Comment: Are Ozone Exceedance Rates Decreasing?

Adrian E. Raftery

In this excellent paper, Professor Smith has synthesized a range of powerful methods for the analysis of extreme values. The point process of cluster peaks over a high threshold provides a remarkable condensation of the massive data set that he analyzes. It involves little loss of relevant information and permits fairly simple analyses. The methodology is sure to find wide application.

Nevertheless, I find it hard to think of physical explanations for the conclusion that there has been a downward trend in the extreme values without any accompanying decrease in the overall levels of the ozone series. Here I try to reassess the evidence in terms of a comparison between competing models for the intensity of a Poisson process. The analysis suggests that there is some evidence for a decreasing trend in exceedance rates but that it is rather weak. If there is a trend, it seems more likely to consist of a fairly abrupt change than a gradual decrease. The possibility that such a change is due to an improvement in measurement technology is discussed. I also consider the possibility of long-memory dependence and discuss the clustering method used.

1. ARE OZONE EXCEEDANCE RATES DECREASING?

The evidence in the paper for decreasing exceedance rates consists mainly of the fact that the estimated trend was downward in all the models that incorporated a trend. However, these models did not appear to fit better than models that did not incorporate a trend. For example, the likelihood ratio test statistic for splitting the data was 16.6 with 18 degrees of freedom.

This may be due more to the large number of degrees of freedom than to the absence of an effect. It might be worth, for example, fitting a model of the form (4.1), but with $\mu_{ij} = \alpha_j + \beta\delta_i$, where $\delta_i = 0$ for 1973–80 and $\delta_i = 1$ for 1981–86. One could then test the hypothesis that $\beta = 0$, which involves only one degree of freedom rather than 18. There are many other parsimonious possibilities.

The computer programs used to carry out the analyses may be obtained from the author by sending electronic mail to raftery@entropy.ms.washington.edu.

Adrian E. Raftery is Associate Professor, Department of Statistics, GN-22, University of Washington, Seattle, Washington 98195.

Nonhomogeneous Poisson Process Models for Exceedances

Professor Smith's conclusion corresponds to a decreasing rate of occurrence in the point processes of exceedances above high thresholds. This process was not fully observed, and the proportion of time monitored varied over the period, increasing gradually but significantly. I therefore expressed times of occurrence in terms of monitored time since the start of the data, rather than calendar time. Also, ozone levels are highly seasonal. I estimated the seasonal effect as piecewise constant within each of the six 61-day periods and deseasonalized the data by transforming the time axis (Cox and Lewis, 1966). The resulting series of events are shown in Tables 1 and 2. I denote by T the period of observation and by $t = (t_1, \cdots, t_n)$ the event times.

If there is no trend, the data in Tables 1 and 2 are very nearly from a homogeneous Poisson process; we denote this model by M_0. This assumes that any short-term correlation has been removed by considering only cluster peaks. An alternative hypothesis is that the exceedance rate has been decreasing smoothly and gradually. This may conveniently be represented by the log-linear Poisson process, M_1: $\lambda(s) = \rho e^{-\beta s}$, where $\lambda(s)$ is the rate of occurrence at time s. Another possibility, suggested by the splitting of the data in the paper, is that the exceedance rate decreased fairly abruptly within a short time period. This may be represented by the change-point Poisson process, M_2: $\lambda(s) = \lambda_1$ if $0 \le s \le \tau$ and $\lambda(s) = \lambda_2$ if $\tau < s \le T$.

Model Comparison

The three competing models, M_0, M_1 and M_2, may be compared using the Bayes factor, or ratio of posterior to prior odds for M_i against M_j, B_{ij}, for each pairwise comparison. It has been argued that Bayes factors are better measures of evidence than P values (Berger and Sellke, 1987), and they are also more readily applicable to the comparison of nonnested models.

TABLE 1
Times of occurrence of exceedances above threshold level 16

62	511	913	1442	1760	2168	2890	3550
71	539	934	1484	1774	2204	2976	3565
88	555	1006	1506	1830	2245	3051	3636
114	585	1022	1527	1893	2363	3079	3685
122	596	1043	1554	1924	2417	3153	3827
146	610	1057	1585	1986	2465	3257	3846
184	627	1138	1617	2007	2535	3302	3891
219	662	1184	1621	2045	2568	3325	
235	740	1231	1624	2076	2608	3346	
285	779	1318	1642	2099	2705	3403	
327	847	1409	1742	2122	2791	3421	

In monitored days from the start of the data, deseasonalized (read down the columns).

TABLE 2
Times of occurrence of exceedances above threshold level 20

71	327	1057	1624	1986	2122	2705	3079
88	555	1442	1642	2045	2417	2791	3550
122	847	1484	1830	2076	2465	2890	3846
285	913	1585	1893	2099	2608	2976	

In monitored days from the start of the data, deseasonalized (read down the columns).

With vague prior information, the Bayes factor for M_0 against the log-linear Poisson process, M_1, is

$$(1)\quad B_{01} = 0.645(n-1) \bigg/ \int_0^\infty e^{-Ry}\left[\frac{y}{1-e^{-y}}\right]^{n-1} dy,$$

where $R = \sum t_i/T$ (Akman and Raftery, 1986a). The Bayes factor against the change-point Poisson process, M_2, is

$$(2)\quad B_{02} = \frac{4\sqrt{\pi}\,\Gamma(n+½)}{\sum_{i=0}^{n} \Gamma(i+½)\Gamma(n-i+½) I_i}$$

(Raftery and Akman, 1986). In (2),

$$I_i = \int_{u_i}^{u_{i+1}} x^{-(i+1/2)}(1-x)^{-(n-i+½)}\, dx,$$

where $u_i = t_i/T$.

A classical test statistic for M_0 against M_1 is $U = (R - ½n)/\sqrt{n/12}$, which has approximately a standard normal distribution under M_0 (Cox and Lewis, 1966). A test of M_0 against M_2 may be based on the quantity

$$(3)\quad \Delta = n^{-1/2} \max_{.01 \le u_i \le .99} \{|g(i-1, u_i)|, |g(i, u_i)|\},$$

where

$$g(i, u) = i\sqrt{\frac{1-u}{u}} - (n-i)\sqrt{\frac{u}{1-u}}.$$

The approximate 5% critical value for this test is 3.29 (Akman and Raftery, 1986b).

TABLE 3
Model comparison results

	Threshold level	
	16	20
$\log_{10}B_{01}$	1.69	1.11
U	−1.02	−0.97
$\log_{10}B_{02}$	−0.44	−0.30
Δ	2.33	2.15

B_{01} is the Bayes factor for the homogeneous Poisson process against the log-linear Poisson process given by (1); U is the corresponding classical test statistic. B_{02} is the Bayes factor for the homogeneous Poisson process against the change-point Poisson process given by (2); Δ is a corresponding classical test statistic given by (3).

In Table 3 there is evidence against a gradual decrease in exceedance rate of the form specified by M_1. The posterior odds for the change-point Poisson process are 2.7:1 and 2:1 at threshold levels of 16 and 20, respectively. In the words of Jeffreys (1961), this constitutes evidence against the homogeneous Poisson process, but it is not worth more than a bare mention. In addition, the result of the classical test based on (3) is not significant.

Checking the Homogeneous Poisson Process

Since the evidence for a trend appears weak, it seems worth checking the homogeneous Poisson process model itself. One way of doing this is to compare the observed evolution with those of several data sets simulated from the model. Since the homogeneous Poisson process is time-reversible, we can do the same with the time-reversed data set. Ripley (1977), who pioneered this approach, used point estimates of the model parameters in the simulations, but this may lead to simulated bands which are too narrow. Here, uncertainty about the parameter λ of the homogeneous Poisson process is incorporated as follows (Rubin, 1984; Raftery, 1988). First, generate a value of λ from its posterior distribution, taken here to be Gamma $(n + ½, T)$, and then proceed as before.

The result is shown in Figure 1. The data do not go outside the simulated bands except very briefly in Figures 1(b) and 1(d). Once again, the evidence against the homogeneous Poisson process does not seem strong. The simulated bands may still be too narrow because they do not take account of uncertainty about the estimated seasonal effect, and so the evidence may be even weaker than it appears.

Analysis of the Change-Point Poisson Process Model

The analysis so far suggests that, if there is a trend, it is better represented by a change-point than by a gradual decrease. The posterior distribution of the

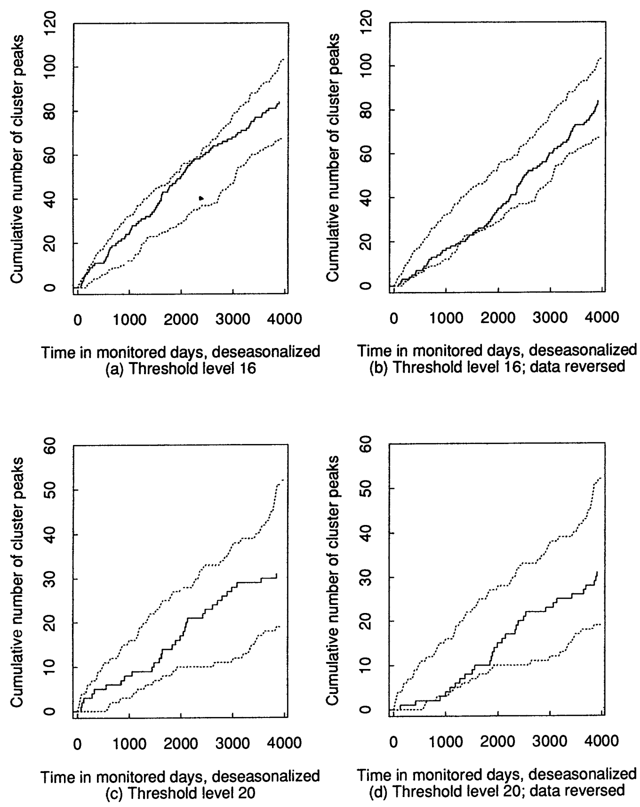

FIG. 1. *Diagnostic checking for the homogeneous Poisson process. In each graph, the solid line represents the data and the dotted lines are the outer envelopes of* 19 *simulations from the model, as described in the text.*

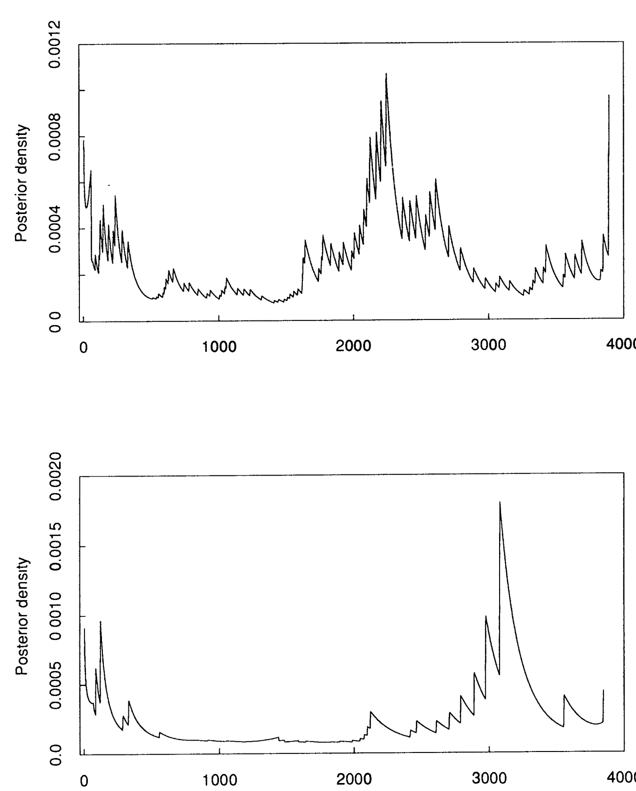

FIG. 2. *Posterior distribution of the change-point in the change-point Poisson process model. Time in monitored days, deseasonalized.* (a) *Threshold level* 16; (b) *Threshold level* 20.

change-point is given by equation (2.3) of Raftery and Akman (1986) and is shown in Figure 2. At threshold level 20, the posterior distribution is less diffuse than at level 16. The posterior mode is April 11, 1984, which is at the beginning of the 1984 "ozone season." This result reflects the fact that there was only one exceedance above level 20 in each of 1984, 1985 and 1986, compared with 28 such exceedances in the previous 10 seasons. The posterior mode at threshold level 16, which is less marked than at level 20, is June 25, 1981.

The analysis here is tentative in many ways. In particular it seems important to include relevant covariates, especially temperature, as emphasized by Davison and Hemphill (1987). It would be interesting to know if there were any events which could have caused an abrupt change, such as legislation, changes in Federal standards, highway development, or changes in data collection, checking and reporting practices.

My colleague Peter Guttorp has suggested the following as a possible explanation. In the past, hourly ozone measurements were often based on 25-minute averages. Now it is more usual for measurement to be continuous, so that hourly measurements are based on 60-minute averages. Such a change in instrumentation would not have changed the overall level of the series, but it might well have reduced its variability, and hence the exceedance rate, because the measurements are based on more data. This would have been an abrupt change rather than a gradual one, and so seems consistent with the results here and in Professor Smith's paper. It would be interesting to know if there was such a change in instrumentation in Houston during the period covered by the data. If so, it would suggest that we are not seeing an improvement in compliance with the Federal standard, but rather a change in measurement technology.

2. OTHER ISSUES

Long-Memory Dependence

Long-memory dependence is known to be a feature of at least some climatic variables (Haslett and Raftery, 1989, and references therein). Climate influences ozone levels, so it is possible that ozone levels may also exhibit long-memory dependence. This is characterized by small but nonnegligible autocorrelations at long lags, an infinite spike at zero in the spectrum or "cycles of all periods," and high variability of the sample mean and other statistics. It is hard to detect but can dramatically affect statistical analyses.

Is there any evidence of long-memory dependence in the ozone data? How could it be detected? Would it affect the analysis of extreme values? If so, how could it be incorporated into the analysis? I would welcome Professor Smith's views on these questions.

The Clustering Method

The method for forming clusters used in the paper is essentially the single link method. This has the possible disadvantage that two clusters six days apart with a single exceedance between them could be merged (Gordon, 1981). An agglomerative sum of squares method might be preferable given that the aim is to obtain compact clusters. Inspection of the dendogram could help with the choice of a cluster interval.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research under Contract no. N-00014-88-K-0265. The author is grateful to Richard L. Smith for providing the ozone data on which the analyses are based, to Peter Guttorp for very helpful discussions and to Michael A. Newton for useful comments.

ADDITIONAL REFERENCES

AKMAN, V. E. and RAFTERY, A. E. (1986a). Bayes factors for non-homogeneous Poisson processes with vague prior information. *J. Roy. Statist. Soc. Ser. B* **48** 322–329.

AKMAN, V. E. and RAFTERY, A. E. (1986b). Asymptotic inference for a change-point Poisson process. *Ann. Statist.* **14** 1583–1590.

BERGER, J. O. and SELLKE, T. (1987). Testing a point null hypothesis: The irreconcilability of P values and evidence. *J. Amer. Statist. Assoc.* **82** 112–122.

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen, London.

GORDON, A. D. (1981). *Classification*. Chapman and Hall, London.

HASLETT, J. and RAFTERY, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (with discussion). *Appl. Statist.* **38** 1–50.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford.

RAFTERY, A. E. (1988). Analysis of a simple debugging model. *Appl. Statist.* **37** 12–22.

RAFTERY, A. E. and AKMAN, V. E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika* **73** 85–89.

RIPLEY, B. D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 172–212.

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172.

Comment

David Fairley

I read Dr. Smith's analysis with great interest. I was unfamiliar with the point-process approach and was impressed by how elegantly it encompasses the other extreme value theory methods. At the same time, I was impressed by how hard it is to apply this theory to the problem at hand. Beyond his expertise in extreme value theory, Dr. Smith has clearly taken great pains to take the practical issues like dependence and seasonality into account. Nevertheless, the results appear somewhat weak given the power of the theory that went into them.

My work at the San Francisco Bay Area Air Quality Management District (affectionately known as the BAAQMD) has given me an applied orientation to the analysis of ozone trends, so most of my comments are directed more toward specific practical problems than at theoretical ones. However, some of the practical considerations involved in ozone trend detection suggest ways in which the theory might be usefully extended.

In Section 2, I will try to define the underlying problem as clearly as possible. This leads to several possible ways to extend the analysis and make it more powerful (Section 3). Beforehand, though, I will suggest several reasons why the data appear to me even messier than Dr. Smith suggests.

David Fairley is the statistician at the Bay Area Air Quality Management District, 939 Ellis Street, San Francisco, California 94109.

Disclaimer: The opinions expressed here are the author's own and not necessarily those of the Board of Directors or staff of the Bay Area Air Quality Management District.

1. MORE COMPLICATIONS

Dr. Smith dealt with some important practical difficulties, including the short-term dependence, seasonality and missing values. Several other factors complicate the picture.

Measurement error is generally overlooked in ozone trend studies, but it can affect the results substantially. The instrument measuring ozone can only measure to the nearest pphm, so the actual data are discrete. There has been more than one method for measuring ozone. The instruments used at the BAAQMD up through the mid '70's measured all