

# Some examples of the communication of risk and uncertainty<sup>‡</sup>

David R. Brillinger<sup>\*,†</sup>

*Statistics Department, University of California, Berkeley, CA 94720-3860, U.S.A.*

The uncertainty is as important a part of the result as the estimate itself... . An estimate without a standard error is practically meaningless. Jeffreys (1967)

## SUMMARY

Definitions are set down and results of analyses of communication of risk and uncertainty are presented for the fields of wildfires, earthquakes and space debris. These are all fields of some societal importance. Also there is discussion of methods of evaluating and displaying estimates of risk estimates' uncertainty. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS: earthquakes; risk; simulation; space debris; wildfires

## 1. INTRODUCTION

Risk is a vague concept for which there are in use both formal and informal definitions. This paper will focus on the formal side and applications. Speaking generally risk analysis refers to formal study of some event of adverse consequences. The result of an analysis is often communicated as a probability/chance or odds. The paper presents aspects for the examples of space debris collisions, earthquakes capable of causing damage, and losses from wildfires near urban facilities. In the last case one might be interested in the occurrence probability of a wildfire or the expected loss in such an event with loss the value of the houses destroyed. The reasons for preparing risk analyses include developing insurance rates, preparing legislation, decision making and forming procedures for risk reduction. In each case an uncertain future event of extreme consequences is of concern. In carrying out risk analyses one seeks to use data from past events, e.g. similar wildfires.

Communication of the results is a basic part of a formal risk analysis. If the study is to have an impact it is crucial that the authors be able to present the results in a rapid, direct, understandable, accurate, convincing manner. Luckily these days there are many visualization tools and there is the World Wide Web with its ability to display updated risk maps continually.

It is basic to provide an indication of uncertainty, see the Jeffreys' quote above. Such indications need to be included in reports and publicity.

---

\*Correspondence to: D. R. Brillinger, Statistics Department, University of California, Berkeley, CA 94720-3860, U.S.A.

<sup>†</sup>E-mail: brill@stat.berkeley.edu

<sup>‡</sup>This article is published in *Environmetrics* as a special issue on TIES 2008: *Quantitative Methods for Environmental Sustainability*, edited by Sylvia R. Esterby, University of British Columbia Okanagan, Canada.

Through the years there have been a number of environmental controversies. The discussion and resolution of these has involved statistical concepts and methods. One can mention: global warming, smoking, agent orange, fluoridation and aldrin. General information may be found in the Encyclopedia of Environmetrics (2002). The entry El-Shaarawi and Hunter (2002) concerns environmetrics, while that of Smith (2002) discusses uncertainty analysis specifically.

The paper's layout involves discussion, and methodology, followed by examples from space debris, earthquakes and urban wildfires, and then discussion and summary.

## 2. SOME PERTINENT METHODOLOGY

Point processes and marked point processes are basic to risk analysis and in particular to each of the examples presented. In two cases there is a temporal and in one case a spatial point process.

A sequence of temporal points,  $\{t_j\}$ , may be described by a counting function.

$$N(t) = \#\{0 < t_j \leq t\}, \quad \text{with the } t_j\text{'s occurrence times}$$

The rate,  $p_N = E\{dN(t)\}/dt$ , is an important descriptive parameter. If, for example,  $N$  is a Poisson process with rate  $p_N$  then  $E\{N(t)\} = t p_N$  and  $\text{var}\{N(t)\} = t p_N$ . One can estimate  $p_N$ , unbiasedly, by  $N(t)/t$  with coefficient of variation  $1/\sqrt{t p_N}$ .

The process  $\{N(t)\}$  becomes a marked point process when each point has a characteristic or mark, say  $L$ , associated with it. A realization can then be written as

$$\{(t_j, L_j)\}$$

where  $L_j$  is the mark associated with the  $j$ th event. Basic parameters are the rate of the point process  $\{t_j\}$  and in the case that  $L$  is real-valued the expected mark,  $E(L)$ .

In the case of a spatial point process with locations  $(x_j, y_j)$  and can work with the count function  $N(Q) = \#\{(x_j, y_j) \text{ in } Q\}$ .

In risk analyses basic probabilities that arise include,  $\text{Prob}\{\text{some } t_j \text{ in } (0, T] | A\} = \text{Prob}\{N(T) > 0 | A\}$  and

$$\text{Prob}\{\text{some } t_j \text{ in } (0, T] \text{ with } L_j \leq l | A\}$$

for some event  $A$  of interest and mark value  $l$ .

In various cases one can set down a likelihood function and use maximum likelihood and common inferential procedures, see Brillinger *et al.* (2002), Guttorp *et al.* (2002), Schoenberg *et al.* (2002). For example a variety of measures of uncertainty become available.

## 3. EXAMPLES

Three examples, from the writer's own experience, are discussed. They are from the fields of space debris, earthquakes and wildfires.

### 3.1. The haystack orbital Debris review

There are thousands of objects orbiting the Earth. The ones of concern here have sizes less than 10 cm but can still cause substantial damage to the International Space Station, the Space Shuttle and orbiting

satellites. NASA has an Orbital Debris Program concerned with this problem. In 1997 they set up a Panel to review the procedures of their Program having viewed the problem as an environmental one. The Panel members were: D. K. Barton, D. R. Brillinger, A. H. El-Sharaawi, P. McDaniel, A. H. Pollock and M. T. Tuley, a mixture of engineers and statisticians.

In particular NASA wished knowledge of the space debris flux characteristics. Here flux is the rate at which debris pieces of a given size pass through a particular cell in the sky, at a given elevation, per unit time. Flux may be viewed as the rate function of the point process of passage times and this aids its study.

NASA assigned the Panel four specific issues to consider. These included:

Issue 1. The number of observations relative to the estimated population of interest.

In other words, whether or not enough observations were taken to adequately characterize the population of interest.

Issue 4. The adequacy of the sample data set referred to characterize the debris population's potential geometry.

A data set based on a satellite mockup was employed to develop a calibration relationship between the strength of a radar reflection and the actual size of an object.

The Panel took a statistical approach to addressing these issues employing data collected by Haystack telescope designed for radio astronomy. During selected time periods Haystack was converted into a radar transmitter-receiver. It was focused on a particular cells in the sky and the times and sizes of passing debris pieces were noted. Haystack thus sampled the population of orbiting objects. The data obtained allowed both flux and risk estimates to be computed.

Figure 1 provides an example of the estimated flux for different (scaled) sizes, at a height of 900 km, in units of objects' count per square meter-year. There were 512 objects in the particular sample employed in preparing Figure 1. In the figure one sees a steady decrease in the flux rate with an increase in object size.

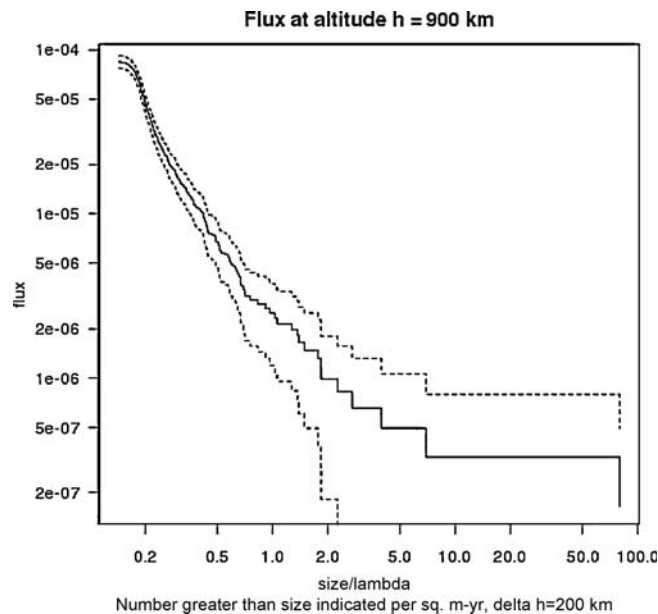


Figure 1. Flux estimate for height 900 km in units of count/m<sup>2</sup>-year vs. object 'size'. The cell reached from 800 to 1000 km in elevation. Also included are approximate  $\pm 2$  s.e. bounds

The estimate of the flux at elevation  $h$  and object size  $s$  was taken to be

$$\hat{p} = n(h, s, T) / [T\Delta A]$$

with  $T$  the length of observation time,  $\Delta A$  the cross-section area of the radar beam and  $n(h, s, T)$  the count of observed objects. The units of the estimate are then counts/m<sup>2</sup>-year. If  $p$  denotes the theoretical flux and the passage process is Poisson then the standard deviation of the estimate is  $\sqrt{\{pT\Delta A\}}$ .

The Panel presented a form of display intended to assist the NASA scientists in the design of future observational studies. It was a plot of the coefficient of variation (CV) as a function of the length of observation period. Assuming an observation time of length  $t$  and a Poisson distribution for the count the CV is  $1/\sqrt{\{pt\Delta A\}}$ .

In preparing the plot an indication of over-dispersion with respect to the Poisson was noted and an extra-Poisson multiplier  $\phi$  was included. The estimated CV was then taken to be

$$\hat{\phi} / \sqrt{\hat{p}t\Delta A}$$

The flux estimate  $\hat{p}$  was computed for four ranges of object sizes. The result is provided in Figure 2 in a log-log plot.

The figure is based on data collected in 1994. There were 840 detections over a 97.22 h observation period

Further details of this work may be found in Barton *et al.* (1998) and in Brillinger (2006).

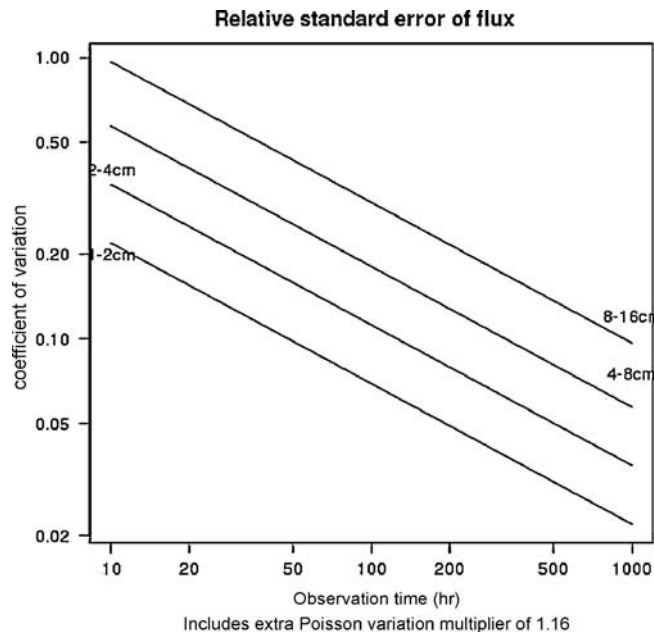


Figure 2. Typical plots of the coefficient of variation of the flux vs. the total hours of observation for objects of different sizes. The graph is a log-log plot.

### 3.2. Seismic risk

Seismic risk assessment may be defined as the process of estimating the probability that certain performance variates at site of interest exceed relevant critical levels within a specified time period, as result of nearby seismic events.

The performance variates referred to could be displacement, acceleration, mercalli intensities, damage, and norms of vector-valued quantities.

Because the negative consequences of earthquakes can be so great it is important to have effective means of communicating the results of seismic risk assessments.

The more general term earthquake potential can be expressed either verbally or numerically. It can be separated into three cases: long-term, intermediate-term or short-term. Speaking generally long-term is the concern of city and regional planners, intermediate the concern of insurers, while short-term is the concern of emergency officials. There are both theoretical and political problems associated with each of these. One reference is Jones (1996).

An example of a long to intermediate study of great earthquakes on the San Andreas Fault is provided in Sieh *et al.* (1989). It made use of data through 1988. The data involved were collected by Kerry Sieh as part of his doctoral thesis at a location called Pallett Creek near Los Angeles. Water had been flowing at that location for many years. The water laid down sedimentation layers. Large earthquakes fractured and realigned these over the years. The fractures could be inferred when a trench was dug across the fault. In many cases dead organic material could be found near a fracture. This was then dated by radiocarbon techniques and one had a sequence of estimated large earthquake dates, and their associated uncertainties.

The intervals between successive events of the point process could be computed and, assuming a chance model, their distribution estimated. Cumulative hazard plotting proved a convenient method for assessing the appropriateness of a distribution, see Brillinger (1989). It led to a Weibull distribution. Figure 3 shows the result. The points are the computed intervals between successive events. The vertical lines about them are their estimated radiocarbon dating errors obtained at M. Stuiver's laboratory. The line sloping up to the right is the result of maximum likelihood fitting. Censoring, a missing value, and measurement error all had to be dealt with in setting down the likelihood function.

The parameter estimates may be employed to estimate probabilities of future events. In the data studied these went back from the well-identified 1857 event. The probabilities provided in Figure 4 have 1988 as their base as this is the year when the analysis presented was carried out.

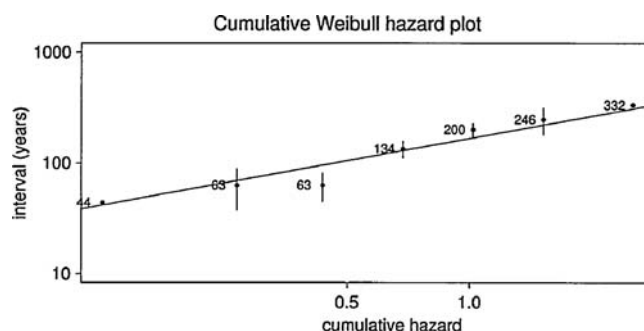


Figure 3. A cumulative hazard plot, see Brillinger (1989), to assess the reasonableness of the Weibull distribution for the intervals between earthquakes at Pallett Creek. The vertical bars indicate approximate plus and minus two standard errors of the estimated intervals

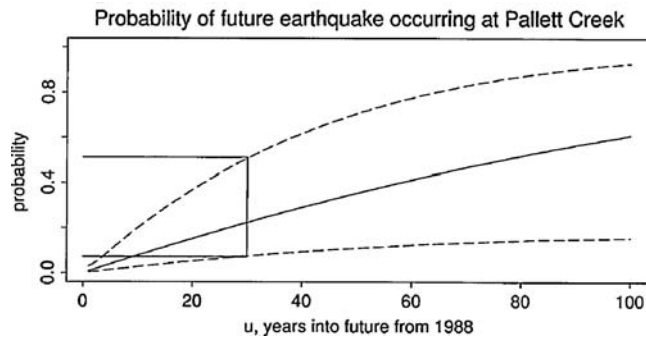


Figure 4. An estimate of the future probability of an earthquake occurring at Pallett Creek within,  $u$ , the indicated number of years. The dashed lines give the upper and lower values of a corresponding approximate 95% confidence interval. The horizontal solid lines provide these values for an earthquake within  $u = 30$  years

The probability estimates employed were plug in estimates. The uncertainty limits in Figure 4 include the uncertainty caused by estimating the parameters of the Weibull. One sees the intervals widening as the time limit increases. The probability result obtained in 1988 was about 30%. In 2008 the USGS estimate had grown to 46%, USGS Newsroom (2008), but no indication of the estimates uncertainty was provided. In their earlier warnings, e.g. USGS Newsroom (1999) there was confusion of odds and probabilities. Such confusion seems common in the communication of scientific results to the public.

The results of Sieh *et al.* (1989) were communicated by professional talks, by interviews and by articles. Brillinger (1989) was a publication for a general audience. The Sieh *et al.* paper has been often referenced, for example National Academy of Science (2003), Akciz *et al.* (2009).

### 3.3. Fires at the wildland-urban interface

The wildland-urban interface may be defined as the place where humans and their development interface with wildland fuel. Fires occurring there can be both deadly and expensive.

Some risk analyses have been done concerning the Cedar Fire of 2003, e.g. ISO Properties (2004) and Kim *et al.* (2006). This fire occurred in Southern California near San Diego and was the largest such fire in California history. The losses included costs of \$12 billion, 2232 homes destroyed, 14 deaths and 280 000 acres burnt over.

A large data set was created by the San Diego community to be available for analysis, see SANGIS (2006). It includes the fire perimeter, house damage, tax assessor records and house locations for San Diego County. Information concerning houses in the City of San Diego and a vegetation map came from scattered sources, see Brillinger *et al.* (2009).

The data studied can be viewed as a segment of a spatial marked point process  $\{(x_j, y_j), S_j, V_j\}$ . In the case at hand  $(x_j, y_j)$  denotes the location of the  $j$ th house,  $S_j$  denotes its square footage and  $V_j$  the dominant vegetation type in its pixel. For risk analysis purposes at the time of the fires a dollar loss to a house destroyed was taken to be its square footage multiplied by \$150.

One model to consider is

$$\text{logit}[\text{Prob}\{\text{house destroyed}|\text{located at } (x, y), v(x, y)\}] = v(x, y) + \beta(x, y) \quad (1)$$

with  $\nu$  a factor with levels corresponding to the available vegetation types and  $\beta$  a smooth function of location,  $(x, y)$ . In the computations  $\beta$  is modelled as a thinplate spline

$$\beta(x, y) = \sum \gamma_j r_j^2 \log r_j$$

where

$$r_j^2 = (x - x_j)^2 + (y - y_j)^2$$

The model was fit employing the function `glm()` of the statistical package R. This function produces estimates of the effects of the factor  $\nu$ , the  $\gamma_j$  and other quantities useful for assessing model fit.

Figure 5 provides an estimate of the boundary of the fire, and the estimated surface  $\hat{\beta}(x, y)$ . The houses with the highest estimated probability of being destroyed are those in the green regions, levels 3, 4.

When estimates of the probability of destruction,  $\hat{\pi}_j$ , are available, as a function of location, various other quantities of interest may be computed. For example the expected square feet to be lost in a similar fire could be estimated by

$$\sum_j \hat{\pi}_j S_j \quad (2)$$

The parameter  $j$  runs over all the houses within the fire boundary. The estimate obtained for the Cedar Fire region is 3 247 666 sq ft or in terms of \$ about 487 million dollars using the \$150 multiplier.

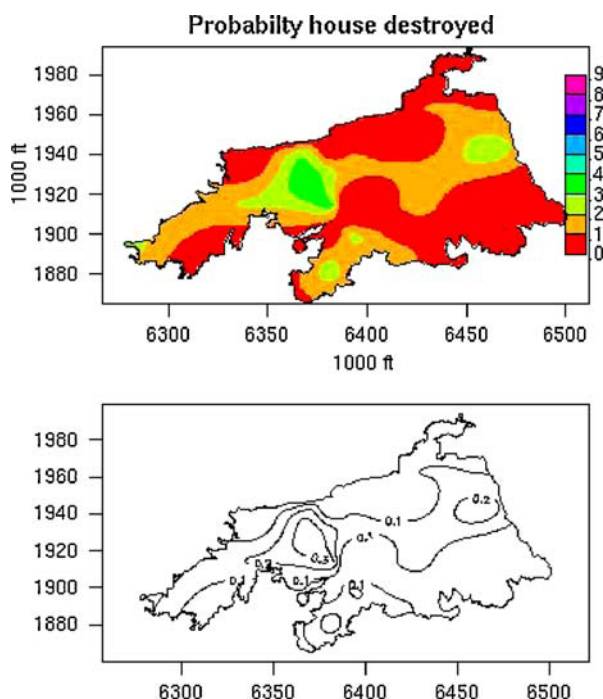


Figure 5. Image and contour plots of the model (1) estimated surface  $\hat{\beta}(x, y)$ . This figure is available in colour online at [www.interscience.wiley.com/journal/env](http://www.interscience.wiley.com/journal/env)

Table 1. Percentiles and standard errors of square footage obtained in 10 000 simulations of the values (2)

Percent	Percentile	s.e.
90	99 620	1264
95	127 095	1205
99	176 794	2738

Insurance companies are interested in the distribution of the difference between a possibly realized total loss and an estimated average total loss. In the present case it was convenient to obtain characteristics of this distribution by simulation. In ruin analysis companies are interested in upper percentiles of what the level of the difference between realized and expected lost square feet might be. Such numbers can help a company to understand the level of reserves that it needs to maintain. For the  $j$ th house within the fire boundary a 1-0 value was generated with probability  $\hat{\pi}_j$  and the corresponding  $L_j$  values summed for the '1' values realized. (The square footage was available from the Assessor Records.) This was repeated 10 000 times.

The estimated 90, 95 and 99th percentiles of the square footage, and their standard errors, are given in Table 1.

The units here are sq ft. If wished the numbers may be multiplied by \$150 to obtain an estimated loss in dollar terms.

When communicating risks the provision of higher percentiles of the distribution of losses both clarifies risk and formalizes the uncertainty.

#### 4. DISCUSSION AND SUMMARY

The paper has presented three examples. The examples each illustrate methods of communicating results and providing uncertainty measures. The focus was on visual displays, tables, published results and estimates based on stochastic models and random processes.

The first example refers to space debris. There are plots of the flux rate versus object size and of the coefficient of variation of the flux estimate versus the length of observation time. The CV itself is a measure of uncertainty. The first plot includes approximate two standard error limits based on a Poisson assumption, while the computations behind the second figure include a multiplier of 1.16 for extra-Poisson variation. The quantities involved in these figures are intermediate results towards obtaining estimates of the risk of collision of a piece of debris with an orbiting object.

The risk in the seismic example involved an assumption that the times between large earthquakes have the same Weibull distribution. Figure 3, a cumulative hazard plot, is a method to assess the Weibull assumption. The fit does not appear unreasonable. Estimates and their uncertainties came from plugging parameter estimates into the risk formula and then employing the delta method.

In the final example, the 2003 Cedar Fire was investigated as a prototype of fires at a wildland-urban interface. There were houses scattered within its boundary, some being destroyed and some not. The probability model involved a marked spatial point process. The smooth rate function estimate,  $\hat{\beta}(x, y)$ , could be substituted into formula (1) to obtain risk estimates as desired. Simulation was employed to evaluate uncertainty and upper percentiles of possible losses. These are provided in Table 1.



## ACKNOWLEDGEMENTS

I thank Abdel El-Sharaawi for involving me with the space debris problem and Peter Guttorp for inviting me to speak on the topic of this paper at TIES Kelowna. I thank the referees for their patience. I also acknowledge support of the NSF Grants DMS-0504162, and DMS-0707157.

## REFERENCES

- Akciz SO, Grant LB, Arrowsmith JR. 2009. Revised dates of large earthquakes along the Carrizo section of the San Andreas Fault, California, since A.D. 1310  $\pm$  30. *Journal of Geophysical Research* **114**: B01313. DOI: 10.1029/2007JB005285
- Barton DK, Brillinger DR, El Sharaawi AH, McDaniel P, Pollock KH, Tuley MT. 1998. Final report of the haystack orbital debris data review panel. *Technical Memorandum 4809 NASA*, Johnson Space Center.
- Brillinger DR. 1989. Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. In *Statistics a Guide to the Unknown*, 3rd edn. Wadsworth: Pacific Grove, CA; 249–260
- Brillinger DR. 2006. Space debris: yet another environmental problem. In *Statistics a Guide to the Unknown*, 4th edn. Thomson: Belmont, CA; 183–226
- Brillinger DR, Autrey BS, Cattaneo MD. 2009. Probabilistic risk modelling at the wildland urban interface: the 2003 Cedar Fire. *Environmetrics* **20**: 607–620.
- Brillinger DR, Guttorp PM, Schoenberg FP. 2002. Point process: temporal. *Encyclopedia of Environmetrics* **3**: 1576–1581.
- El-Shaarawi AH, Hunter JS. 2002. Environmetrics, an overview. In *Encyclopedia of Environmetrics*, Vol. 2. Wiley: New York; 698–702.
- Encyclopedia of Environmetrics. 2002. El-Sharaawi AH, Piegorsch WW (eds). Wiley: Chichester.
- Guttorp PM, Brillinger DR, Schoenberg FP. 2002. Point process: spatial. *Encyclopedia of Environmetrics* **3**: 1571–1577.
- ISO Properties. 2004. Pinpointing insured losses in the Southern California wildfires of 2003. Available at: [www.isoproperty-resources.com/dloads/CA\\_Fire.pdf](http://www.isoproperty-resources.com/dloads/CA_Fire.pdf)
- Jeffreys H. 1967. Statistical methods in seismology. In *International Dictionary of Geophysics*, K Runcorn (ed.), Pergamon: London; 1398–1401.
- Jones LM. 1996. Earthquake prediction: the interaction of public policy and science. *Proceedings of the National Academy of Sciences*, 3721–3725.
- Kim TH, Cova TJ, Brunelle A. 2006. Exploratory map animation for post-event analysis of wildfire protective action recommendations. *Natural Hazards Review*, 1–11.
- SANGIS. 2006. *San Diego Firestorm 2003*. Compact disc. San Diego, CA. [www.sangis.org](http://www.sangis.org)
- Schoenberg FP, Brillinger DR, Guttorp PM. 2002. Point process: spatial-temporal. *Encyclopedia of Environmetrics* **3**: 1573–1577.
- Sieh K, Stuiver M, Brillinger DR. 1989. A more precise chronology of earthquakes produced by the San Andreas Fault in Southern California. *Journal of Geophysical Research* **94**: 603–623.
- Smith E. 2002. Uncertainty analysis. In *Encyclopedia of Environmetrics*, Vol. 4. Wiley: New York; 2283–2296.
- USGS Newsroom. 1999. 70 Percent odds for large earthquake. Available at: [www.usgs.gov/newsroom/article.asp?ID=121p](http://www.usgs.gov/newsroom/article.asp?ID=121p)
- USGS Newsroom. 2008. New study shows odds high for big California quakes. Available at: [www.usgs.gov/newsroom/article.asp?ID=1914](http://www.usgs.gov/newsroom/article.asp?ID=1914)