CHAPTER 5: CALCULATING EARTHQUAKE PROBABILITIES FOR THE SFBR

Introduction to Probability Calculations

The first part of the calculation sequence (**Figure 2.10**) defines a regional model of the long-term production rate of earthquakes in the SFBR. However, our interest here is in earthquake probabilities over time scales shorter than the several-hundred-year mean recurrence intervals of the major faults. The actual time periods over which earthquake probabilities are calculated should correspond to the time scales inherent in the principal uses and decisions to which the probabilities will be applied. Important choices involved in engineering design, retrofitting homes or major structures, and modifying building codes generally have different time-lines. Accordingly, we calculate the probability of large earthquakes for several time periods, as was done in WG88 and WG90, but focus the discussion on the period 2002-2031. The time periods for which we calculate earthquake probability are the 1-, 5-, 10-, 20-, 30- and100-year-long intervals beginning in 2002.

The second part of the calculation sequence (**Figure 5.1**) is where the time-dependent effects enter into the WG02 model. In this chapter, we review what time-dependent factors are believed to be important and introduce several models for quantifying their effects. The models involve two inter-related areas: recurrence and interaction. Recurrence is concerned with the long-term rhythm of earthquake production, as controlled by the geology and plate motion. Interaction is the syncopation of this rhythm caused by recent large earthquakes on faults in the SFBR as they affect the neighboring faults.

The SFBR earthquake model allows us to portray the likelihood of occurrence of large earthquakes in several ways. This flexibility allows the various users of the information to select the expression of earthquake likelihood most applicable to their needs. We express the likelihood of occurrence of one or more **M**≥6.7 earthquakes in the SFBR in these time periods in the following five ways.

- The probability for each characterized large earthquake rupture source.
- The probability that a particular fault segment will be ruptured by a large earthquake.
- The probability that a large earthquake will occur on any of the seven characterized fault systems.
- The probability of a background earthquake (i.e., in the SFBR, but not on one of the seven characterized fault systems).
- The probability that a large earthquake will occur somewhere in the region.

We also determine the probability of 6.0≤M<6.7 earthquakes in the region.

Calculating Probabilities

To make these forecasts, a *probability model* for earthquake occurrence is required. The probability model describes how earthquakes are distributed in time, given a set of assumed conditions. It uses the rupture source mean occurrence rate (**Table 4.8**) as primary input[1]. We employ a suite of probability models that take into account various amounts of physics (from none to some), various views on the relative importance of certain observables (e.g., date of the last rupture; recent seismicity rates; or slip in the 1906 earthquake), and various degrees of confidence in how well we are able to model the behavior of the region and predict its future. We did not restrict ourselves to methods used by previous Working Groups, nor did we embrace all models that have been proposed. Rather, we selected five models that—in our view—represent a range of valid approaches and give quantitative results that span the range of plausibility. None in the Working Group believes that the regional 30-year probability lies outside the range ~0.4 to ~0.9 – the range resulting from the combined set of probability models used – but few in the group would reject the possibility that the answer could lie anywhere within that range. The five models used are shown in **Figure 5.1** and described below.

Each probability model begins with an estimate of the long-term rate of either rupture sources or fault segment ruptures (**Figure 5.2**). We determine the probability that an earthquake will occur on the rupture source, regardless of where it initiates. The resulting probabilities are aggregated to yield the probabilities that an earthquake will occur on each fault. Finally, these are combined with the probability of background earthquakes to calculate the probability for the region as a whole (**Figure 5.1**). Also calculated is the probability that each fault segment experiences an earthquake (regardless of the rupture source). Probabilities are calculated relative to a threshold magnitude M_T.

We model earthquakes that rupture a fault segment as a *renewal process* (Cornell and Winterstein, 1988). In a renewal process, the times between successive events are considered to be independent and identically distributed random variables. This is to say that the expected time of the next event does not depend on any of the details of the last event (except the time it occurred). When a rupture occurs on the segment, it resets the renewal process to its initial state. The model is further simplified by reducing the complex physical process of the "earthquake machine" to a point process model (Cox and Lewis, 1966). That is, complexities involved in the physics of rupture initiation, propagation and stopping are ignored, and earthquakes are considered to be "events" characterized only by an occurrence time, location, and magnitude.

We employ two point process probability distributions: the Exponential distribution in the Poisson model, which yields time-independent probabilities, and the *Brownian Passage Time* (BPT) distribution, which is based on a new time-dependent model (Matthews et al., in press). These two probability distributions are combined with other data and assumptions to make five probability models. The simplest probability model we consider is the *Poisson*, which assumes

[1] In the past, a variety of probability distribution functions have been used as probability models, including the Lognormal (Nishenko and Buland, 1987; WGCEP, 1988; WGCEP, 1990); Weibull (Hagiwara, 1974; Sykes and Nishenko, 1984); Gamma (Utsu, 1984); Gaussian (Rikitake, 1974; Lindh, 1983); Exponential; Double Exponential (Utsu, 1972); and Brownian Passiage Time (BPT) (WG99).

that the probability does not vary in time and is thus fully determined by the long-term rate of occurrence of the rupture source. We also consider an *Empirical* model, a variant of the Poisson, in which the constant long-term mean rate for a rupture source is modulated by an extrapolation of the recent regional rate of earthquakes. The Empirical model uses recent regional earthquake activity as a proxy for physical time-dependent processes.

We also consider time-varying probability models—the *BPT* , *BPT-step*, and *Time-Predictable*—that take into account information about the last earthquake. The BPT model shares many features of other probability models used for earthquake forecasting, including the Lognormal model used by WG88, WG90 and WG95. In the BPT model, the failure condition of the fault is described by a state variable that rises from a ground state to the failure state during the earthquake cycle. Evolution of the point process model toward failure is governed by a deterministic parameter that pertains to the reloading rate of a specific segment, and a stochastic parameter. The model provides a statistical/mechanical description of the earthquake cycle. The use of a state variable in the BPT model permits explicit incorporation of the "stress shadow" effects of the 1906 earthquake (Harris and Simpson, 1992) and the 1989 Loma Prieta earthquake into the regional probability estimates, as the BPT-step model.

When both the time and amount of slip in the most recent event are known, we also employ the *time-predictable model* (Shimazaki and Nakata, 1980). In the time-predictable model, the expected time of the next event is equal to the time required to restore the fault to the same state it was in when the preceding earthquake occurred. Again, a point process model is used to represent the physics of the earthquake and loading cycle. In this report, the time-predictable model is applied only to the segments of the San Andreas Fault that ruptured in the 1906 earthquake because this is the only historical rupture for which there is a detailed model of the slip distribution along the entire fault system. While reliable estimates of slip in the 1868 **M**6.8 earthquake on the Hayward fault segment HS are available (Yu and Segall, 1996), neither the slip in or times of the last ruptures of the other two segments (HN, RC) of the Hayward-Rodgers Creek fault system are available.

All of the models have the same underlying mathematical model for the calculation of probabilities. The probability of rupture for each fault segment in the SFBR models is specified by a *probability density function* (pdf), *f(t),* that defines the chance that failure will occur in the interval from t to $t+\Delta t$, where t is time measured from the date of the most recent earthquake (**Figure 5.3**). The area under *f(t)* between $t = 0$ and $t = \infty$ is 1, as it is assumed that there will be a next earthquake. The area under *f(t)* between $t = T$ and $t = \infty$ or right tail of the distribution defines the *survivor function*, $F(T)$, which gives the probability that at least time T will elapse between successive events.

$$F(T) = \int_{T}^{\infty} f(t)dt \tag{5.1}$$

For any probability model, $F(0)=1$ and $F(\infty) = 0$. A more useful quantity for many purposes is the *hazard function, h(t),* defined by the ratio $h(t) = f(t)/F(t)$. It gives the instantaneous rate of failure at time t conditional upon no event having occurred up to time t, or more plainly, the

chance that anything surviving to time *t* will fail immediately. The hazard function provides a convenient way to compare the properties of probability distributions.

The probability values presented in this report are *conditional probabilities*. These give the probability that one or more earthquakes will occur on a rupture source of interest during an interval of interest (e.g., the 30-year interval from 2002 through 2031), conditional upon it not having occurred by the year 2002. The conditional probability is determined by dividing the area under the density function in the interval of interest by the area of the density function at times equal to or greater than the start of the interval of interest (**Figure 5.3**). The conditional probability is defined by

$$P(T \le t \le T + \Delta T \mid t > T) = \frac{F(T) - F(T + \Delta T)}{F(T)} \tag{5.2}$$

Probability Models Used in the Calculations

Five probability models are considered, with each forming a branch (with weights assigned based on expert opinion) in the calculation sequence (**Figure 5.1**). Calculations for each of the five probability models are carried out independently all the way to the computation of the combined earthquake probabilities for the San Francisco Bay region. Each model and its application to calculating probabilities are described below. The assignments of weight to the probability models (and why some WG02 members chose to put weight on a given model) are described toward the end of this chapter.

Poisson Model

The *Poisson model* describes the distribution of times between successive events for a homogeneous Poisson process (random occurrence) and is specified by the pdf

$$f_{Exp}(t) = \lambda e^{-\lambda t} \tag{5.3}$$

where λ is the mean rate of events per unit time (reciprocal to the mean interval between events). In our calculations, λ is the mean rupture rate of each rupture source, as determined by the long-term model. The Poisson distribution has the important property that the hazard function is constant, $h_{\mathrm{Exp}}(t) = \lambda$ (**Figure 5.4**). Thus, it has no "memory" of the time of the most recent event. An earthquake is just as likely to occur on a fault segment one day after the most recent event as it is to occur on a day two hundred years later. The conditional probability is also independent of the time of the most recent event.

The Poisson model is the standard model for probabilistic seismic hazard analyses, having been used most recently in the National Earthquake Hazard maps (for example, Frankel et al., 1997) and in the Seismic Hazard Maps of California (Petersen and others, 1996). This model is appropriate when no information other than the mean rate of earthquake production is known; it

can be viewed as the "least-informative" or "most-conservative" model, depending on one's perspective. However, it fails to incorporate the most basic physics of the earthquake process, whereby the tectonic stress released when a fault fails must rebuild before the next earthquake can occur at that location. Evidence for this deficiency is seen in the timing of three series of paleoseismic occurrence times in central Japan, including some 29 earthquakes in all, where the Poisson model provides a significantly poorer fit to the data than do any of the Lognormal, Gamma, Weibull and Double Exponential distributions (Ogata, 1999). The Poisson model also fails to account for the observation that the level of earthquake activity in the SFBR dropped sharply after 1906 and has remained low most of the 20th century, presumably as a result of a regional stress relaxation caused by the 1906 earthquake. Nevertheless, we include the Poisson model to provide a conservative estimate of the probability on faults for which one suspects that the time-dependent models are either too poorly constrained or missing some critical physics of the system (e.g., interactions). The model provides a "baseline" probability calculation that reflects only the long-term rates of earthquakes in the SFBR earthquake model.

Application of the Poisson Model

In the Poisson model, conditional probabilities for a specified time interval depend only on the length of the interval, Δt, and the long-term rate of rupture for each source**.** The conditional probability for each source is given by

$$1 - e^{-\lambda \Delta t} \qquad (5.4)$$

where Δt (in this study) is 5, 10, 20, 30, or 100 years.

Because the expected magnitudes of earthquakes are themselves probabilistic, an additional step is needed to compute probabilities for earthquakes above a given magnitude threshold M_T (e.g., **M**≥6.7 events). For each source, the rate of **M**≥M_T events is determined from the magnitude probability density function, as described in Chapter 4 (**Figure 4.3**). The rate at which the rupture source produces earthquakes exceeding M_T is computed from equation (4.10), and assigned to λ in the exponent of (5.4).

Empirical Model

The empirical model (Reasenberg et al., 2002) modifies the Poisson model earthquake probability (**Figure 5.5**) based on the rate of occurrence of moderate size earthquakes (**M**≥5.5) in the SFBR since 1906. It complements the other models because the record of historical (**M**≥5.5) earthquakes (**Figure 5.6**) is not used in the BPT, Poisson or time-predictable probability models. Complementing a theoretical model with an extrapolation of empirical observation is a common forecasting technique, often employed in a method of Bayesian inference. For example, a complex atmospheric model for weather prediction might be combined with regional observations of recent weather activity (e.g., looking out the window or viewing satellite imagery). Our use of an empirical model is motivated by concern (discussed below) that the effect of the regional stress change produced by the 1906 earthquake may not be adequately represented by the elastic dislocation calculations and their application in the BPT model.

We are aware that other methods (post-seismic viscoelastic stressing and rate-and-state friction models) for calculating stress changes (and their effects) produced by the 1906 and 1989 earthquakes based on more complex physical processes are available (e.g., Kenner and Segall, 1999; Parsons, 2002a; Pollitz et al., 1998). However, these models were considered by their authors and by the WG02 Oversight Committee to be insufficiently vetted for incorporation into the current SFBR probability calculations.

Based on observations and theoretical considerations, Tocher (1959), Ellsworth and others (1981), Jaumé and Sykes (1996), and Bakun (1999), concluded that the seismicity rate in the SFBR varies throughout a regional earthquake cycle that is defined by recurrences of the largest earthquake in the region, perhaps spending little if any time near its long-term average rate (Ellsworth and others, 1981; Mogi,1982). The empirical model embodies this conception of a regional earthquake cycle.

In the empirical model, the time-dependence of earthquake activity is not described by a pdf, as it is for the fault segment-based probability models. Rather, earthquake activity is represented by a non-stationary Poisson process with varying rate $\lambda(t)$, which is estimated for the region from the historical earthquake record (**M**≥3.0 since 1942 and **M**≥5.5 since 1906) and normalized by the long-term mean rate λ (**Figure 5.6**).

$$\gamma(t) = \lambda(t)/\lambda \qquad (5.6)^2$$

To estimate the probability for a future period $(t, t+\Delta t)$, $\gamma(t)$ is extrapolated forward in time using a variety of methods and the extrapolated rates are averaged. The probability P_i^{Emp} for fault i with mean rate λ_i is given by

$$P_i^{Emp} = 1 - e^{-\bar{\gamma}\lambda_i \Delta t} \qquad (5.7)$$

where $\bar{\gamma}$ is the average of $\gamma(t)$ over $(t, t+\Delta t)$.

In the empirical model, as in the Poisson model, the timing of individual segment ruptures is not specified. Instead, the empirical model assumes that the rates of all ruptures in the regional fault model vary in time (and in unison) about their respective long-term means (**Figure 5.7**). The empirical model specifies only time-dependence, preserving the magnitude distribution of the rupture sources in the fault characterization. Hence, the geological framework, paleoseismic history, fault segmentation, slip rates, aseismic slip factors and magnitude-area relations are preserved.

This model property says that there is at least some spatial coherence in the earthquake activity, but says nothing about whether this coherence is due to stress interactions, time-varying driving forces, or anything else. This is to say that $\gamma(t)$, which is estimated for the region, is assumed to represent the time dependence of activity on each fault.

[2] There is no equation (5.5).

The Empirical model is a proxy for realistic physical models of earthquake occurrence and interaction not fully represented in the analytical time-dependent models (BPT and Time-Predictable) described below. The underlying, time-dependent processes controlling earthquake activity on the faults and in the region are built in to the empirical model in the sense that their effects are assumed to be reflected in the historical rate of seismicity. The empirical model is expected to provide an accurate forecast to the extent that (1) fluctuations in the rate of **M**≥3.0 and **M**≥5.5 earthquakes reflect fluctuations in the probability of larger events; (2) fluctuations in rate on individual faults are correlated; and (3) the rate function $\gamma(t)$ can be sensibly extrapolated forward in time (5, 10, 20, or 30 years). Assumption (2) is uncertain because the calculated effect of the 1906 earthquake stress shadow is not homogeneous in space (Harris and Simpson, 1992; Kenner and Segall, 1999). However, observations suggest that the stress shadow has affected seismicity on all major faults in the region (see Figure 8 in Reasenberg et al., 2002).

Application of the Empirical model.

Reasenberg et al. (2002) estimated $\lambda(t)$ for the SFBR from **M**≥5.5 earthquakes since 1850 (Bakun, 1999) and from **M**≥3.0 earthquakes since 1942 in the CNSS catalog (**Figure 5.6**). They proposed a variety of methods for extrapolating $\gamma(t)$ forward in time, from which we selected six for the period 1906-2000 (**Table 5.1**). The mean regional annual rate (**M**≥6.7) for the six branches is 0.016±0.004, where the uncertainty corresponds to the range of extrapolated rates among the six branches. Weights of 0.1, 0.5, 0.4 are assigned to these values to reflect our view that the up-trending extrapolations are more likely correct (i.e., that the region is coming out of the stress shadow). The weighted mean rate is $\bar{\gamma}$ = 0.018, with corresponding 30-year probability 0.42.

Table 5.1. Empirical model extrapolated regional rate estimates.

Branch	Data	Time Period	Method	Mean rate[1] (**M**≥6.7) in 2002-2031	Mean[2] of $\gamma(t)$ in 2002-2031
A	**M**≥3.0	1942-1998	mean	0.014	0.45
B	**M**≥3.0	1984-1998	mean	0.016	0.52
C	**M**≥5.5	1906-2000	mean	0.011	0.35
D	**M**≥5.5	1979-2000	mean	0.020	0.65
E	**M**≥3.0	1942-1998	trend	0.016	0.52
F	**M**≥3.0	1970-1998	trend	0.020	0.65

[1]Rates are extrapolated to **M**≥6.7 by assuming a Gutenberg-Richter relation with b=0.9.
[2]Based on a long-term regional rate (M≥6.7) of 0.031.

To de-aggregate the empirically estimated regional probability onto the faults, we assume that the rates of all earthquake sources vary in time with $\gamma(t)$ about their respective long-term means, with each source contributing a constant proportion of the regional rate (Figure 5.7). Support for this assumption is found in Reasenberg et al. (2002), who modeled $\gamma(t)$ for each fault using the Wesson et al. (in prep.) fault associations. They found a common decrease in rate on nearly every fault in the SFBR in the 20th century, and a common increase during the second half of the 19th century, both relative to the respective estimated long-term mean rates. We are more confident of the empirical regional probability model than the de-aggregated probabilities for the individual faults. This situation appears to be an intrinsic aspect of regional and fault-specific earthquake probability estimates, reflecting nothing more than the relative amounts of

information available.

Brownian Passage Time Models: BPT and BPT-step

In contrast to the Poisson model, a time-dependent renewal process model embodies the expectation that after one earthquake on a fault segment, another earthquake on that segment is unlikely until sufficient time has elapsed for stress to gradually re-accumulate (Lindh, 1983, Sykes and Nishenko, 1984, Buland and Nishenko, 1988, Ellsworth, 1995; Ogata, 1999). Such models require a minimum of two parameters, and typically include knowledge of the time of the most recent rupture. One is the mean recurrence interval, $\mu = 1/\lambda$, and the other describes the variability of recurrence intervals and can be related to the variance, σ^2 of the distribution. (For the Poisson distribution, $\sigma=\mu$.) We define this variability of recurrence times as the aperiodicity, $\alpha=\sigma/\mu$.

The Brownian Passage Time (BPT) model (Matthews et al., in press; Ellsworth et al., 1999; Kagan and Knopoff, 1987) is a renewal model that describes the statistical distribution of rupture times. The BPT distribution is also known as the inverse Gaussian distribution (Seshadri, 1993). The probability density is defined by

$$f_{BPT}(t) = \sqrt{\frac{\mu}{2\pi\alpha^2 t^3}} e^{-\frac{(t-\mu)^2}{2\mu\alpha^2}} \qquad (5.8)$$

The hazard function (instantaneous failure rate), $h_{BPT}(t)$, is always zero at t = 0 (Figures 5.4 and 5.9). It increases to achieve a maximum value at a time greater than the mode of $f_{BPT}(t)$, and from there decreases toward an asymptotic value of $h_{BPT}(t) = 1/(2\mu\alpha^2)$. Thus, a BPT process always attains a finite quasi-stationary state in which the failure rate is independent of elapsed time. When the aperiodicity $\alpha = 1/\sqrt{2}$, the asymptotic failure rate is $1/\mu$, which equals the asymptotic failure rate for a Poisson process with the same μ. In practice, the behavior of a BPT model is similar to that of a delayed Poisson process, for which the failure rate is zero up to a finite time following an event and then steps up to an approximately constant failure rate at all succeeding times.

The behavior of a BPT model depends strongly on the value of α. For smaller values of α, $f_{BPT}(t)$ is more strongly peaked and remains close to zero longer (Figure 5.9a). For larger values, the "delay" or "dead time" becomes shorter, $f_{BPT}(t)$ becomes increasingly Poisson-like, and its mode decreases. The hazard function in the quasi-stationary state (Figure 5.9b) increases with decreasing values of α and becomes Poisson-like with increasing values.

Matthews et al. (in press) derived the BPT distribution by adding Brownian perturbations to a steady tectonic loading process in which rupture occurs when a critical failure threshold is reached. The earthquake relaxes the stress to a characteristic ground level, beginning a new failure cycle. This model of a Brownian relaxation oscillator provides a connection between the time intervals separating events and a formal state variable that reflects the macroscopic mechanics of stress accumulation and release. It enables the calculation of the probabilities following the perturbation of the stress state by an external event, such as the regional effects of

the stress release in the great 1906 earthquake. The influence of such interactions on recurrence times is transient and strongly depends on when in the loading cycle the perturbation occurs.

Earlier Working Groups used the Lognormal model (Nishenko and Buland, 1987) as a generic recurrence model for earthquake forecasting. The shape of the pdf of the Lognormal model is very close that of the BPT model. However, there are some important differences between the behaviors of the models. The hazard function of the Lognormal distribution, $h_L(t)$ initially increases with t, beginning from $h_L(0) = 0$, but always goes to zero as $t \rightarrow \infty$ (**Figure 5.4**). Although the properties of h_L for $0 \leq t \leq \mu$ match our geological intuition, the long time behavior does not, and thus does not constitute a fully satisfactory model of the earthquake renewal cycle (Matthews et al., in press).

BPT-step model

The BPT model allows us to explicitly account for the effects of stress changes caused by other earthquakes on the segment under consideration, such as the effect of the 1906 "stress shadow" on the regional fault model, or the effect of the 1989 Loma Prieta earthquake on the nearby segments of the San Andreas Fault. This variation of the BPT model is referred to as the *BPT-step model*. Interaction in the BPT-step model occurs through the state variable. A decrease in the average stress on a segment lowers the probability of failure, while an increase in average stress causes an increase in probability or triggers a rupture. The effects on probability strongly depend on when in the loading cycle the step is applied. The effects are strongest when the segment is near failure (**Figure 5.11**). The probabilities evolve over time following the application of the step, and decay to the probabilities given by the simple "clock change" method of adjusting the reset time used by WG90 and WG99. The rate at which probability evolves is inversely proportional to the time elapsed since the step, or in other words, in accord with Omori's law (Matthews et al., in press).

Application of the BPT model

The BPT and BPT-step models are applied to each of the 18 segments using the mean rates of segment rupture in **Table 4.6 (see Figure 5.8)**. Separate branches of the calculation sequence either ignore or apply stress changes from the 1906 and 1989 earthquakes when calculating the probability of an earthquake involving the segment. Segment rupture probabilities are then partitioned onto the rupture sources according to the relative rates of their occurrence in the long-term model. (An exception is made for the "floating" rupture sources. Because these rupture sources do not have reset times associated with them, the Poisson model is used to calculate probabilities for these ruptures when the BPT or BPT-step model is specified.) The additional assumption is made that the likelihood of a rupture starting on a segment (in the case of a multi-segment rupture source) is proportional to the fractional length of the segment. With this assumption it is straightforward to map the segment rupture probabilities into the sources that produce them. This construction of source probabilities is needed in order to aggregate probabilities for faults and the region as a whole. The BPT model is expected to provide accurate estimates of earthquake probability to the extent that (1) the model represents the statistics of recurrence intervals for segment rupture, and (2) the time of the most recent event is known or

constrained. The BPT-step model depends on (1) and (2), and on the assumption that (3) the effects of interactions, especially those of the 1906 earthquake, are properly characterized.

Time of most recent rupture.

A key input to the BPT and BPT-step models is the time of the most recent rupture—the time at which the renewal model is "reset". Estimates of these times come from several lines of evidence, including paleoseismic observations, dates of historical ruptures, and historical seismicity, as described in **Chapter 3**. On the San Andreas fault segments and on the southern Hayward fault segment (HS), these dates are known (1906 and 1868, respectively). For all other fault segments, however, these observations can only constrain the date of the last rupture to be within an interval that may range in length from 63 years (segment CS) to 666 years (segment GS) (**Table 3.9** – need to add geologically-constrained ranges of dates of last ruptures to table). The calculation sequence deals with these cases by sampling the range of possible dates (**Figure 5.8**).

Aperiodicity parameter (α)

The BPT distribution (5.8) is fully described by two parameters. One is the mean rate of events, μ. The other, aperiodicity (α), is a measure of the irregularity of the length of the intervals between successive events. A perfectly regular sequence has α=0. The aperiodicity is equal to the coefficient of variation (cov) of the distribution, defined as the standard deviation of the intervals divided by the mean rate. Because common practice has been to use cov to refer to a value of the coefficient of variation estimated from data, we use aperiodicity to refer to the BPT model parameter.

Aperiodicity was estimated by Ellsworth et al. (1999) from 37 sequences of repeating earthquakes. These sequences range in magnitude from **M** –0.7 to **M** 9.2, range in length from 3 to 13 events (median=6 events), include (in some cases) open intervals, and occurred in a wide variety of tectonic settings **(Table 5.2**). The values of α estimated for these sequences do not depend in any obvious way on their magnitude, length or tectonic setting (Ellsworth et al., 1999). We consider this data set sufficiently robust for inferring a representative distribution of values for α to use in our models.

Ellsworth et al. (1999) noted that the estimated values of α are biased to the low side because of the finite length of the sequences. To get around this problem, they used the actual number of intervals in each sequence to generate sets of synthetic earthquake sequences for various α-values, and then used the same computational procedure to estimate α in the synthetic sequences that was used for the observed sequences. Their results show that sets of synthetic sequences generated with a fixed value of α produce a distribution of estimates of α whose shape is similar to that obtained for the observed data. Further, they showed that a "generic" aperiodicity of 0.5 was consistent with the 37 sequences.

WG02 chose sampling points for α (with relative weights) of 0.3 (0.2), 0.5 (0.5), 0.7 (0.3). These values and weights were also used in WG99, and are similar to the cov of 0.5±0.2 used by

WG95. These sample points and weights define a pdf for α similar to the distribution observed by Ellsworth et al. (1999) for the 37 observed sequences (**Figure 5.10**).

Earthquake Interactions

The occurrence of an earthquake redistributes stress throughout the region. Depending on fault geometry, one earthquake may move a nearby fault segment closer to failure if the stress increases, or further from failure if the stress decreases. We consider the effect of the 1906 San Francisco earthquake (which ruptured the San Andreas fault) on segments of the other SFBR faults, and the effect of the 1989 Loma Prieta earthquake on the two nearby San Andreas fault segments (SAS and SAP). (Limitations in the Fortran program allow us to incorporate only one perturbation per segment; therefore we do not model the small effect of the 1989 earthquake on faults other than the San Andreas.) We calculate the stress changes using an elastic dislocation model. The average Coulomb stress change on a fault segment is converted into an equivalent time change by dividing stress change by estimated loading rate. For example, a drop in stress of 1 bar on a fault segment that is being loaded at a rate of 0.1 bar/year produces a "clock change" of –10 years, as 10 years will be required to bring the fault back to the same state.

Table 5.2. Recurrent Earthquake Sequences and Their Estimated Parameters for the Brownian Passage Time Model

Location	M	Last	N	μ	α	$\mu_{0.5}$	Reference
Copper River Delta, USA	9.2	1964	9	683	0.23	753	Plafker and Rubin, 1994.
Willipa Bay, USA	9.0	1700	7	526	0.53	530	Atwater and Hemphill-Haley, 1997.
Wairarapa fault, N.Z.	8.2	1855	5	1551	0.18	1355	Van Dissen and Berryman, 1996.
Nankaido, Japan	8.1	1946	9	158	0.40	166	Ishibashi and Satake 1998.
Tonankai, Japan	8.1	1944	7	210	0.75	192	Ishibashi and Satake 1998.
Pallett Creek, USA	7.8	1857	10	146	0.97	115	Sieh et al., 1989.
Wrightwood, USA	7.8	1857	6	150	0.71	138	Biasi and Weldon, 1998.
Pitman Canyon, USA	7.8	1812	6	180	0.96	144	Seitz, Weldon and Biasi, 1997.
Miyagi-Oki, Japan	7.5	1978	11	36	0.27	40	Utsu, 1984.
Brigham City, USA	7	-130	6	1476	0.31	1645	McCalpin and Nishenko, 1996.
Tanna fault, Japan	7.0	1930	7	972	0.65	866	Tanna Fault Trenching Research Group, 1983.
Irpinia fault, Italy	6.9	1980	5	2058	0.58	2042	Pantosti et al., 1993.
Parkfield, USA	6.4	1966	6	25.0	0.44	26.5	Bakun and Lindh, 1995.
Stone Canyon (San Andreas fault)							
- Set 2	5.0	1995	4	14.6	0.40	18.6	Ellsworth, 1995.
- Set 3	4.7	1986	3	20.3	0.37	24.4	"
- Set 1	4.2	1995	5	14.7	0.29	16.2	"
- Set 10	4.1	1995	7	10.2	0.25	11.3	"
- Set 5	4.0	1990	6	10.6	0.32	12.4	"
- Set 8	4.0	1990	5	12.3	0.35	14.7	"
- Set 9	4.0	1990	5	13.0	0.42	14.7	"
Parkfield (San Andreas fault)							
- PK1	1.4	1994	9	1.12	0.16	1.25	Ellsworth, 1995.
- S46	0.9	1993	5	1.3	0.22	1.5	Nadeau and Johnson, 1998.
- S44	0.9	1995	6	1.7	0.24	1.9	"
- S40	0.8	1995	5	1.6	0.23	1.8	"
- S39	0.7	1993	7	0.99	0.55	0.99	"
- S35	0.7	1994	5	1.8	0.26	2.0	"
- S34	0.6	1993	5	1.6	0.47	1.7	"
- S33	0.5	1992	5	1.4	0.87	1.2	"
- S27	0.5	1992	9	0.54	0.62	0.52	"
- S25	0.4	1996	6	1.6	0.43	1.7	"
- S22	0.4	1992	5	0.83	0.43	0.87	"
- S21	0.3	1995	8	1.1	0.50	1.2	"
- S20	0.3	1995	6	1.5	0.59	1.5	"
- S18	0.3	1992	5	1.3	0.33	1.4	"
- S07	0.0	1992	9	0.64	0.11	0.72	"
- S05	-0.1	1995	13	0.73	0.32	0.78	"
- S01	-0.7	1995	9	0.95	0.40	1.0	"

Notes to Table. Location - geographic placename of paleoseismic site, name of fault, or fault segment. M - magnitude of typical event. Last - calendar year of last event; negative dates for B.C. N - number of events in earthquake series. μ - estimated mean recurrence interval in years. α - estimated aperiodicity of earthquake series. $\mu_{0.5}$ - estimated mean recurrence interval in years for $\alpha = 0.5$. Reference - primary data source for this earthquake series. [Table 2 of Ellsworth et al., 1999]

Calculating the BPT-step perturbation requires information about the time of the perturbation producing the step in state and about the size of the state change. Although other approaches are possible, we converted the clock change to a state step using the recurrence time of the event. If, for example the mean recurrence time were 200 years, a –10 year clock change would amount to a –5% change in state. Details of the BPT-step calculation are provided in **Appendix F**. Stress perturbations associated with the 1906 earthquake were large and affected faults throughout the region; the range of values shown in **Appendix F** (**Table 6**) reflect the variety of parameter choices and model geometries used in the calculation of the Coulomb stress changes. The **M**6.9 Loma Prieta earthquake in 1989 had a more local effect on the nearby Santa Cruz Mountains and Peninsula segments of the San Andreas fault (**Appendix F, Table 6**). Because of the proximity of these two segments to the earthquake, the stress effects vary considerably over the segments, yielding a broad range of clock change values within the 90% confidence bounds. However, the probability of failure on these two segments was already low given the relatively small time that has elapsed since 1906, and the perturbation from 1989 did not change this low probability by much.

The Loma Prieta earthquake is considered here to have occurred in the "background", near but not on the San Andreas fault (Shaw and Suppe, 19xx; Dietz and Ellsworth, 1990, 1997). Because this interpretation is widely but not universally accepted, we examined the probability implications of this choice by calculating probabilities (using the BPT-step model) for the case in which the Loma Prieta earthquake is on the SAS segment of the San Andreas fault. In this comparison, the reset date on the SAS segment was changed to 1989 and no state steps were applied to any of the San Andreas fault segments. The results are only slightly different in this case, as shown in **Table 5.3**.

Table 5.3. Comparison to the case in which the Loma Prieta earthquake is considered to have occurred on the San Andreas fault. Results of 10,000 trials using the BPT-step probability model.

	Mean Probability of M≥6.7 Earthquake [95% uncertainty bounds]	
	Loma Prieta earthquake is on the SAS segment	Loma Prieta earthquake is in the background
Segment SAS	0.078 [0.025 - 0.163]	0.099 [0.004 - 0.272]
Segment SAP	0.091 [0.005 - 0.190]	0.101 [0.005 - 0.233]
Segment SAN	0.106 [0.008 - 0.220]	0.110 [0.009 - 0.230]
Segment SAO	0.105 [0.006 - 0.220]	0.109 [0.007 - 0.232]
Entire San Andreas fault	0.146 [0.015 - 0.314]	0.235 [0.029 - 0.524]
SFBR Region	0.673 [0.534 - 0.855]	0.682 [0.538 - 0.864]

Other methods for calculating stress changes (and their effects) produced by the 1906 and 1989 earthquakes based on more complex physical processes were considered. These involve models

of post-seismic viscoelastic stressing and rate-and-state friction. However, these approaches were judged to be insufficiently vetted for incorporation into the current SFBR probability calculations. These approaches are discussed in **Chapter 8**.

Time Predictable Model

The final probability model is the *time predictable* model (Shimazaki and Nakata, 1980). In this model, the next earthquake will occur when tectonic loading restores the stress released in the most recent earthquake. Typically, dividing the slip in the most recent earthquake by the fault slip rate approximates the expected time to the next earthquake. The time predictable model does not say anything about the size of the next earthquake, only when it will occur.

WG88 and WG90 used the time predictable model to make their 30-year probability calculations. They assumed that the time of the next earthquake varied about the expected time according to a lognormal distribution. In this report, we use the BPT distribution to describe the dispersion of times about the expected time. The differences between the lognormal and BPT results are small compared to uncertainty associated with other parameters that go into this model.

We determine time predictable probabilities only for the San Andreas fault, because the information available for the remainder of the SFBR faults was either lacking or judged to be too uncertain. Application of the time predictable model to the San Andreas fault also requires several extensions to the procedures employed by the earlier Working Groups:

1. We model the SFBR as a network of faults. The tectonic loading rate at a point in the model is the sum of the contributions from all of the loading processes. In general, in the loading model used (see **Appendix F**), the inter-seismic loading rate at any point on a fault is slightly greater than what it would be if the fault were considered in isolation. This is because there are many sub-parallel faults, and the tectonic loading of one fault by slip at depth in the model also loads the nearby faults, at least to a small degree. Other loading models may yield different results.
2. The time predictable model strictly gives the probability that a rupture will start on a segment. Earlier Working Groups did not consider multi-segment ruptures, and so for those reports the probability of rupture starting was the same as the probability of rupture. This is not the case in our model, as a segment may fail in a rupture that starts on another segment.
3. Fault segments in our model can rupture in more than one way. A segment can fail by itself, fail in combination with other segments, or produce a floating earthquake. The combination rules that give the chance that failure of one segment will propagate into its neighboring segments should depend on how close those neighboring segments are to failure, at least in the time predictable model.
4. The earlier Working Groups incorporated all of the uncertainty in the renewal time into the standard deviation of the lognormal distribution. We use Monte Carlo sampling of the parent distributions to propagate uncertainty through the model.

Application of the time-predictable model.

The time predictable model is implemented by following a six-step calculation sequence.

Step 1: Slip in the most recent event. Determine the slip on each of the four segments of the San Andreas fault by drawing a random sample from the slip model for the 1906 earthquake by Thatcher (1997).

Step 2: Slip rate of the segment. Determine the slip rate on each segment by drawing a random sample from slip rates determined for an elastic 3-D dislocation model of the fault system. In detail, the 3-D elastic model gives a stressing rate in bars, which is converted into a slip rate by multiplying it by the amount of slip needed to produce a 1 bar stress drop on the segment (**Appendix F**).

Step 3: Expected time of the next rupture of the segment. Set the time predictable return time for the segment by dividing the slip from Step 1 by the slip rate from Step 2.

Step 4: Probability of a rupture starting on the segment (ignoring stress interaction neffects). Calculate the 30-year probability of an earthquake *starting* on the segment using the BPT model and a randomly drawn aperiodicity from the weighted values listed earlier. These are "epicentral probabilities".

For the Santa Cruz and Peninsula segments, the stress effects of the 1989 Loma Prieta earthquake are used to modify the probability. This is done by adjusting the *state* of the probability model, instead of by a clock change, as was done by WG90. A randomly drawn value for the average stress change on the SAS and SAP segments is first converted to a clock change by dividing it by the same factor used in Step 2. The probability is then computed using the same procedures as were used to model the stress effects of the 1906 earthquake on the rest of the faults in the BPT-step branch of the calculation sequence.

Step 5: Convert epicentral probabilities into earthquake probabilities. From Step 4 we have the probability that an earthquake will start on each of the four San Andreas Fault segments in the next 30 years. In Step 5, a probability table is created that gives the probability that an epicenter on a segment will lead to one of the 14 possible San Andreas fault sources (10 segment combinations + 4 floating earthquakes). Several assumptions are required to make this table. They are 1) the long-term distribution of the epicenters for each of the sources; 2) the rate at which the rupture sources occur; and 3) the influence of the 1906 earthquake on multi-segment ruptures.

The distribution of epicenters is assumed to be proportional to the length of the segment. This means that if a multi-segment rupture is considered, the chance that rupture will begin on any particular segment is the length of the segment divided by the total length of the rupture source. Floating earthquakes are apportioned onto the segments in the same manner. Other possible methods would be to apportion according to segment seismogenic area or moment rate.

Our long-term earthquake source model gives the unconditional rate of rupture sources. It is straightforward to partition these rates into the rate at which its epicenters occurs on a given segment. It is simply the fraction of the time that the epicenter is on the segment (ratio of segment length to source length) times the rate of the source. If the 1906 earthquake was believed to have no influence on these rates, then this table of rates could be normalized to give the probability that when an earthquake starts on a given segment it will become any of its possible sources. If we follow this logic, then the next San Andreas event is most likely to be a full-fault rupture, as this source is the most common in the long-term model. Perhaps this is not unreasonable, as the time predictable model does not inform us about the size of the next rupture.

However, many who have thought about the behavior of the northern San Andreas fault have come to the conclusion that the likelihood of a full-fault rupture is low (Lindh, 1983; Sykes and Nishenko, 1984; WGCEP, 1988; WGCEP, 1990). This is because the San Andreas fault slipped substantially more in the 1906 earthquake on the SAN and SAO segments, compared to the SAP and SAS segments. The question is how to implement this hypothesis in the context of our model.

We begin by examining the full four-segment rupture of the fault. In the 95 years since the 1906 earthquake, seismic moment has been accumulating at a rate of 3.7×10^{23} dyne-cm/year on the four fault segments. The accumulated moment is equivalent to a **M**7.67 earthquake. This is a substantial amount of stored moment, but is it enough to produce a full-fault rupture? If the fault was "reset" by the 1906 earthquake, then we can frame this question in probabilistic terms. Specifically, we ask the probability that the stored moment is sufficient to produce a full fault rupture. According to the long-term model, the magnitude of a full fault rupture is **M**7.90±0.12. The probability of a full fault rupture being no larger than **M**7.67 is 0.03. Under the assumption that the next earthquake will be produced by the strain energy accumulated since the last one, we would conclude that a full fault rupture is unlikely at the present time.

Consider now the likelihood that a floating earthquake (**M**6.9 ± 0.12) could occur on the SAN. The moment stored since 1906 on any part of the SAN that is long enough to produce a floating earthquake is equivalent to **M**6.98, and the probability that this is sufficient moment to produce a floating earthquake is 0.75. Thus, we might expect that if a large earthquake were to begin on the SAN, it would more likely be a floating earthquake than a full fault rupture, if these were the only two options.

The operating assumption here that the strain energy available for release is the strain energy stored since the last rupture is simply the slip predictable model (Shimazaki and Nakata, 1980). The empirical evidence provides more support for this model than it does for the time predictable model (Ellsworth, 1995). We will adopt this hypothesis for the purpose of modulating the long-term probability that an earthquake that starts on one segment will result in the failure of any of its possible rupture sources.

In these calculations, the rates from the SFBR model are the beginning point. The slip predictable assumption is used to ask the question "Has enough moment accumulated on this rupture source since 1906 to make the earthquake?" This is posed as a probability problem, and

the answer (a number between 0 an 1) is used to modulate the probability of the rupture source *given* an epicenter on the segment.

To implement this model, we multiply the long-term rate for each rupture source by the probability that the moment stored since 1906 is sufficient to produce the earthquake. The adjusted rates are then converted into the probability that an epicenter on a segment will produce a given source.

Table 5.4 is the result of this procedure. The columns correspond to the locations of the epicenter. The rows correspond to the rupture sources. The entries in the table are the probability, *given an epicenter on the segment*, that it will become a particular rupture source. The individual columns must sum to unity, because the occurrence of the epicenter on the segment is a given. For example, the probability that an earthquake that starts on SAP segment will *not* propagate into either the SAS or SAN segments is 0.809 (0.462+0.347). The probability that a SAP rupture will grow into a full rupture of the fault is 0.019.

The most likely rupture to result from a earthquake initiating on the SAN or SAO segments is a floating earthquake, a result controlled by the slip-predictable assumption, the small magnitude of the floating earthquakes relative to the single-segment rupture sources SAN and SAO, and the large, unrecovered moment deficit on those segments from 1906. Conversely, earthquakes initiating on the SAS or SAP segments are less likely to be floating earthquakes (but the likelihood is still significant) because the mean magnitudes for these single-segment sources are closer to the magnitude of the floating earthquake and much of the 1906 moment deficit has been recovered.

Table 5.4. Probability that an epicenter on a given segment will become a specific rupture source.

Resulting Rupture Source		Location of Epicenter			
	Mean Mag	SAS	SAP	SAN	SAO
SAS	7.03	0.682	0.000	0.000	0.000
SAP	7.15	0.000	0.462	0.000	0.000
SAN	7.45	0.000	0.000	0.071	0.000
SAO	7.28	0.000	0.000	0.000	0.227
SAS + SAP	7.42	0.106	0.170	0.000	0.000
SAP+ SAN	7.65	0.000	0.000	0.000	0.000
SAN + SA0	7.70	0.000	0.000	0.122	0.103
SAS + SAP + SAN	7.76	0.000	0.001	0.001	0.000
SAP + SAN + SAO	7.83	0.000	0.001	0.002	0.002
Full fault rupture	7.90	0.012	0.019	0.030	0.026
SAS floating earthquake	6.90	0.200	0.000	0.000	0.000
SAP floating earthquake	6.90	0.000	0.347	0.000	0.000
SAN floating earthquake	6.90	0.000	0.000	0.773	0.000
SAO floating earthquake	6.90	0.000	0.000	0.000	0.643

Step 6: Compute 30-year source probabilities. Step 4 gives us the probability of an epicenter on each of the four San Andreas Fault segments. To convert these epicentral probabilities into

source probabilities it is only necessary to multiply the corresponding column of **Table 5.4** from Step 5 by the epicentral probability. This operation apportions probability among exclusive alternatives while preserving the total probability.

Finally, the probability of a rupture of any segment is computed from the rupture source probabilities in the standard way for combining probabilities:

$$P_{total} = 1 - (1 - P_1)(1 - P_2)...(1 - P_N) \qquad (5.9)$$

where P_i are rupture source probabilities.

The ability of the time-predictable model to provide accurate estimates of earthquake probability depends on the accuracy of the input data on slip in 1906; of the loading rate of the faults; of the stress effects of the 1989 Loma Prieta earthquake; and of the slip-predictable assumption for rupture source probability.

Final calculation steps

Probabilities for Fault Segments and Fault Systems

The earthquake probability for a given fault segment is obtained by combining the probability that the segment will rupture in any of the fixed rupture sources that involve that segment, or in an earthquake of the exponential tail (for sufficiently small threshold magnitude M_T) (**Figure 5.1**). This probability is then combined with a probability contribution from the fault's "floating" earthquake, in proportion to the segment's relative moment rate. To calculate the probability of an earthquake occurring anywhere on the fault, the source probabilities are aggregated. Probabilities are aggregated using equation (5.9).

Probabilities for Earthquakes in the Background

We use the exponential model to describe the distribution of times between successive **M** ≥ 6.7 earthquakes in the background. Two G-R relations (log N/yr = $a - b$**M)** describing the magnitude-frequency distributions of earthquakes in the background were developed in Chapter 4: one for the 1951-1998 period with a = 3.67 (3.60 to 3.74 at 95% confidence) and b = 0.89; and the other for the 1836-2001 period with a = 3.94 (3.62 to 4.31 at 95% confidence) and b = 0.89.

The 1951-1998 relation provides a rate based on recent earthquake activity, while the 1836-2001 relation represents the longer-term activity. We use both models to estimate the 30-year conditional probability of **M** ≥ 6.7 earthquakes in the background. WG02 voted on the relative weights of the 2 models, assigning 0.458 to the long-term rate estimates and 0.542 to the estimate based on recent seismicity. Each model is split into 3 branches based on the mean rate, the +2σ rate and the -2σ rate, with weights of 0.74, 0.13, and 0.13 respectively.

Weighting alternative probability models

Each of the five probability models described above offers an alternative method for calculating 30-year earthquake probabilities for the SFBR fault segments, faults, rupture sources, and region as a whole. The decision of how to weight these alternative methods represents a major source of epistemic uncertainty in the SFBR probabilities.

After considerable discussion of the pros, cons, implications, machinery, and quantitative results of each probability model, each member of WG02 was asked to divide unit weight among the four (or five, in the case of the San Andreas fault) probability models. Different weight distributions were permitted on each fault. For each model and fault the 13 fractional weights were averaged, and those mean weights were used to scale the contribution of each probability model to each fault system. **Table 5.5** summarizes the mean branch weights thus determined.

The BPT, BPT-step and time-predictable models accounted for approximately half or more of the assigned weight on all faults (except on the Mt. Diablo thrust, for which the geologic data are the most limited). On the San Andreas fault, for which we know the most about past earthquake occurrence, these models account for 72% of the assigned unit weight. The Poisson and Empirical models account on average for 21% and 29% of the unit weight, respectively. In our discussions, WG02 members often cited two issues that influenced their judgement in assigning these weights: the relative amount and quality of geological data on each fault (with more and better data generally favoring more weight on recurrence models); and the modeling of the stress interaction effects of the 1906 earthquake (with lower confidence in the accuracy of the calculations for interaction effects favoring more weight on the Empirical and Poisson models).

The division of weight among the probability models is displayed graphically in **Figure 5.12** for each of the seven fault systems. Because the weights assigned to any particular model differ between the fault systems and the Time Predictable model is only used for the San Andreas Fault, a mixture of probability models must occur in the random sampling of branches by the code, at least some of the time.

Table 5.5. Mean expert weights for probability models applied to the SFBR fault systems.

Fault system	Poisson	Empirical	BPT	BPT-step	Time-predictable
San Andreas	0.100	0.181	0.154	0.231	0.335
Hayward/Rodgers Creek	0.123	0.285	0.131	0.462	-
Calaveras	0.227	0.315	0.142	0.315	-
Concord/Green Valley	0.246	0.277	0.123	0.354	-
San Gregorio	0.196	0.292	0.115	0.396	-
Greenville	0.231	0.288	0.131	0.350	-
Mt. Diablo thrust	0.308	0.396	0.092	0.204	-

In order to maximize the extent to which the choice of probability model is correlated across the region, the input probability model weights were arranged in the order shown in **Figure 5.12**. (It will be seen that this is in order of increasing mean earthquake probability, except in the case of the San Andreas fault.) On each realization, the code draws a single random number (0 to 1) to select the probability model branch for each fault system. Expressed graphically, this random

number corresponds to a horizontal position in **Figure 5.12** and a vertical line drawn at that position specifies the probability model branch tip. Inspection of the figure shows that for most realizations the choice of model will be the same across the six faults, excluding the San Andreas. For example, the Empirical model weights are similar across the suite of faults, and so there are many realizations for which the probabilities for all faults are calculated using the Empirical model. When the Time Predictable model is operative on the San Andreas Fault, the BPT and/or BPT-step model is operative on most of the others.

This sampling strategy does not alter the mean probability for any fault or for the region as a whole. However, it does affect the variance of the regional probability due to the correlation of probability models between fault systems. Had independent random draws been used for each fault system, the range of regional probability values would have been smaller. Members of the working group held differing views on this question. Some of us felt that the model weights should ideally be correlated across the region, because probability models are alternative views of the physics of the earthquake generation process and as such apply to all the faults, since fault physics doesn't change from place to place. A single realization of the model should, in this view, have correlated physics across the region. (Note that we adhere to this principle of regional uniformity when it comes to fault width: our interpretation of microseismicity depth may be giving us a biased measure of width, but that bias is presumed to be regional, not different on every fault; therefore, the width branch choice is correlated over the faults.)

In contrast, some members of the working group believed that the application of a probability model should depend on particulars about each fault, such as the amount and quality of information available for the fault, their trust in the calculation of the effects of the 1906 earthquake near to and far from the San Andreas fault, their degree of faith in the characterization of the faults, etc. In the end, however, correlating the application of probability models over the faults would have required time-predictable calculations for all faults, which were not judged to be appropriate. So, as a practical matter, correlating the probability models was not possible.

Probabilities for the SFBR model

The process of computing source, segment, fault and fault system probabilities is repeated for each branch tip in the calculation sequence. Recall that each of the 10,000 computed branch tips represents a viable model for the entire fault system. The distribution of probability for any element of the model, or aggregation, captures the epistemic uncertainty in probability. In other words, the distribution of probability arises from a lack of knowledge about each of the parameters in the model or applicability of the models themselves. These distributions are summarized by their mean value and the probabilities for models at the 2.5% and 97.5% points in the range of the values.

Alternative renewal models not used in this study

As mentioned above, we had a choice of several renewal models, including the Lognormal, Normal, Weibull, Gamma, and BPT. These models all perform similarly for elapsed times near the mean recurrence time, and available data cannot distinguish between them.

We checked this assertion directly by substituting a Lognormal calculation for the BPT model in the WG02 code. The results show that given the same input parameters and distributions with the same mean and coefficient of variation, the two distributions give nearly identical probabilities. This is true for both the case of non-interacting faults and the case in which the effect of the1906 earthquake is represented by a clock change on the other faults. Thus, we conclude that the choice of renewal model is not a significant source of epistemic uncertainty.

The observation that these two models give nearly the same results is not a reason for retaining the Lognormal model used by previous working groups. The Lognormal model, which has seen much use in the literature and been used by previous Working Groups, has not been demonstrated to be superior to other pdf's. The BPT model, however, offers the opportunity to include the stress step from the 1906 and 1989 earthquakes directly in the probability calculations, whereas the Lognormal does not.

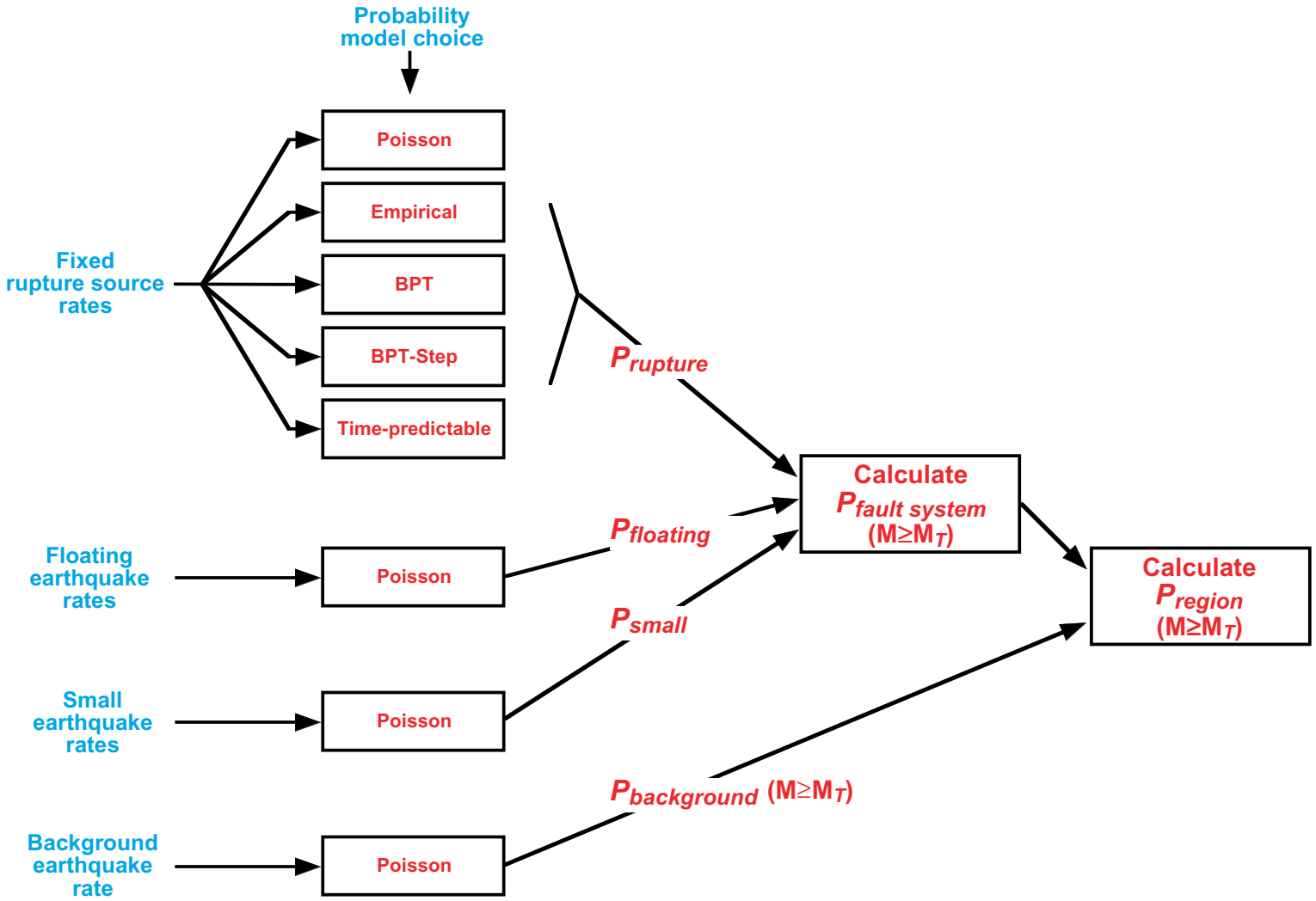

Figure 5.1. Final steps in the calculation sequence, in which time-dependent effects enter into the WG02 model. Long-term rates of fixed rupture sources, floating earthquakes, small earthquakes and background earthquakes are input to a suite of five probability models. Fixed rupture sources are input to all probability models, according weights determined by expert opinion. The other earthquake sources are input only to the Poisson model. Resulting probabilities for the fixed rupture sources, floating sources and small earthquakes are combined for each fault system, for events with magnitude $\mathbf{M} \geq M_T$. Finally, the probability of background earthquakes is combined with the probabilities for fault systems to give the regional earthquake probability.

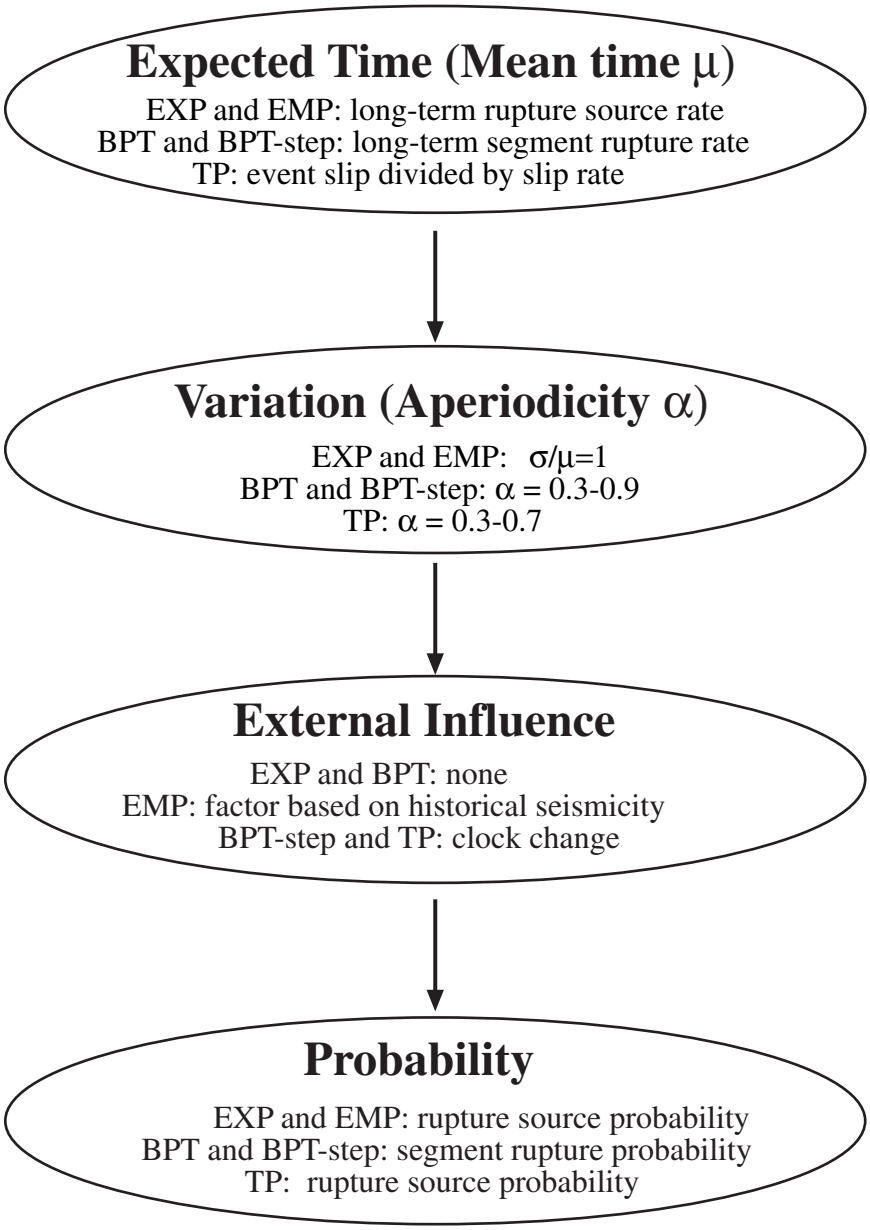

Figure 5.2. Summary of information used and assumptions made by each probability model. WG02 used five probability models: Exponential or Poisson (EXP); Empirical (EMP): Brownian Passage Time (BPT); BPT with stress-step interactions (BPT-step); and Time Predictable (TP).

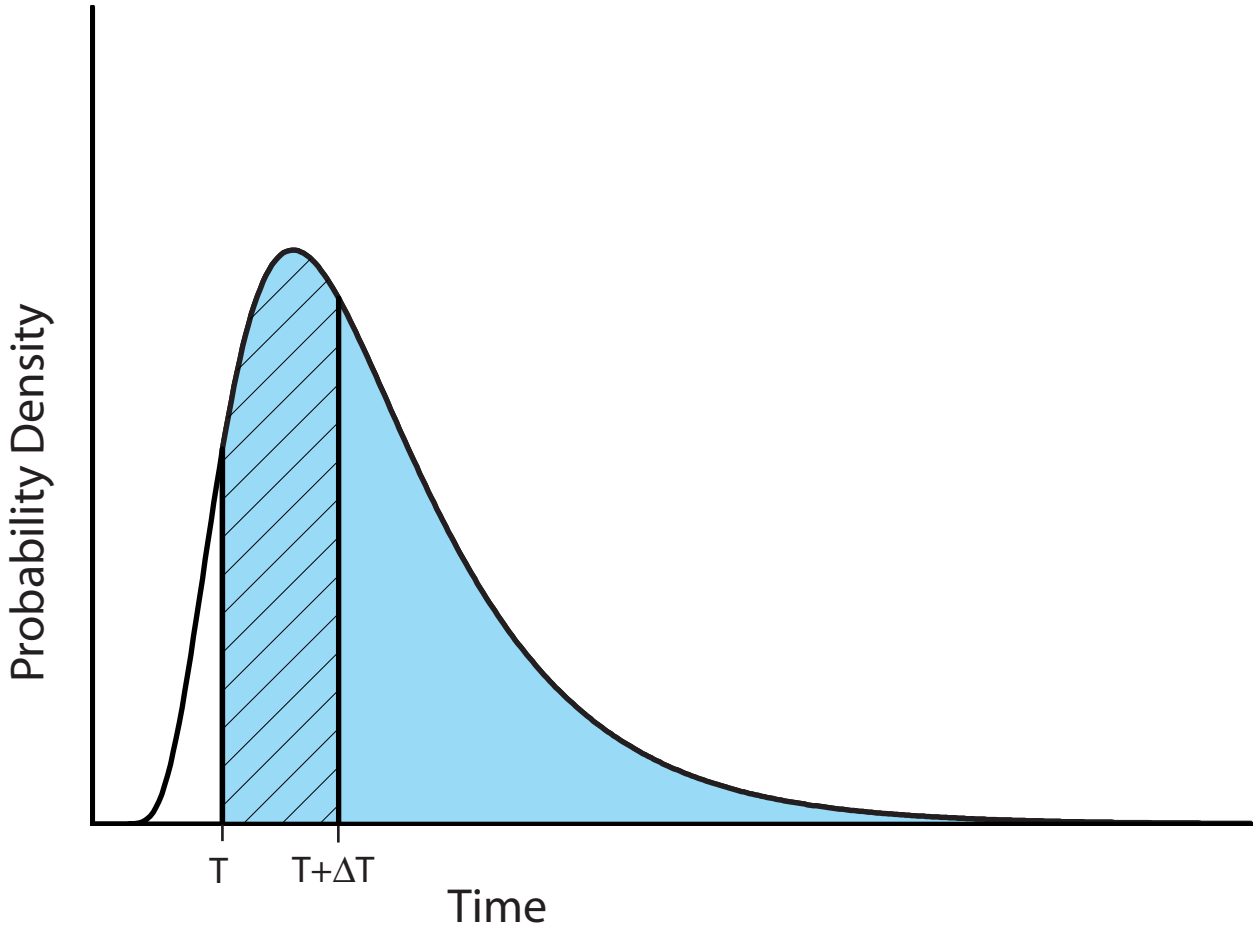

Figure 5.3. Diagram illustrating the calculation of conditional probability from a probability density function. The time interval of interest (exposure time) is from T (the present time) to T+ΔT (hatchured area). The survivor function at time T is equal to the shaded area. The conditional probability is the ratio of these two areas.

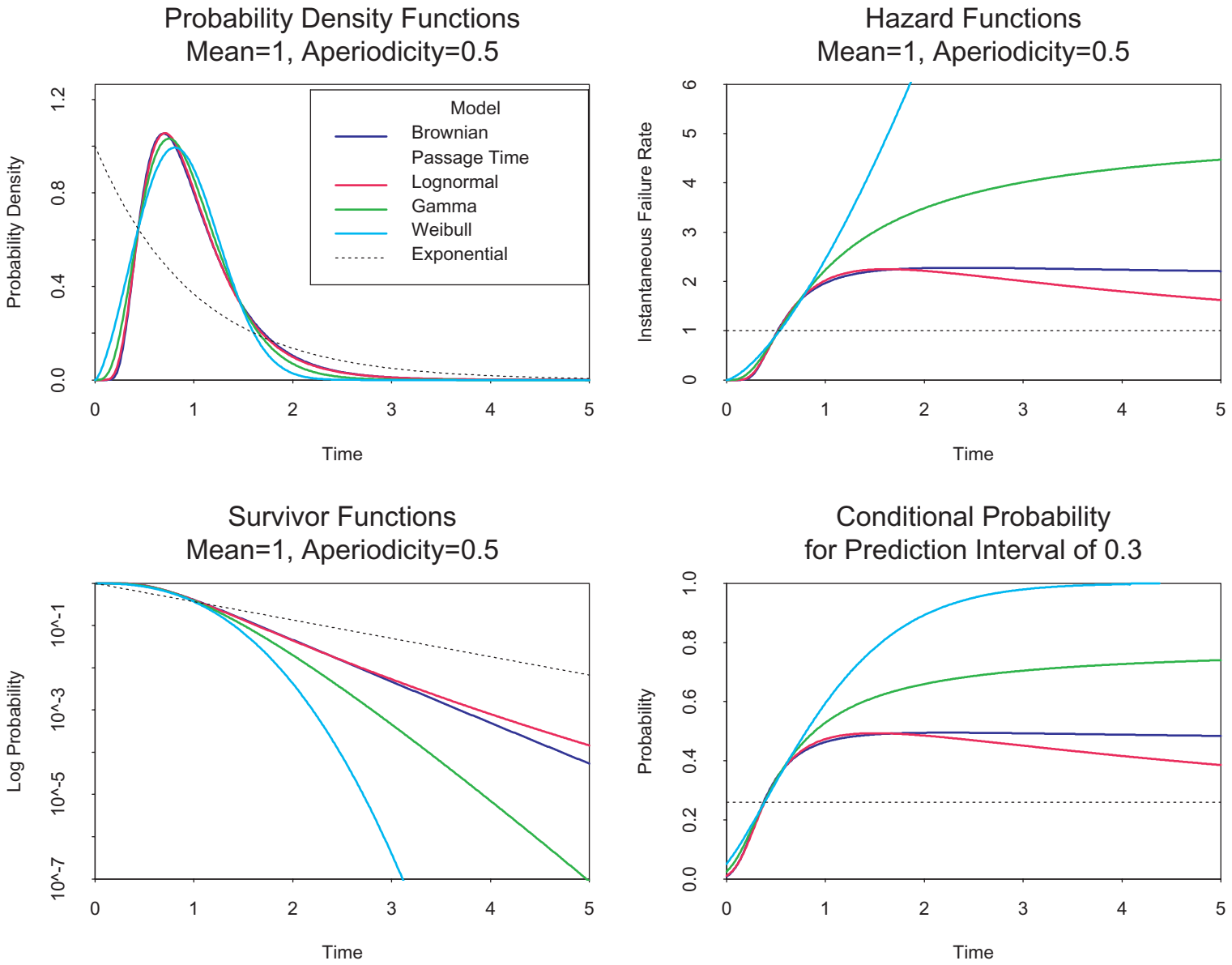

Figure 5.4. Comparison of several probability models used in long-term earthquake forecasting. All models have a mean of 1, and with the exception of the Exponential (Poisson) model, an aperiodicity (coefficient of variation) of 0.5. Probabiity density, hazard, survivor, and conditional probability functions are shown.

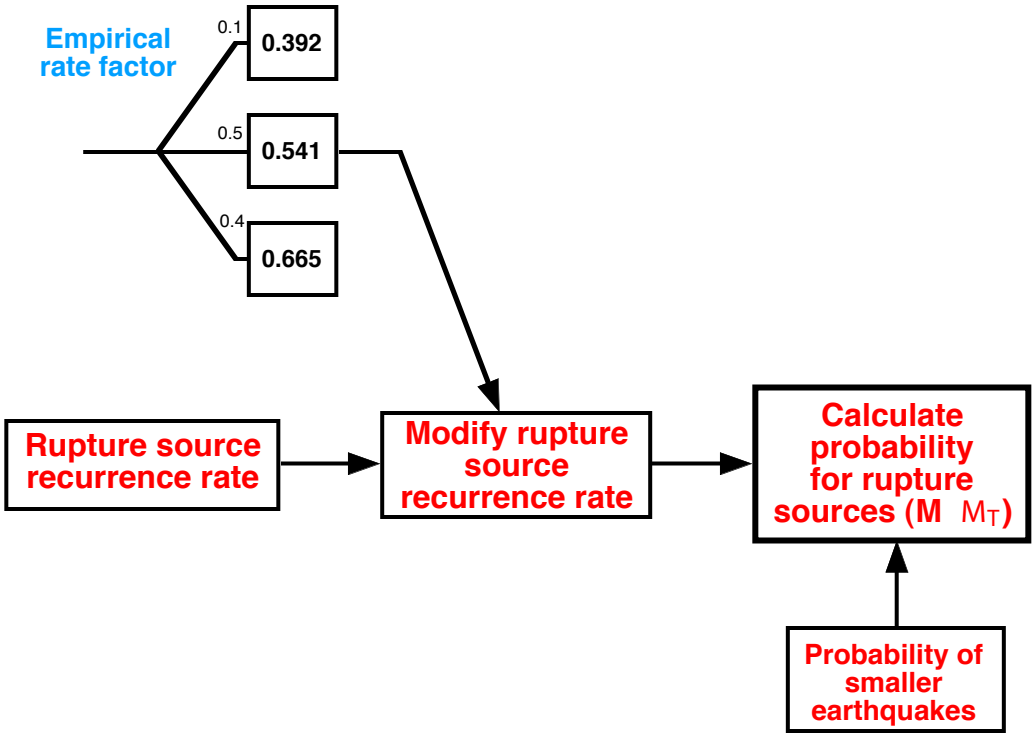

Figure 5.5. Method of calculating rupture source probabilities using the Empirical model. This model requires values for the long-term recurrence rate, plus a rate factor (derived from observations of seismicity rate) that modifies the rupture source recurrence rates.

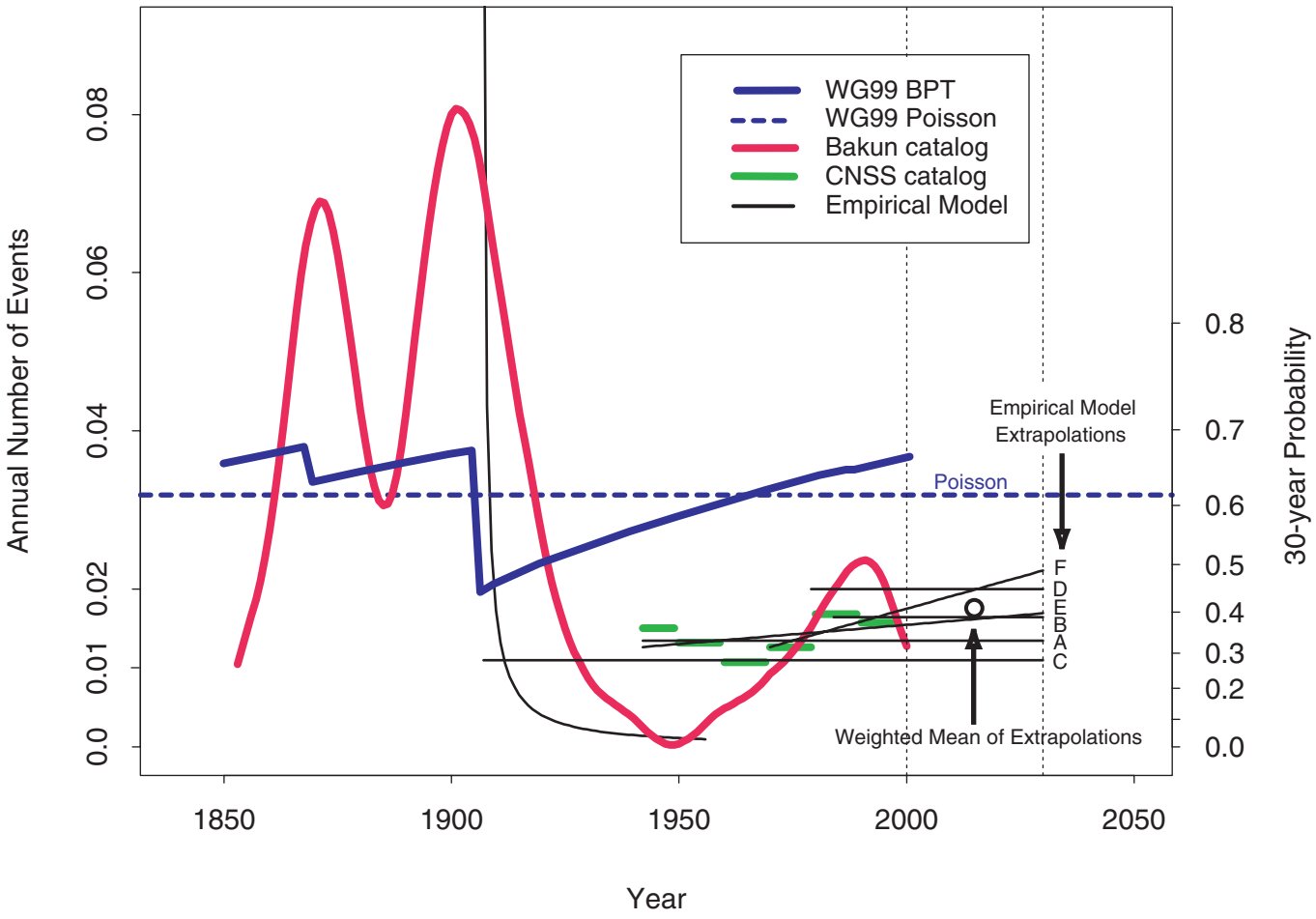

Figure 5.6. Summary of rates and 30-year probabilities of earthquakes (M≥6.7) in SFBR calculated with various models. Dark blue line shows BPT model probability calculations (for 30 years forward) with the effects of the 1906 and 1989 stress perturbations modeled with the state-step method. Dashed blue line is long-term (Poisson) regional rate in WG02 model. Red line is a smoothed representation of the rate of M≥5.5 earthquakes (Bakun, 1999) based on a sliding 15-year-wide averaging window. Green lines are mean rates of (M≥3.0) earthquakes in approximately 10-year periods in the CNSS catalog. Black lines are extrapolations of the M≥3.0 and M≥5.5 earthquake rates made using a variety of extrapolation methods (see text and Table 5.1). Black circle is weighted mean of extrapolated rates for 2002-2031. Curved light black line is generic aftershock sequence for 1906 earthquake.

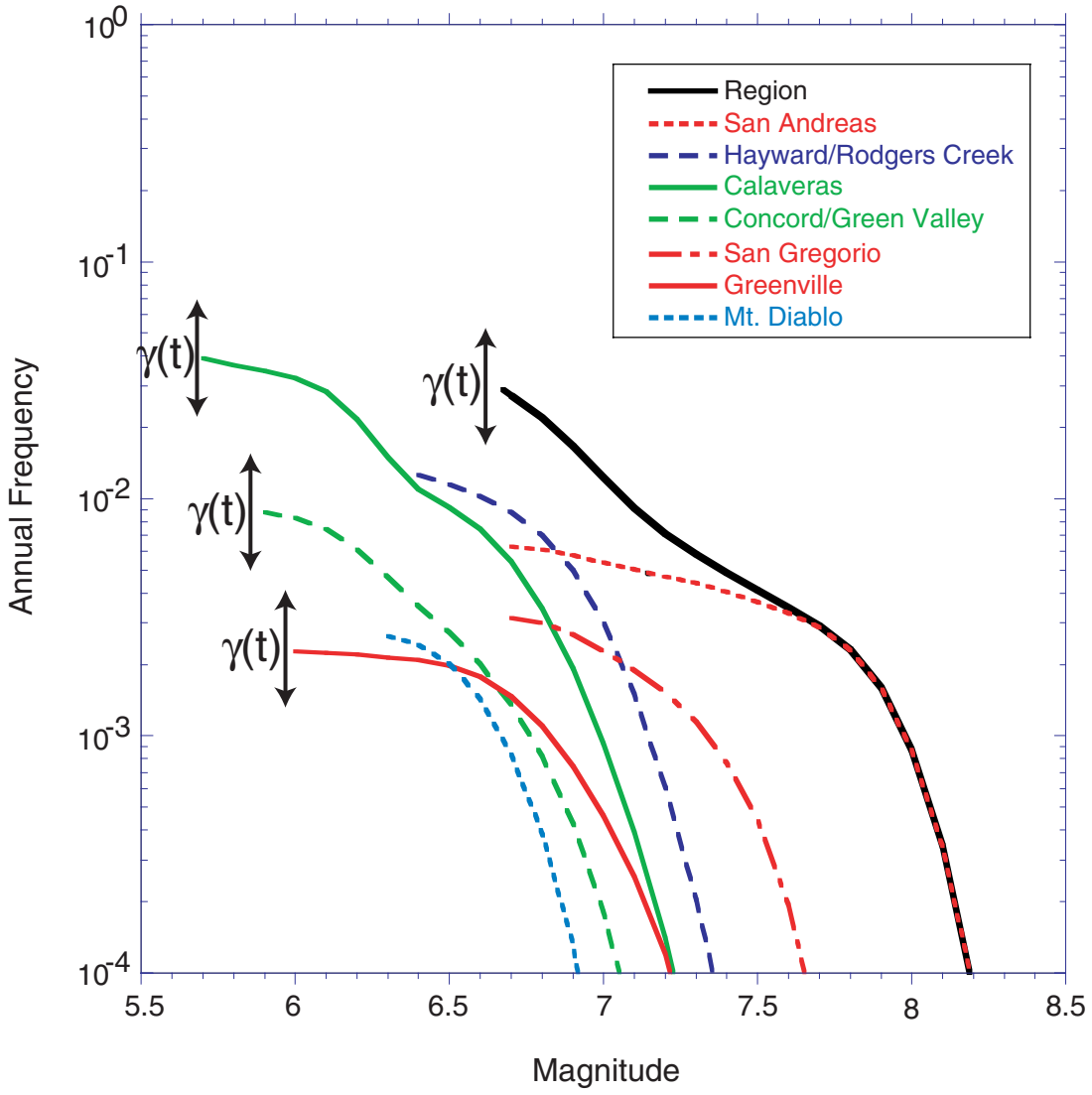

Figure 5.7. In the Empirical model, fluctuations in earthquake activity are represented by $\gamma(t)$, which modulates the rates on the faults and in the background relative to their respective long-term means (colored lines). The rates are modulated in unison. The shape of the magnitude distribution on each fault, as defined by the fault characterization model, remains unchanged; the whole distribution moves up and down in time.

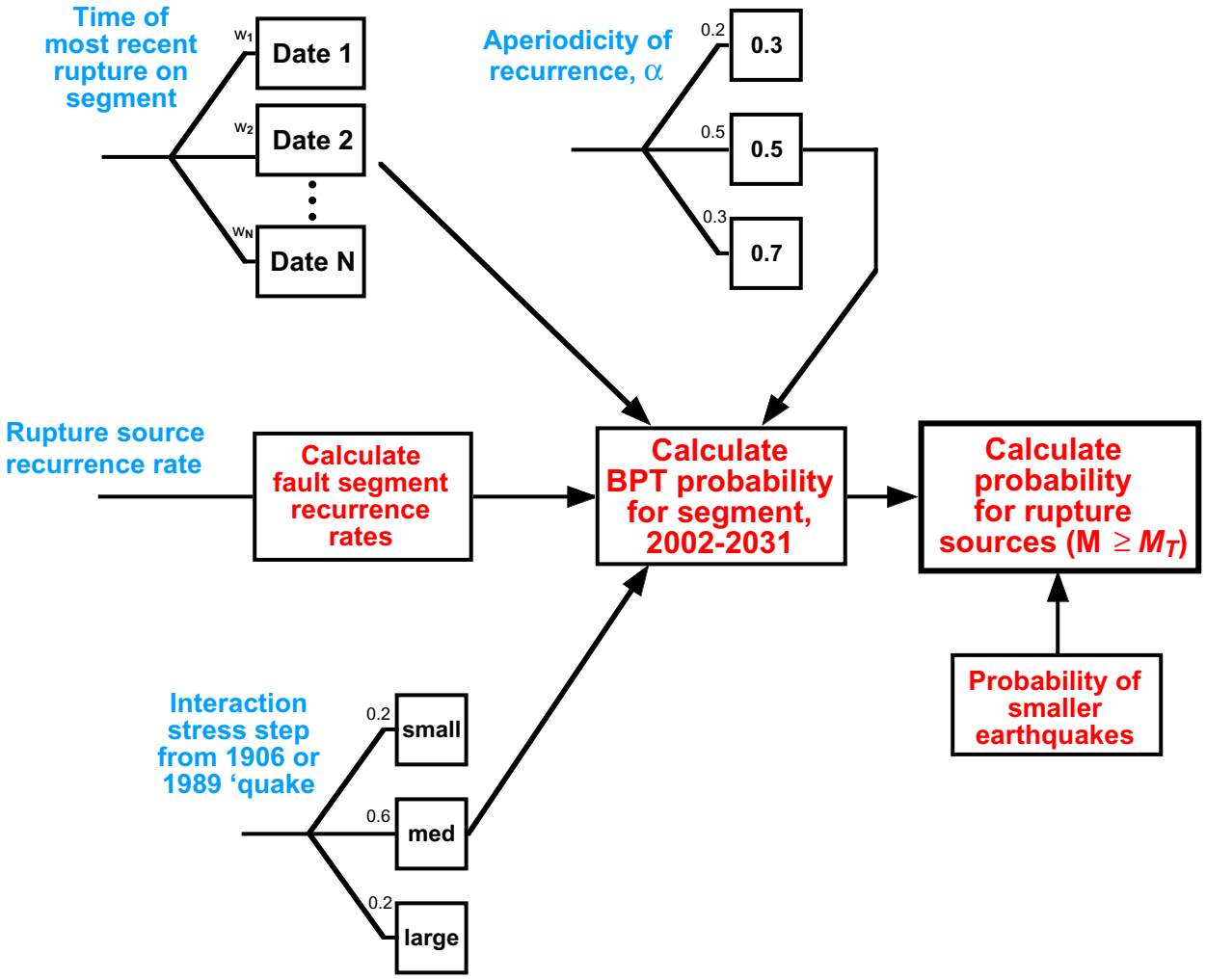

Figure 5.8. Method of calculating time-dependent probabilities using the BPT model. This model requires values for the long-term recurrence rate, the elapsed time since the last earthquake, and the aperiodicity. The modifying effect of stress steps from the 1906 and 1989 earthquakes can be accounted for.

Paleoseismic data on elapsed time are more readily available for fault segments than for rupture sources; therefore, the probability is calculated in three steps. First, the rate of failure is calculated from the rate of failure of ruptures that include it. Next, the probability of failure of each fault segment is calculated. Finally, segment probabilities are combined to determine the probability of failure of each rupture source, for earthquake magnitudes above a threshold magnitude M_T.

If M_T is small, the rupture source probability may include that of smaller earthquakes (those of the exponential tail).

A similar procedure is used in the case of the time-predictable model. In that case additional information about the slip in the last event is required.

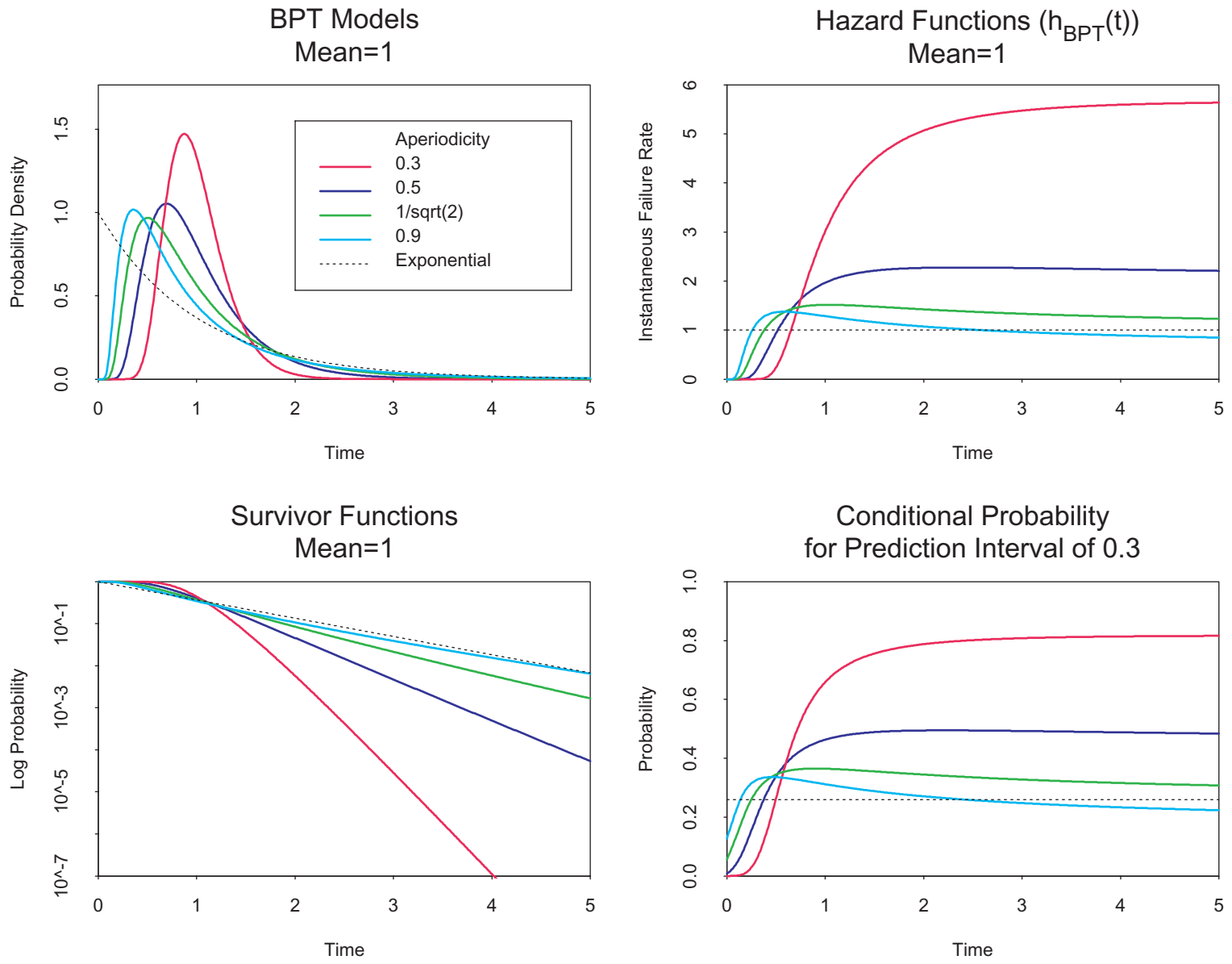

Figure 5.9. Effect of aperiodicity on the behavior of the Brownian Passage Time model. All distributions have a mean of 1. Exponential model shown for comparison. Probability density, hazard, survivor, and conditional probability functions are shown.

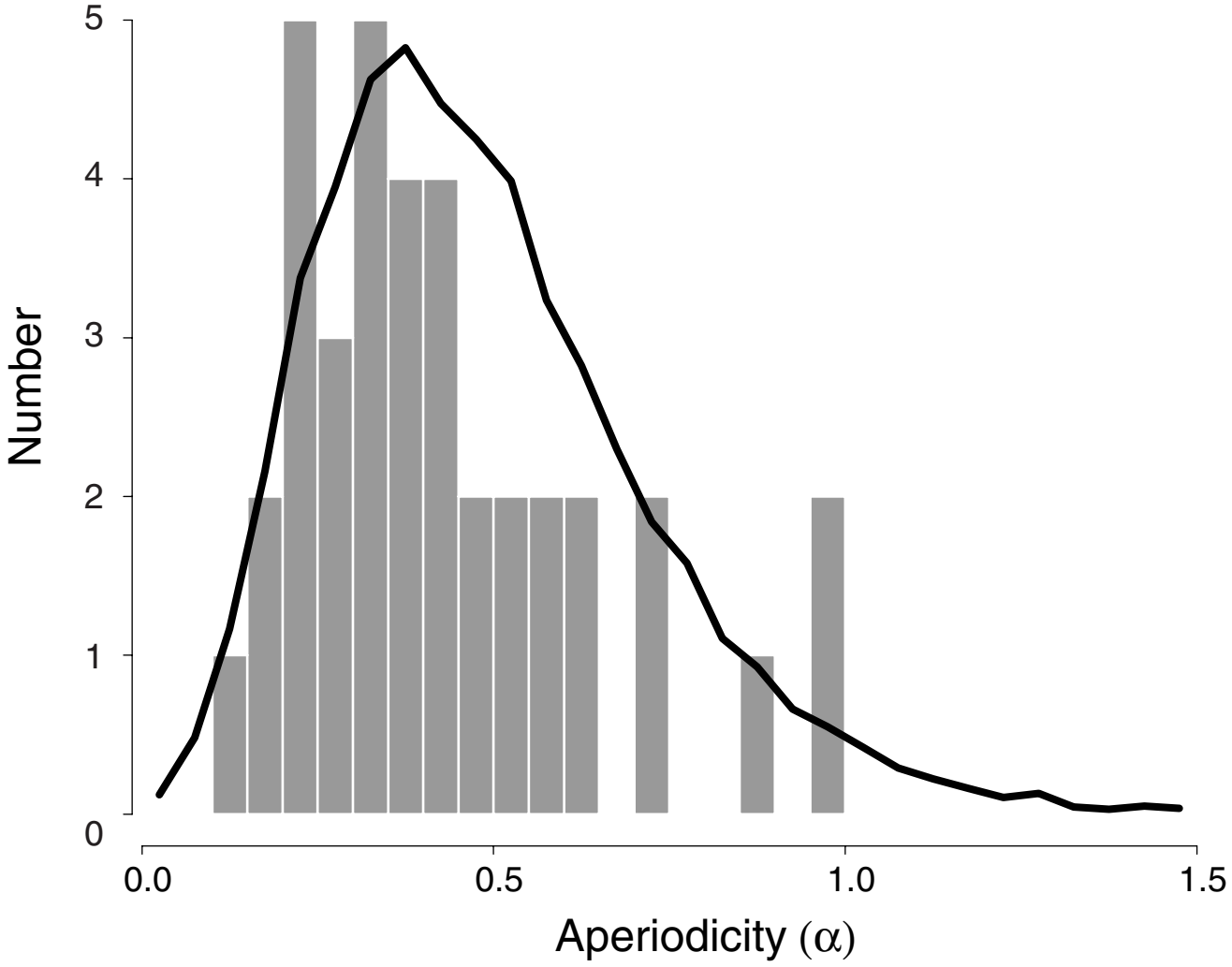

Figure 5.10. Estimates of aperiodicity (α) obtained by Ellsworth et al. (1999) for 37 earthquake sequences (histogram) and the WG02 model (solid line). The WG02 model for α is represented here by the distribution of estimates of α obtained for a set of 10,000 synthetic sequences constructed from it. The model values for α (and sampling weights) are 0.3 (0.2), 0.5 (0.5), and 0.7 (0.3).

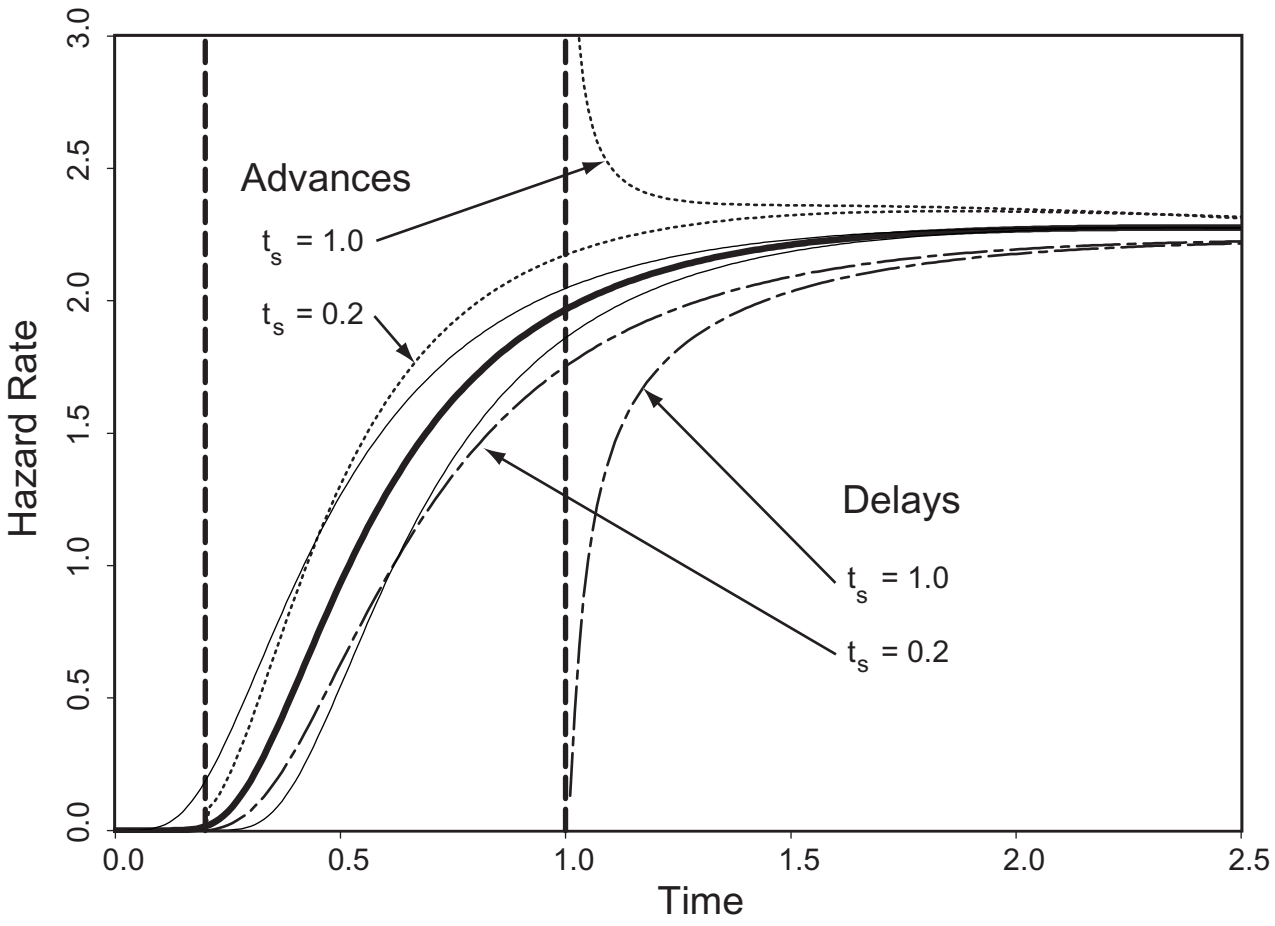

Figure 5.11. Effects on hazard rate of steps in the state of the BPT model, shown as a function of time normalized by mean recurrence time. Solid heavy curve is the hazard rate function for the unperturbed state. Dashed and dotted curves are hazard rates when an interaction step is applied early in the recurrence cycle (phase $t_S = 0.2$) and at the mean recurrence time (ts = 1.0). The absolute size of the step is 1/10 the stress drop for the rupture. Dotted curves correspond to positive steps, dashed curves to negative steps. Solid light lines, shown for comparison, are hazard rates for a "clock advance" (upper) and "clock delay" (lower) of $\pm$0.1 times the mean recurrence time, which are simply horizontal translations of the unperturbed hazard rate.

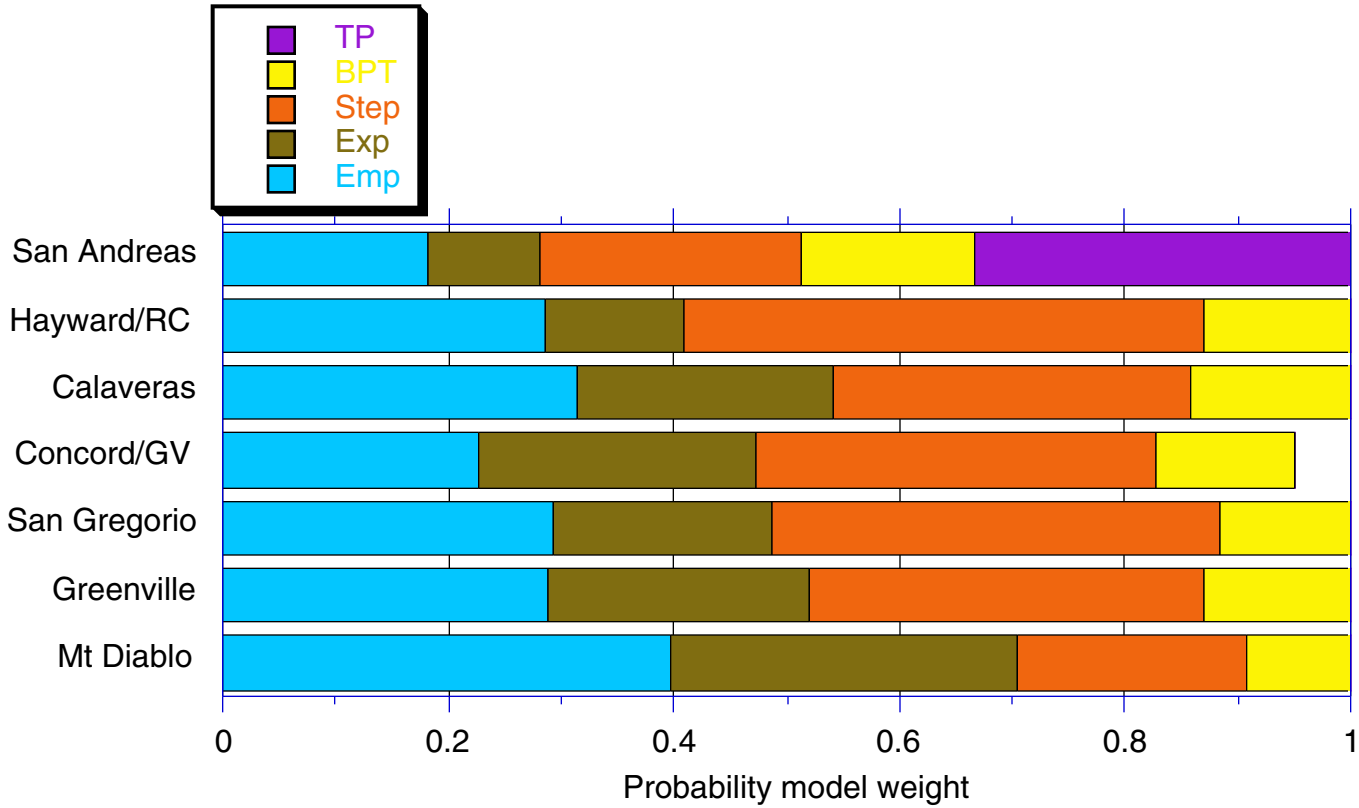

Figure 5.12. Division of weight assigned to each probability model for each fault, as determined by expert opinion and given in Table 5.5. Time predictable (TP) model was applied only to the San Andreas fault.