Renewal models of seismic recurrence applied to paleoseismological data

I. Mosca[1], R. Console[2,3] and G. D'Addezio[2*]

[1]Utrecht University, The Netherlands

[2]Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

[3]Centro di Geomorfologia Integrata per l'Area del Mediterraneo, Potenza, Italy

Abstract

Because paleoseismology can extend the record of earthquakes back in time up to several millennia, it represents a great opportunity to study how earthquakes recur through time and thus provide innovative contributions to seismic hazard assessment.

A worldwide compilation of a database of recurrence from paleoseismology was developed in the frame of the ILP project "Earthquake Recurrence Through Time", from which we were able to extract five sequences with 6 and up to 9 dated events on a single fault. By using the age of the paleoearthquakes with their associated uncertainty we have tested the null hypothesis that the observed inter-event times come from a uniform random distribution (Poisson model). We have made use of the concept of likelihood for a specific sequence of observed events under a given occurrence model. The difference *dlnL* of the likelihoods estimated under two hypotheses gives an indication of which between the two hypotheses

[*] Corresponding author. Tel.: +39 06 51860482; fax: +39 06 51860507
E-mail addresses: ilaria_mosca@hotmail.it, console@ingv.it, daddezio@ingv.it.

fits better the observations. To take into account the uncertainties associated to paleoseismological data, we used a Monte Carlo procedure, computing the average and the standard deviation of *dlnL* for 1000 inter-event sets randomly obtained by choosing the occurrence time of each event within the limits of uncertainty provided by the observations. Still applying a Monte Carlo procedure, we have estimated the probability that a value equal to or larger than each of the observed *dlnL*s comes by chance from a Poisson distribution of inter-event times. These tests have been carried out for a set of the most popular statistical models applied in seismic hazard assessment, i.e. the Log-normal, Gamma, Weibull and Brownian Passage Time (BPT) distributions. In the particular case of the BPT distribution, we have also shown that the limited number of dated events creates a trend to reducing both the observed mean recurrence time and the coefficient of variation for the studied sequence which can possibly bias the results. Our results show that a renewal model, associated with a time dependent hazard, and some kind of predictability of the next large earthquake on a fault, only for the Fucino site, out of the five sites examined in this study, is significantly better than a plain time independent Poisson model. The lack of regularity in the earthquake occurrence for three of the examined faults can be explained either by the large uncertainties in the estimate of paleoseismological occurrence times or by physical interaction between neighbouring faults.

Keywords: earthquake forecast; paleoseismological data; statistical tests; inter-event time.

1. Introduction

In the last decades the use of probabilistic, time-dependent models of earthquake occurrence has grown up in the context of seismic hazard analysis. The basic idea for time-dependent models is to consider the earthquake occurrence as a quasi-periodic process (Shimazaki and Nakata, 1980): stresses which cause earthquakes are slowly built up by plate movements until the stress or deformation energy reaches a critical value, at which a rupture occurs. This idea has been worked out in the characteristic earthquake model (Schwartz and Coppersmith, 1984). According to this model, strong earthquakes have a general inclination to repeat themselves along the same fault segment or plate boundary. The occurrence of a characteristic earthquake ruptures the entire segment and relieves tectonic stress within the segment. The same idea is also the basis of the

seismic gap hypothesis (McCann *et al*., 1979), according to which the earthquake hazard is small immediately following the previous large earthquake and increases with time since the latest event on a certain fault or plate boundary. Kagan and Jackson (1991, 1995) tested this hypothesis for earthquakes on the circum-pacific belt using an ensemble of seismic zones because of the shortness of the seismic record within a single zone. Their result did not support the seismic gap hypothesis, as the fault segments that had experienced higher activity in the former period of time appeared to be more active also in the following period of test.

In order to assess the validity of the seismic gap hypothesis for its possible application to seismic forecasting, a probabilistic approach is used for comparison with a null hypothesis. Earthquake occurrence is regarded as a point process, and the inter-event time is modelled by a probability density function (*pdf*). In this respect, the null hypothesis is that for which the earthquake process has no memory (described by a uniform Poisson model). For a uniform Poisson model, whose *pdf* is a negative exponential function, only one parameter, the inter-event time, is necessary for a complete description. Conversely, the gap hypothesis needs a more complicated model, named renewal model, whose *pdf* contains a further free parameter, conditioning the shape of the distribution in terms of its periodicity. The *pdf* for a renewal model exhibits a maximum for inter-event times close to its expected recurrence time.

The present paper belongs to the kind of studies that aim to the evaluation of the seismic gap hypothesis through statistical methods. The two compared hypotheses of earthquake recurrence are the gap and null hypothesis. In this respect this work can be considered a development of the paper published by Console *et al.*, (2002). Their method was based on the comparison of the coefficient of variation observed for real seismic sequences with the distribution of the same parameter computed from a large number of simulations obtained from the Poisson hypothesis. In this case the comparison is made on the likelihood function computed for the real and simulated sequences.

2. Method

In the context of seismic forecasting one of the most popular investigation methods is the hypothesis testing through a stochastic procedure. We compare two hypotheses: the first one represents the reference model, commonly accepted; and the second is the alternative hypothesis. In this study the reference model (the null hypothesis) is described by the exponential distribution in the continuous domain, while the alternative model (the gap hypothesis) is described by renewal models (Log-Normal, Gamma, Weibull and Brownian Passage Time distributions).

The comparison between these models and the Poisson model has been carried out introducing the concept of likelihood, *L*, of a given realisation of a stochastic process under a given assumption. The function *L* is defined as the hypothetical probability that a set of events would yield a specific outcome under a specific hypothesis.

The log-likelihood function is evaluated for both the null hypothesis ($ln\ L_P$) and the gap hypothesis ($ln\ L_G$). Regarding the first one, described by the exponential distribution, $ln\ L_P$ is defined as:

$$\ln L_P = (N-1)\ln\left(\frac{1}{Trm}\right) - \frac{t(N)}{Trm}$$

where N is the number of observed events, $t(N)$ is the occurrence time of the most remote earthquake of the sequence, and *Trm*, is the mean inter-event time (or recurrence time).

The $ln\ L_G$ of the gap hypothesis, described by renewal models, is:

$$\ln L_G = \sum_{j=1}^{N-1} \ln\left\{\frac{f(Tr(j))}{1 - F(t \le Tr(j))}\right\} - \frac{t(N)}{Trm}$$

where $Tr(j)$ is the time difference, or inter-event time, between the j-th and the (j+1)-th event and $f(Tr(j))$ is the *pdf*.

As stated in the introduction, in this study we consider four kinds of statistical families, i.e. the Log-normal, the gamma, the Weibull and the Brownian passage time (BPT) distributions.

The *pdf* of the Log-normal distribution is :

$$f(Tr(j)) = \frac{1}{\alpha\sqrt{2\pi}Tr(j)} \exp\left\{-\frac{[\ln Tr(j) - \mu]^2}{2\sigma^2}\right\}$$

where μ and σ are the mean and standard deviation of the logarithm of the inter-event time.

The Log-normal statistical family has been evaluated estimating the shape parameter σ both from the data and assuming the fixed value $\sigma = 0.4$ (Wells and Coppersmith, 1994): this value describes the shape parameter of a quasi-periodic seismic sequence. The increasing of σ represents the decrease of the periodicity of the seismic sequence. When $\sigma \sim 1$ earthquakes occur at random over the time. Evaluating σ from data, the following equation is used:

$$\sigma = \sum_{j=1}^{N-1} \frac{[\ln Tr(j) - \mu]^2}{N}.$$

The *pdf* of the Gamma distribution is:

$$f(Tr(j)) = \frac{1}{\beta\Gamma(\gamma)} \left[\frac{Tr(j)}{\beta}\right]^{\gamma-1} \exp\left\{-\frac{Tr(j)}{\beta}\right\}$$

where γ and β are the shape and the scale parameters of this statistical family, respectively:

$$\gamma = \left[\frac{Trm}{sd(Trm)}\right]^2, \quad \beta = \frac{[sd(Trm)]^2}{Trm}$$

where *Trm* and *sd(Trm)* are the mean and the standard deviation of the inter-event times in the sample:

$$Trm = \frac{1}{N}\sum_{j=1}^{N-1} Tr(j), \quad sd(Trm) = \sqrt{\sum_{j=1}^{N-1} \frac{[Tr(j) - Trm]^2}{N}}.$$

The *pdf* of the Weibull distribution is:

$$f(Tr(j)) = \frac{\gamma}{\mu}\left[\frac{Tr(j)}{\mu}\right]^{\gamma-1} \exp\left\{-\left(\frac{Tr(j)}{\mu}\right)^{\gamma}\right\}$$

where γ and μ are the shape and the scale parameters, respectively. In particular the scale parameter of the Weibull distribution is coincident with the mean value of the inter-event times, *Trm*. Instead γ is the inverse of the coefficient of variation, or aperiodicity, defined as the ratio between the standard deviation and the mean of the observed inter-event time.

The *pdf* of the BPT distribution is (Ellsworth *et al*., 1999; Matthews *et al*., 2002):

$$f(Tr(j)) = \left[\frac{Trm}{2\pi C_v^2 Tr(j)} \right]^{1\backslash 2} \exp\left\{ -\frac{\left[Tr(j) - Trm\right]^2}{2C_v^2 Trm Tr(j)} \right\}$$

where *Trm* is the mean value of the inter-event time and C_v is the coefficient of variation (or aperiodicity).

The difference between the log-likelihood for the null hypothesis and the gap hypothesis is defined as:

$$d \ln L = \ln L_G - \ln L_P .$$

A positive *dlnL* means that the sequence is better described by the gap hypothesis than by the null hypothesis.

To take into account the effect of the uncertainties of paleoseismological data on the estimate of *dlnL* values, we have used a Monte Carlo procedure. So, we have computed the average and the standard deviation of *dlnL* from a thousand inter-event sets randomly obtained by choosing the occurrence time of each event within the limits of uncertainty provided by the observations. In this procedure it is assumed that the real occurrence time has a uniform probability distribution within such time limits.

In order to check the statistical significance of the *dlnL* results we have followed a standard procedure. This procedure consists in finding out the confidence level by which a hypothesis can be rejected with respect to the other. According to a standard practice, we can reject one of the two hypotheses only if the confidence level is higher than 95%. In this test we are interested in testing if the null hypothesis of the Poisson model can be rejected in light of the available paleoseismological data for any of the observed sites. Still making use of a Monte Carlo procedure, we have built up a thousand synthetic sequences based on a uniform Poisson distribution for the same number of events and the same total time covered by the observed data for each fault. Then we have computed the desired confidence level from the percentile corresponding to the real *dlnL* value in the synthetic distribution. It corresponds to the probability that a value equal or smaller than the observed *dlnL* comes by chance from casual fluctuations of a uniform random distribution (Console *et al*., 2002).

3. Data

Sequences of events on a single structure are quite infrequent to observe because the time interval covered by historical and instrumental catalogues is often too short when compared to the average recurrence time of individual faults. Since paleoseismology can extend the record of earthquakes of the past back in time up to several millennia, it represents a great opportunity to study how seismic events recur through time and thus provide innovative contributions to seismic hazard assessment (Figure 1).

Based on these considerations, for the present study we have used data from the Database of "Earthquake recurrence from paleoseismological data" developed in the frame of the ILP project "Earthquake Recurrence Time" (Pantosti, 2000). One of the main aims of this database is to resume the information concerning the recurrence through time of strong earthquakes occurred along seismogenic faults by means of paleoseismological study. It includes information about the analyzed sites (fault, segmentation, location, kinematics, slip rates) as well as the definition of paleoearthquakes (type of observation for event recognition, type of dating, age, size of movement, uncertainties). The database contains prevalently faults for which more than two dated events (one inter-event) exist.

In this work we have considered sites whose seismic sequence is composed of at least six events. Focusing the attention on the Mediterranean area, we have extracted five sequences of earthquakes: the Fucino fault in Central Italy, the Irpinia and the Cittanova fault in Southern Italy, the Skinos fault in Central Greece and El-Asnam fault in Northern Algeria (Figure 2). Each paleosismological site was investigated by scientists who proposed an interpretation of the seismic sequence combining the instrumental and historical earthquake records with paleoseismological study (Galadini and Galli, 1999; Galli and Bosi, 2002; Pantosti *et al*., 1993; Collier *et al*., 1998; Meghraoui and Doumaz, 1996).

It is evident that in all the sites only the youngest events are characterized by an exact occurrence time because they are instrumental or historical, instead most of them are paleoseismological, thus their age is affected by uncertainty (Figure 3 and Table 1). These uncertainties are related to the availability of chronological constrain in the stratigraphic sequence and to the uncertainty that affect the single radiocarbon date.

4. Results and discussion

By using the ages of paleoearthquakes with their associated uncertainties, we have compared the renewal (gap hypothesis) and the uniform Poisson (null hypothesis) models. We have considered the Log-normal distribution in two ways: first using the shape parameter σ obtained from the real sequences, and then fixing it at the value 0.4 considered appropriate by Wells and Coppersmith (1994). The comparison has been made between the log-likelihoods of the observed sequences under each model to test which of them fits better the observations for a number of studied sites. The other renewal models considered here are the Gamma, the Weibull and the BPT distributions.

For each of the five fault sites we have computed the mean inter-event time *Trm* with its uncertainty (see the first column of Tables 2 and 3). Then, for every renewal model, we computed also the shape parameter with its error and the difference *dlnL* between the log-likelihood obtained from the renewal and the Poisson models. The values of these parameters are reported in Tables 2 and 3, where each pair of columns refers to each renewal model separately. These results are discussed in the following subsections.

4.1 Log-normal distribution

Looking at the values referring to the shape parameter σ when this parameter is obtained from the observations (Table 2), we see that only for the Fucino fault σ (0.206 ± 0.021) is smaller than the standard value *0.4* adopted by Wells and Coppersmith (1994). This means that only the seismic sequence of this site exhibits a significantly high periodicity, while the earthquake occurrence of other sequences is characterized by less regularity and more casuality. The largest σ value belongs to the Skinos site (σ =1.07 ± 0.44) suggesting that this sequence follows the Poisson model reasonably well.

Since all the *dlnL* values are positive, we could apparently infer that the seismic sequences are characterized by a non random behaviour (Table 2). However, looking at the *dln*L values with their uncertainties, it is easy to notice that only for the Fucino fault this value is clearly higher than zero, while for the others the errors are comparable with the corresponding *dlnL* values.

As said earlier, the statistical significance of the comparisons has been investigated by means of a Monte Carlo procedure. One thousand synthetic sequences have been simulated under the Poisson model, and the *dlnL* values so obtained have been sorted out in increasing order. Figure 4 shows the cumulative distributions of the synthetic *dlnL* values for the five faults of the Mediterranean area. For each site, the plots show the comparison between the $\sigma \neq$ *const* Log-normal distribution and the Poisson distribution.

We can observe that in correspondence of the zero value of the x-axis (*dlnL=0*) all the plots cross a value pretty close to 50%. It means that the simulations yield approximately the same number of positive and negative results for *dlnL*. The percentage of simulations that fall below the observed *dlnL* value indicates the level of confidence by which the null hypothesis can be rejected. Table 4 shows these results in terms of the confidence level, α, for each of the models and each of the sites considered in this study. Only for the Fucino fault we can reject the null hypothesis with α >95%.

Looking at the plots for the σ=0.4 Log-normal distribution we can notice that the range of the x-axis is much wider (-150;10) and the *dlnL*=0 value is between the 70 and 80 percentile (Figure 5). The *dlnL* values of Skinos and El Asnam faults are negative and thus their earthquake sequences appear characterized by random occurrence of seismic events, rather than by quasi-periodical behaviour. Although the *dlogL* of the other sites are positive, we can't reject the null hypothesis by the 95% confidence level criterion in any case, even for the Fucino fault.

The evident difference between the Log-normal distribution with $\sigma \neq const$ and σ =0.4 is caused by the capacity of the former to adjust the shape parameter to the data. Thus, computing σ from the data, the shape parameter improves the performance of the renewal model with respect to the Poisson hypothesis.

4.2 Gamma distribution

Considering the Gamma distribution from a theoretical point of view, one can expect that its behaviour should not be so different from that of the $\sigma \neq const$ Log-normal distribution, because the *pdf*s of these renewal models are similar to each other and the only difference is the more or less prominent peak.

Indeed, the $ln\ L_G$ values, and so the *dlnL*s, of the Gamma distribution are comparable with those of the $\sigma \neq const$ Log-normal family (Table 2). However, the corresponding confidence levels are different from each other: for the Gamma model αs are higher than those of the other renewal model. The reason of the remarkable difference is the wider range of *dlnL* that is (-15;15). Its meaning is the following: when we introduce the experimental *dlnL* in the plot of the synthetic simulations, this is considerably shifted right-ward and so its confidence level results higher. For the Gamma distribution the probability that a *dlnL* value is smaller than or equal to the observed one comes from a random distribution, is higher than the same probability evaluated with the Log-normal statistical family.

The point $dlnL=0$ is in the middle of the x-axis (Figure 6), thus the number of the simulations with $dlnL>0$ and to the number of the simulations with $dlnL<0$ are balanced. Then the probability of having a *dlnL* value higher or less than 0 from a random distribution is similar.

The smallest *dlnL* belongs to the Skinos fault whose shape parameter γ is equal to 2.1 ± 1.6 as shown for the Log-normal distribution. Instead, the Fucino fault is characterized by a high shape parameter, $\gamma = 26.4 \pm 7.2$, and high *dlnL*, so for this fault we can reject the null hypothesis with $\alpha = 99.45\% \pm 0.70\%$ (Table 4).

4.3 Weibull distribution

Regarding the Weibull distribution, when its shape parameter γ is higher than 1, the sequence of earthquakes has a quasi-periodic behaviour; instead, when $\gamma < 1$ the seismic sequence is clustered. The studied Mediterranean sites have shape parameter larger than 1, thus this points out the regular behaviour of these seismic sequences.

We could draw the same conclusion also looking at the values of *dlnL* (Table 3). Indeed also for this renewal model the *dlnL*s are positive even if we take into account the fact that their uncertainties have the same order of magnitude of the average values.

From Tables 3 and 4 we can immediately see that the quasi-periodic behaviour of the Fucino fault, already observed from the previous statistical families, is even more evident in this case: its shape parameter is 4.23 ± 0.55, its *dlnL* is 9.70 ± 0.29 and the confidence level, by which the Poisson distribution can be rejected, is almost equal to 100%.

Moreover, another feature of this statistical model is that the values of γ, *dlnL* and α for the Irpinia and Cittanova faults are almost equal to each other. This suggests that according to the Weibull distribution, these two seismic sites have a similar behaviour of recurrence. Also, this renewal model shows that the most random sequence is that of the Skinos fault, even if its shape parameter is equal to the γ of the El Asnam site.

Figure 7 shows plots for the comparison between the Weibull and Poisson distributions. We can clearly observe that the point $dlnL=0$ is close to the 30 percentile, thus the number of simulations with positive *dlnL* is much higher than the number of simulations with $dlnL<0$.

All the features considered for the Weibull model (shape parameters γ higher than 1, positive *dlnL*s, and the location of the point $dlnL=0$) confirm that this renewal model fits well the real seismic sequences.

However, the superiority of this renewal model on the Poisson distribution is not effective because when we compare the observed *dlnL* with those of simulations, we can reject the Poisson distribution with α >95% only for the Fucino fault, as shown with the Log-normal and Gamma distributions.

4.4 BPT distribution

The BPT distribution confirms the general trend of the previous renewal models. The Fucino fault is still featured by the highest value of *dlnL* with a low uncertainty (Table 3). Instead, the Cittanova, Skinos and El Asnam faults have negative *dlnL*s, which denote a random behaviour of these seismic sequences.

Looking at the shape parameters C_v they are similar to the σ values for the $\sigma \neq const$ Log-normal distribution. In detail, the C_v of the Skinos fault, C_v=0.76 ± 0.18, is lower than that of the other renewal model, σ =1.07 ± 0.44. This suggests that considering the BPT distribution as renewal model, the observations show behaviour a bit closer to a quasi-periodical one. Even if the C_vs of the BPT family are comparable with those of the $\sigma \neq const$ Log-normal distribution, the *dlnL*s of the BPT are more similar to the *dlnL* values of the σ=0.4 Log-normal distribution.

Regarding the confidence level, only the α value of the Fucino fault is higher than 95% (Table 4), allowing the rejection of the null hypothesis just in this case.

Looking at the plots of the BPT model we can see that also the zero of the x-axis is between the 60 and 80 percentile as in the plots of the σ =0.4 Log-normal distribution (Figure 8). This means that the number of simulations with negative value of *dlnL* is higher than the number of simulations with *dlnL*>0.

For this renewal model we have carried out a further test. For each fault we have built up, through a Montecarlo procedure, synthetic BPT distributions characterized by the same number of events and the same total time covered by the observed data. The computer code allows for the arbitrary choice of the inter-event time *Trm* (input) and the coefficient of variation C_v (input). For each of these synthetic distributions, the corresponding *Trm* (output) and C_v (output) were computed. Repeating the procedure 1000 times, we have obtained an average *Trm* (output) and C_v (output), which are not necessarily the same as the respective input parameters. By means of a trial and error procedure, it was easy to find which pair of input parameters *Tr*m (input) and C_v (input) would provide the same output values as observed from the real seismic sequence.

The results of these simulations, reported in Table 5, show that both the inter-event time *Trm*(output) and the coefficient of variation C_v (output) are systematically smaller than the respective *Trm* (input) and C_v (input). In some cases, as for the Skinos and El Asnam faults, such difference is substantial.

We have then iterated the analysis carried out for the other renewal models, using *Trm* (input) and C_v (input) as they were the parameters estimated directly from the observations. In this way we have obtained new *dlnL* values and the relative significance level α for all the five faults (Table 6). These results don't change the conclusions which have been achieved directly from the real observations, though they show a smaller α value for both the Fucino and Irpinia faults.

5. Conclusions

In this paper we have tested the seismic recurrence of earthquake sequences to assess their characteristics of random or regular occurrence. With the only exception of the Fucino fault, whose regularity is a statistically significant feature, the analyzed seismic sequences appear characterized by irregular behaviour. We can show this conclusion by the comparison between the Log-normal, Gamma, Weibull and BPT distributions, and the exponential distribution.

Our analysis has pointed out a slight superiority of the Weibull model with respect to the others, as it can fit the data with a larger value of $dlnL$. However, the difference is not clearly significant. Indeed, a clear difference among these distributions could be easily noted only for high values of the recurrence time and for high values of their shape parameters.

We should take into consideration the uncertainty inherent in the paleoseismological data, because geological expressions of the past earthquakes are not easily discernible. Moreover, uncertainties affect the age estimates of the paleoearthquake due to both the dating methods and to the availability of dating evidence in the stratigraphic sequences. In a few cases these uncertainties may be comparable to or even larger than one seismic cycle. Only for the Fucino fault, the observed occurrence times for all the events are constrained within very narrow ranges. A rigorous statistical approach to the problem of the uncertainties in the observations of recurrence times for seismic hazard assessment has been introduced by Rhoades *et al.*, (1994) and Rhoades and Van Dissen (2003). In this study we have made use of the Monte Carlo method for dealing with such uncertainties.

The lack of regularity in the earthquake occurrence may be explained either by non-deterministic fault behaviour or by interaction between different faults. Indeed, closely spaced faults are characterized by a stress field that affects each other, possibly interacting with failure triggering processes. Consequently, for a more precise study of the seismic forecasting it will be necessary to consider the stress transfer between neighbouring faults in somehow more deterministic way.

Acknowledgements

This research started as the Master thesis of I.M. and was financially supported by the Istituto Nazionale di Geofisica e Vulcanologia. We thank Professor M. Caputo and D. Pantosti for valuable discussion and encouragement.

References

Collier, R., D. Pantosti, G. D'Addezio, P.M. De Martini, E. Masana and D. Sakellariou (1998), Paleoseismicity of the 1981 Corinth earthquake fault: seismic contribution to the extentional strain in Central Greece and implications for the seismic hazard. *J. Geophys. Res.*, 103, 30001-30019.

Console, R., D. Pantosti and G. D'Addezio (2002), Probabilistic approach to earthquake prediction. *Annals of Geophys.,* Vol.45, N.6, 723-731.

Ellsworth, W.L., M.V. Matthews, R.M. Nadeau, S.P. Nishenko and P.A. Reasenberg (1999), A physically-based recurrence model for estimation of long-term earthquake probabilities. *U.S. Geological Survey Open-File Report*, 99- 522.

Galadini, F. and P. Galli (1999), The Holocene paleoearthquakes on the 1915 Avezzano earthquake faults (Central Italy): Implications for active tectonics in Central Apennines. *Tectonophysics*, 308, 142-170.

Galli, P. and V. Bosi (2002), Paleoseismology along the Cittanova fault: Implications for the seismotectonic and earthquake recurrence in Calabria. *J. Geophys. Res*., 107, 142-170.

Kagan, Y.Y. and D.D. Jackson (1991), Seismic gap hypothesis: ten years later. *J.Geophys. Res*., 96, 21419-21431.

Kagan, Y.Y. and D.D. Jackson (1995), New seismic gap hypothesis: five years later. *J. Geophys. Res*., 100, 3943-3959.

Matthews, M.V., W.L. Ellsworth and P.A. Reasenberg (2002), Brownian model for recurrence earthquakes. *Bull. Seismol. Soc. Am*., 92, 2233-2250.

Mc Cann, W.R., S.P. Nishenko, L.R. Sykes and J. Krause (1979), Seismic gaps and plate tectonics: seismic potential for maior boundaries. *Pure Appl. Geophys.*, 117, 1082-1147.

Meghraoui, M. and F. Doumaz (1996), Earthquake induced flooding and paleoseismicity of the El Asnam, Algeria, fault-related fold. *J. Geophys. Res.*, 101, 17617-17644.

Pantosti, D. (2000), Earthquake recurrence through the time. *Proceeding of the Hokudan International Symposium and School on Active Faulting, Awaij Island, Hyogo, Japan, 17th-26th January 2000*, 363-365.

Pantosti, D., D.P. Schwartz and G. Valensise (1993), Paleoseismicity along the 1980 surface rupture of the Irpinia fault: Implications for earthquake recurrence in Southern Apennines, Italy. *J. Geophys. Res.,* 98, 6561-6577.

Rhoades, D.A., R.J. Van Dissen and D.J. Dowrick (1994), On the handling of uncertainties in estimating the hazard of rupture on a fault segment. *J. Geophys. Res.*, 99, B7, 13701-13712.

Rhoades, D.A. and R.J. Van Dissen (2003), Estimates of the time-varying hazard of rupture of the Alpine Fault, New Zealand, allowing for uncertainties. *New Zealand Journal of Geology and Geophysics*, 46, 479-488.

Schwartz, D.P. and K.J. Coppersmith (1984), Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas Fault Zones. *J. Geophys. Res.*, 89, 5681-5698.

Shimazaki, K. and T. Nakata (1980), Seismic gap hypothesis: ten years later. *Geophys. Res. Lett*., 7, 279-282.

Wells, D.L. and K.J. Coppersmith (1994), New empirical relationships among magnitude, rupture lenght, rupture width, rupture area, and surface displacement. *Bull. Seismol. Soc. Am.,* 84, 974-1002.

Figure captions

Figure 1 - Distribution of events along a hypothetical seismic structure compared with the length of instrumental, historical and paleoseismological catalogues of seismicity. We may have short time window within which we can observe how earthquakes recurred in the past even using historical record that may span between a few centuries to a couple of millennia. Paleoseismology can extend the record of past earthquakes back in time up to several millennia representing a good opportunity to investigate how strong earthquakes recur through time.

Figure 2 - Location of the sites of the Mediterranean area considered in this study.

Figure 3 - Time distribution of earthquakes for each sequence. The most recent ages of the events of each sequence are mostly historical and are indicated by single solid lines, whereas for the paleoseismic events the ages are indicated by mean ages (solid lines) and the associated uncertainties (shaded areas).

Figure 4 - Cumulative distributions of *dlnL*s for a thousand synthetic sequences compared with the observed *dlnL* with its uncertainty, for each studied fault of the Mediterranean area. The ordinate of the real *dlnL* in the synthetic distribution gives

the probability that the observed *dlnL* comes by chance from a random distribution. These plots show the comparison between the $\sigma \neq const$ Log-normal and the Poisson distribution.

Figure 5 - As in Figure 4, for the comparison between the $\sigma = const$ Log-normal and the Poisson distribution.

Figure 6 - As in Figure 4, for the comparison between the Gamma and the Poisson distribution.

Figure 7 - As in Figure 4, for the comparison between the Weibull and the Poisson distribution.

Figure 8 - As in Figure 4, for the comparison between the Brownian Passage Time and the Poisson distribution.

Figure 9 - As in Figure 4, for the comparison between the "modified" BPT and the Poisson distribution for each analysed site.

Table 1 - Age of the events for the sites analyzed in the present study. Historical earthquakes are indicated by a single date whereas for the paleoseimic events the age is characterized by a more or less wide range of uncertainty.

Table 2 - Results of studied seismic sequences for the comparison of the $\sigma \neq const$ and σ=0.4 Log-normal, and Gamma, with the Poisson distribution. For each site it is shown: the mean inter-event time *Trm*, the shape parameter σ, the difference of log-likelihood between the renewal model and the Poisson distribution *dlnL*. Every value is shown with its uncertainty, except for the fixed shape parameter σ=0.4, which comes from the literature.

Table 3 - Results for the comparison between the Weibull and Brownian Passage Time renewal models, and the Poisson distribution.

Table 4 - Significance levels with their uncertainties by which the Poisson distribution can be rejected. These values refer to the five Mediterranean sites and to different renewal models.

Table 5 - The shape parameter C_v and the mean value of the inter-event times *Trm* of a Poisson distribution given as input, compared with the corresponding C_v and *Trm* values of the BPT model obtained as output. All these values are shown for all the Mediterranean seismic sequences.

Table 6 - Results (difference of log-likelihood between the two hypothesis, *dlnL*, and confidence level α with their uncertainties) for the comparison between the

"modified" BPT and the Poisson distributions. "Modified" BPT model means that as input values, C_v*(input)* and *Trm*, we have provided the parameters estimated from the specific analysis carried out for the BPT model (see Table 5).

Abstract

Because paleoseismology can extend the record of earthquakes back in time up to several millennia, it represents a great opportunity to study how earthquakes recur through time and thus provide innovative contributions to seismic hazard assessment.

A worldwide compilation of a database of recurrence from paleoseismology was developed in the frame of the ILP project "Earthquake Recurrence Through Time", from which we were able to extract five sequences with 6 and up to 9 dated events on a single fault. By using the age of the paleoearthquakes with their associated uncertainty we have tested the null hypothesis that the observed inter-event times come from a uniform random distribution (Poisson model). We have made use of the concept of likelihood for a specific sequence of observed events under a given occurrence model. The difference *dlnL* of the likelihoods estimated under two hypotheses gives an indication of which between the two hypotheses fits better the observations. To take into account the uncertainties associated to paleoseismological data, we used a Monte Carlo procedure, computing the average and the standard deviation of *dlnL* for 1000 inter-event sets randomly obtained by choosing the occurrence time of each event within the limits of uncertainty provided by the observations. Still applying a Monte Carlo procedure, we have estimated the probability that a value equal to or larger than each of the observed *dlnL*s comes by chance from a Poisson distribution of inter-event times. These tests have been carried out for a set of the most popular statistical models

applied in seismic hazard assessment, i.e. the Log-normal, Gamma, Weibull and Brownian Passage Time (BPT) distributions. In the particular case of the BPT distribution, we have also shown that the limited number of dated events creates a trend to reducing both the observed mean recurrence time and the coefficient of variation for the studied sequence which can possibly bias the results. Our results show that a renewal model, associated with a time dependent hazard, and some kind of predictability of the next large earthquake on a fault, only for the Fucino site, out of the five sites examined in this study, is significantly better than a plain time independent Poisson model. The lack of regularity in the earthquake occurrence for three of the examined faults can be explained either by the large uncertainties in the estimate of paleoseismological occurrence times or by physical interaction between neighbouring faults.

Figure 1

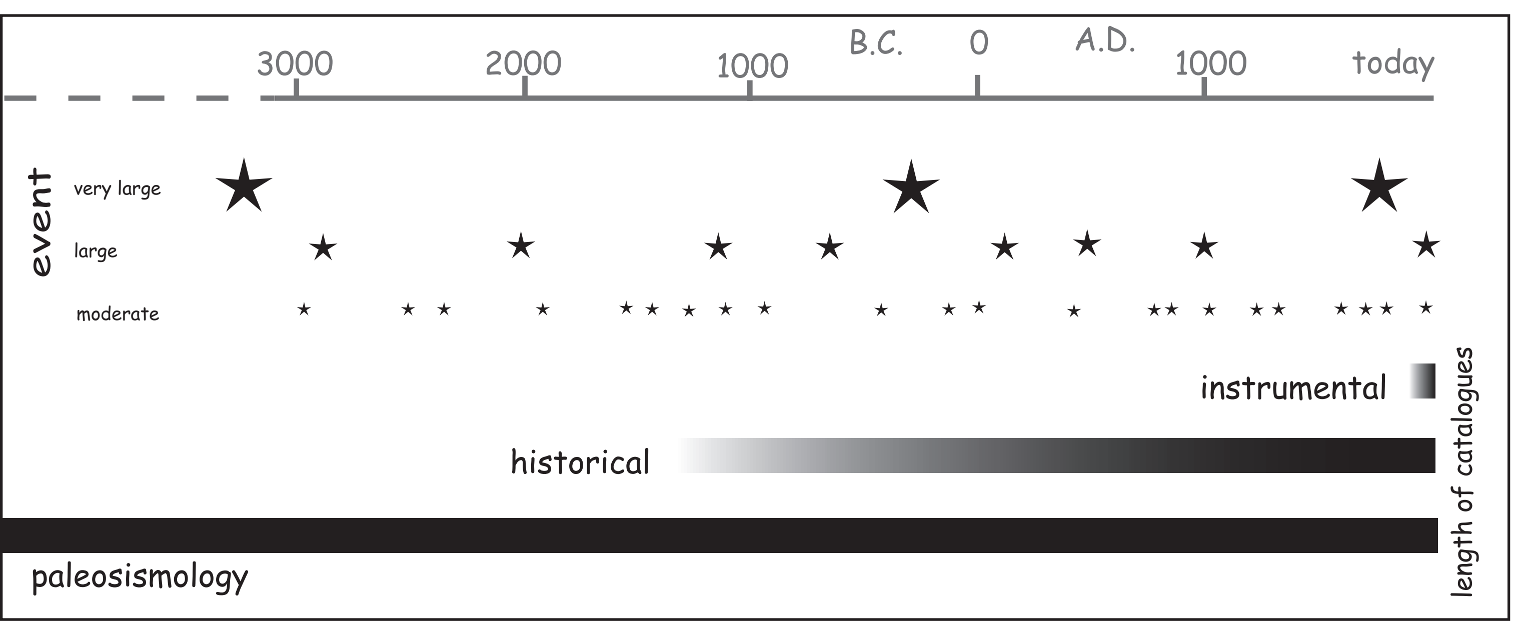

Figure 1

Figure 2

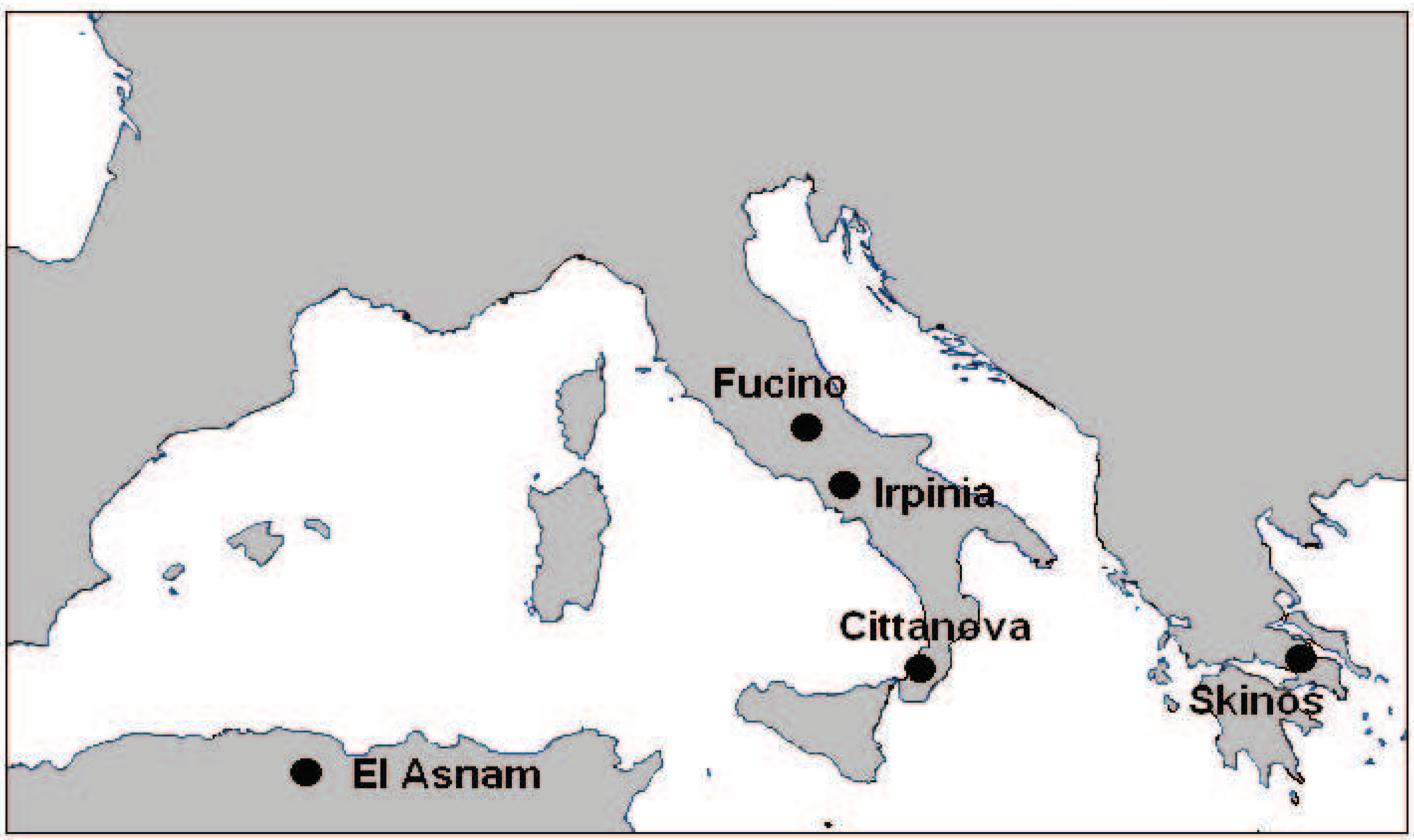

Figure 2

Figure 3

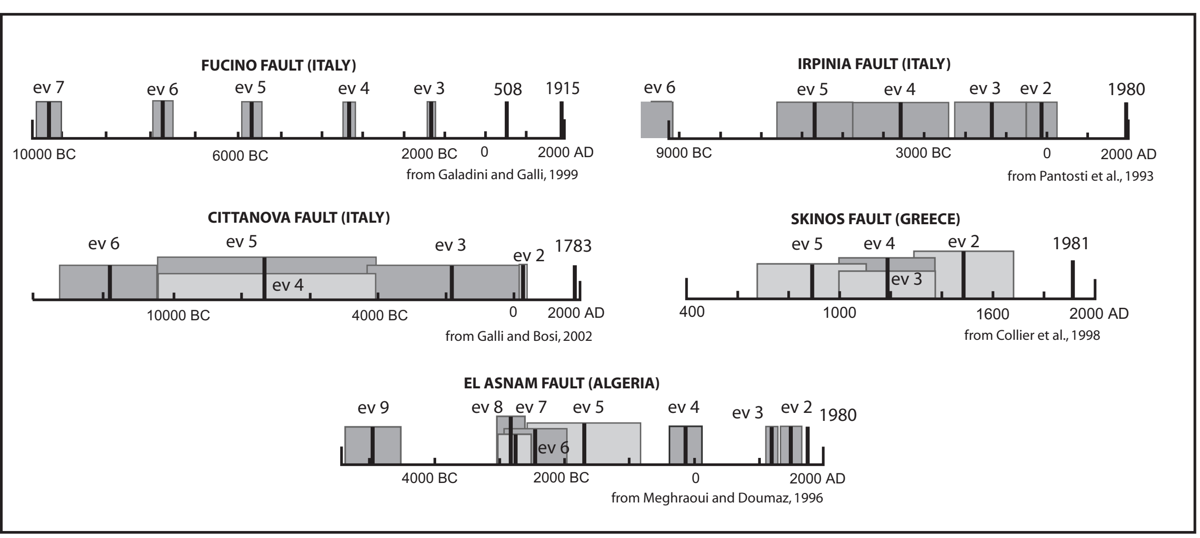

Figure 3

Figure 4

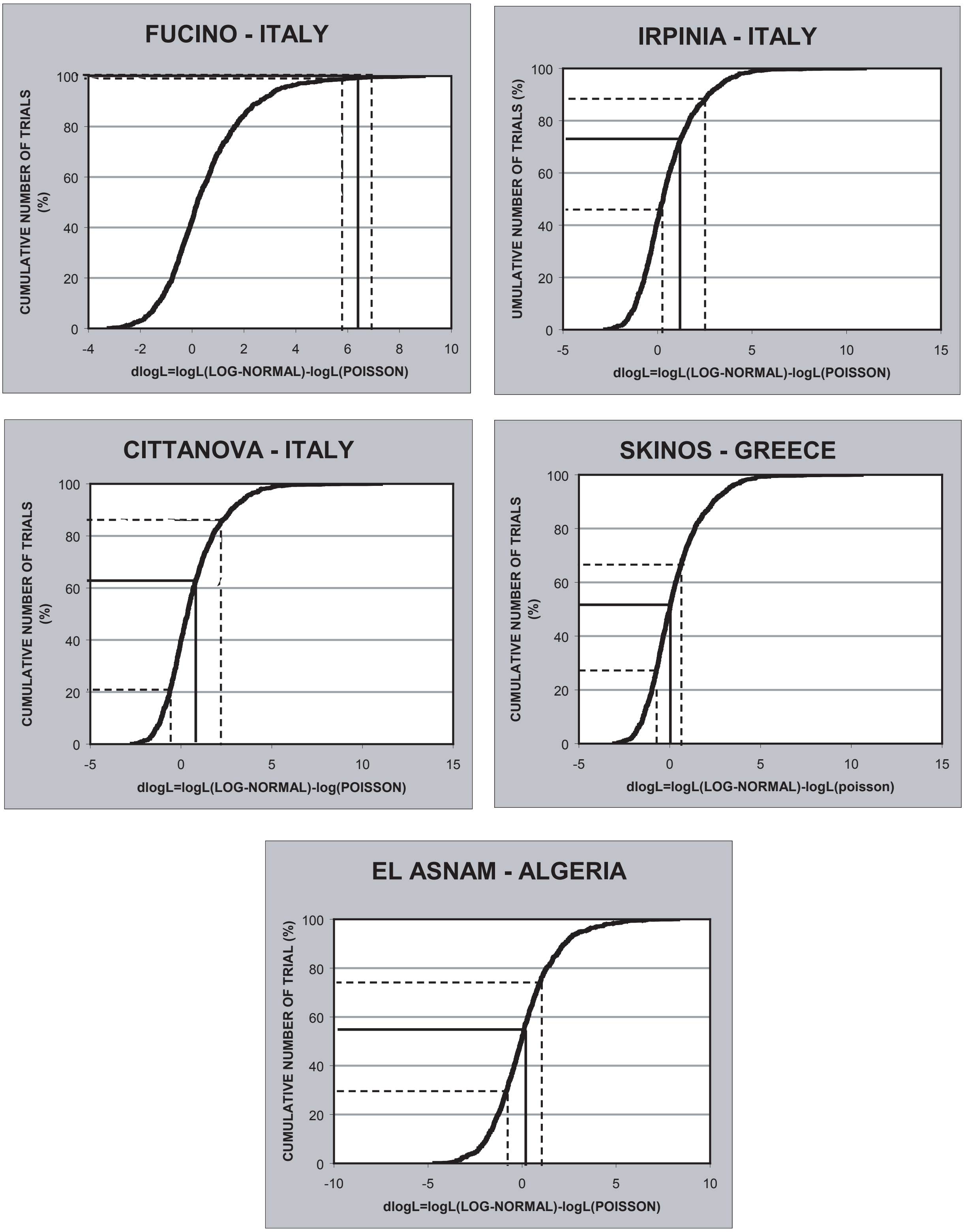

Figure 4

Figure 5

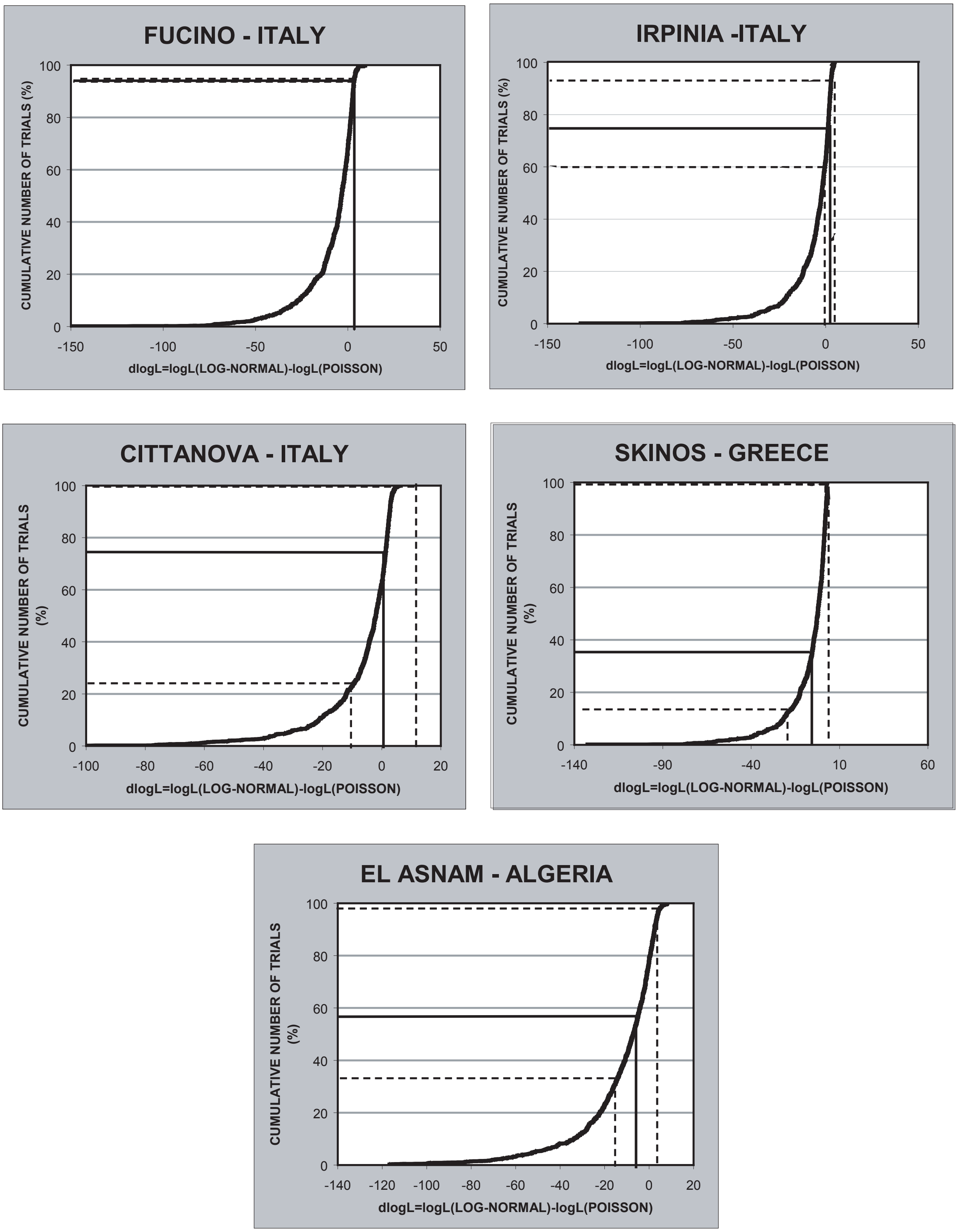

Figure 5

Figure 6

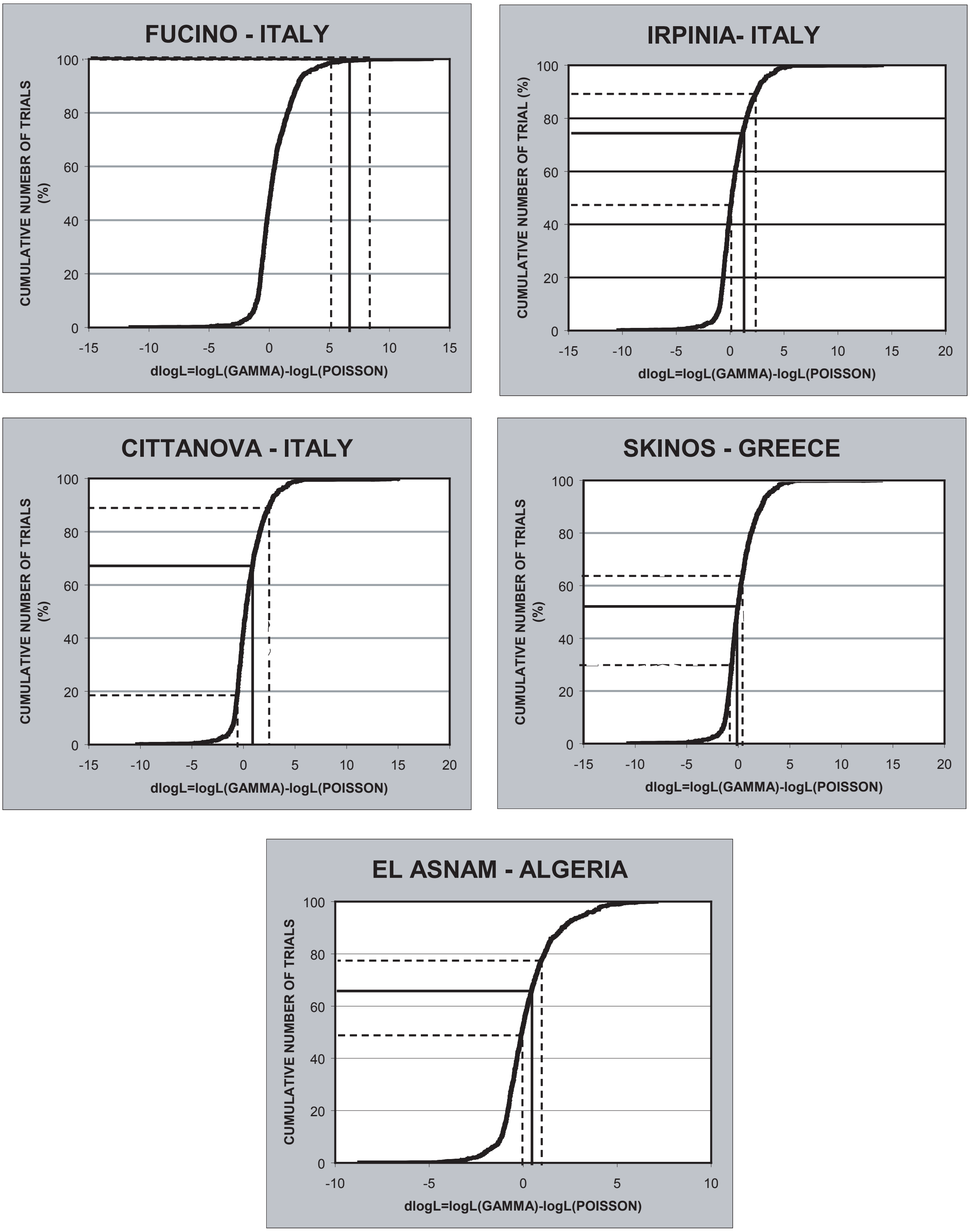

Figure 6

Figure 7

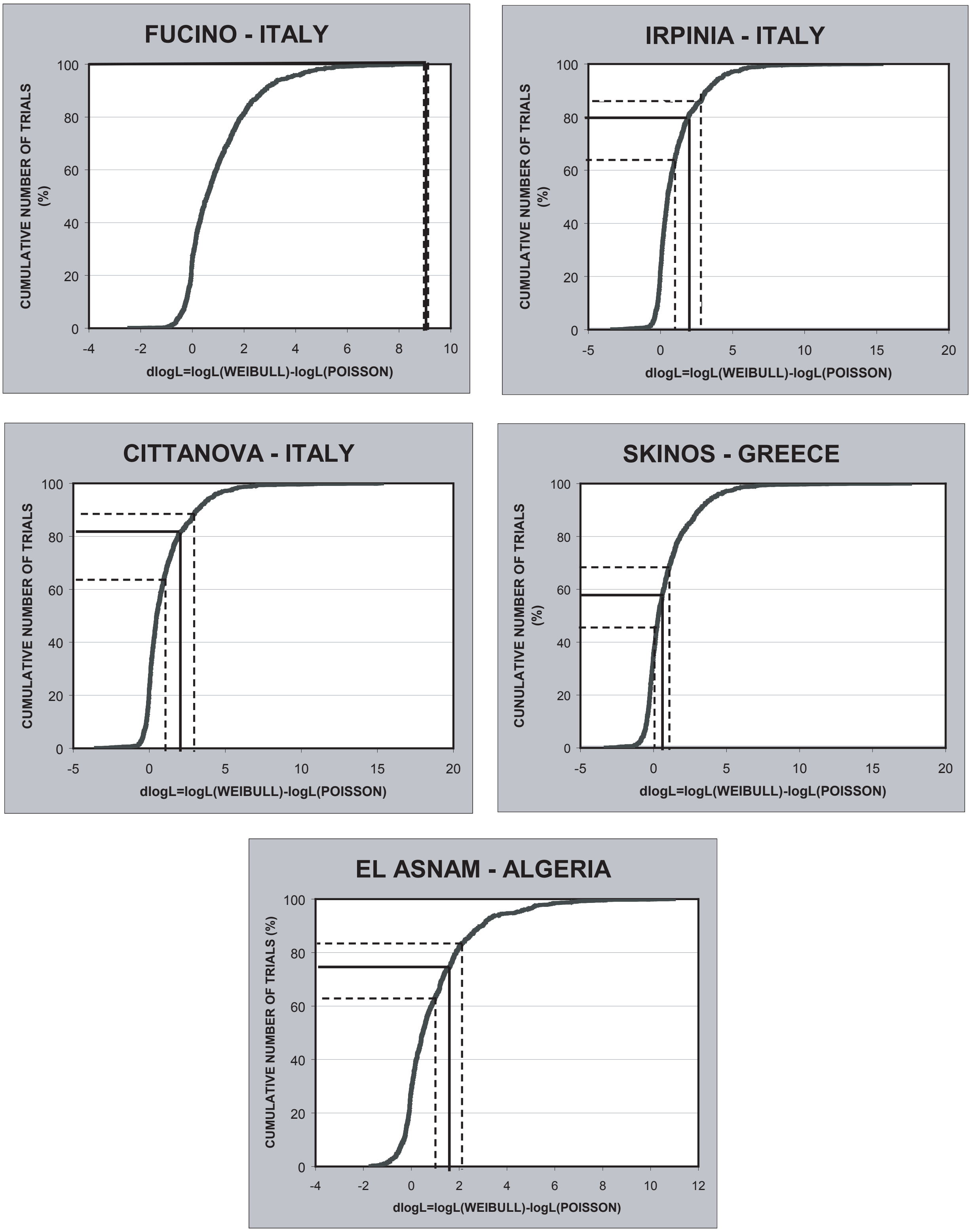

Figure 7

Figure 8

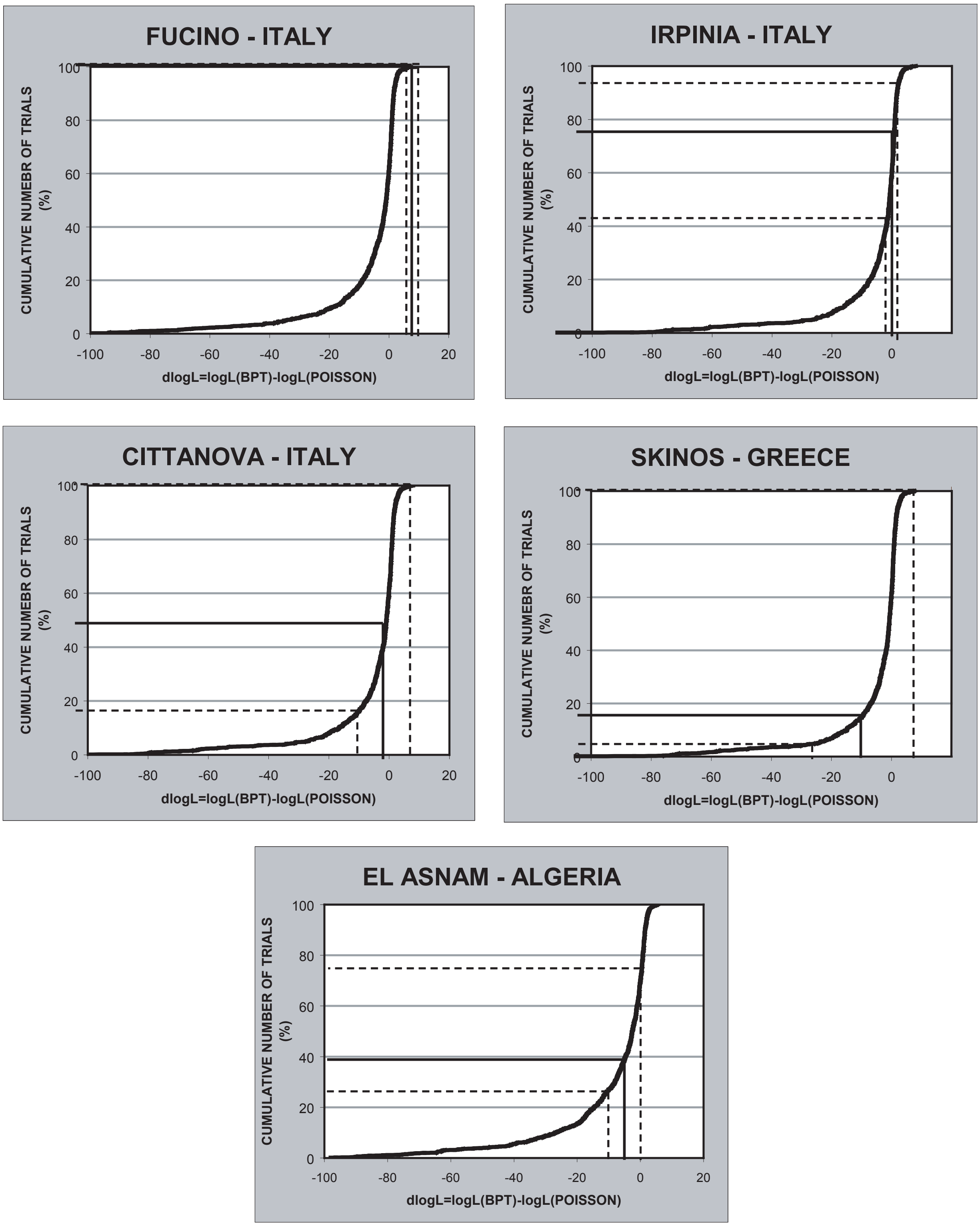

Figure 8

Figure 9

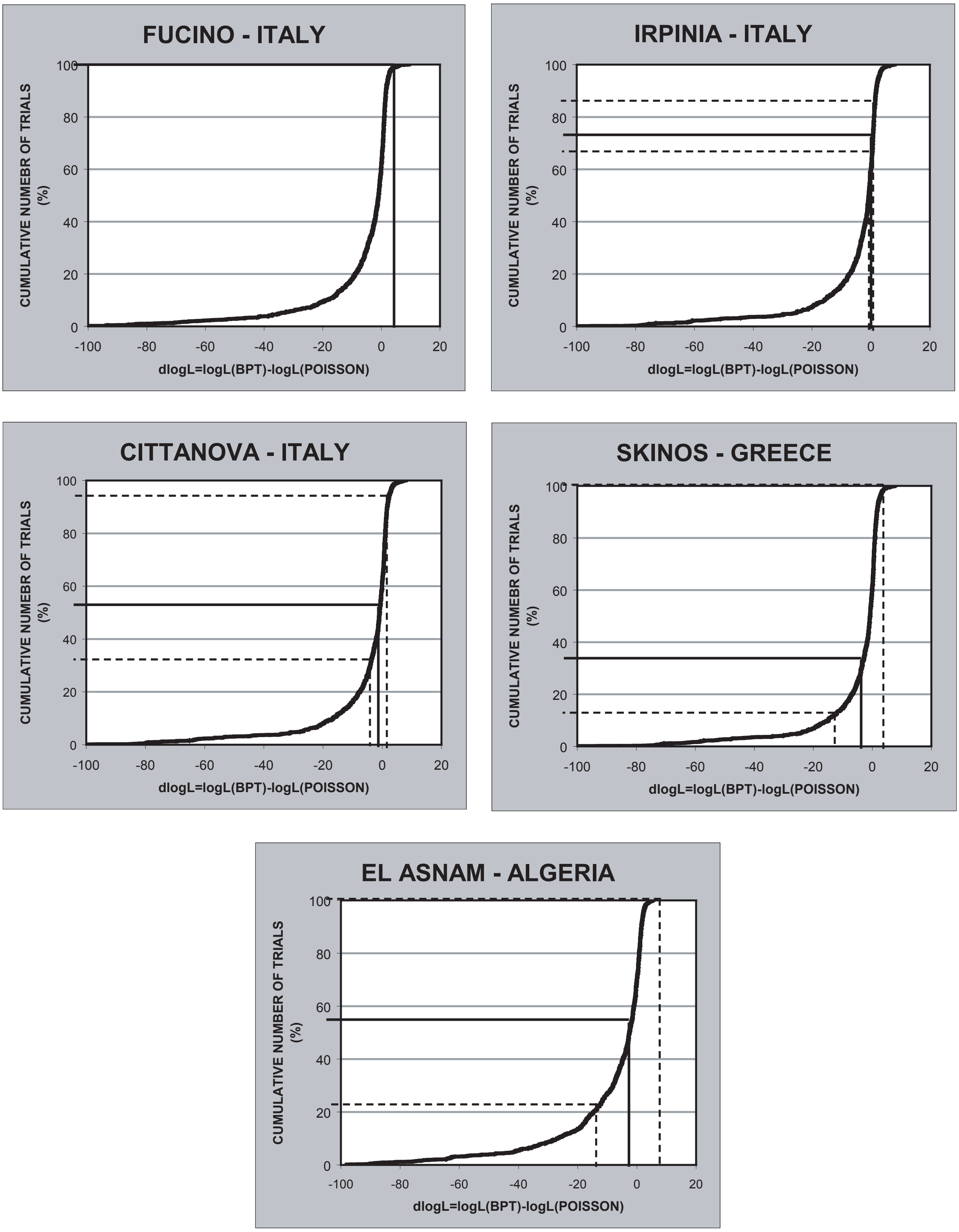
FUCINO - ITALY
CUMULATIVE NUMEBR OF TRIALS (%)
0
20
40
60
80
100
-100
-80
-60
-40
-20
0
20
dlogL=logL(BPT)-logL(POISSON)
IRPINIA - ITALY
CUMULATIVE NUMEBR OF TRIALS (%)
0
20
40
60
80
100
-100
-80
-60
-40
-20
0
20
dlogL=logL(BPT)-logL(POISSON)
CITTANOVA - ITALY
CUMULATIVE NUMEBR OF TRIALS (%)
0
20
40
60
80
100
-100
-80
-60
-40
-20
0
20
dlogL=logL(BPT)-logL(POISSON)
SKINOS - GREECE
CUMULATIVE NUMEBR OF TRIALS (%)
0
20
40
60
80
100
-100
-80
-60
-40
-20
0
20
dlogL=logL(BPT)-logL(POISSON)
EL ASNAM - ALGERIA
CUMULATIVE NUMBER OF TRIALS (%)
0
20
40
60
80
100
-100
-80
-60
-40
-20
0
20
dlogL=logL(BPT)-logL(POISSON)

Table 1
Click here to download Table: table1.doc

	FUCINO	IRPINIA	CITTANOVA	SKINOS	EL ASNAM
Event 1	1915 AD	1980 AD	1783 AD	1981 AD	1980 AD
Event 2	508 AD	230 AD-620 BC	300-370 AD	1295-1680 AD	1329-1630 AD
Event 3	1442 BC	620-2330 BC	390 AD-4300 BC	990-1390 AD	1040-1280 AD
Event 4	3230-3450 BC	2460-4790 BC	4060-10770 BC	990-1390 AD	90 AD-400 BC
Event 5	5570-5894 BC	4790-6650 BC	4060-10770 BC	670-1165 AD	830-1256 BC
Event 6	7526-7929 BC	9230-13050 BC	10710-13770 BC	670-1165 AD	1985-2559 BC
Event 7	10053-10729 BC				2509-3040 BC
Event 8					2509-3040 BC
Event 9					4510-5350 BC

Table 1

Table 2
Click here to download Table: table2.doc

		Log-normal with $\sigma \neq$ *const*		Log- normal with $\sigma =$ *const*		Gamma	
Sequence	*Trm* (yr)	σ	*dlnL*	σ	*dlnL*	γ	*dlnL*
Fucino	2051 ± 32	0.206 ± 0.021	6.50 ± 0.58	0.4	3.36 ± 0.11	26.4 ± 7.2	7.1 ± 2.0
Irpinia	2263 ± 167	0.61 ± 0.19	1.4 ± 1.2	0.4	1.4 ± 1.7	3.7 ± 2.9	1.2 ± 1.2
Cittanova	2802 ± 177	0.76 ± 0.34	0.9 ± 1.4	0.4	1.3 ± 9.7	3.9 ± 2.7	0.8 ± 1.4
Skinos	229 ± 28	1.07 ± 0.44	0.03 ± 0.67	0.4	-5 ± 11	2.1 ± 1.6	0.03 ± 0.47
El Asnam	862 ± 28	0.94 ± 0.25	0.15 ± 0.93	0.4	-4.3 ± 8.9	2.00 ± 0.46	0.52 ± 0.51

Table 2

Table 3
Click here to download Table: table3.doc

		Weibull		Brownian Passage Time	
Sequence	*Trm* (yr)	γ	*dlnL*	Cv	*dlnL*
Fucino	2051 ± 32	4.23 ± 0.55	9.70 ± 0.29	0.255 ± 0.037	8.01 ± 0.81
Irpinia	2263 ± 167	1.79 ± 0.70	1.96 ± 0.89	0.63 ± 0.19	0.8 ± 1.8
Cittanova	2802 ± 177	1.70 ± 0.57	2.1 ± 1.1	0.66 ± 0.21	-0.9 ± 7.7
Skinos	229 ± 28	1.40 ± 0.42	0.71 ± 0.43	0.76 ± 0.18	-8.0 ± 16
El Asnam	862 ± 28	0.94 ± 0.25	1.61 ± 0.51	0.720 ± 0.0.81	-4.6 ± 5.2

Table 3

Table 4
Click here to download Table: table4.doc

	α (%), Log-normal with $\sigma \neq$ const	α (%), Log-normal with $\sigma =$ const	α (%) Gamma	α (%) Weibull	α (%), Brownian Passage Time
Fucino	99.30 ± 0.30	94.6 ± 1.6	99.45 ± 0.70	100.00 ± 0.30	99.96 ± 0.10
Irpinia	76 ± 20	77 ± 17	75 ± 21	81 ± 10	76 ± 24
Cittanova	65 ± 31	74 ± 33	66 ± 34	82 ± 13	51 ± 41
Skinos	52 ± 19	38 ± 42	54 ± 15	60 ± 11	18 ± 47
El Asnam	56 ± 23	31 ± 59	68 ± 13	75.1 ± 9.2	40 ± 25

Table 4

Table 5
Click here to download Table: table5.doc

	Cv input	Trm input yr	Cv output	Trm output yr
Fucino	0.31	2051 ± 32	0.25	1913 ± 488
Irpinia	1.0	2263 ± 167	0.63	1858 ± 1212
Cittanova	1.1	2802 ± 177	0.66	2010 ± 1382
Skinos	1.55	229 ± 30	0.76	159 ± 127
El Asnam	1.03	862 ± 28	0.72	679 ± 500

Table 5

Table 6

	Cv input	*Trm* yr	*dlnL*	α
Fucino	0.31	2051 ± 32	5.76 ± 0.11	99.400 ± 0.080
Irpinia	1.0	2263 ± 167	0.72 ± 0.68	73 ± 13
Cittanova	1.1	2802 ± 177	-0.56 ± 3.2	54 ± 32
Skinos	1.55	229 ± 30	-2.57 ± 8.4	36 ± 43
El Asnam	1.03	862 ± 28	-1.5 ± 9.7	55 ± 37

Table 6