very preliminary data. The initial underestimation might be due to both the provisional nature of data and to some inefficiencies of the ETAS modeling (Figure 2).

- The parameter sensitivity of the ETAS model is negligible respect to impact of the occurrence history (Figure 2, 3 and 4).

- The inefficiency of the ETAS model to predict the occurrence of sequences may be due to poor background modeling. By a mathematical point of view, the ETAS model identifies a branching struc-

ture for the earthquakes. So, the background events are the earthquakes not triggered by previous ones. To give a seismological interpretation to this definition of background is not straightforward. Specifically, the instrumental catalogs could cover a too small temporal window to infer a reliable background spatial distribution, unless we assume a time-dependence of the background. But this last hypothesis is contrary to Poisson (time-independent) modeling of background itself, adopted by ETAS models.

- The SA algorithm used here allows the quantification of parameters uncertainties. These last represent both the ability of the algorithm to identify the best model from data and the correlation of parameters giving indistinguishable models (by using the maximum log-likelihood

criterion). This point puts in question the physical significance of the ETAS model and will be further investigated.

- To provide more accurate forecasts soon before the sequences occurrence, we need to improve the stochastic models as ETAS. A correct integration of seismological, geological and geodetic data is likely the best way to improve the spatio-temporal resolution of stochastic forecast models.

- The present study is a preliminary analysis of the first 20 days of a very complex sequence, marked by three mainshocks in two months of moment magnitude 6.0 (August 24), 5.9 (October 26) and 6.5 (October 30), respectively. The check of the ability of ETAS modeling to explain the spatio-temporal evolution of the sequence deserves more investigation and will be the topic of future work.

References

[Castello et al., 2007] Castello B. , Olivieri M. and Selvaggi G.. (2007). Local and duration magnitude determination for the Italian earthquake catalogue (1981-2002). *Bull. Seism.Soc.Am.*, 97(1B), 128-139.

[Lombardi, 2015] Lombardi A.M. (2015). Estimation of the parameters of ETAS models by Simulated Annealing *Sc. Rep.*, Feb 12;5:8417. doi: 10.1038/srep08417.

[Lombardi, 2016a] Lombardi A.M. (2016). SEDA a software package for the Statistical Earthquake Data Analysis: a tutorial application to the 2009 L' Aquila and the 2012 Emilia (Italia) sequences *Ann. Geoph.*, submitted.

[Lombardi, 2016b] Lombardi A.M. (2016). SEDA a software package for the Statistical Earthquake Data Analysis *Sc. Rep.*, submitted.

[Marzocchi et al.,, 2012] Marzocchi, W., A. Amato, A. Akinci, C. Chiarabba, A.MLombardi, D. Pantosti and E. Boschi (2012). A ten-year earthquake occurrence model for Italy. *Bull. Seism. Soc. Am.*, 102, 1195-1213.

[Marzocchi et al.,, 2014] Marzocchi W, Lombardi A.M. and E. Casarotti, (2014) The Establishment of an Operational Earthquake Forecasting System in Italy, *Seism. Reaserch Lett,*, 85, 961-969, doi: 10.1785/0220130219.

[Ogata, 1998] Ogata, Y. (1998). Space-Time Point-Process models for Earthquake Occurrences. *Ann Inst Stat Math*, 50, 379-402.

[Rovida et al., 2016] Rovida A., Locati M., Camassi R., Lolli B., Gasperini P. (2016). CPTI15, the 2015 version of the Parametric Catalogue of Italian Earthquakes. *Istituto Nazionale di Geofisica e Vulcanologia. doi:http://doi.org/10.6092/INGV.IT-CPTI15*