STEFFENS, F.E. Spatial distribution of probabilities. APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries. Volume 3: *Geostatistics*. Johannesburg, SAIMM, 1987. pp. 83–92.

Spatial Distribution of Probabilities

F.E. STEFFENS

Department of Statistics, University of South Africa, Pretoria, South Africa

The application of kriging in assessing the area rainfall in a weather modification experiment is complicated by the presence of many zero rainfall recordings on a given day, and the fact that the variograms seem to be quite different for the rainfall on different days. Indicator variograms are shown to be directly related to two sets of conditional probabilities. Probability kriging is used in estimating the probability of rainfall amounts above any cutoff at any point

Introduction

The Bethlehem Precipitation Research Project (BPRP) was undertaken with a view to determining the effect of rainfall stimulation on the microphysics of cumulus clouds during the summer months. The project area is centred at Bethlehem in the Orange Free State, and includes an area roughly within a 100 km radius of Bethlehem. While the experiment presently involves studying fairly small clouds and cloud complexes, and the effect of seeding is measured in terms of ice formation in the cloud, it would be of interest to be able to estimate the area rainfall over an area below the cloud.

There are presently about 230 raingauges in the project area. These are scattered somewhat randomly, with fewer raingauges at the outer edges of the project area. Raingauges are usually placed at farmsteads in order to have them read every day.

The interest of the project is restricted to days on which convection occurs, thus excluding fine weather days and days on which general rain occurs. The days of interest usually lead to scattered rain, with rainfall recorded at some stations but not at all of them. This leads to many zeroes in the rainfall data, which complicates the fitting of variograms.

Semivariograms of daily rainfall

Experimental semivariograms[2] for the daily rainfall on three different days are shown in Figures 1, 2 and 3. There were no noticeable anisotropies in these and any of the semivariograms discussed later, and therefore omnidirectional semivariograms are discussed throughout.

The experimental semivariogram for 15 November 1984 (Figure 1), a day on which 78% of the stations recorded rain, shows some structure, although with a fairly high apparent nugget effect. This could easily be fitted by a spherical semivariogram. If the zeroes are excluded, the result is a semivariogram which has much more structure and which is more complex. The sill is also much lower.

The semivariogram for 20 November 1984 (Figure 2) shows very little structure, as does the semivariogram based on the positive values only. On this day 50% of the stations

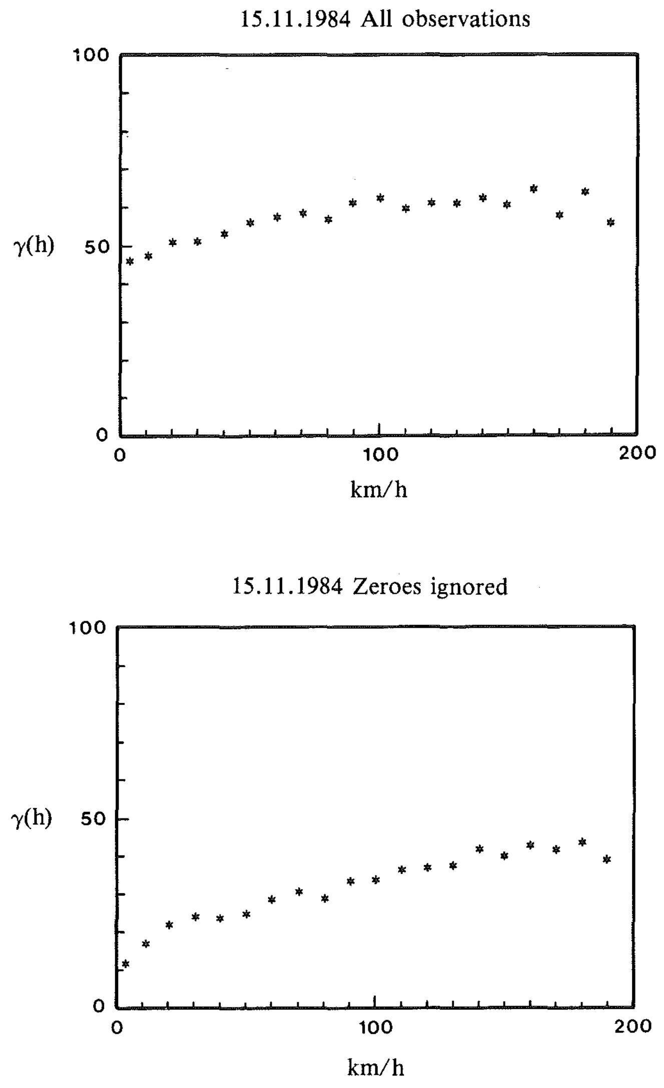

FIGURE 1. Experimental semivariograms for the rainfall recorded on 15 November 1984

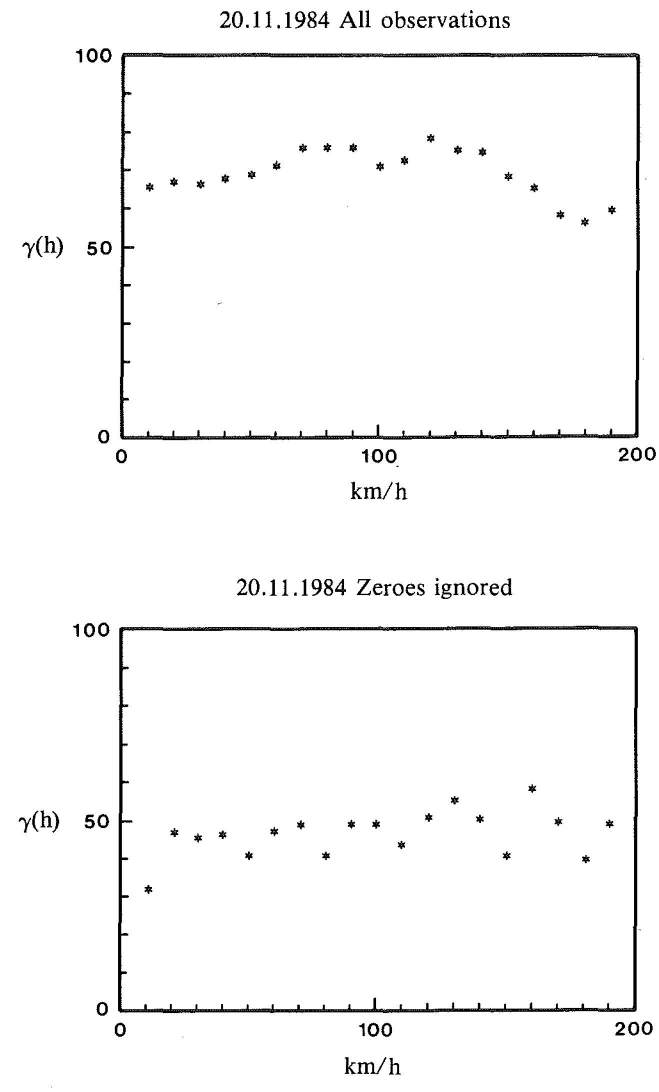

FIGURE 2. Experimental semivariograms for the rainfall recorded on 20 November 1984

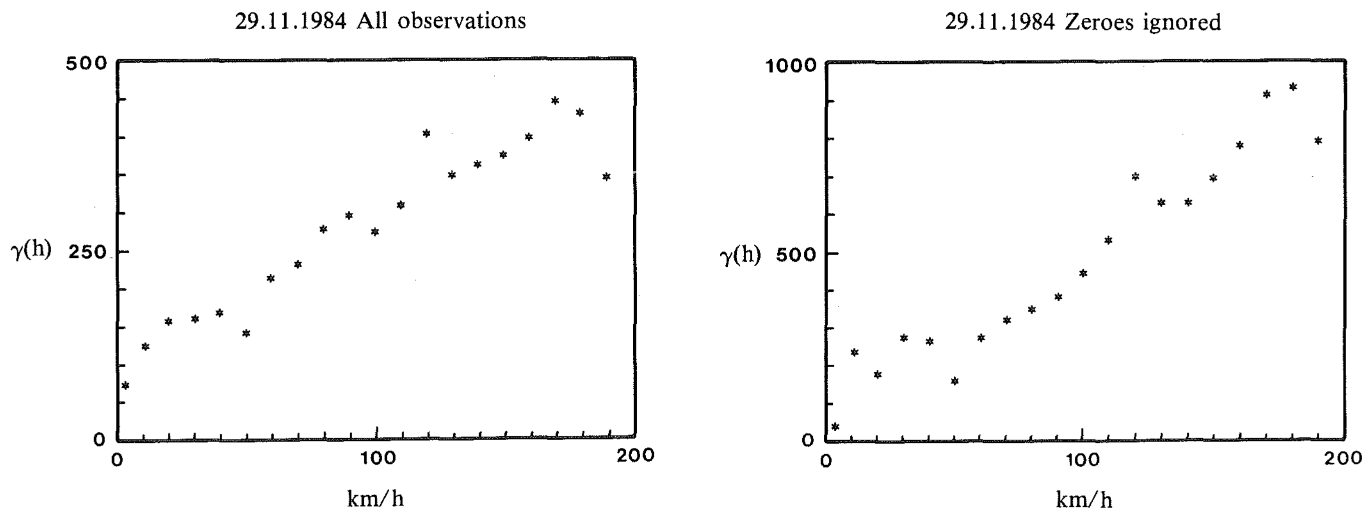

FIGURE 3. Experimental semivariograms for the rainfall recorded on 29 November 1984

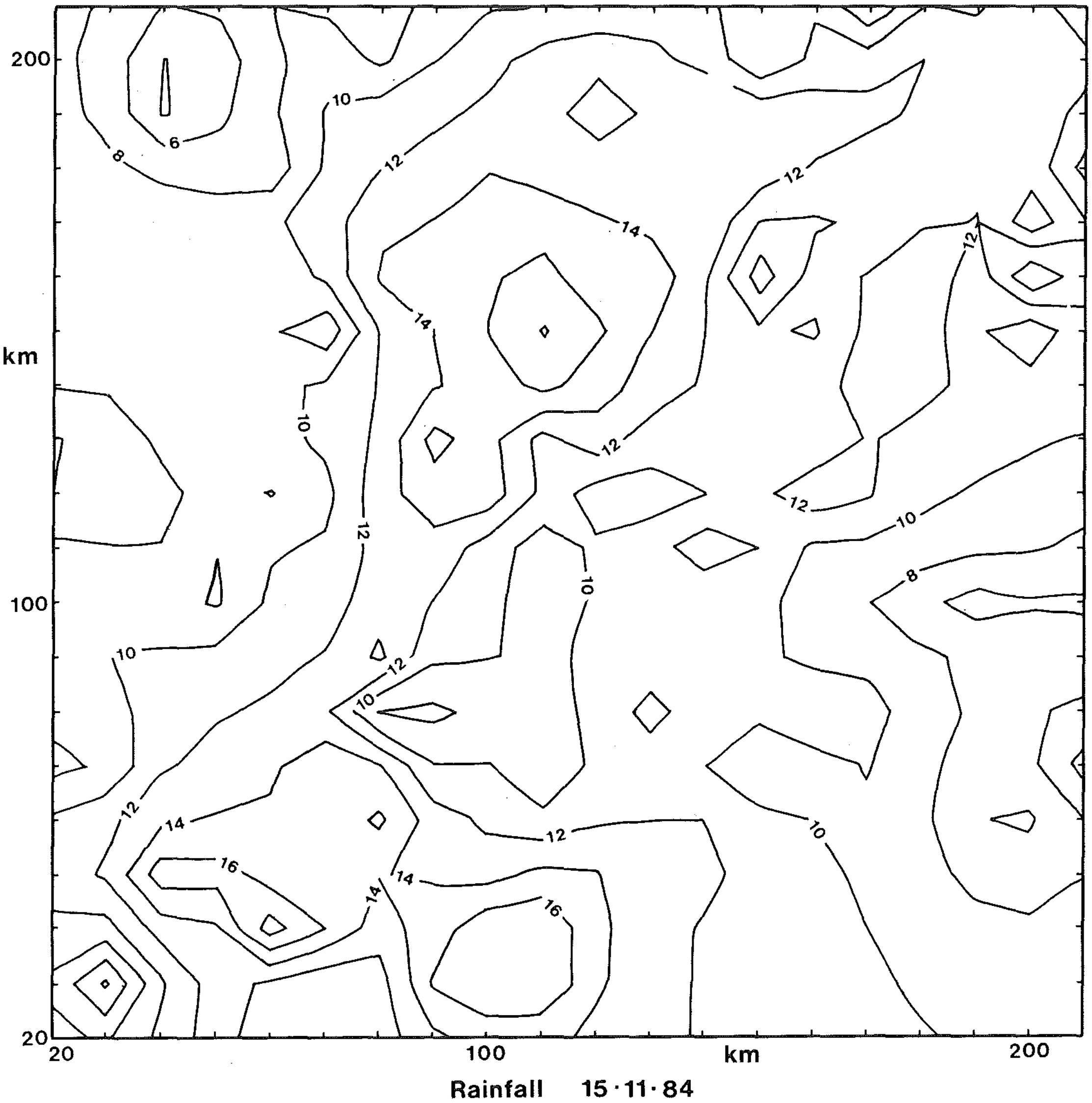

FIGURE 4. Contour map of the rainfall (mm) recorded on 15 November 1984

recorded rain.

On 29 November 1984 52% of the stations recorded rain, and in this case both experimental semivariograms show a marked linear trend. In this case the semivariogram (Figure 3) is much higher when the zeroes are ignored.

Contour maps of the actual rainfall on these three days are shown in Figures 4 - 6. Figure 4 (15 November) shows a fairly even distribution of rainfall throughout the area. Figure 5 (20 November) shows low rainfall in the south-west and pockets of high rainfall in the central parts. Figure 6 (29 November) shows a fairly even distribution of 0 - 10 mm over most of the area, but high rainfall in the south-west (146,4 mm at one station).

The histograms of these rainfall amounts (excluding the zeroes) are shown in Figures 7 and 8. In Figure 7 an attempt was made to fit normal, lognormal and Weibull distributions to the histogram of the rainfall on 20 November. The Weibull distribution seems to give the best fit, but the fit is by no means excellent. It is clear, at any rate, that none of the distributions are normal or lognormal. The summary statistics are given in Table 1.

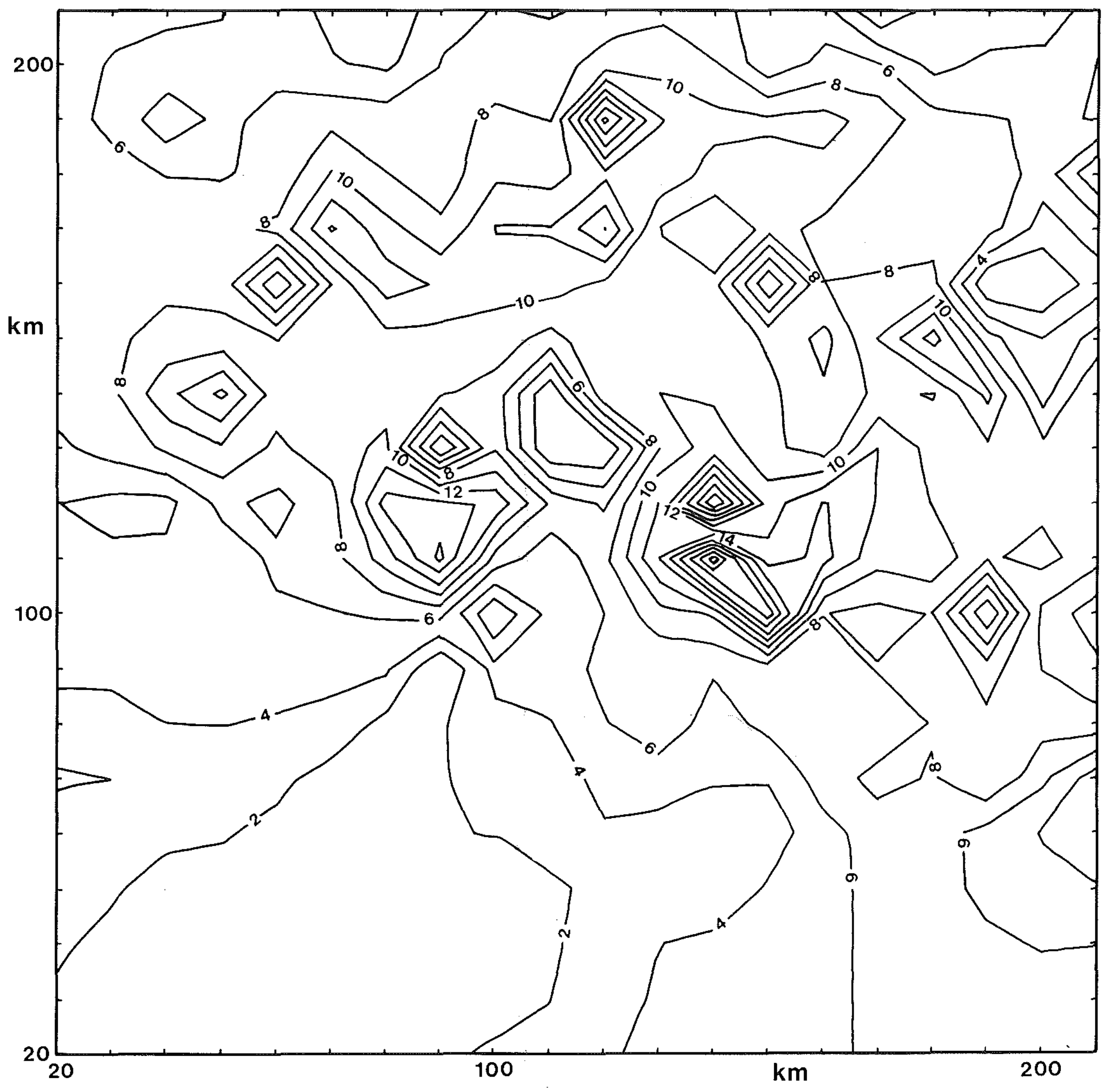

FIGURE 5. Contour map of the rainfall (mm) recorded on 20 November 1984

TABLE 1.

Summary statistics for the rainfall amounts

	15.11.84	20.11.84	29.11.84
n	230	230	230
Zeroes	49	114	119
Means	10,8	6,9	7,8
*	13,8	13,7	14,9
Medians	12,0	1,2	0,9
*	14,0	13,3	5,0
Standard dev.	7,6	8,4	17,0
*	5,7	6,8	21,2
Skewness	0,1	1,0	3,8
*	0,3	0,7	2,8
Kurtosis	-0,7	0,3	20,8
*	0,4	1,5	11,9

* zeroes excluded

Indicator semivariograms

Following Lemmer[2] indicator semivariograms were studied to see what the effect of the zeroes is. If the rainfall recorded at location x is $R(x)$, then the indicator function $I(x)$ is defined as follows:

$$I(x) = 1 \text{ if } R(x) = 0; \quad = 0 \text{ otherwise.}$$

The indicator semivariogram is defined as

$$\gamma(h) = \tfrac{1}{2} E[I(x) - I(x+h)]^2.$$

Define the following probabilities:

$$p_1 = P[I(x) = 1] = P[R(x) = 0]$$
$$p_0 = P[I(x) = 0] = P[R(x) > 0]$$
$$p_{ij}(h) = P[I(x) = i;\ I(x+h) = j],\ i;j=0;1.$$

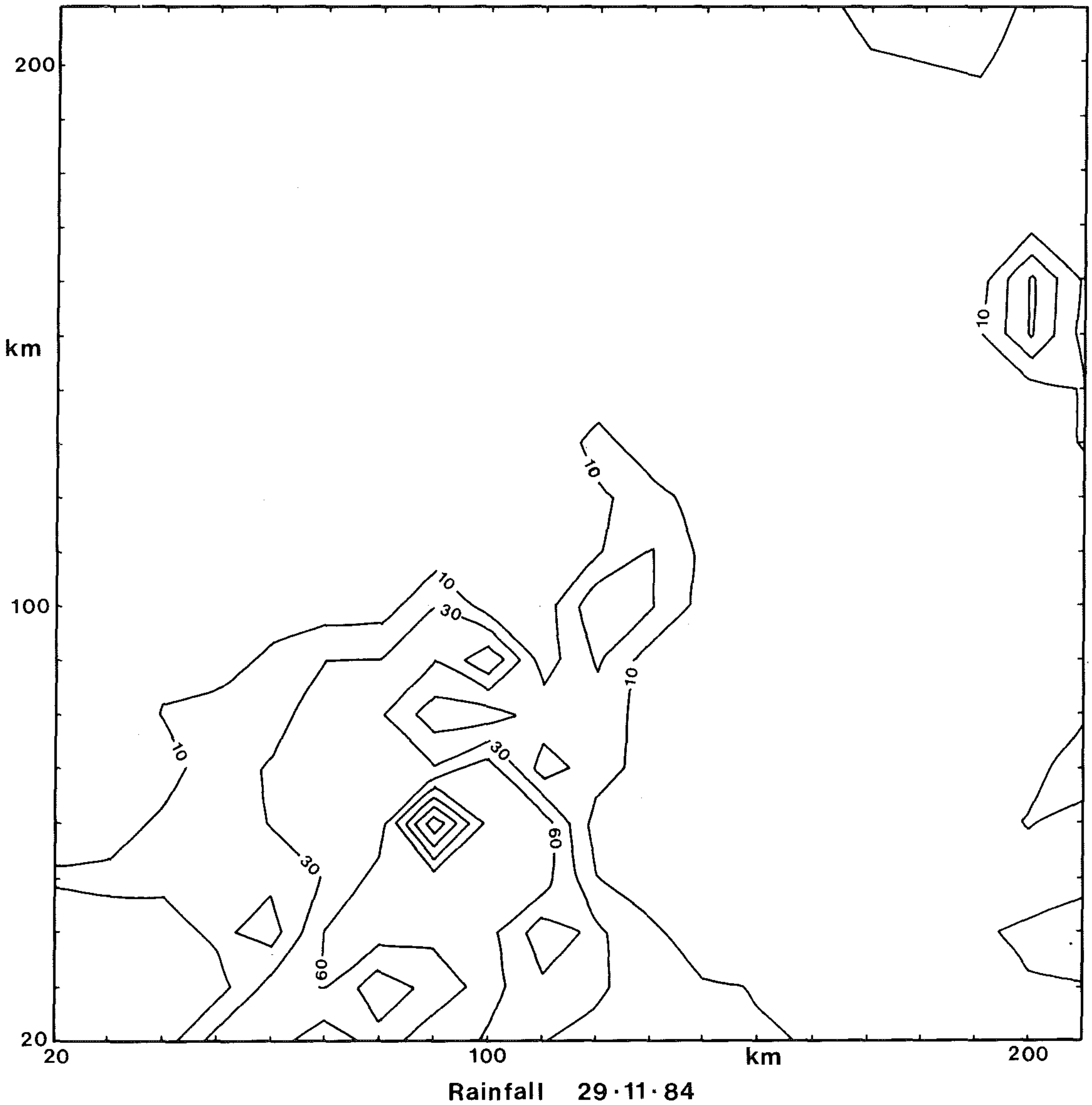

FIGURE 6. Contour map of the rainfall (mm) recorded on 29 November 1984

These probabilities are summarized in the following table:

	$I(x+h)=0$	$I(x+h)=1$	Total
$I(x)=0$	$p_{00}(h)$	$p_{01}(h)$	p_0
$I(x)=1$	$p_{10}(h)$	$p_{11}(h)$	p_1
Total	p_0	p_1	1

It is assumed that $p_{10}(h) = p_{10}(-h)$
$= p_{01}(h)$, i.e. that there is no preferential direction. It is also assumed that there is no drift, i.e.

$$p_0 = P[I(x) = 0] = P[I(x+h) = 0],$$

and similarly for p_1.

Then

$$2\gamma(h) = p_{01}(h) + p_{10}(h) = 2p_{01}(h). = P[I(x) \neq I(x+h)].$$

The indicator variogram thus represents the probability that the two indicators at x and x+h are different, and

$$1 - 2\gamma(h) = p_{00}(h) + p_{11}(h) = P[I(x) = I(x+h)],$$

which indicates that the indicator variogram implicitly lumps together the event that both indicators are equal to 0 and the event that both indicators are equal to 1. In some instances these two events may be of interest in themselves, and it could also be that the two events, when pooled, show little

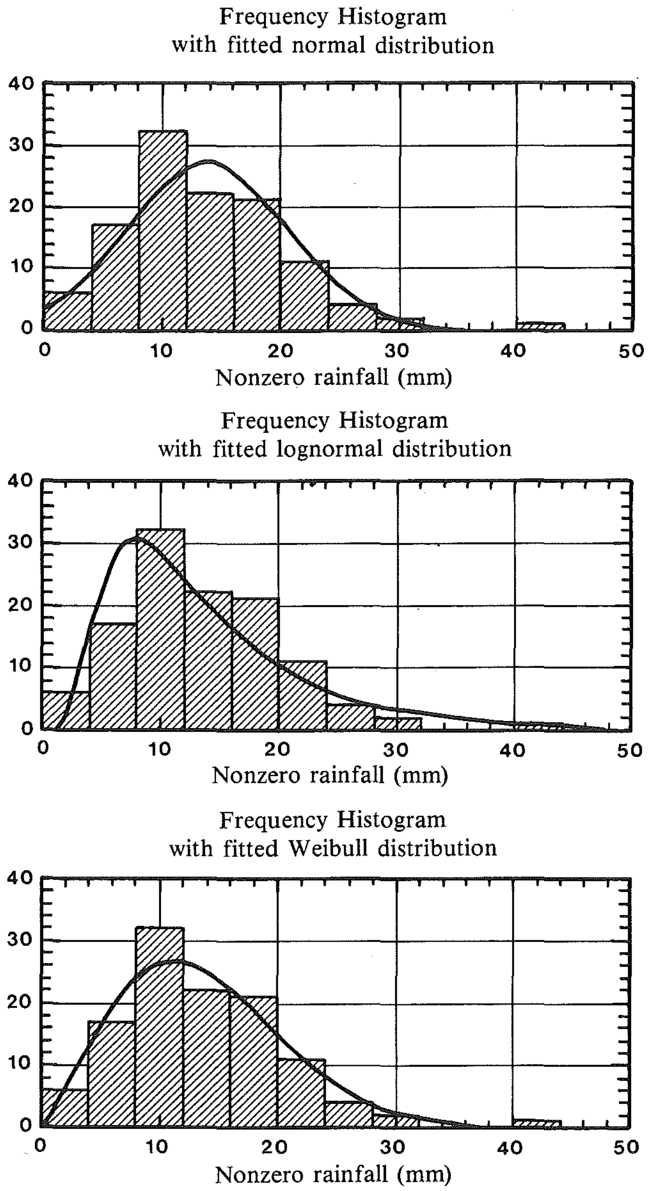

FIGURE 7. Frequency histogram of rainfall recorded on 20 November 1984 with fitted normal (top) lognormal (middle) and Weibull distributions (bottom)

dependence on h while one or both may well show considerable structure.

To study this phenomenon further, consider the two conditional probabilities

$$
\begin{aligned}
p_{0|0}(h) &= P[I(x) = 0 \mid I(x+h) = 0] \\
&= p_{00}(h)/p_0, \\
p_{1|1}(h) &= P[I(x) = 1 \mid I(x+h) = 1] \\
&= p_{11}(h)/p_1.
\end{aligned}
$$

The relationship between these conditional probabilities and the indicator variogram is as follows:

$$\frac{1-2\gamma(h)}{2\gamma(h)} = \frac{1}{2}\left\{\frac{p_{0|0}(h)}{1-p_{0|0}(h)} + \frac{p_{1|1}(h)}{1-p_{1|1}(h)}\right\}.$$

The term on the right-hand side is the average of two (conditional) odds.

The assumption of independence (the value of the indicator at x is independent of the value of the indicator at x+h) would imply

$$p_{ij}(h) = p_i p_j \text{ for all } i \text{ and } j,$$

in which case

$$
\begin{aligned}
\gamma(h) &= p_0 p_1 \\
p_{0|0}(h) &= p_0 \\
p_{1|1}(h) &= p_1.
\end{aligned}
$$

These expressions provide base lines against which the dependence of the indicator variogram and the conditional probabilities on h may be judged.

In terms of the observations, let

$n_{ij}(h)$ = Number of pairs of points separated

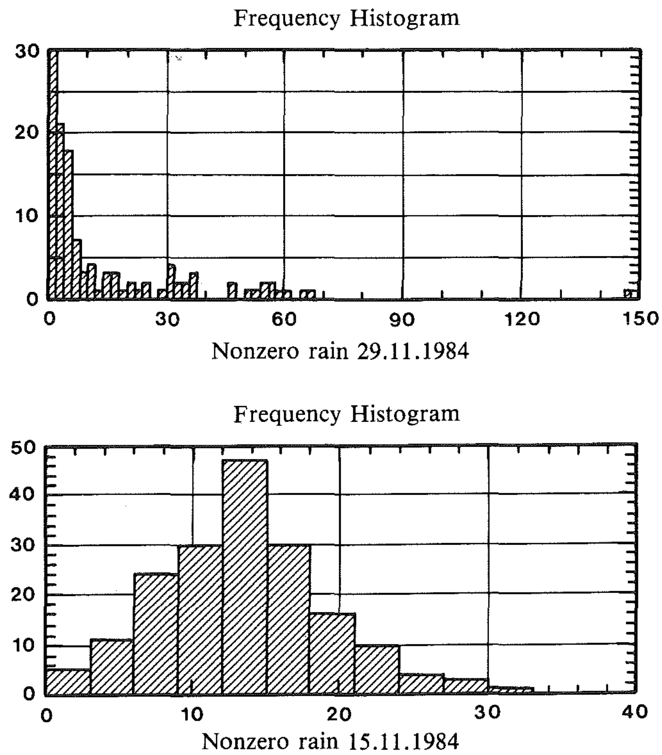

FIGURE 8. Frequency histograms of rainfall recorded on 29 November 1984 (top) and 15 November 1984 (bottom)

by h for which $I(x) = i$ and $I(x+h) = j$; $i;j = 0;1$.

$n_i(h) = n_{i0}(h) + n_{i1}(h)$
$\quad = n_{0i}(h) + n_{1i}(h)$ by assumption.

$n(h) = n_0(h) + n_1(h)$
$\quad$ = number of pairs of points separated by h.

The contingency table is then as follows:

	I(x+h)=0	I(x+h)=1	Total
I(x)=0	$n_{00}(h)$	$n_{01}(h)$	$n_0(h)$
I(x)=1	$n_{10}(h)$	$n_{11}(h)$	$n_1(h)$
Total	$n_0(h)$	$n_1(h)$	$n(h)$

The assumption of lack of directional preference implies that $n_{01}(h) = n_{10}(h)$.

The experimental semivariogram is given by

$$\gamma^*(h) = \frac{1}{2}\,\frac{n_{01}(h) + n_{10}(h)}{n(h)} .$$

The conditional probabilities are likewise estimated by

$$P^*_{0|0}(h) = \frac{n_{00}(h)}{n_0(h)} ; \qquad P^*_{1|1}(h) = \frac{n_{11}(h)}{n_1(h)} .$$

To illustrate these ideas, the experimental semivariograms and conditional probabilities were computed for the rainfall in the BPRP project area on the three days discussed previously. These are illustrated in Figures 9 - 11. For the rainfall on 15 November 1984 the conditional probabilities as well as the indicator semivariogram show no stucture at all. For the rainfall on 20 November 1984, for which the ordinary semivariograms showed a very diffuse structure, the conditional probability $P_{0|0}(h)$ shows a very strong dependence on h, while the other conditional probability and the indicator semivariogram do not show such a strong dependence on h. There is no

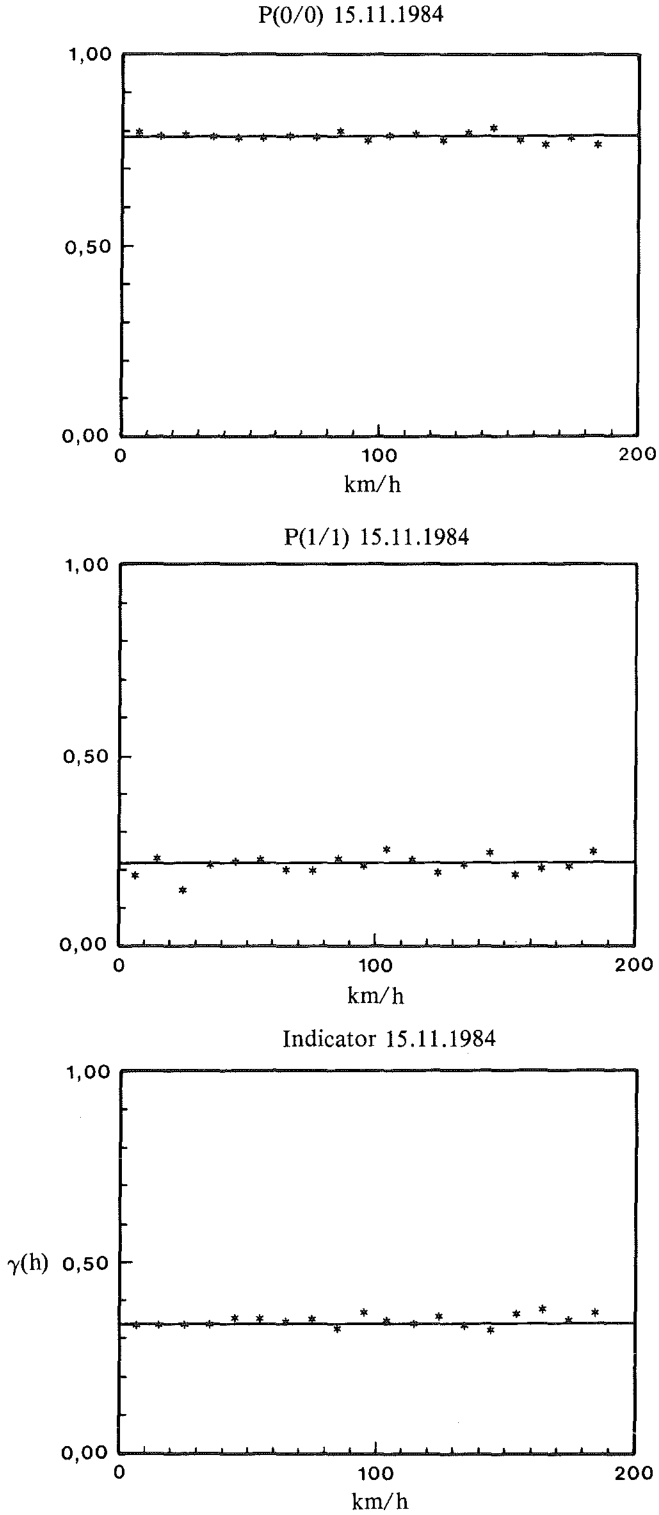

FIGURE 9. Conditional probabilities and indicator semivariogram for 15 November 1984

logical reason why the conditional probabilities would show an increase with h, and the indicator semivariogram should logically be non-decreasing. The opposite tendencies at short distances in the last two graphs in figure 10 are therefore ascribed to

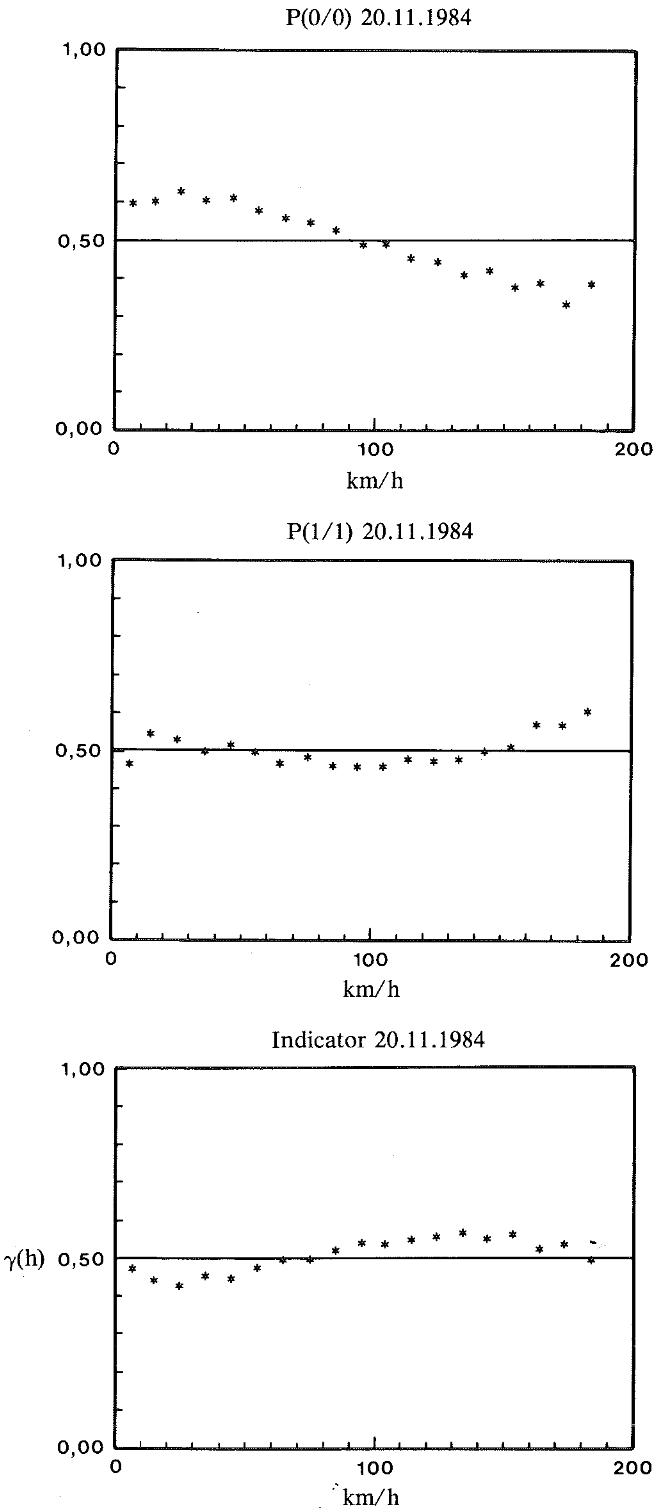

FIGURE 10. Conditional probabilities and indicator semivariogram for 20 November 1984

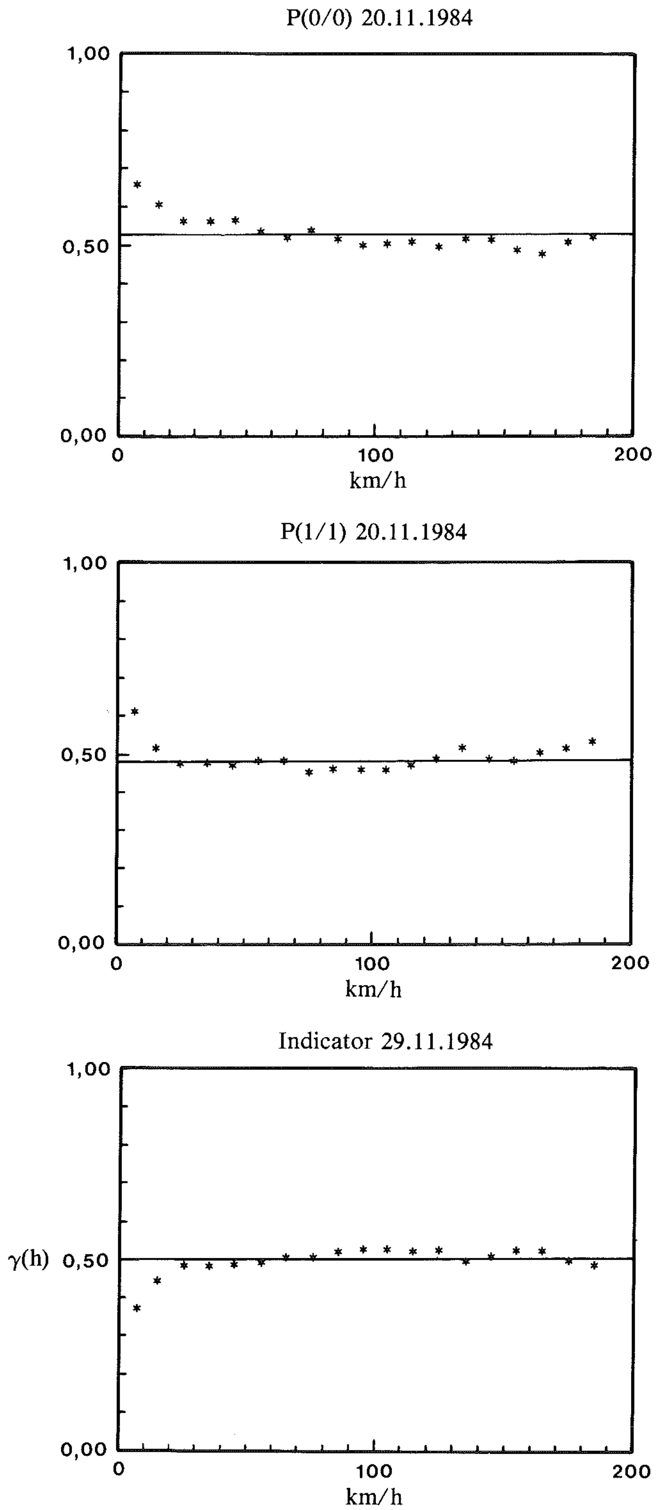

FIGURE 11. Conditional probabilities and indicator semivariogram for 29 November 1984

sampling fluctuations. The graphs for 29 November 1984 show a strong structure at short distances (up to about 20 km), and stable values beyond this distance.

The above discussion was conducted in terms of the probabilities of zero and positive rainfall. It is easily generalised to probabilities below and above any cutoff c by letting

$$I_c(x) = 1 \text{ if } R(x) \leq c$$
$$= 0 \text{ if } R(x) > c.$$

These notions are applied in the next section.

Application of probability kriging

Probability kriging (see Sullivan[3]) may be used to estimate the probability of rainfall above any cutoff at any point in the project area or in any subregion in the area. This was done for 20 November 1984 using a microcomputer package developed at Stanford University (Isaacs[4]). Semivariograms had to be fitted for the indicators for a series of nine cutoff values, for the uniform transform of the rainfall, as well as cross-semivariograms between the uniform variable and the indicators. Some care has to be taken in modelling these semivariograms in order to preserve order relations. In this case the semivariograms were modelled as a combination of a nugget effect and spherical semivariograms with ranges 10 and 140 km. For the uniform variable an additional spherical term with range 200 km had to be added. Some of these fitted models are illustrated in Figure 12. The overall fit of the whole family of semivariograms and the order relations had to be taken into account in the process of fitting these.

The result of the probability kriging is a wealth of information, some of which is illustrated in Figures 13-15. These show the probabilities of positive rainfall at various points in the project area and, for comparison, the rainfall at stations within 30 km of these points. Such estimated cumulative probabilities for points as well as subregions are of great potential interest to the meteorologists.

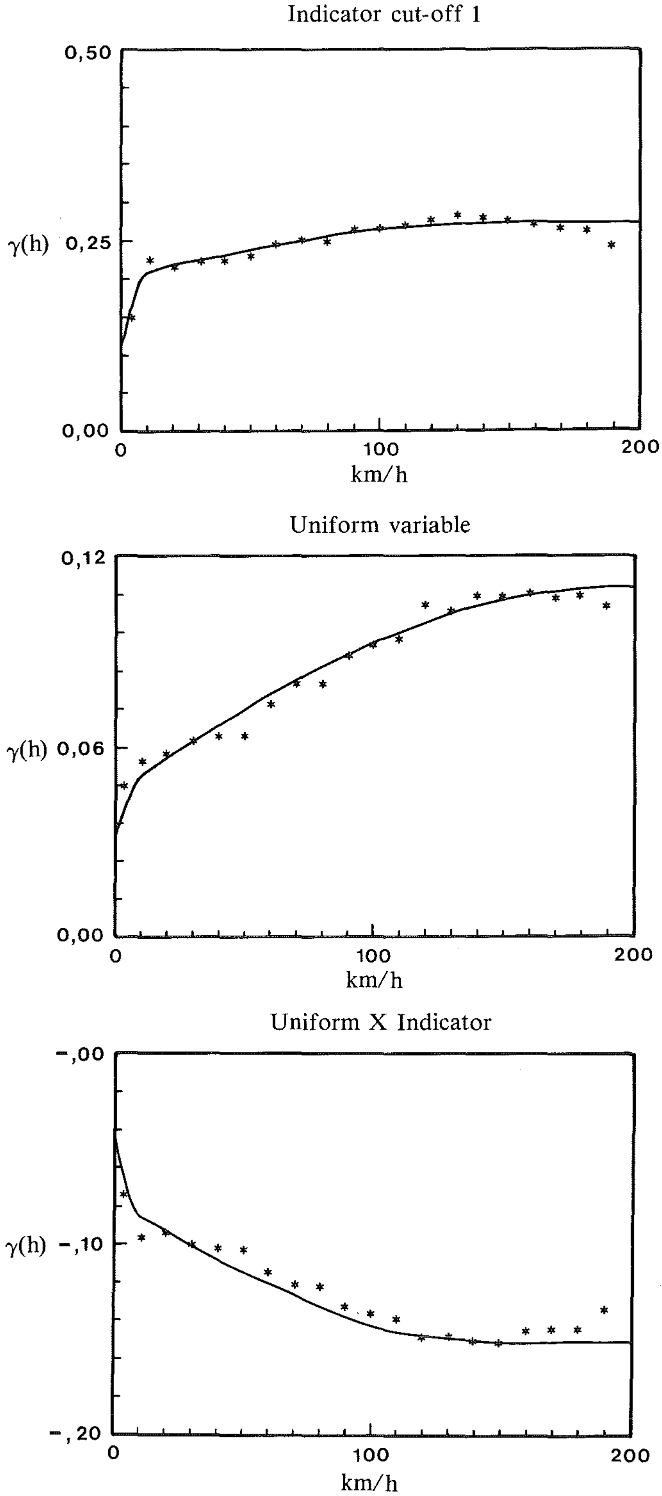

FIGURE 12. Semivariograms used in probability kriging (cut-off 0). 20 November 1984

References

1. JOURNEL, A.G. and HUIJBREGTS, Ch. J. *Mining Geostatistics*. London, Academic Press, 1978.

2. LEMMER, I.C. Estimating local recoverable reserves via indicator kriging. In: *Geostatistics for Natural Resources Characterization, Part 1, Verly et al.* NATO ASI Series. Dordrecht, Reidel Publishing Co, 1983. pp. 349-364.

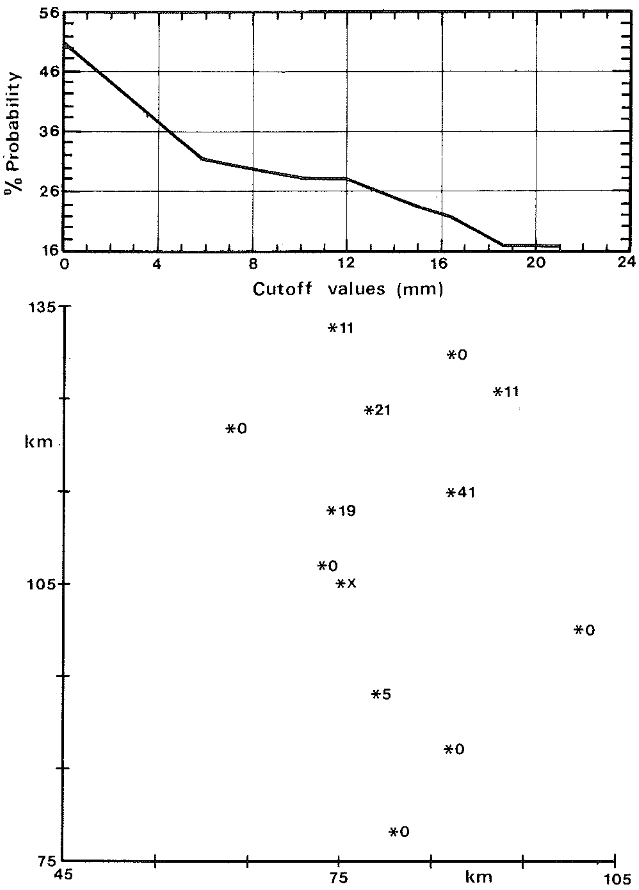

FIGURE 13. Estimated exceedance probabilities (top) and rainfall (mm) at stations within 30 km of location (75:105) (bottom) 20 November 1984

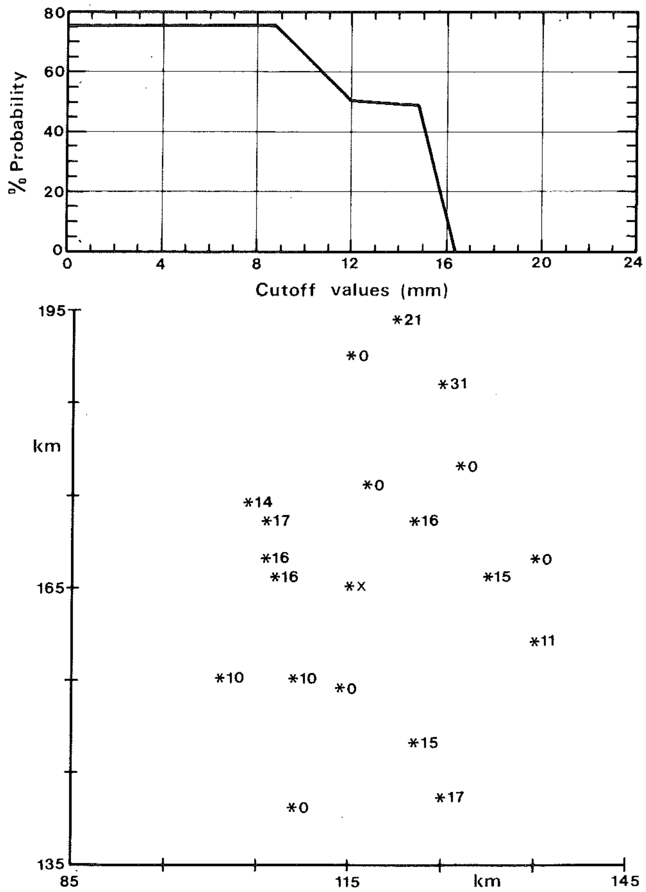

FIGURE 14. Estimated exceedance probabilities (top) and rainfall (mm) at stations within 30 km of location (115:165) (bottom) 20 November 1984

3. SULLIVAN, J. Conditional recovery estimation through probability kriging. In: *Geostatistics for Natural Resources Characterization, Part 1, Verly et al.* NATO ASI Series. Dordrecht, Reidel Publishing Co, 1983. pp. 365-384.

4. ISAACS, E.H. *User Manual for Probability Kriging Software.* Applied Earth Science Department, Stanford University, California, 1985.

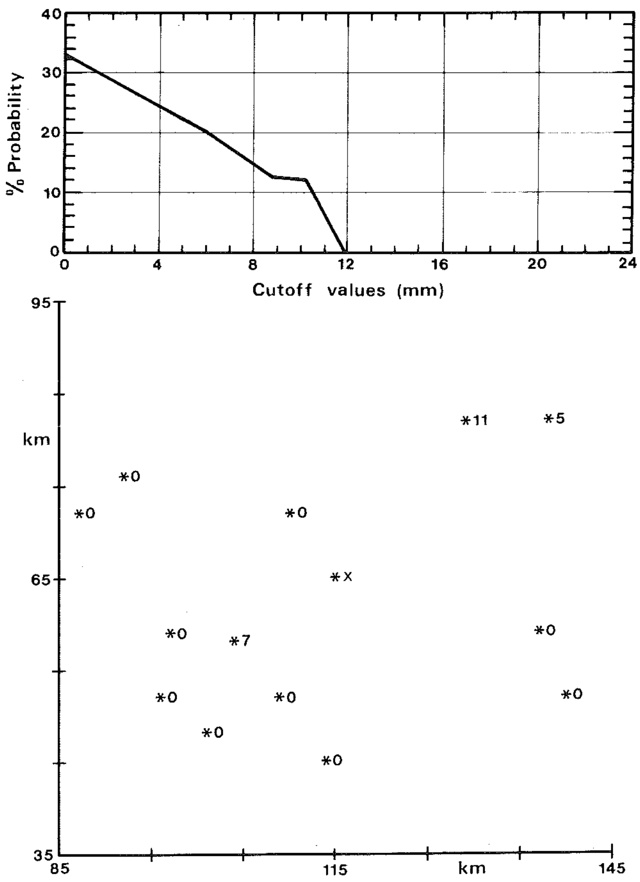

FIGURE 15. Estimated exceedance probabilities (top) and rainfall (mm) at stations within 30 km of location (115:65) (bottom) 20 November 1984