INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY OF NORTH CAROLINA

ON THE SEASONAL CYCLE OF HURRICANE FREQUENCY

by

Robert Lund

September 1991

Mimeo Series #2064

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

ON THE SEASONAL CYCLE OF HURRICANE FREQUENCY

Robert Lund

Department of Statistics

The University of North Carolina at Chapel Hill

Chapel Hill, NC 27599-3260

August 21, 1991

Abstract

Hurricanes and tropical storms occur in the North Atlantic Ocean and the Gulf of Mexico throughout the year. They are most prevalent in September and almost nonexistent from January through April. The objective of this paper is to quantify their annual arrival cycle in a meaningful mathematical form suitable for engineering computations.

A Poisson process with a time-varying intensity function is used to model the intraseasonal variations in storm arrivals. A concrete definition of the model is presented, interpreted, and consequences of the model are examined. The kernel method is introduced as an effective method of estimating the time-varying intensity function. This model is applied in the analysis of a data set containing all recorded hurricanes and tropical storms occurring in the North Atlantic Ocean and the Gulf of Mexico during the years 1871-1990 inclusive.

The methods yield a graph of the estimated intensity function. From this graph, we are able to infer: (1) the peak of hurricane season is around September 12; (2) an early season flare up of activity exists around June 19; and (3) a possible "Indian summer" flare up of activity exists around October 12. More importantly, the methods yield an estimate of the distribution for the number of storms occurring over any time period.

1. Introduction

The official hurricane season in the North Atlantic Ocean and the Gulf of Mexico runs from June 1 through November 30. Unfortunately, hurricanes and tropical storms frequently occur outside of the official season. Tropical events have been recorded in every month of the year except April since records were kept in 1871 (Neumann et al. 1981). However, the frequency at which these tropical events occur varies greatly over the course of the year: tropical events occur much more frequently during September than they do in June.

The objective of this paper is to quantify these intraseasonal variations in a mathematical manner. A probabilistic model is developed and applied to a data set that contains all recorded tropical storms and hurricanes occurring in the North Atlantic Ocean and the Gulf of Mexico during the 120 year time period 1871-1990. Inferences about hurricane season are made from the results.

2. Methods

The time when tropical events occur can not be predicted exactly in advance. In order to model this type of phenomenon, a probabilistic model that allows events to occur randomly on the time axis is employed: the Poisson process.

a. The Poisson Process with Constant Intensity

A constant intensity Poisson process with intensity parameter $\lambda > 0$ is defined as follows. Events occur on the time axis according to the rules: if $(t, t+h)$ is a small time interval, then

(A) $P[$ Exactly one event occurs in $(t,t+h)\ ] = \lambda h + o_1(h)$;

(B) $P[$ No events occur in $(t,t+h)\ ] = 1 - \lambda h + o_2(h)$;

(C) $P[$ More than one event occurs in $(t,t+h)\ ] = o_3(h)$;

(D) The number of events occurring in disjoint sets of time are independent;

where $P[\ \ ]$ denotes the probability of the event in brackets and $o_i(h)$ denotes an order function satisfying $h^{-1}o_i(h) \to 0$ as $h \downarrow 0$ for $i = 1, 2, 3$. The $o_i(h)$ sum to zero for any h and are interpreted as functions that take on small values for h close to zero. It is conventional to drop the subscripts from the order functions; this convention will be followed.

These properties indicate that either no events occur or one event occurs in $(t, t+h)$ with high probability. They also imply that any interval of length h is equally likely to see one event occur, thus the term constant intensity. Notice that the larger λ is, the more likely it is that an event will occur in $(t, t+h)$. The Poisson process is directly related to the Poisson distribution: let $N(a, b)$ denote the total number of events occurring in the interval (a, b). Then

$$P[\ N(a,b) = k\ ] = \frac{e^{-\lambda(b-a)}[\lambda(b-a)]^k}{k!}, \quad k = 0, 1, 2, \ldots \tag{1}$$

Thus, $N(a, b)$ has a Poisson distribution with parameter $\lambda(b-a)$.

b. The Poisson Process with Time-Varying Intensity

The constant intensity Poisson process has many convenient properties; however, it will not model the storm arrival data set accurately. This is because the constant intensity assumption is false; it is much more likely that a storm will occur in September than in January. One easy approach in bypassing this problem is to let the parameter λ depend on time: $\lambda(t)$. A Poisson process with time-varying intensity function $\lambda(t)$ may be defined via:

(A) $P[$ Exactly one event occurs in $(t,t+h)\ ] = \lambda(t)h + o(h)$;

(B) $P[$ No events occur in $(t,t+h)\ ] = 1 - \lambda(t)h + o(h)$;

(C) $P[$ More than one event occurs in $(t, t+h)\ ] = o(h)$;

(D) The number of events occurring in disjoint sets of time are independent;

where $(t, t+h)$ is a small interval of time and $o(h)$ is defined as before. We will make the assumption that $\lambda(t)$ is a continuous, bounded function of t. The above properties provide a convenient interpretation for the intensity function: the probability an event occurs in the small time interval $(t, t+h)$ is approximately $\lambda(t)h$. Again, let $N(a,b)$ denote the number of events occurring in (a,b). As before, $N(a,b)$ follows the Poisson distribution:

$$P[\ N(a,b) = k\ ] = \frac{e^{-\Lambda(a,b)}\ \Lambda(a,b)^k}{k!}, \quad k = 0,\ 1,\ 2,\ \ldots \qquad \text{where } \Lambda(a,b) = \int_a^b \lambda(u)\ du, \qquad (2)$$

and it is termed that $N(t) = N(0,t)$ is a Poisson process with time-varying intensity function $\lambda(t)$.

The Poisson process with time-varying intensity function was suggested as a model for monsoons in the Bay of Bengal by Thompson and Guttorp (1986). Solow (1989a) used a monthly discrete version of the time-varying Poisson process in analyzing storm data from the mid-Atlantic coast for the years 1942-83.

Now suppose that random events occur on the time axis according to a Poisson process with time-varying intensity function $\lambda(t)$, but that some of the events are not counted. Assume that each event is counted with probability p and that the counting decisions are made independently at all points of occurrence. It is easy to show that the number of counted events is also a Poisson process with time-varying intensity function $p\Lambda(t)$. This "thinned" Poisson process will arise later as a model for tropical events that occurred but went unrecorded. Solow (1989b) previously suggested this methodology for missing storm data.

c. Estimation of the Intensity Function

An estimate of $\lambda(t)$ from an observed sequence of event occurrence times must be developed. The method to be employed is called the kernel method and has been used to model lake freeze-over times by Solow (1991). The kernel method traces its roots to the nonparametric estimation of probability density functions (Rosenblatt 1956).

Suppose n tropical events occurred at the ordered times $x_1, x_2, \ldots, x_n$. Let $\hat{\lambda}(t)$ denote the estimate of $\lambda(t)$. Then $\hat{\lambda}(t)$ should be large for values of t where many events occur close by, and small for values of t where few events occur close by. One way of constructing such an estimate is to put a symmetric probability density function, denote it $K(x)$, over each observation. $K(x)$ is called the kernel function and $\hat{\lambda}(t)$ is defined as the sum of each "kernel density" evaluated at t:

$$\hat{\lambda}(t) = \sum_{i=1}^{n} K(t - x_i). \tag{3}$$

Fig. 1 demonstrates the idea. The occurrence time of each event is marked with an X on the time scale. The dashed curve is the sum of all individual kernels and is the intensity estimate. The kernel function chosen for the graph was the standard normal density function

$$K(x) = \frac{\exp[-x^2/2]}{\sqrt{2\pi}}, \quad -\infty < x < \infty. \tag{4}$$

In general, assume that $K(x)$ is a mean zero symmetric probability density function with a finite second moment. For added flexibility, a "spreading factor" h that controls the spread or variance of the kernel function $K(x)$ will be introduced. This is parametrized as follows: for $h > 0$, let $K_h(x) = h^{-1}K(x/h)$. Then $K_h(x)$ is also a mean zero symmetric probability density function. The parameter h is called the bandwidth and has the following interpretation. If h is large, the variance of the kernel function $K_h(x)$ is large; if h is small, $K_h(x)$ has a small variance. Fig. 2 graphs two normal kernel functions, one with $h = 1$, the other with $h = 5$.

The equation for $\hat{\lambda}(t)$ with kernel function $K_h(x)$ is

$$\hat{\lambda}(t) = \sum_{i=1}^{n} K_h(t - x_i) = \frac{1}{h} \sum_{i=1}^{n} K\left(\frac{t - x_i}{h}\right). \tag{5}$$

Statistical properties of $\widehat{\lambda}(t)$ can be found in Brooks (1991), Diggle (1985), and Diggle and Marron (1988). Solow (1991) discusses confidence functions for $\lambda(t)$; the interested reader is referred there for more details.

In applications, the kernel function $K(x)$ and the bandwidth h must be selected. Many methods have been proposed in tackling this problem. Brooks (1991), Diggle and Marron (1988), and Solow (1991) have suggested choosing h by a procedure known as cross-validation; alternate methods are suggested in Silverman (1978). No method proposed to date seems to work well in all cases; however, there is wide agreement that the selection of h is much more critical than the choice of $K(x)$ (Silverman 1986; Solow 1991). Because of its many convenient statistical properties, the standard normal density function will be used for $K(x)$ in all work that follows. This choice gives

$$K_h(x) = \frac{\exp[-x^2/2h^2]}{h\sqrt{2\pi}}, \qquad -\infty < x < \infty. \tag{6}$$

If the selected bandwidth is too small, the intensity estimate will be a jagged curve (undersmoothed); if the selected bandwidth is too large, the intensity estimate may smooth away modes and other true features in the data (oversmoothed). This dilemma, commonly referred to as a smoothing problem, is analogous to selecting the best cell width when constructing a histogram from data. Perhaps the most common method of selecting h is the "eyeball method": plot the intensity estimate for different h and choose an h that gives a reasonable graph - not too smooth and not too jagged. Since bandwidth selection is not a major theme of this paper, the "eyeball method" is used in what follows.

3. Results

a. The Data

The data set to be analyzed consists of 955 recorded tropical storms and hurricanes occurring in the North Atlantic Ocean and the Gulf of Mexico during the years 1871-1990 inclusive. Data for the years 1871-1980 can be found in Neumann et al. (1981). The data for the years 1981-1990 can be found in the annual tropical summaries published in *Monthly Weather Review.* Storms achieving a maximal status of tropical depression have been excluded from the analysis.

There is evidence to suggest that a few tropical storms and hurricanes went undetected during the early years of this data set. A plot of the number of storms occurring in each year over the duration of the data set is provided in Fig. 3. The plot shows that the average number of storms per year is relatively stationary over the separate time periods 1871-1930 and 1931-1990; however, the average number of yearly storms are 6.450 and 9.483 respectively for these two periods. This is a marked increase. Dotted lines are drawn in Fig. 3 to enhance visual clarity of this "changepoint". Neumann et al. (1981) conjectured that this anomaly is due to improvements in sensing techniques, such as the emergence of aircraft, and is not due to any type of natural phenomenon. On the surface, this explanation seems quite plausible.

The data for the period 1871-1930 will be modelled with a thinned Poisson process; consequently, we assume that any tropical event that occurred during the years 1871-1930 was detected and recorded with probability p. Most tropical events occurring during the years 1931-1990 were probably detected and this data is modelled with an unthinned Poisson process. From the averages given in the last paragraph, one can estimate the value of p with $\widehat{p} = 6.450/9.483 \approx .680$.

It is natural to assume that the unthinned intensity function $\lambda(t)$ is periodic with a period of 365 days. Under this assumption, one would like to combine all 120 years of data into one large data set and compute an intensity estimate on the interval [0,365]. Conversion to a yearly intensity estimate is obtained by dividing the combined data set intensity estimate by 120. However, one must be careful to account for the thinned observations during the years 1871-1930. Since the intensity function during a thinned year is $p\lambda(t)$, each observation during a thinned year should be weighted by p^{-1} to cancel this bias. Using $\widehat{p}$ to estimate p and combining this with (5), the yearly intensity estimate takes the form

$$\hat{\lambda}(t) = \frac{1}{Nh} \sum_{i=1}^{n} \gamma_i K\Big(\frac{t - x_i}{h}\Big), \quad 0 \le t \le 365. \tag{7}$$

Here, N denotes the number of years of data (120 for our data) and γ_i is the weighing factor defined by $\gamma_i = \hat{p}^{-1}$ if storm i occurred during the years 1871-1930 and $\gamma_i = 1$ if storm i occurred during 1931-1990. One sees that more complicated thinning assumptions will only change the weighing factors. Solow (1989b) explored a weighing scheme for Australian tropical cyclones that uses the incomplete beta ratio.

The arrival date for each tropical event is assigned to be the average of the event's birthdate and deathdate. For example, a hurricane occurring from September 1 through September 5 yields a point at September 3. Recall that the normal kernel function $K_h(x)$ as defined in (6) is being used.

b. The Intensity Estimates

Fig. 4 shows the yearly intensity estimate of our data set with a bandwidth of $h = 2.5$. This curve is too jagged. Fig. 5 shows the yearly intensity estimate with the bandwidth $h = 15.0$ - this is too smooth. After a few more plots were made with varying bandwidths, the bandwidth $h = 5.0$ was selected as being reasonable. The intensity estimate for this bandwidth is plotted in Fig. 6. The bandwidth selected by cross-validation was around $h = 16.5$ - much too large. For other examples of data sets oversmoothed by the cross-validation bandwidth, see Brooks (1991).

c. Remarks

Fig. 6 shows that the peak of hurricane season is around day 255 (September 12). Fig. 6 also shows a definite early season flare up of activity around day 170 (June 19) followed by a calmer period until about day 195 (July 14). An examination of Figures 4 and 6 indicate a smaller possible increase of activity around day 285 (October 12). Whether or not this "Indian summer effect" is actually there is debatable.

In this data set, only a few storms occurred near the yearly boundaries at day 0 and day 365. Problems can arise when a significant number of events occur close to a boundary. An example of this problem and a solution is presented in Diggle and Marron (1988).

Many times, the researcher is interested in a subset of the data only - e.g., all class 4 and above hurricanes or all Cape Verde tropical events. The preceding analysis could be repeated after all unimportant observations are deleted from the original data set.

4. Consequences of the Model

a. The Number of Storms Occurring in a Time Period

If it is assumed that the arrival times of tropical events are statistically governed by a Poisson process with the time-varying intensity function $\lambda(t)$, then the distribution for the random number of tropical events, denote it by X, occurring in any set $S \subseteq [0,365]$ is explicitly given:

$$P[\, X = k \,] = \frac{e^{-\alpha}\alpha^k}{k!} \quad k = 0, 1, 2, \ldots \qquad \text{where } \alpha = \int_S \lambda(u)\, du. \tag{8}$$

Equation (8) says that X has a Poisson distribution with parameter α. An estimate of α, call it $\hat{\alpha}$, can be obtained by numerically integrating $\hat{\lambda}(t)$ in Fig. 6 over the set S. A simple expression can be derived for $\hat{\alpha}$ without numerical integration when the normal kernel function is used as defined in (6) and the set S is an interval of the form (a,b):

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} \gamma_i \left(\Phi\!\left(\frac{b - x_i}{h}\right) - \Phi\!\left(\frac{a - x_i}{h}\right) \right)}{N}, \qquad \text{where} \qquad \Phi(x) = \int_{-\infty}^{x} \frac{\exp[-u^2/2]}{\sqrt{2\pi}}\, du. \tag{9}$$

Here, $\Phi(x)$ is the cumulative distribution function of the standard normal random variable and is easily evaluated by many computer software packages. The denominator of N in (9) indicates the use of a yearly intensity function.

Table 1 applies this idea by computing $\hat{\alpha}$ for each month of the year. The value $h = 5.0$ was chosen to be compatible with Fig. 6. Also included in Table 1 is the estimated probability that no tropical events occur in each month of the year, $e^{-\hat{\alpha}}$.

By choosing $S = [0,365]$, an estimate for the distribution of the number of tropical events occurring per year is obtained:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} \gamma_i \left(\Phi\left(\frac{365 - x_i}{h}\right) - \Phi\left(\frac{-x_i}{h}\right) \right)}{N} \approx \frac{\sum_{i=1}^{n} \gamma_i}{N}. \tag{10}$$

The approximation in (10) is based on the fact that only a few storms occurred close to the boundaries at day 0 and 365, and on the property that the standard normal random variable contains most of its probability mass close to the origin. For our data set, the exact value of $\hat{\alpha}$ is 9.472; the approximated value of $\hat{\alpha}$ is 9.475. Thom (1960) discusses the Poisson distribution and the negative binomial distribution as models for the number of tropical events occurring in a year.

b. Nonencounter Probabilities

One desirable feature of the Poisson process model is that it can be used to compute nonencounter probabilities. Suppose that tropical events occur on the time axis according to a Poisson process with time-varying intensity function $\lambda(t)$. Further suppose that the maximum windspeed of each tropical event is recorded. Label the windspeed of the ith storm W_i, and assume that the sequence $\{W_i\}$ is independent and identically distributed and that W_i is independent of the number of tropical events occurring. Let the cumulative distribution function of W_i be given by $F_W(w) = P[\ W_i \leq w\ ]$ and define $W_{max}(t)$ to be the maximum windspeed observed up to time t (note that we need not restrict ourselves to $t \leq 365$ here). The distribution of $W_{max}(t)$ is sought and the law of total probability provides

$$P[\ W_{max}(t) \leq w\ ] = \sum_{n=0}^{\infty} P[\ W_{max}(t) \leq w \mid N(t) = n\ ]\ P[\ N(t) = n\ ]. \tag{11}$$

The conditional probability term in the sum is $F_W(w)^n$ by the independence assumptions, and (2) evaluates the term $P[\ N(t) = n\ ]$. Putting these together and using $\Lambda(t)$ to denote $\Lambda(0, t)$ provides

$$P[\ W_{max}(t) \leq w\ ] = \sum_{n=0}^{\infty} F_W(w)^n \frac{e^{-\Lambda(t)}\ \Lambda(t)^n}{n!} = \exp[\,-\Lambda(t)(1 - F_W(w))\,] \tag{12}$$

by the Taylor expansion for e^x. The practical application of (12) requires the estimation of both $\Lambda(t)$ and $F_W(w)$.

One can obtain an estimate of $\Lambda(t)$ by integrating $\hat{\lambda}(u)$ over the interval $(0,t)$:

$$\hat{\Lambda}(t) = \int_0^t \hat{\lambda}(u)\ du. \tag{13}$$

The distribution function $F_W(w)$ is typically estimated with the classical methods of extreme value theory (Isaacson and MacKenzie 1981; Borgman and Resio 1977). An alternative scheme of estimating $F_W(w)$ is the points-over-threshold model discussed in Smith (1989).

5. Conclusions

We have seen that a Poisson process with a time-varying intensity function effectively models the intraseasonal variability in the arrival times of tropical events. One consequence of the Poisson process model is that the distribution of the number of storms occurring in any time period has a Poisson distribution; nonencounter probabilities also take on a very simple form.

The kernel method can be used to obtain an estimate of the time-varying intensity function. From the intensity estimate, one can easily estimate nonencounter probabilities and the Poisson parameter for the distribution of the number of storms occurring over any time period.

Acknowledgements. The author wishes to thank Steve Preissler, Lynne Seymour, Jason Brown, and Mike Sherman for their useful comments.

REFERENCES

Borgman, L. E., and D. T. Resio, 1977: Extremal Prediction in Wave Climatology, *In:* Proceedings *Ports, 1977*, 1, 394-412.

Brooks, M. M., 1991: *Bandwidth Selection Methods for Kernel Estimators of the Intensity Function of a Nonhomogeneous Poisson Process*, Ph.D. Dissertation, The University of North Carolina at Chapel Hill, 119pp.

Clevson, M. L., and J. V. Zidek, 1989: Bayes Linear Estimators of the Intensity Function of the Nonstationary Poisson Process, *J. Amer. Statist. Assoc.,* **72**, 112-120.

Diggle, P., 1985: A Kernel Method for Smoothing Point Process Data, *Appl. Stat.,* **34**, 138-147.

------, and J. S. Marron, 1988: Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation, *J. Amer. Statist. Assoc.,* **83**, 793-800.

Isaacson, M., and N. G. MacKenzie, 1981: Long-Term Distributions of Ocean Waves: A Review, *J. Waterway, Port, Coastal and Ocean Division,* **107, No. WW2**, 93-109.

Neumann, C. J., G. W. Cry, E. L. Caso, and B. R. Jarvinen, 1981: *Tropical Cyclones of the North Atlantic Ocean, 1871-1980,* National Hurricane Center, National Oceanic and Atmospheric Administration, 174pp.

Rosenblatt, M., 1956: Remarks on Some Nonparametric Estimates of a Density Function, *Annals Math. Stat.,* **27**, 832-835.

Silverman, B. W., 1978: Choosing the Window Width When Estimating a Density, *Biometrika,* **65**, 1-11.

——, 1986: *Density Estimation for Statistics and Data Analysis,* Chapman and Hall, New York, 175pp.

Smith, R. L., 1989: Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone, *Statistical Science,* **4**, 367-393.

Solow, A. R., 1989a: Statistical Modeling of Storm Counts, *J. Climate*, **2**, 131-136.

——, 1989b: Reconstructing a Partially Observed Record of Tropical Cyclone Counts, *J. Climate*, **2**, 1253-1257.

——, 1991: The Nonparametric Analysis of Point Process Data: The Freezing History of Lake Konstanz, *J. Climate*, **4**, 116-119.

Thom, H. C. S., 1960: The Distribution of Annual Tropical Cyclone Frequency, *J. Geophys. Res.,* **65**, 213-222.

Thompson, M. L., and P. Guttorp, 1986: A Probability Model for Severe Cyclonic Storms Striking the Coast around the Bay of Bengal, *Mon. Wea. Rev.*, **114**, 2267-2271.

Table 1: Monthly Poisson Parameters

Month	$\hat{\alpha}$	Estimated Monthly Nonencounter Probability
Jan	0.00772	.99231
Feb	0.00675	.99327
Mar	0.01126	.98880
Apr	0.00136	.99865
May	0.11110	.89485
Jun	0.56243	.56982
Jul	0.62360	.53601
Aug	1.95550	.14149
Sep	3.42702	.03248
Oct	2.20679	.11005
Nov	0.48733	.61427
Dec	0.07093	.93153

Fig. 1: The Kernel Method Idea

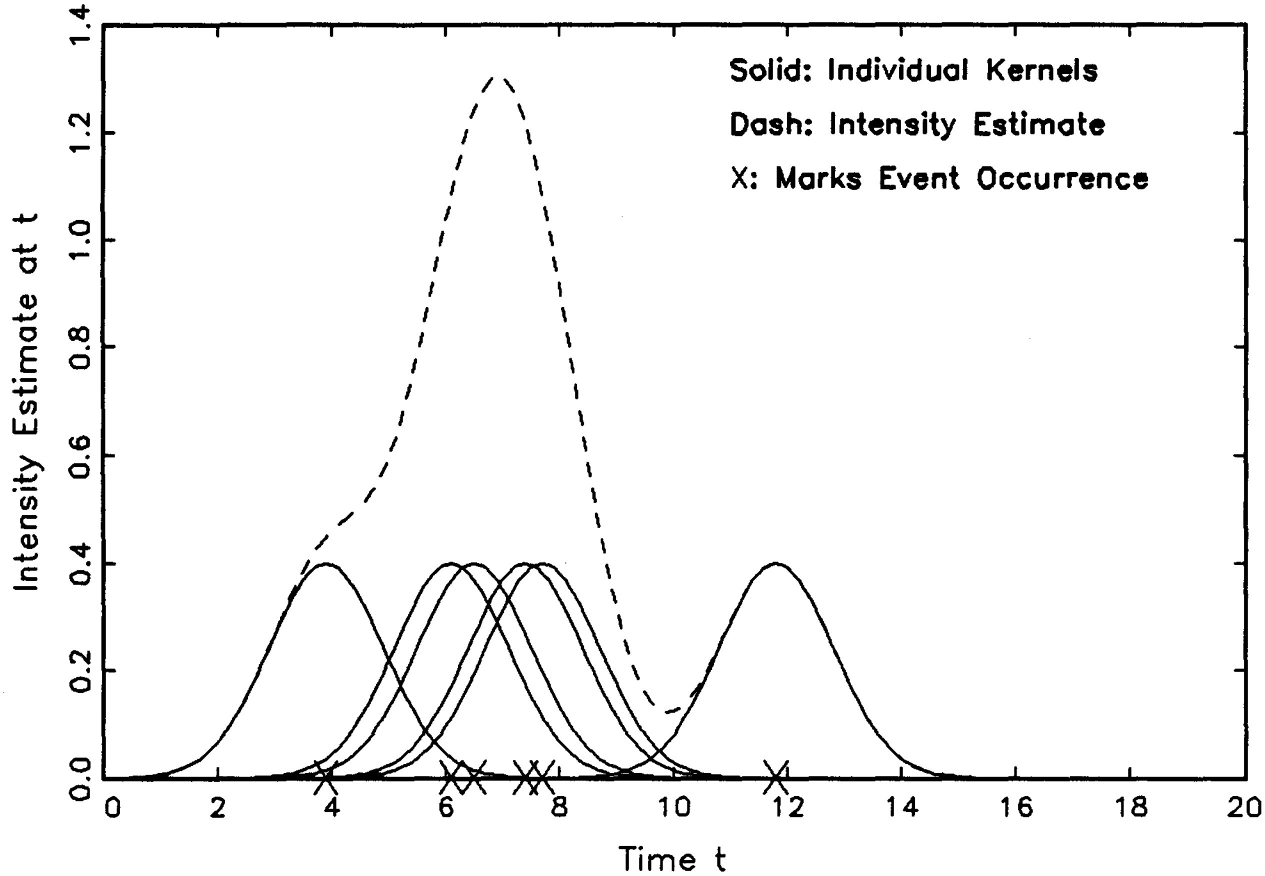

Fig. 2: Comparison of Two Normal Kernels

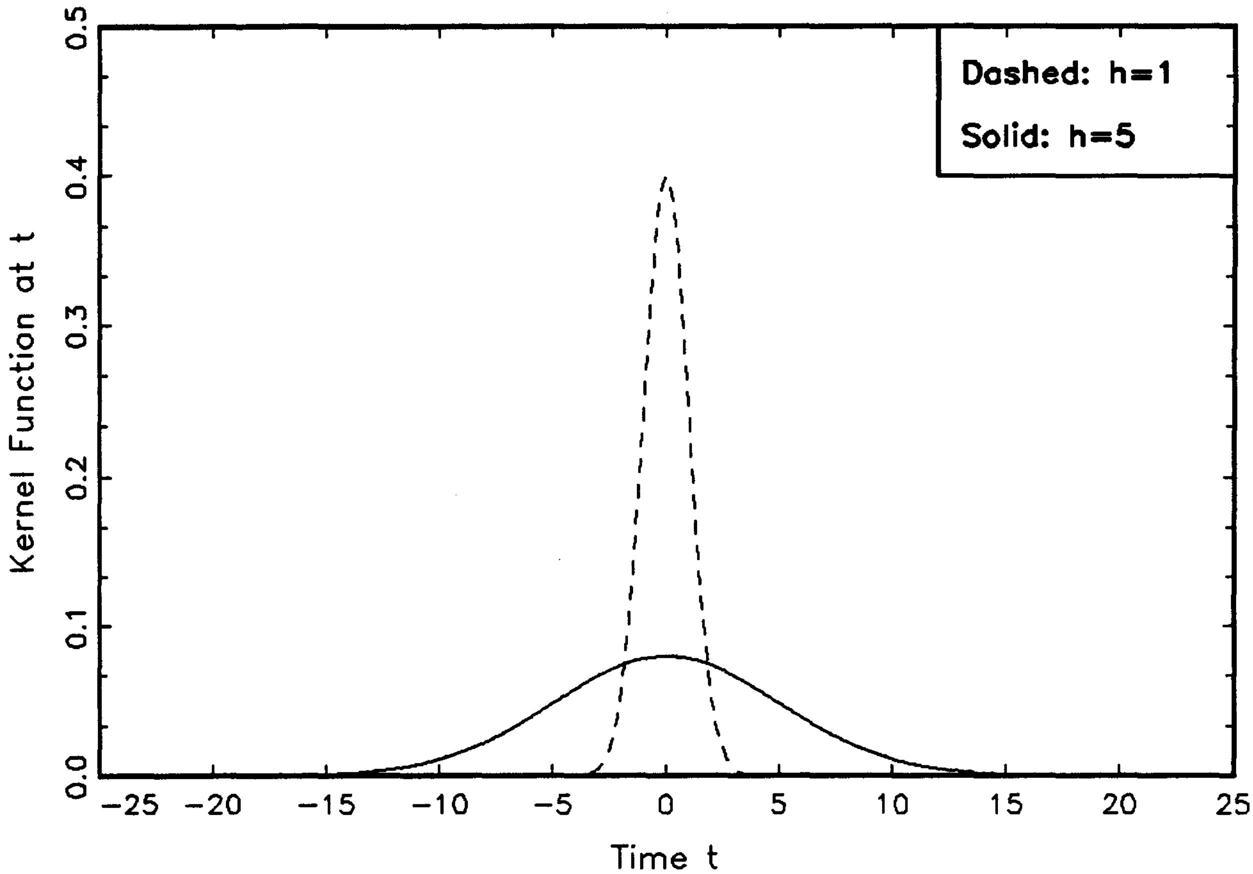

Fig. 3: The Observed Number of Annual Tropical Events

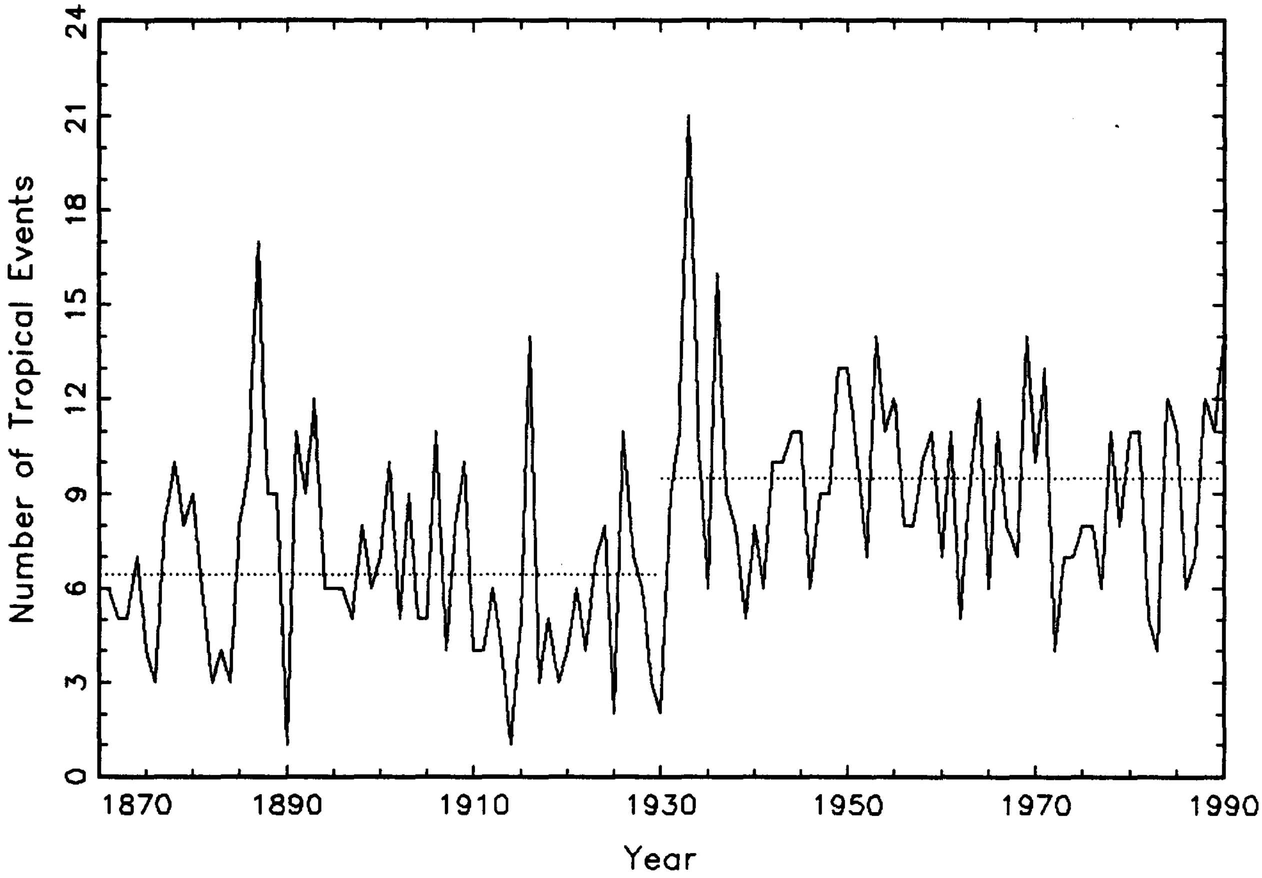

Fig. 4: Yearly Intensity Estimate with h=2.5

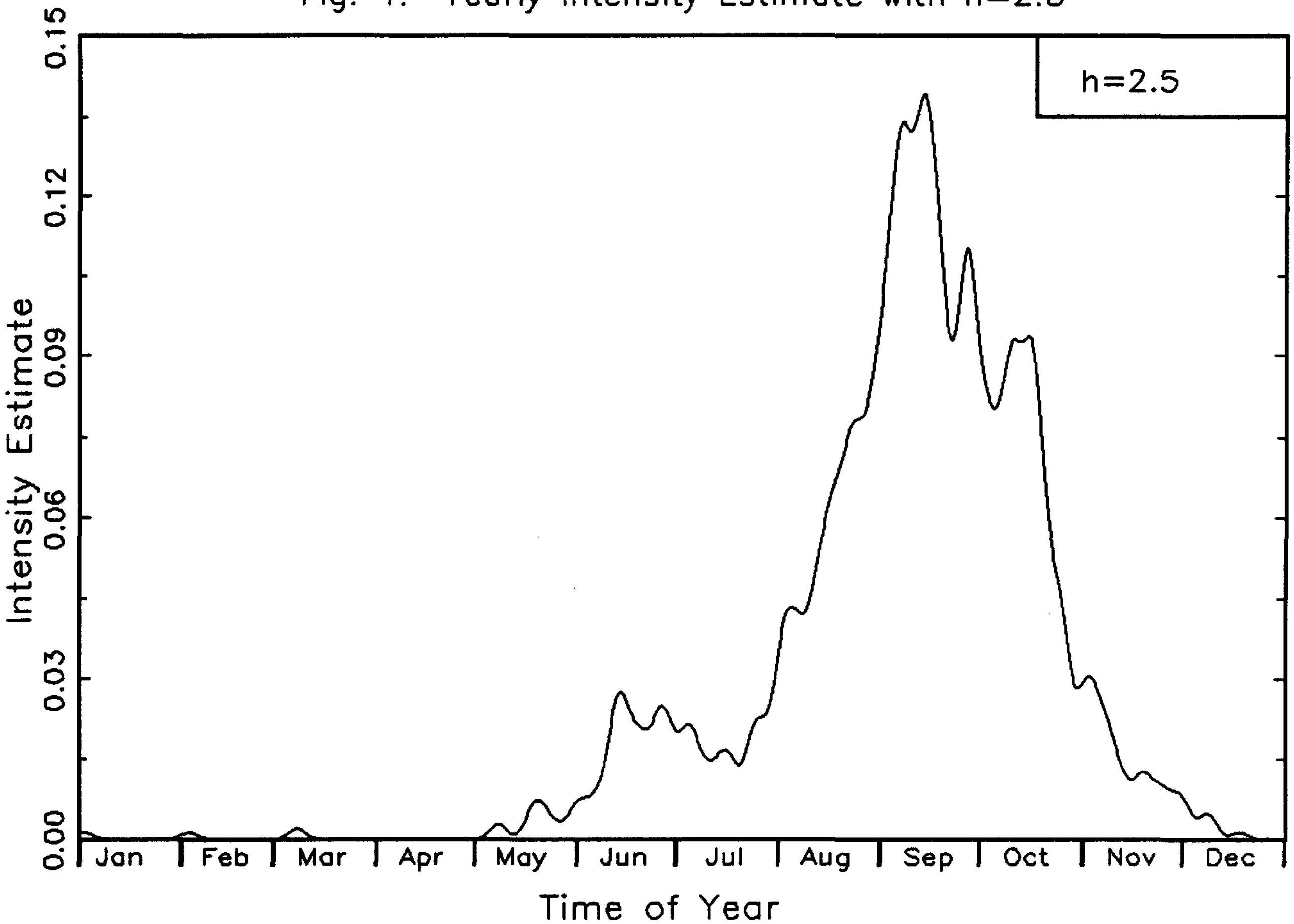

Fig. 5: Yearly Intensity Estimate with h=15.0

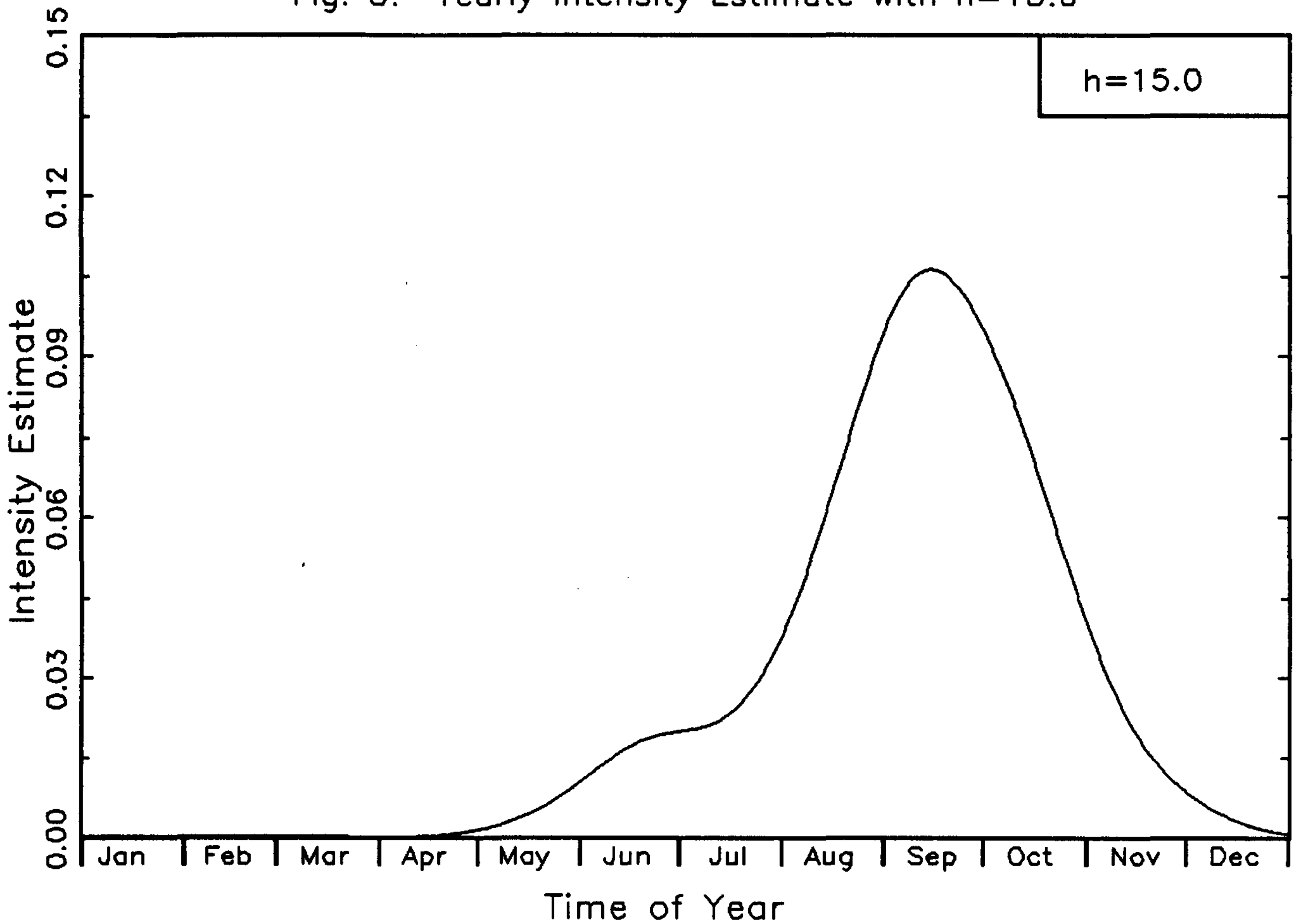

Fig. 6: Yearly Intensity Estimate with h=5.0