The Beginnings of Some Exploratory Environmental Risk Analyses

David R. Brillinger
University of California, Department of Statistics
367 Evans
Berkeley, California, 94720-3860, USA
brill@stat.berkeley.edu

Introduction

These days there are many demands for risk analyses that is: estimating the probabilities of unpleasant, high implication events. These analyses are useful for: risk management, insurance premium computations, formulating regulations and laws, design of structures, and advancing science and technology generally. Exploratory data analysis (EDA) is a mainstay. This paper presents beginnings of some exploratory environmental risk analyses.

Bayes Theorem and box and arrow diagrams are basic to risk analyses. They decompose the problem. EDA helps discover expressions for what goes on in the boxes.

In this brief paper we chose to focus on parallel boxplots as an effective tool for starting risk studies. They are especially useful for examining batches of data.

Boxplots are displayed in the lefthand panel of Figure 1. They are now appearing throughout the literature of science. Using a boxplot one can quickly pick out: location, spread, skewness, tail length, outliers in a batch of data. Parallel boxplots refers to graphing boxplots against a treatment variable, preparing a boxplot for each batch of items with the same treatment. See Emerson and Strenio (1983) for details concerning boxplots.

Four Examples

1. **Flooding of the Amazon**

The first example concerns the risk of floods on the Amazon River at the city of Manaus. Data for the years 1892-2002 are employed. Manaus is a city well up the Amazon River in Central Brazil. The data are actually for

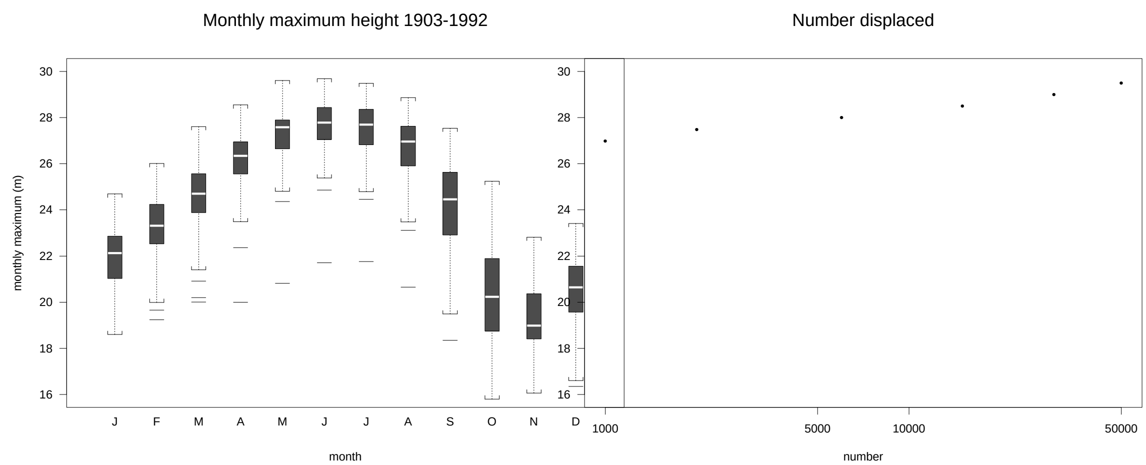

Figure 1

the Rio Negro, but that river becomes the Amazon not far downstream from Manaus. The Rio Negro's height has been recorded daily since 1903. A concern is whether the risk of flooding is increasing, see Sternberg (1987).
The lefthand panel of Figure 1 provides parallel boxplots with the x-variable month. One sees a pronounced annual effect, flooding occurs in June through August. One notes that the outliers are on the low, rather than high, side.
A so-called damageability matrix is given in La Rovere and Crespo (2002). It provides estimates of the numbers of persons affected by a flood as a function of the maximum height flood waters reach. Figure 1 provides this information. The righthand panel in Figure 1 has been set up parallel to the lefthand one so that one can read off the number of people possibly displaced as the peak gets higher.
More detail may be found in Brillinger (2003).

2. **Wildfires in Oregon**
Data are available for the sizes of fires taking place on the Federal Lands in the State of Oregon during the years 1989 to 1996, see Brillinger et al. (2003). The concern in that, and later, papers is to estimate the risk of a fire as a function of explanatories such as: location, day of the year and various "fire indices". Parallel boxplots have been prepared for the fire months. The dependent variable is the square root of the acres burnt, for fires burning

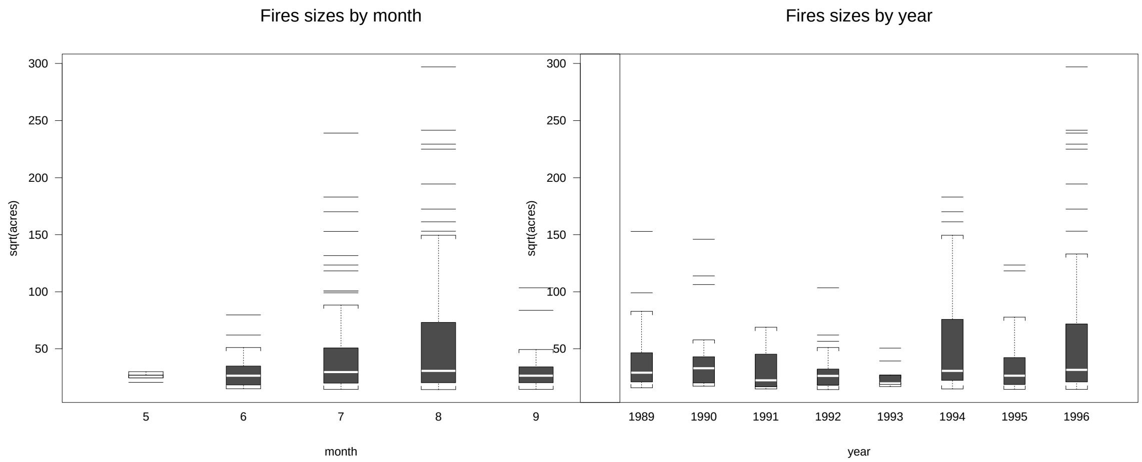

Figure 2

over 200 acres. The concern is large fires. The boxplots are provided in the lefthand panel of Figure 2. One sees a definite seasonal effect peaking in August. There are also pronounced outliers. The righthand panel has year as the x-variate. Some unusual years stand out. This display is useful for developing time series models.
More detail may be found in the above reference.

3. **Space debris risk**

Among the goals of space debris research is estimating the risk of a piece of space debris hitting things like the International Space Station, satellites, or the space shuttle. The data of this example were provided by NASA. They consist of the radar cross sections (RCS) and the elevations of pieces observed crossing through an observation cell of the Haystack telescope during several time periods.
Parallel boxplots were prepared and are in the lefthand panel of Figure 3. The x-axis batches correspond to elevation intervals 300 to 400km, 400 to 500km and so on. There are varying numbers of debris pieces in these intervals. One sees something of an increase in size of the objects as the elevation rises. There are also outliers on the larger side.
For interests sake we also present a plot that NASA has found very useful for modelling breakup events and their long-term environmental effects. It is the

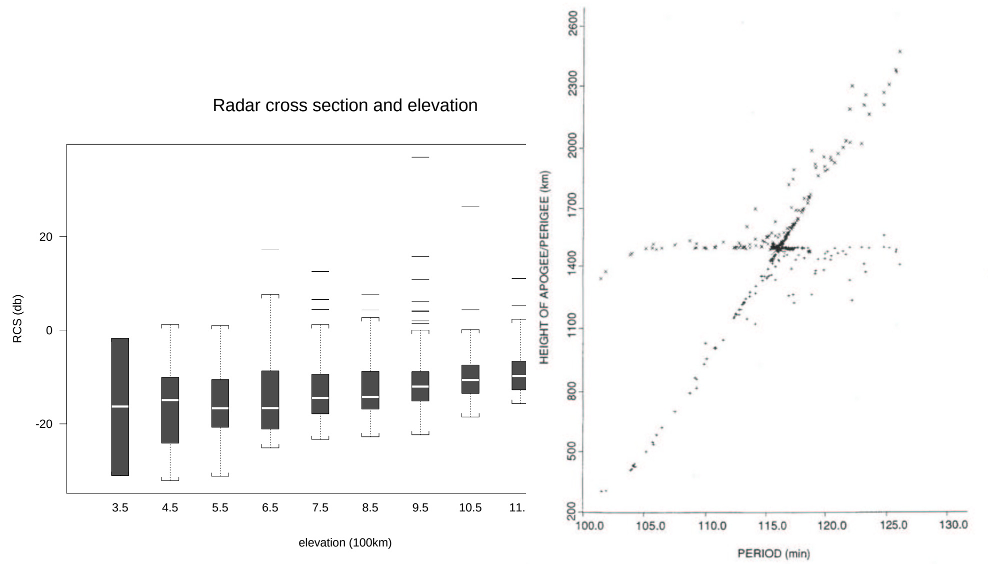

Figure 3

so-called Gabbard diagram. This display plots perigee and apogee altitudes for pieces produced in on-orbit breakups as a function of orbital period. The figures is taken from Portree and Loftus (1999).
More detail on such data and the space debris risk problem may be found in Barton et al. (1998).

4. **The Northridge earthquake**

The Northridge event occurred 17 January 1994. It was situated about 30 km NW of Los Angeles, California. Its size, as measured by magnitude, was 6.7. There were 57 deaths, 1500 serious injuries, and 12,500 structures moderately to severely damaged. The damage cost was estimated as 12.5 $US billion. There are 554 observations to be used in the analysis.
There are a variety of ways to describe and estimate earthquake damage. An old and elementary one, yet an important one, uses the Modified Mercalli Intensity (MMI). One reason for this scale's importance is that sometimes values may be derived from historic accounts. A second is that it refers to damage directly. MMI values are given by roman numerals I to XII (and sometimes 0 referring to nothing felt or noticed.) The scale is ordinal

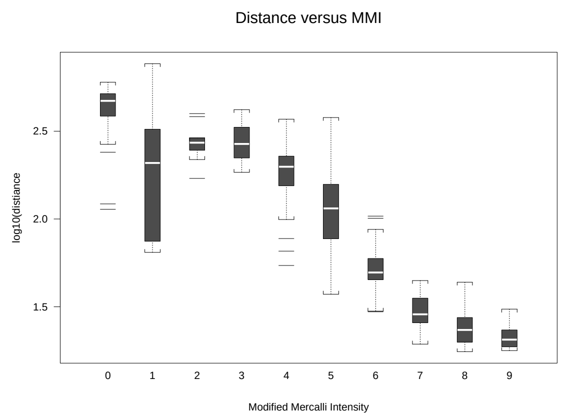

Figure 4

increasing with growing severity of damage. For example the definition of MMI $VIII$ includes: “Damage slight in specially designed structures; considerable in ordinary substantial buildings; ...” while that of MMI IV includes “Dishes, windows, doors disturbed; walls make creaking sound; ...”, Bullen and Bolt (1985). The highest intensity recorded was IX.
Figure 4 graphs log10 of distance, from the epicenter of the event to the location of the damage, against MMI. One sees that the distance is less the greater the MMI.
More details of such analyses may be found in Brillinger (2003).

Discussion

In parallel boxplots the batches can contain varying numbers of observations without affecting the interpretability. This is in contrast to the case of scatter plots where regions with many points catch the eye easily.
We might have used notched boxplots to indicate uncertainty, see Emerson and Strenio (1983), but the emphasis of this paper is non probability methods.

Acknowledgements

The Manaus data came from H. O’R. Sternberg, of the University of California, Berkeley. The Oregon wildfire data were provided by H. K. Preisler,

of the Pacific Southwest Research Station, U. S. Foresty Service. The space debris data were provided by M. Matney, Johnson Space Center, Houston. The Northridge intensity data came from J. Dewey of the United States Geological Survey.
The work was supported by the NSF Grant DMS 02-03921 and US Forest Service Contract 02-JV-11272165-020.

REFERENCES

Barton, D. K. et al. (1998) Final Report of the Haystack Orbital Debris Data Review Panel. Technical Memorandum 4809 NASA, Johnson Space Center.

Brillinger, D. R. (2003). Three environmental probabilistic risk problems, Statistical Science, 412-421

Brillinger, D.R., Preisler, H. K. and Benoit, J. W. (2003) Risk assessment: a forest fire example. Pp. 177-196 in Science and Statistics. Lecture Notes in Statistics 40, Institute of Mathematical Statistics.

Bullen, K. E. and Bolt, B. A. (1985). An Introduction to the Theory of Seismology, Fourth Edition. Cambridge, Cambridge.

La Rovere, A. L. N. and Crespo, S. (2002). Relatório Urbano Ambiental Integrado - Informe GEO - Manaus. IBAM/ ISER/REDEH, Brazil.

Emerson, J. D. and Strenio, J. (1983). Boxplots and batch comparison. Pp. 58-96 in Understanding Robust and Exploratory Data Analysis. (Eds. D. C. Hoaglin, F. Mosteller and J. W. Tukey). Wiley, New York.

Portree, D. S. F. and Loftus, J. P. (1999). Orbital Debris: a Chronology. NASA Report NASA/TP-1999-208856. Johnson Space Center, Houston.

Sternberg, H. O'R. (1987). Aggrevation of floods in the Amazon River as a consequence of deforestration? Geografiska Annaler 69A, 201-219.

RÉSUMÉ

L'analyse des données est très important quand on étude les risques de l'environnement. Cet article présent quatre exemples. Chacum des exemples emploie les boxplots.