

CONTAMINATION IN SEED CROPS

III. Relation with Isolation Distance

A. J. BATEMAN

John Innes Horticultural Institution, Merton

Received 7.vi.47

CONTENTS

- | | |
|-------------------------------------|--------------------------|
| 1. Introduction | 4. Wind-pollinated crops |
| 2. Behaviour of pollinating insects | (a) Pollen dispersal |
| Observations | Theoretical |
| Closeness of foraging | Fitting the data |
| Theoretical | (b) Contamination |
| Fitting the data | 5. Conclusions |
| The effect of several flights | 6. Summary |
| 3. Insect-pollinated crops | References |
| Theoretical | |
| Empirical | |
| Fitting the data | |

I. INTRODUCTION

A NUMBER of results are now available from systematic experiments on natural cross-pollination between varieties of crops grown for seed, *e.g.* on radish by Crane and Mather (1943); on radish and turnip, using various planting arrangements by Bateman (1947*a*); and on the wind-pollinated crops, beet and maize, by Bateman (1947*b*). These data are in a consistent form which makes comparison and generalisation possible.

The conclusion that may be drawn from these experiments is that whatever the absolute level of contamination or the range of distance involved in any experiment, the shape of the curve relating contamination to isolation distance is the same. It is such that the rate of decrease of contamination per unit increase of isolation distance itself decreases with that increase. At first, increases in isolation distance rapidly reduce contamination, but at greater distances the contamination, though small, becomes persistent and in one instance (Bateman, 1947*a*) there was no detectable decrease in contamination when the isolation distance was increased from 50 yards to 200 yards.

The consistency of these results justifies the present attempt to derive a general mathematical expression which, by the substitution of appropriate values for the constants, could express the variation of contamination with distance under any conditions and for any crop.

In the following sections (the first of which contains new observations on the pollinating methods of bees) the argument is roughly divisible into three parts:—

(i) The discussion of the causation of the particular phenomena from which are derived one or more possible fundamental equations.

(ii) The choice of suitable simple equations which give an approximation to the results anticipated from the fundamental equations.

(iii) The testing of these empirical equations against experimental data.

To the pure mathematician this may be no more satisfactory as a method than a purely empirical one. There are, however, many possible empirical formulæ to fit a given set of data. Some of these would immediately prove unsuitable when applied to another set of data using a different range of factors. There would still be a large residue whose unsuitability would only slowly be exposed by accumulating evidence. The method herein adopted does, if the fundamental equation is satisfactory, enable one to make at the start a good choice of empirical equations, which must pass the severe test of agreement with the fundamental equation.

Since an attempt is made to discuss each phenomenon according to its causation, insect-pollination and wind-pollination have to be considered apart at first and, only at the end, taken together.

For the sake of consistency the same symbols have been used throughout for the same variables. This has meant transposing the formulæ of other authors.

The main symbols used are as follows :—

F = the proportion of contamination

D = the distance

x = the power of D

n = the insect density

p = the pollen or spore density

2. BEHAVIOUR OF POLLINATING INSECTS

In insect-pollinated crops the effect of distance on cross-pollination is intimately connected with the behaviour of the pollinating insects. The connection is simplest in self-incompatible crops since with them all the functioning pollen has to be brought by insects. The proportion of contamination in the seed will then be equal to that in the available pollen on the insect visitor. There will of course be pollen in the baskets on the bee's legs and on some other parts of its body which will not be available for pollination. In self-compatible crops the contamination of the seed will be diluted to a varying extent according to the amount of automatic self-pollination in the particular crop.

The proportion of contaminant pollen on an insect must depend on the number of visits made after it has left the contaminating variety. The effect of distance on this depends on two factors: the flight habits of pollinating insects when at work, and (more difficult to observe) the rate of replacement per visit of contaminant by non-contaminant pollen.

Observations

The first feature observed about the flight of a pollinating insect is that there are two distinct kinds of flight. This is most noticeable in bees, which are the most important pollinators. Most of the insects' movement takes the form of slow flights which are easy for the observer to follow, and during which the insect hardly moves above the level of the flowers of the crop being visited. But every now and then an insect will soar up suddenly from the crop and disappear from sight. In hive bees these soaring flights probably mean the end of the forage and a return to the hive. By analogy, with bumble bees and solitary bees (such as *Andræna*), these flights would mean a return to the nest. In hover flies there is more doubt as to whether the two kinds of flight have distinct functions, but these insects are relatively unimportant in pollinating most crops.

In general, it would seem that pollen carried forward from one forage to the next is not likely to be responsible for much of the seed set because a single forage includes a large number of flights; and although there does not appear to be any evidence as yet on the amount of pollen carried forward from one forage to the next, after the bee has cleaned itself in the hive, this carry forward is unlikely to be great. Most of the effective pollen will be collected and deposited during a single forage.

If information is to be obtained which will assist in understanding the process of contamination it is therefore the activity of the bee during the forage which needs most attention.

For this purpose observations were made on hive bees, solitary bees and hover flies visiting a turnip plot in bloom. The plants were spaced 6 in. apart each way. Individual insects appeared to move at random over the plot as judged by the paths traced for single insects during a forage. There was apparent disorderliness, the same plant often being revisited at irregular intervals.

Other observations made on bumble bees on a radish plot enable us to make a statistical test of randomness. Here the path of each bee was projected on to a line at right angles to the rows. Each flight was recorded in terms of its component parallel to this line. When bees did not cross from one row to the other the flight length was recorded as zero. Other flights were recorded as so many rows to the left or right. Three forages were analysed with respect to the direction of consecutive directed flights, *i.e.* those whose direction was recorded. The results are summarised on the following page.

Like follows like in 115 instances out of a total of 199. If the flights were completely random the expectation would be 99.5. There is therefore an excess of cases where consecutive directed flights were in the same direction. $\chi^2 = 4.829$ for one degree of freedom with a probability of 0.02. In none of the three forages was there an excess of flights in one direction over the other. This means that, though

over longer periods the bees did not appear to move in any particular direction, over shorter periods the bees tended to have an over-all

		Subsequent flight	
		Left	Right
First flight	Left	55	42
	Right	42	60

direction first one way and then the other. The deviation from complete randomness, however, though statistically significant, is not very great.

The flight lengths of insects visiting the turnip plot were recorded in the manner illustrated in fig. 1. As the plants were 6 in. apart

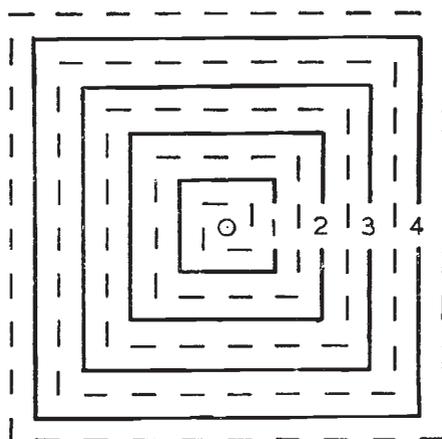


FIG. 1.—The method of recording the flight lengths of pollinating insects in two dimensions. For explanation see text.

each way the unit of distance was taken as 6 in. Any plant just visited is taken as being at position 0 in fig. 1. The next plant visited can be taken as being on a square with plant 0 as the centre. The figures 1, 2, 3 and 4 denote progressively larger concentric squares whose sides are distances of 1, 2, 3 and 4 respectively from the centre. If the insect lands on a plant on square 2 the flight will be recorded as of length 2. Fig. 2 shows the distribution of flight lengths on turnip, of hive bees, solitary bees and hover flies. The distributions in the graphs appear similar for all the species. No record was kept of the length of the forages, but it can be stated

that hive bees foraged longest of the three insect types, and hover flies shortest.

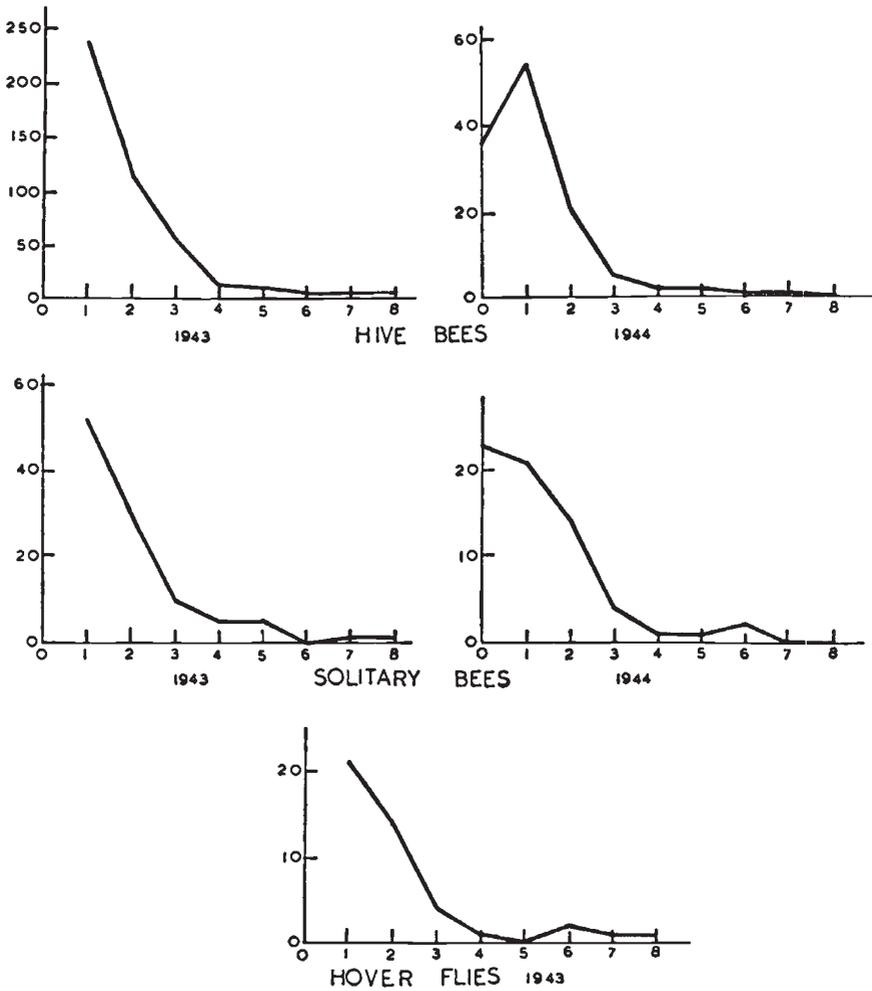


FIG. 2.—The frequency distribution of the two dimensional flight lengths of various insects on turnip. *D* is expressed in units of 6 inches.

Bumble bees do not visit turnip, possibly because the slender pedicels will not take their weight, but it would appear from observations on the radish that bumble bees forage longer than hive bees.

Closeness of foraging

The closeness with which insects forage is reflected in the distribution of the points which they reach by single flights, about the plant from which the flights start. The closer the foraging, the narrower this frequency distribution will be ; and the wider the foraging, the

wider the distribution. The variance of the distribution can be taken, therefore, as a measure of foraging behaviour. An adjustment must, however, be made to allow for the two dimensional nature of flights, because the variance is measured relative to one dimension only.

This is done by dividing the frequency of flights, T , by a number proportional to the area of the zone into which the flight takes the insect. In Fig. 1 the boundaries of the zones are shown by dotted lines. The area of zone 4 is the area of the square enclosed by the outer zone boundary minus the area enclosed by the inner zone boundary. This is $9^2 - 7^2 = 32$. In general the area of a zone is given by $(2D+1)^2 - (2D-1)^2 = 8D$ where D is the number of the zone (*i.e.* the distance of the plant from plant 0). This formula applies for all zones except zone 0 whose area is 1. If for all the other zones we use D as the correction factor for area, then the factor for zone 0 is $\frac{1}{8}$. The corrected value for flight distribution is termed n which is the density of insects round a source after one flight. The value of n for $D = 0$ is halved for estimation of variance, since the other values of D represent only half the distribution.

The bumble bee data from the radish plot with plants one foot apart gives the distribution in one dimension and therefore needs no correction. The observed frequencies give directly the distribution of n .

The mean squares for the distribution of insects after one flight are calculated as $\frac{\Sigma(nD^2)}{\Sigma n}$. Except in the bumble bee data, there is no

means of estimating the mean, which is therefore taken as zero. D is therefore the deviation from the mean and no correction for the mean is required. As the mean is assumed, no degree of freedom is utilised in its estimation and the number of degrees of freedom is Σn .

In 1943 the frequency of flights within zone $D=0$ was not recorded. The "mean square" calculated from these observations is thus not a true mean square but is inflated by the omission of the class closest to the assumed mean. A comparison of these false mean squares will show, however, any differences in the degree of dispersal of the flights.

Table 1 shows the values for T and n for six sets of observations with the calculations of the mean squares, true and false. The first three columns are strictly comparable, being based on the same plots at the same time. The next two differ from the first three only in pertaining to the following year. The true mean square for the bumble bee is six times that of the insects in columns 4 and 5. Here not only the years were different but also the crop, the spacing, and the method of recording, so little significance can be attached to the comparison.

The first five sets of observations suggest that if all cross-pollination is carried out during forages and none between forages, the dispersal of contaminating pollen on a given crop will be independent of the type of visitor. It is possible, however, that some pollen is carried forward from one forage to the next. If at the same time one forage

of a visitor does not commence where its previous forage finished, conditions will be such that the shorter the forage the more wide-

TABLE I

The dispersal of insect flights

D=Flight length.

T=Frequency of flights according to length in two dimensions.

n=Frequency of flights according to length in one dimension.

The mean squares have been corrected for grouping.

The frequencies actually recorded are in *italic* type : derived frequencies are in ordinary type.

	Hive bees		Solitary bees		Hover flies		Hive bees		Solitary bees		Bumble bees
	1943		1943		1943		1944		1944		1945
<i>D</i>	<i>T</i>	<i>n</i>	<i>T</i>	<i>n</i>	<i>T</i>	<i>n</i>	<i>T</i>	<i>n</i>	<i>T</i>	<i>n</i>	<i>n</i>
0	<i>36</i>	144·00	<i>23</i>	92·00	<i>506</i>
1	<i>238</i>	238·00	<i>52</i>	52·00	<i>21</i>	21·00	<i>54</i>	54·00	<i>21</i>	21·00	...
2	<i>118</i>	59·00	<i>30</i>	15·00	<i>14</i>	7·00	<i>21</i>	10·50	<i>14</i>	7·00	<i>169</i>
3	<i>56</i>	18·67	<i>10</i>	3·33	<i>4</i>	1·33	<i>5</i>	1·67	<i>4</i>	1·33	...
4	<i>16</i>	4·00	<i>5</i>	1·25	<i>1</i>	0·25	<i>2</i>	0·50	<i>1</i>	0·25	<i>20</i>
5	<i>12</i>	2·40	<i>5</i>	1·00		0·00	<i>2</i>	0·40	<i>1</i>	0·20	...
6	<i>5</i>	0·83		0·00	<i>2</i>	0·33	<i>1</i>	0·17	<i>2</i>	0·33	<i>7</i>
7	<i>5</i>	0·71	<i>1</i>	0·14	<i>1</i>	0·14	<i>1</i>	0·14
8	<i>5</i>	0·63	<i>1</i>	0·13	<i>1</i>	0·13		0·00	<i>5</i>
9	<i>1</i>	0·11		0·00	<i>1</i>	0·11		0·00
10		0·00		0·00		0·00	<i>1</i>	0·10
11		0·00	<i>1</i>	0·09		0·00
12		0·00		0·00	<i>1</i>
13	<i>1</i>	0·08		0·00
14	<i>1</i>	0·07	<i>1</i>	0·07
15		0·00	<i>1</i>	0·07
16	<i>2</i>	0·13
.											
22	<i>1</i>
Total for all <i>D</i>	<i>123</i>	211·48	<i>66</i>	122·11	<i>709</i>
Total excluding <i>D</i> =0 . . .	<i>460</i>	324·63	<i>105</i>	72·94	<i>47</i>	30·43	<i>87</i>	67·48	<i>43</i>	30·11	...
True— Sum of squares Mean square	152·01 0·6355	81·85 0·5870	2196·0 2·9306
False— Sum of squares Mean square .	940·45 2·8137	213·04 2·8375	130·41 4·2023	152·01 2·1694	81·85 2·6351

spread will be the contamination. As noted above, the average duration of a forage varies greatly with the type of visitor. For example, under these conditions, hover flies would produce more widespread contamination than hive bees.

Theoretical

The frequencies in table 1 may be used to test the validity of various formulæ which might express the flight of foraging insects. It is first necessary, however, to decide which formulæ are sufficiently appropriate to warrant consideration. Various formulæ have been previously proposed for similar phenomena, some on empirical grounds, some on theoretical grounds and some, after being derived theoretically, have been tested against actual data.

An example of the purely empirical approach is that of Wadley and Wolfenbarger (1944) who observed the dispersal of the Smaller European Bark Beetle from a centre. They found that the regression equation $n = a + b_1 \log D + \frac{b_2}{D}$ gave a good fit with the data.

In contrast, is the theoretical method of Pearson and Blakeman (1906) who considered the distribution of mosquitoes round a point source after r flights, on the assumption that all flights were of equal length (l) though random in direction. They found that as r increased the distribution approached a normal distribution. When r was greater than 7 the theoretical formula would be $n = \frac{Q}{\pi r l^2} \cdot e^{-\frac{D^2}{r l^2}}$

where Q is the total number of insects.

The same relationship was assumed by Frampton, Linn and Hansing (1942) who were concerned with the spread of viruses of the yellows type by means of leaf-hoppers. They likened the movement of leaf-hoppers over a crop to the two dimensional diffusion of a gas according to kinetic theory. The distribution of leaf-hoppers round a point source would then be a normal distribution the variance of which increased with time. No direct test of the validity of this hypothesis was made, but they derived from it a formula for the spread of virus which showed good fit with their observations.

Brownlee (1911) offered various criticisms of Pearson and Blakeman's theoretical formula for insect flight. On the basis of these he proposed a formula of his own $n = ge^{-tD^x}$. This formula was tested against data for the distribution of water-fleas and winkles at intervals after their liberation from a point source, and also for the distribution of epidemics both in space and time (substituting t for D). All these divers phenomena appeared to agree with his formula.

From the above, it would seem that two basic formulæ merit serious consideration to decide their suitabilities for accounting for the above observations on the flight of foraging insects. These can be expressed in a common form, viz. $n = ge^{-tD^x}$. According to Pearson and Blakeman, and Frampton *et al.*, $x = 2$; according to Brownlee, $x = 1$. Other values of x might also be considered. On empirical grounds it was felt desirable to test the fit when $x = \frac{1}{2}$.

These formulæ can be tested against the data by the method of regression. This is done by taking logs throughout.

$$\begin{aligned} \text{Thus } \log_{10} n &= \log_{10} g - c \log_{10} e \cdot D^x \\ \text{or } \log n &= a + bD^x. \end{aligned}$$

We can now estimate the regression of $\log n$ on D^x for various values of x . The goodness of fit of several regressions can be compared by subtracting the sum of squares due to the regression from the total sum of squares for $\log n$. The smaller the remainder sum of squares (which have the same degrees of freedom) the better the fit of the regression. The ratios between the remainders are treated as variance ratios and their significance estimated by reference to the usual tables.

Fitting the data

Subject to a few necessary modifications of the data, it is now possible to proceed with the estimation of the regressions for each of the six sets of observations given in table 1. These modifications are required owing to the fact that when $n = 0$, $\log n = -\infty$, which

TABLE 2

The regression of n, the frequency distribution of insect flights, on D², D and D¹

The sums of squares are the remainder sums of squares after the estimation of the regression. The smaller the remainder sums of squares the better the fit of the regression. The variance ratios are the sum of squares for the regression divided by the sum of squares of the regression of $D^{\frac{1}{2}}$.

Regression of log n on	D ²		D		D ¹		N
	Sum of squares	Variance ratio	Sum of squares	Variance ratio	Sum of squares	Variance ratio	
Hive bees . 1943	1·475948	7·22	0·456561	2·23	0·204384	1·00	8
Solitary bees 1943	1·437808	4·35	0·556384	1·68	0·330863	1·00	7
Hover flies . 1943	2·551971	2·74	1·467272	1·58	0·930856	1·00	8
Hive bees . 1944	2·793632	7·28	0·744612	1·94	0·383599	1·00	7
Solitary bees 1944	2·484920	5·66	0·860717	1·96	0·438819	1·00	6
Bumble bees 1945	1·881934	7·63	0·494034	2·00	0·246597	1·00	6
Weighted mean variance ratio of all above series	...	5·734	...	1·895	...	1·00	42
	P. very small		P. 0·05				

is useless for the purposes of a regression. Only the smaller values of D , therefore, can be used. As will be seen from table 1 there are instances where $n = 0$ and adjacent values of n on both sides are positive. Arbitrary values have been given to n of 0·10 in columns 2 and 3, and 1 in column 6. The bias introduced cannot be serious

as all six sets of data gave similar regressions. These are given in table 2. A correction is required in the bumble bee data where all the frequencies except for $D = 0$ are the sum of two classes left and right. The frequency for $D = 0$ has therefore to be doubled before calculating the regressions.

In all six sets of data the regression of $\log n$ on $D^{\frac{1}{2}}$ gives the best fit. A graphical representation of the fit of the three regressions can be made by calculating the means of the observed and estimated values of $\log n$ for the first five sets of data over the range of D from 1.6 inclusive. (The bumble bee data do not give values of n over this range.) The means are then converted to their antilogarithms. The values are shown in table 3 and

TABLE 3
Comparison of observations with the expectations according to the regressions of $\log n$ on (a) D^2 , (b) D , (c) $D^{\frac{1}{2}}$

The values of $\log n$ shown are the mean values for all data shown in table 1 except the bumble bee. The values of n are the antilogs of the mean $\log n$'s. Expectation (c) is seen to be closest to the observed values of both $\log n$ and n . Discrepancies between expectation and observation at low values of D are exaggerated when n is calculated.

		D	1	2	3	4	5	6
Log n	Observed	.	1.694	1.132	0.453	1.839	1.657	1.437
	Expected	(a)	1.138	1.003	0.779	0.465	0.061	1.567
		(b)	1.464	1.061	0.658	0.255	1.852	1.450
		(c)	1.516	0.937	0.493	0.119	1.789	1.491
n	Observed	.	49.4	13.6	2.84	0.690	0.454	0.274
	Expected	(a)	13.7	10.07	6.01	2.92	1.15	0.369
		(b)	29.1	11.51	4.55	1.80	0.711	0.282
		(c)	32.8	8.65	3.11	1.32	0.615	0.310

graphically in fig. 3. Fig. 3*b* exaggerates the discrepancy between expectation and observation for $D = 1$ in regressions on D^2 and D . The regression on D^2 corresponds to a normal distribution as proposed by Frampton *et al.* and Pearson and Blakeman. The regression on D is according to Brownlee's formula. The magnitude of the peak at short distances appears to be even higher than suggested by Brownlee and is fitted still better by the regression on $D^{\frac{1}{2}}$.

The effect of several flights

In the bumble bee data each flight was recorded in strict sequence and the direction was also observed. This permits a further analysis which has bearing on contamination. If the flights were completely at random the variance of the distribution after r flights should be r times the variance after one flight. It has been shown above that

the bumble bee flights were not entirely random in direction. It is obviously desirable to know to what extent this non-randomness will affect the variance after r flights, which the bumble bee observations allow us to estimate. The data consisted of three portions each consisting of observations on a single bee during one forage. The first portion was very small. The second and third were of approxi-

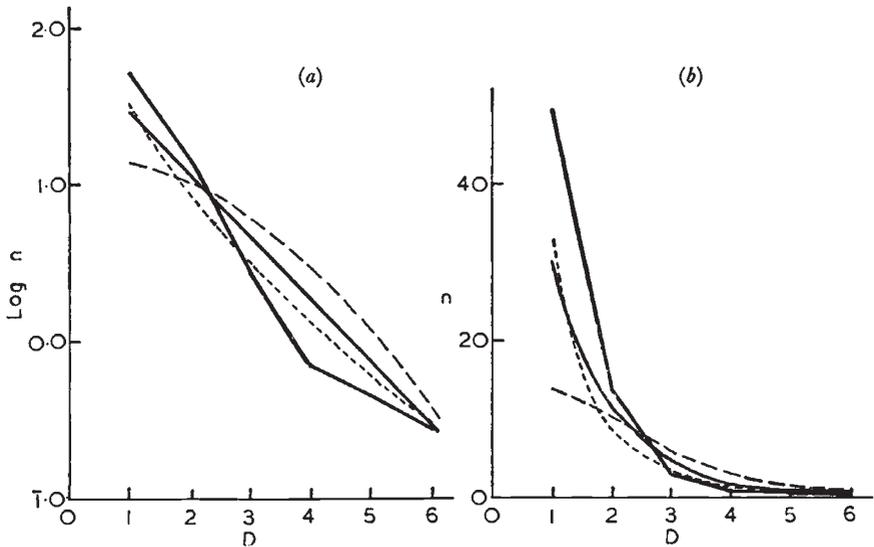


FIG. 3.—Fitting insect flight to a formula. (a) $\log n$, (b) n plotted against D , where n is the frequency distribution in one dimension. The thick black line represents the pooled data of table 1, except for the bumble bees. Other lines represent the expectations according to the regression of $\log n$ on D^2 (broken line), D (entire line) and $D^{\frac{1}{2}}$ (dotted line).

mately equal size comprising 295 and 397 flights respectively. It was evident that the degree of dispersal of the second bee was much greater than that of the third. The two portions were therefore considered separately for their distributions of n_r , given in table 4.

The mean square is calculated as $\frac{\sum n D^2}{\sum n}$ with Shephard's correction for grouping (-0.0833). The number of degrees of freedom is $\sum n$, as before.

The ratio $\frac{MS}{r}$ is fairly constant within each series, but showing a definite upward trend with increasing r . The correlation coefficient between MS and r is actually 0.9982 in the second series and 0.9984 in the third series. The tendency for $\frac{MS}{r}$ to increase with r is a consequence of the positive correlation already detected between the direction of successive directed flights. It will also be seen that as r increases n for $D = 0$ becomes smaller than n for $D = 1$. This would

not happen if flights were completely random. The general effect is, however, a relation between *MS* and *r* according to expectation on the assumption of randomness.

TABLE 4
Distribution (irrespective of direction) of bumble bees after r flights
 The mean squares have been corrected for grouping

<i>D</i> in feet		0	1	2	3	4	5	6	7	8	9	10	11	12	Mean square	<i>MS/r</i>		
2nd Series	<i>r</i>							1					1		1.3487	1.3487		
	1	195	73	12	7	5									2.8587	1.4294		
	2	135	102	28	12	11	2		1					1	4.7934	1.5978		
	3	93	114	40	17	15	6	1	2	1				1	6.6143	1.6536		
	4	73	112	43	25	17	9	3	1	3	2			3	8.8064	1.7613		
Correlation coefficient between mean square and <i>r</i> = 0.9982																		
<i>D</i> in feet		0	1	2	3	4	5											
3rd Series	<i>r</i>																	
	1	301	89	6													0.2021	0.2021
	2	233	144	18													0.4635	0.2318
	3	188	173	27	6											0.7670	0.2557	
	4	161	177	42	13									1.0923	0.2731			
5	138	175	60	17	2							1.4473	0.2895					
Correlation coefficient between mean square and <i>r</i> = 0.9984																		

This enables us to obtain from the expressions relating *n* and *D* a more general expression concerning the distribution of *n*. The usual expression for a normal distribution is

$$m = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The formula herein used for the normal distribution when *r* = 1 is

$$n = ge^{-cD^2}$$

As it has been shown above that in bee flight the variance, σ^2 , is proportional to *r*, it will be seen that there is the following correspondence between the symbols used in the two expressions.

$$\begin{aligned} (x-\mu) &: D \\ 2\sigma^2 &: \frac{r}{c} \\ \sigma &: \sqrt{\frac{r}{2c}} \\ g &: \sqrt{\frac{c}{\pi}} \end{aligned}$$

Thus to derive a formula for an n_r , which is normally distributed, we obtain

$$n_r = \frac{g}{r^{\frac{1}{2}}} e^{-\frac{cD^2}{r}}$$

Similarly, allowance for r can be made in the general formula

$$n_1 = g e^{-cD^2}$$

Then

$$n_r = \frac{g}{r^{\frac{1}{2}}} e^{-\frac{cD^2}{r}}$$

The effect of the factor $\frac{1}{r^{\frac{1}{2}}}$ is to compensate for the change in slope of the distribution with varying numbers of flights, thus keeping the total number of insects constant.

3. INSECT-POLLINATED CROPS

Theoretical

Formulae have been derived which appear to express adequately the distribution of foraging insects. It is now necessary to see how these can be applied to the distribution of contamination.

The general formulae for the distribution of insects after r flights from the source can be taken also to express the relative frequency with which insects visiting a plant at distance D from a contaminating plant have taken r flights to cover the distance. If the symbol for this frequency is N_r , then ΣN_r for all visits to a given plant will equal unity and it is necessary to adjust the factor g accordingly in the expression for n_r given above. If there are many contaminant plants there will be an increase in the magnitude of N_r for small values of r at the expense of the larger values of r . Thus g will be increased and c decreased in such a manner as to keep ΣN_r equal to unity.

At each visit an insect will deposit some pollen already on its body and replace it with a fresh supply. In this way the contaminant portion of pollen carried by an insect will decrease with every visit subsequent to its leaving the contaminant plot. If v is the proportion of contaminant pollen on an insect leaving the contaminant plot, we can consider it to decrease to a fraction w at every subsequent visit. The proportion of contaminant pollen on an insect r flights away from the plot will therefore be vw^{r-1} . In self-incompatible crops the effective r will be the number of flights up to the first visit to the plant now being visited, for successive visits to flowers on the same plant, though producing a decrease in the amount of compatible pollen carried, will not alter the proportion of that pollen which is contaminant. In self-compatible crops, on the other hand, r will be the total number of flights. It will include or exclude the last, according to whether the insect picks up pollen from a flower before or after touching the stigma.

The expressions for the frequency of flights and the proportion of contaminant pollen on an insect can be combined to give the following expression :—

$$p_r = \frac{vgw^{r-1}}{r^x} e^{-\frac{cD^x}{r}}$$

p_r is the contaminant portion of the pollen on a stigma deposited by insects which have travelled in r flights from the source of contamination.

The contaminant portion of all insect-deposited pollen will be Σp_r for all values of r from one to infinity. There is frequently, however, some pollen deposited on the stigma of the same flower without first being picked up on to an insect's body. The proportion of contamination in the seed, F , will therefore be $(1-q)\Sigma p_r$ where q is the proportion of compatible self-pollen deposited without the aid of insects. In self-incompatible species and many self-compatible ones with an out-breeding mechanism such as protandry, $q = 0$. In crops such as the French bean q is almost equal to unity. The expression for contamination then becomes, by removing all factors, independent of r to outside the Σ :

$$F = vg(1-q)\Sigma \left(\frac{w^{r-1}}{r^x} e^{-\frac{cD^x}{r}} \right)$$

$r = 1 \rightarrow \infty$

The part of this expression included within the Σ sign cannot be simplified, but for any given set of conditions the expression $vg(1-q)$ can be treated as a single constant. The manner in which F varies with D can be observed if we attribute arbitrary values to the constant, calculate the contributions to F by various values of r , and sum them.

Let $v = g = (1-q) = c = x = 1$
 Let $w = 0.5$

Then $F = \Sigma \frac{e^{-\frac{D}{r}}}{2^{r-1}r}$

Table 5 gives the values of p_r and $\log p_r$ for the range of r from 1 to 8 and of D from 0 to 25. From these are calculated Σp_r and $\log \Sigma p_r$ for r ranging from 1 to 8. The values of $\log p_r$ and $\log \Sigma p_r$ which is equivalent to $\log F$, are shown graphically in fig. 4.

It will be seen that though the relation between any $\log p_r$ and D gives a straight line, the line for $\log \Sigma p_r$ is a continuous curve becoming parallel successively to each $\log p_r$ curve.

So far all the formulæ have been obtained on the assumption that the distributions of insects and contaminant pollen can be expressed as continuous curves of infinite extent. In other words it

has been assumed that in practice there are no maximum values of r and D . Are these assumptions true, and if not do the derived formulæ lose their validity?

TABLE 5

Expected values for pollen distribution according to the formula $p_r = \frac{1}{2^{r-1r}} e^{-\frac{D}{r}}$ and its equivalent $\log p_r = -\left[(r-1) 0.3010 + \log r + \frac{D}{r} 0.4343\right]$

D		0	5	10	15	20	25
	r						
Log p_r	1	0	3.829	5.657	7.486	9.314	11.143
	2	1.398	2.312	3.227	4.141	5.055	7.969
	3	2.921	2.196	3.475	4.749	4.024	5.303
	4	2.495	3.952	3.409	4.866	4.324	5.781
	5	2.097	3.663	3.228	4.794	4.360	5.926
	6	3.717	3.356	4.992	4.631	4.271	5.906
	7	3.349	3.041	4.728	4.420	4.107	5.798
	8	4.990	4.716	4.447	4.173	5.904	5.631
p_r	1	1	0.0067	0.02005	0.03000	0.04000	0.05000
	2	0.2500	0.0205	0.02169	0.03198	0.04114	0.05093
	3	0.0834	0.0157	0.02298	0.03562	0.03106	0.04201
	4	0.0313	0.0090	0.02257	0.03735	0.03211	0.04603
	5	0.0125	0.0046	0.02169	0.03622	0.03229	0.04842
	6	0.0052	0.0023	0.02098	0.03428	0.03187	0.04805
	7	0.0022	0.0011	0.02053	0.03263	0.03128	0.04629
	8	0.0010	0.0005	0.02028	0.03149	0.03080	0.04427
	range of r up to						
Σp	2	1.250	0.0273	0.02173	0.03139	0.04114	0.05093
	4	1.366	0.0519	0.02728	0.02144	0.03328	0.04814
	6	1.383	0.0588	0.02995	0.02249	0.03743	0.03246
	8	1.386	0.0604	0.0108	0.02290	0.03951	0.03352
Log Σp	2	0.097	2.436	3.238	4.142	5.055	7.969
	4	0.135	2.715	3.862	3.157	4.515	5.911
	6	0.141	2.769	3.998	3.395	4.871	4.391
	8	0.142	2.781	2.032	3.462	4.978	4.546
	∞	0.15	2.80	2.05	3.53	3.07	4.65

The values for $\log \Sigma p$ for r up to infinity are obtained by visual extrapolation of graphs showing the effect of increases in r on $\log \Sigma p$. The index numbers to the 0's signify the number of times they are to be repeated.

It is possible that there is a maximum length to single flights, *i.e.* that there is a maximum range of D when $r = 1$. When, as in some wild species, plants are very widely spaced, this maximum is likely to limit the distance over which cross-pollination takes place. In a seed crop, however, the closeness of foraging is so great, that the frequency of such a maximum flight length being exceeded, if the distribution were continuous, is insignificant. Where plants are growing densely, therefore, the discrepancy between expectation and

fact due to the existence of a maximum flight length would be negligible.

The existence of a maximum value of r beyond which contaminant pollen would not be carried depends on whether pollen is carried forward from one forage to the next. If pollen is not carried forward then the maximum value of r will be set by the number of flights per forage. This is already known to vary according to the species of

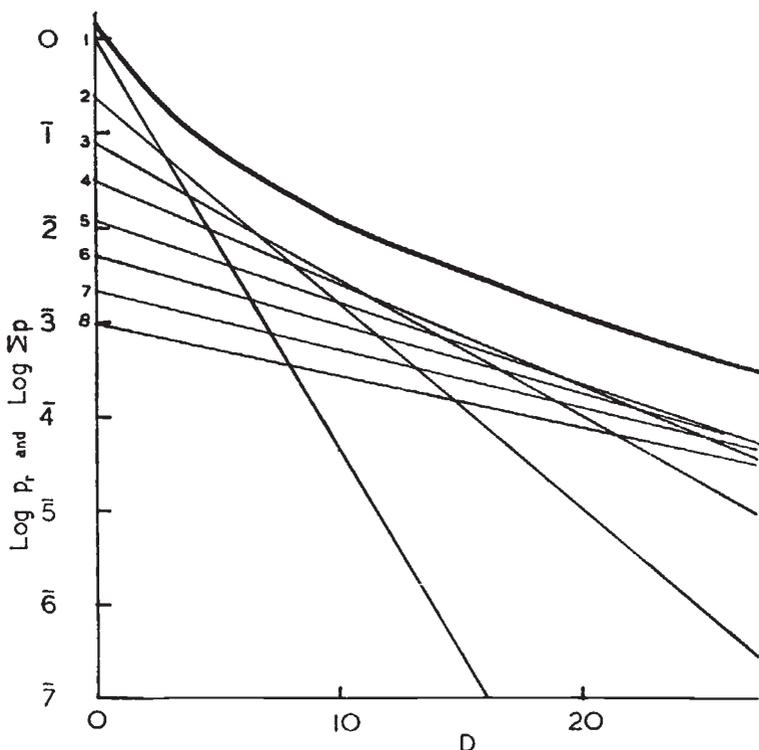


FIG. 4.—Variation of $\log p_r$ and $\log F$ with D according to the equation

$$p_r = \frac{1}{2^r - 1} e^{-\frac{D}{r}}$$

Curves for p_r are shown for values of r from 1 to 8. The $\log F$ curve (thick line) is that estimated as $\log \Sigma p_r$ for r ranging from 1 to ∞ .

insect and is probably affected by other factors such as the amount of nectar obtained per visit. The effect of a maximum r would be that the curve relating $\log F$ to D would become parallel to the p_r curve for the maximum r . If successive forages by an insect continue from where its previous ones finished, and there is a complete carry forward of pollen from one forage to the next there will be no maximum r . If there is an incomplete carry forward of pollen the contribution due to the higher values of r will be correspondingly

reduced. Even if there should be a maximum r it is likely to be relatively large for most pollinating insects.

A further possibility is that consecutive forages are incompletely correlated in position or even uncorrelated. This could only have an effect if pollen were carried forward from one forage to the next. The result would be a flattening of the curve relating $\log F$ and D for the higher values of D , the degree of flattening being increased by decrease in the spatial correlation between forages and increase in the amount of pollen carried forward. If there were no correlation between forages the $\log F$ curve would eventually become parallel to the D axis resulting in a minimum contamination. This, however, is an impossible extreme as there must be some correlation between forages if only because bees will remain within a certain distance of their hive. The presence of wandering insects (not foraging closely) as postulated by Butler *et al.* (1943) would also flatten the curve at greater values of D .

It has been implied by Butler (*ibid.*) that there is a further limitation of insect flight—that foraging tends to be restricted to a given area. This would involve a maximum D , equal to the diameter of the foraging area, irrespective of the number of flights, r . There does not appear to be any evidence of this phenomenon and it need not therefore be considered further.

There is a possibility of secondary contamination arising through insects picking up contaminant pollen from a contaminated flower. The effect would merely be to reduce the slope of the $\log F$ curve at all distances, and the appropriate formula would be the same.

Even though all the above considerations involve speculation about circumstances not yet understood, none necessitate any radical changes in the formula for F as originally proposed. This can be restated in a simpler form, using K as a constant for any particular pair of contaminant and contaminated plots.

$$F = K \sum_{r=1}^{\infty} \frac{w^{r-1}}{r^x} e^{-\frac{cD}{r}}$$

$$r = 1 \rightarrow \infty$$

The evidence of the insect flight data is that the most appropriate value for x is 1 or $\frac{1}{2}$.

Empirical

The next problem is to decide on an empirical formula which will closely parallel the effects of the above fundamental formula. As can be seen from fig. 4, when the theoretical $\log F$ is plotted against D a curve is obtained which deviates by upward curvature from a straight line, the deviation being the effect of successively higher values of r . Fig. 3 shows a very similar curve corresponding to the regression of $\log r$ on $D^{\frac{1}{2}}$ which is derived from the equation $n = ge^{-cD^{\frac{1}{2}}}$.

The corresponding expression for F is $y e^{-kD^{\frac{1}{2}}}$. A similar upward curvature would be a feature of the regression equation

$$\log F = a - bD - \log D. \quad \text{This corresponds to } F = \frac{y e^{-kD}}{D}.$$

These empirical formulæ can be tested for conformity with the theoretical formula in the following way. From the values in table 5 one can calculate the values of $\log \Sigma p$ for the following ranges of r : 1-2; 1-4; 1-6; and 1-8 for values of D from 5 to 25. ($D=0$ is an impossible value for actual cross-pollination and the corresponding value of $\log F$ is in another sense unreal, being positive, which would make $F > 1$.) It is then possible to construct graphs for each value of D showing the effect of increasing the range of r on $\log \Sigma p$. By extrapolation of these graphs we obtain the values of $\log \Sigma p$ to be expected when the range of r is complete (*i.e.* $1 \rightarrow \infty$). This method is obviously subject to personal error in the extrapolation. However, it is now possible to treat these estimates of $\log \Sigma p$ as "observed" $\log F$ and to test the goodness of fit of the two empirical equations by the method of regression.

For the equation, $F = y e^{-kD^{\frac{1}{2}}}$ the appropriate regression is that of $\log F$ on $D^{\frac{1}{2}}$, for the values of D from 5 to 25. The total sum of squares for $\log F$ is 2.8288 whilst the sum of squares due to the regression is 2.8279, leaving a remainder sum of squares for 3 degrees of freedom of only 0.0009. The fit between the fundamental and empirical equations is thus very good, though the extremely small remainder may be deemed a coincidence.

Similarly, for the equation $F = \frac{y e^{-kD}}{D}$ the appropriate regression is that of $(\log F + \log D)$ on D . The total sum of squares of $(\log F + \log D)$ is 1.2875, whilst the regression sum of squares is 1.2816, leaving a remainder of 0.0059 for 3 degrees of freedom. This still represents a good fit. The difference between the two remainders is not significant. It must be concluded that both empirical equations give a good fit with the fundamental equation and no choice can be made between them at this stage.

Comparison of these empirical formulæ with the original shows that the coefficient y corresponds to the part outside the Σ sign, namely $vg(1-q)$. These are respectively the proportion of contaminant pollen on an insect leaving the contaminant plot; a constant derived from the insect density formula which decreases with increasing dispersal of the insect flights; and the proportion of compatible pollen which is deposited by insects. The compound coefficient y expresses the contamination at zero isolation distance.

The coefficient k corresponds to those parts inside the Σ sign, namely w , c and x . These are respectively the diminution of contaminant pollen with each successive forage; the inverse of the degree of dispersal of insects (*i.e.* the closeness of foraging); and the

power of D which expresses most closely the effect of D on insect foraging. The degree of dispersal of the insects, $\frac{I}{c}$, is determined largely by the way in which the plants are arranged, whether densely or sparsely, for example. The compound coefficient k expresses the rate of decrease of contamination with distance.

Before proceeding further to test these formulæ by comparison with the data on contamination, attention should be paid to any formulæ hitherto proposed, which might have a bearing on the question. From their formula for the distribution of leaf-hoppers, Frampton *et al.* derived a formula for the spread of virus disease. The assumption was that the infection I was directly proportional to time t , and to the number of insects feeding. On this basis the following formula was obtained :—

$$I = yte^{k^2t - kD}$$

This situation must inevitably differ from cross-pollination, for an insect vector can continue to produce a new infection with each fresh bite, whereas in the process of contamination the pollinating insect steadily loses its ability to contaminate. The time factor is therefore of no importance (provided, of course, that the contaminant and contaminated varieties are synchronous in flowering period). If we are to apply the formula for virus infection to contamination by cross-pollination we must adjust the expression to $F = ye^{-kD}$.

This formula is supported by the experimental evidence of Currence and Jenkins (1942), who observed the effects of distance on contamination in a cross-pollinating variety of tomato. They found that a good fit with the results was obtained by using the regression equation $\log F = a + bD$ where b was negative. This is merely the logarithmic form of the above modification of Frampton's formula.

There appear, therefore, to be three equations which merit testing for their agreement with observation :

$$F = ye^{-kD}$$

$$F = ye^{-kD^2}$$

$$F = \frac{ye^{-kD}}{D}$$

Fitting the data

We are now in a position to test these formulæ. Fig. 5 shows $\log F$ plotted against D for a varied collection of contamination experiments described elsewhere (Crane and Mather, 1943 ; Bateman, 1947 *a* and *b*). As zero contamination gives an infinitely negative logarithm it has often been necessary to pool the data in groups of adjacent distances so as to give a finite logarithm. The mean D for the group is then used in the graph.

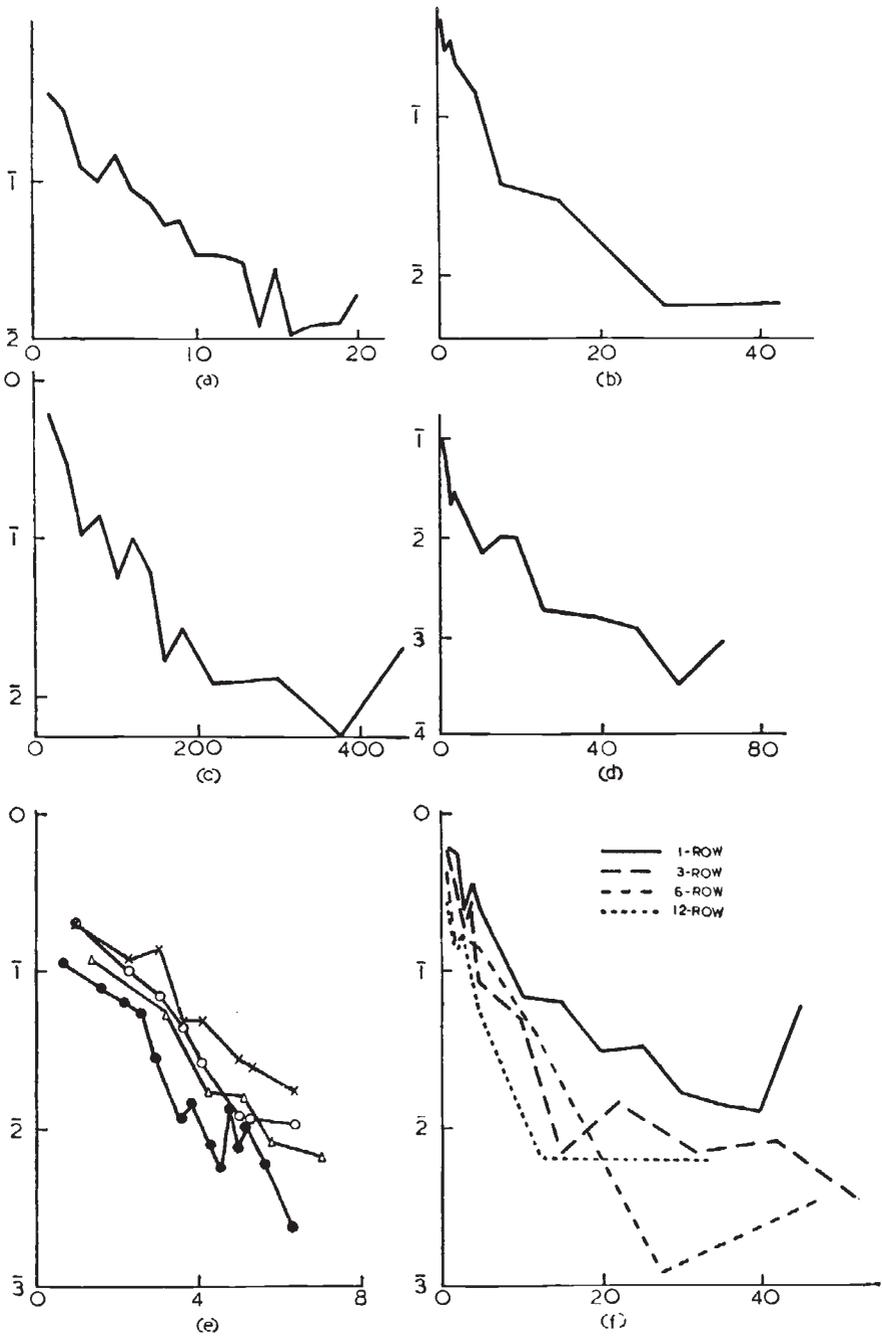


FIG. 5.—Contamination in insect-pollinated crops. Log F is plotted against D in feet. (a) Radish (Crane and Mather, 1943), (b) turnip, and (c), (e) and (f) radish (Bateman, 1947a), (d) beet (Bateman, 1947b).

It will be noticed that fig. 5 (*d*) concerns beet, which is generally regarded as wind-pollinated. It has been included because in the experiments it was found that the beet flowers were visited for the nectar discs by numerous flies. In large fields insect-pollination of beet is not likely to be important.

The suitabilities of the expressions ye^{-kD} , $ye^{-kD^{\frac{1}{2}}}$, and $\frac{ye^{-kD}}{D}$ can be compared by the method of regression. On taking logs these become $\log y - k \log e \cdot D$, $\log y - k \log e \cdot D^{\frac{1}{2}}$, and $\log y - k \log e \cdot D - \log D$ respectively. We can therefore estimate the following regressions:—

$$\begin{array}{l} \log F \text{ on } a - b_1 D \\ \log F \text{ on } a - b_1 D^{\frac{1}{2}} \\ \text{and } \log F \text{ on } a - b_1 D - b_2 \log D \end{array}$$

where a is an estimate of $\log y$, and b_1 is an estimate of $k \log e$. According to hypothesis the coefficient of $\log D$ in the third regression is unity. Furthermore a comparison of regressions can only be made if they are based on the same number of degrees of freedom, in this case, one. The third regression is therefore modified to that of $(\log F + \log D)$ on $a - bD$. The sum of squares due to each regression is subtracted from the total sum of squares of $\log F$ or $(\log F + \log D)$ and the smaller the remainder sum of squares the better the goodness of fit obtained by the regression.

The remainder sum of squares for $\log F$ is equivalent to the $\Sigma(\log F \text{ observed} - \log F \text{ estimated})^2$.

The remainder sum of squares for the regression of $(\log F + \log D)$ on D is therefore $\Sigma[(\log F \text{ observed} + \log D) - (\log F \text{ estimated} + \log D)]^2$. The $\log D$'s cancel out giving $\Sigma(\log F \text{ observed} - \log F \text{ estimated})^2$. The three remainder sums of squares are therefore strictly comparable. Table 6 shows the remainder sums of squares of the three regressions for all the data shown in fig. 4. There is never any significant difference between the second and third regressions. Except in series (*e*), which concerns distances of 7 feet and less, the second and third regressions have a distinct advantage over the first. It has been postulated above that the higher values of r (which produce the upward curvature of the log curve) become of increasing importance as D increases. It is not surprising, therefore, that the straightest curves (those fitted best by the regression of $\log F$ or D) are obtained for the shortest distances.

The data on the effects of distance on the spread of contamination are thus in conformity with two hypotheses:—

(1) That bees and other pollinating insects forage in a manner

which is described by the expression $\frac{ge^{-\frac{cD^x}{r}}}{r^{\frac{1}{x}}}$. A better fit is given

when $x = \frac{1}{2}$ than when $x = 1$.

(2) The amount of contaminant pollen carried by an insect decreases regularly with successive visits to flowers of the other variety.

TABLE 6

The remainder sums of squares for three types of regression for the contamination data represented in fig. 4

In all series the variance ratio is calculated, for the sake of an easier comparison, using the sum of squares remaining after the regression of $\log F$ on $D^{\frac{1}{2}}$ as the denominator. The variance ratios for (a)-(d), (e) as a whole and (f) as a whole are the weighted means of the variance ratios of the parts. The probability for the ratio of the "D" remainder to the "D $^{\frac{1}{2}}$ " remainder is given for these pooled ratios.

Regression	Degrees of freedom	Log F on D		(Log F + log D) on D		Log F on D $^{\frac{1}{2}}$	
		Sum of squares	Variance ratio	Sum of squares	Variance ratio	Sum of squares	Variance ratio
(a)	18	0.427978	1.41	0.361499	1.19	0.304014	1.00
(b)	7	0.567747	3.22	0.238632	1.35	0.176307	1.00
(c)	11	1.452668	1.64	0.699272	0.79	0.887747	1.00
(d)	12	1.166501	2.28	0.460055	0.90	0.511037	1.00
(a) to (d) .	48	P=0.01	1.944		1.049		1.000
(e) EA . . .	12	0.286398	0.82	0.375506	1.08	0.347283	1.00
(e) EB . . .	6	0.060359	0.74	0.101670	1.25	0.081334	1.00
(e) WB . . .	6	0.070510	1.03	0.043749	0.64	0.068648	1.00
(e) WC . . .	4	0.050497	1.22	0.043852	1.06	0.041465	1.00
(e) as a whole .	28	P large	0.905		1.019		1.000
(f) W	10	0.399517	2.89	0.158750	1.15	0.138109	1.00
(f) N	9	1.544541	1.97	0.516786	0.66	0.782717	1.00
(f) Si	6	1.272825	1.69	0.754156	1.00	0.753198	1.00
(f) Ei	4	0.794073	1.84	0.286966	0.66	0.432088	1.00
(f) as a whole .	29	P=0.01	2.211		0.899		1.000

In accordance with these hypotheses the relation between contamination F and distance D can be expressed empirically by either of the equations $F = ye^{-kD^{\frac{1}{2}}}$ or $F = \frac{ye^{-kD}}{D}$. When the distance is not more than a few feet the relation is equally well expressed by $F = ye^{-kD}$.

4. WIND-POLLINATED CROPS

(a) Pollen dispersal

Theoretical

A detailed discussion of the mechanics of the dispersal of air-borne fungal spores has recently been published (Gregory, 1945). Basing the arguments on formulæ proposed by Sutton (1932) for the dispersal

of a smoke cloud from a point source, a formula has been obtained for spore dispersal. The main feature of the discussion is the importance attached to atmospheric turbulence as a dispersing agent. The following formulæ are given :—

- (a) for deposition when the rate of deposition has a negligible effect on the amount of spores in suspension.

$$p = \frac{zQ}{\pi^{\frac{1}{2}}CD^{\frac{1}{2}(m+2)}}$$

Where p = the mean number of spores deposited per unit area in all directions from a point source.

z = the proportion of suspended spores deposited as a spore cloud passes over unit area.

Q = the total number of spores.

D = the distance from the point of liberation of the cloud.

C = an atmospheric constant.

m = a factor varying between 1.24 when turbulence is at a minimum and 1.76 when turbulence is at a maximum.

- (b) for the change of total suspended spores as deposition proceeds

$$Q_D = Q_o e^{\left[\frac{-2zD^{(1-\frac{1}{2}m)}}{\pi^{\frac{1}{2}}C^{(1-\frac{1}{2}m)}} \right]}$$

Combining the two expressions we obtain the complete formula for the dispersal of air-borne spores.

Modification of this formula are deduced to account for deposition down wind from a point source (p_w) and deposition down wind from a line source (p_{lw}).

These are
$$p_w = \frac{2zQ_D}{\pi C^2 D^m}$$

and
$$p_{lw} = \frac{2zQ_D}{\pi^{\frac{1}{2}}CD^{\frac{1}{2}m}}$$

where Q_D is derived from Q_o as above.

Doubt has been thrown by Bosanquet and Pearson (1936) on the practical importance of eddy diffusion or atmospheric turbulence. They assert that when there is a continuous source and observations are made over a period of time which allows for considerable variation in wind velocity, wind variation predominates over eddy diffusion in determining the resultant dispersal. If attention is paid to dispersal in all directions no account need be taken even of wind direction. Over a given period, then, during which Q pollen grains are liberated, ignoring loss through deposition, the grains will pass outwards in two dimensions under the influence of wind so that the number of grains passing over unit circumference at distance D from the source will be

$\frac{Q}{2\pi D}$. Allowance can be made for deposition by assuming that the amount of suspended grains is reduced by a certain proportion per unit distance travelled from the source.

Then $p = \frac{ke^{-cD}}{2\pi D}$ or in general form $\frac{ge^{-cD}}{D}$.

This formula is the same as is obtained from Gregory's expression by making m equal to zero. Taking logs this equation becomes of the form $\log p = a - b_1 D - \log D$. This means that when $\log p$ is plotted against D a line will be obtained curving upward from a straight line of negative slope.

Another formula for eddy diffusion in a stationary cloud has been proposed by Schmidt (1925). This is only the normal distribution in another form.

$$p = \frac{Q}{2\sqrt{\frac{A\pi t}{\rho}}} e^{\left(\frac{-D^2}{\frac{4At}{\rho}}\right)}$$

where p is the spore density at distance z from the source at time t , Q is the total number of spores, and A and ρ are atmospheric constants. It may be written as $p = ge^{-c_1 D^2}$.

If Schmidt's expression is used and allowance is also made for deposition rate

$$p = ge^{-c_1 D - c_2 D^2}$$

If $\log p$ is plotted against D , this equation will give a line curving downward from a straight line with negative slope ($\log p = a - b_1 D - b_2 D^2$). Reference to fig. 5 shows that where definite curvature from a straight line is discernible it is always in the upward direction. This equation may therefore be dismissed.

We are left then with the formulæ proposed by Gregory and the modification which ignores the effects of turbulence (assumes $m = 0$).

Gregory's formula can be simplified in the following manner :—

$$p = \frac{ge^{\frac{-cD^{(1-\frac{1}{2}m)}}{(1-\frac{1}{2}m)}}}{D^{(1+\frac{1}{2}m)}}$$

where g varies with the amount of pollen liberated and the rate of deposition, c varies with rate of deposition and m varies with the turbulence.

Fitting the data

If observations are available for a sufficient number of distances, so that the number of degrees of freedom will allow a comparison, it is possible to compare the adequacy of formulæ for p by the method

of regression. Gregory's formulæ can be made suitable in form for regressions by taking logs.

$$\text{Thus } \log p = \log g - (1 + \frac{1}{2}m) \log D - cD^{(1-\frac{1}{2}m)}.$$

$$\text{Hence } \log p + (1 + \frac{1}{2}m) \log D = a - bD^{(1-\frac{1}{2}m)}$$

Similarly

$$\log p_w + m \log D = a - bD^{(1-\frac{1}{2}m)}$$

$$\text{and } \log p_{tw} + \frac{m}{2} \log D = a - bD^{(1-\frac{1}{2}m)}$$

The maximal and minimal values for m (1.76 and 1.24 respectively) can be inserted and the regression of the left-hand side on the right-hand side calculated. The corresponding regression equations on the assumption that turbulence is without effect can be obtained by making $m = 0$.

These are

$$\log p + \log D = a - bD$$

$$\text{and } \log p_w = \log p_{tw} = a - bD$$

As has been shown earlier, the sum of squares of the deviations of observed ($\log p + \log D$) from expected ($\log p + \log D$) is the same as the sum of squares of the deviation of observed $\log p$ from expected $\log p$, as the $\log D$'s cancel out. The remainder sums of squares for any of these regressions applied to the same data are therefore strictly comparable.

Incidentally the method of regression makes possible the estimation of the values of $\frac{z}{C}$ and Q_o in Gregory's equation corresponding to any assumed value of m provided that all measurements are expressed in the same units.

$$\text{For } b = \frac{2z \log e}{\pi^{\frac{1}{2}} C (1 - \frac{1}{2}m)} \quad \therefore \quad \frac{z}{C} = \frac{b \pi^{\frac{1}{2}} (1 - \frac{1}{2}m)}{2 \log e}.$$

$$\text{Similarly } a = \log \left(\frac{z Q_o}{\pi^{\frac{3}{2}} C} \right) \quad \therefore \quad Q_o = \frac{\text{antilog } a \cdot 2 \pi^{\frac{3}{2}} \cdot \log e}{b \pi^{\frac{1}{2}} (1 - \frac{1}{2}m)}.$$

Material suitable for the estimation and comparison of regressions is available in the maize data from a previous paper in this series (Bateman, 1947*b*) as shown in fig. 6*a*, and the data of Jensen and Bøgh (1941) for rye (fig. 6*b*) and cocksfoot (fig. 6*c*). A great deal of data for fungal spores is presented by Gregory, but it is mostly unsuitable for our purpose as either the number of observations is too few or the distances are expressed too vaguely.

In the maize data p is the total deposition in three directions at right angles to one another over a period during which the wind varied considerably. The most appropriate regression should there-

fore relate to dispersal in all directions from a point source. The following are the estimated regressions :—

	Remainder sum of squares (5 degrees of freedom)	Variance ratio	Probability
Not down-wind, point source—			
$m=0 \log p + \log D = 3.870 - 0.23104 D$	0.121035	4.82	0.05
$m=1.24 \log p + 1.62 \log D = 4.662 - 0.85024 D^{0.38}$	0.028951	1.15	...
$m=1.76 \log p + 1.88 \log D = 6.573 - 2.72272 D^{0.12}$	0.025129	1.00	...

If m is given the values 1.24 or 1.76 a better fit is obtained than if eddy diffusion is ignored. The differences between the expectations of p and $\log p$ when m varies between 1.24 and 1.76 are very slight as can be seen in table 7. The apparent importance of turbulence is the more remarkable in that the actual data suggested otherwise for nearly all pollen deposited was on the slide facing towards the source rather than that facing away from it.

All the data presented by Jensen and Bøgh concern the down-wind distribution of pollen. It is not clear, however, whether the field was large compared to the distances at which pollen deposition was

Assumptions		Rye equations	Remainder sum of squares (5 degrees of freedom)	Variance ratio	Probability
Circumstances	m				
Down-wind from point or line source	0	$\log p = 1.723 - 0.10961 D$	0.029504	2.24	0.2
Not down-wind from point source	0	$\log p + \log D = 1.722 + 0.02293 D$	0.017378	1.32	High
Down-wind, line source	1.24	$\log p + 0.62 \log D = 1.860 - 0.15189 D^{0.38}$	0.013164	1.00	...
Down-wind, line source	1.76	$\log p + 0.88 \log D = 1.509 + 0.20775 D^{0.12}$	0.013817	1.05	High

recorded. Let us first consider the rye data as down-wind distribution from a point source (p_w). The left-hand side of the regression equation is then $\log p + m \log D$. A preliminary examination of the rye data showed that if the value of 1.24 or 1.76 was given to m in this regression, the left-hand side increased with D ; in other words the

coefficient b is positive, implying multiplication of the pollen after liberation, instead of diminution. If $m = 0$, however, the regression remains plausible. If, on the other hand, the distribution is considered as being down-wind from a line source at right angles to the wind, the left-hand side of the regression equation becomes $\log p + \frac{1}{2}m \log D$ and is plausible for values of m between 1.24 and 1.76. We may further query the constancy of wind direction during the experiment (strict p). The coefficient of $\log D$ then becomes $\frac{1}{2}(m+2)$ and the only value of m which then fits the data is zero.

The possible regression equations for the dispersal of rye pollen are shown on page 329.

It will be noticed that in two regressions the b coefficients are positive. When both coefficients are compared to their standard errors, however, they give non-significant t 's. This means that according to the particular regressions, loss of pollen through deposition is insignificant.

In the cocksfoot data also, on the same grounds there are four possible regression equations :—

Assumptions		Cocksfoot equations	Remainder sum of squares (7 degrees of freedom)	Variance ratio	Probability
Circumstances	m				
Down-wind from point or line source	0	$\text{Log } p = 1.582 - 0.12511 D$	0.145988	1.53	High
Not down-wind from point source	0	$\text{Log } p + \log D = 1.678 - 0.02212 D$	0.099631	1.04	...
Down-wind from line source	1.24	$\text{Log } p + 0.62 \log D = 2.044 - 0.40188 D^{0.38}$	0.095554	1.00	...
Down-wind from line source	1.76	$\text{Log } p + 0.88 \log D = 2.604 - 0.93394 D^{0.12}$	0.106150	1.11	...

It will be seen in table 7 that for both rye and cocksfoot the last three regressions give very similar expectations of p .

Summarising the pollen distributions in their relation to possible regressions the following statements can be made :—

- (a) With the data as a whole the best fit is obtained using regressions based on the assumption that eddy diffusion is the basis of pollen dispersal. Whether the atmospheric turbulence is assumed to be at a maximum ($m = 1.76$) or at a minimum ($m = 1.24$) the fit is equally good. The

corresponding values of z , the rate of deposition and Q_0 , the number of pollen grains liberated, vary greatly, however, according to the assumed value of m , as can be seen by the differences in the estimated values of a and b corresponding to the extreme values of m .

TABLE 7

Comparison of observation of log p and expectation according to several assumptions (see text)

D (arbitrary units)	1	2	3	4	5	6	7	8	10
Maize pollen									
Observations	3.865	3.031	2.533	2.185	2.083	1.748	1.477		
Expectations									
Not down-wind, point source $m=0$	3.639	3.107	2.700	2.344	2.016	1.706	1.408		
" " " " " $m=1.24$	3.812	3.068	2.599	2.247	1.962	1.722	1.513		
" " " " " $m=1.76$	3.850	3.050	2.570	2.226	1.953	1.734	1.546		
Rye pollen									
Observations	1.704	1.488	1.308	1.198	1.247	1.061	0.986		
Expectations									
Down-wind, point or line source $m=0$	1.613	1.504	1.394	1.285	1.175	1.065	0.956		
Not down-wind, point source $m=0$	1.745	1.467	1.313	1.211	1.137	1.081	1.037		
Down-wind, line source $m=1.24$	1.708	1.476	1.334	1.230	1.147	1.078	1.018		
" " " " " $m=1.76$	1.717	1.470	1.326	1.224	1.146	1.082	1.027		
Cocksfoot pollen									
Observations	1.538	1.425	1.263	1.080	0.644	0.740	0.700	0.617	0.473
Expectations									
Down wind, point or line source $m=0$	1.457	1.331	1.206	1.081	0.956	0.831	0.706	0.581	0.331
Not down-wind, point source $m=0$	1.656	1.333	1.135	0.988	0.868	0.767	0.678	0.598	0.457
Down-wind, line source $m=1.24$	1.642	1.334	1.138	0.991	0.870	0.768	0.678	0.598	0.460
" " " " " $m=1.76$	1.670	1.324	1.118	0.971	0.855	0.760	0.680	0.610	0.493

The observed values and those expected values which accord most closely with them are shown in clarendon type

- (b) In the two sets of data concerning down-wind distribution an equally good fit is obtained if turbulence is ignored and if the wind is assumed to be variable in direction.
- (c) Though in every instance the worst fit is given in comparable regressions when turbulence is ignored, the fit obtained might be good enough for many purposes.

Before a final decision can be reached on the most appropriate regression, which might indicate the most fundamental equation,

it will be necessary to obtain more accurate data over a sufficiently wide range of distances. At the present stage the evidence is in favour of Gregory's assumptions regarding the importance of atmospheric turbulence.

(b) Contamination

The derivation of F , the proportion of contamination, from p_o , the concentration of contaminant pollen, is much simpler than the equivalent derivation of F from n the frequency distribution of insects. We can assume that the concentration of non-contaminant pollen, p_i , is independent of the distance D from the contaminant.

Now $F = \frac{p_o}{p_o + p_i}$ which can also be expressed as $\frac{p_o}{p_i} = \frac{F}{1-F}$. If p_i is constant $\frac{p_o}{p_i}$ and $\frac{F}{1-F}$ will vary with p_o . We may therefore write $\frac{F}{1-F} = \frac{ye^{-kD}}{D^{(\tau+\frac{1}{2}m)}}$ for contamination in all directions from the source,

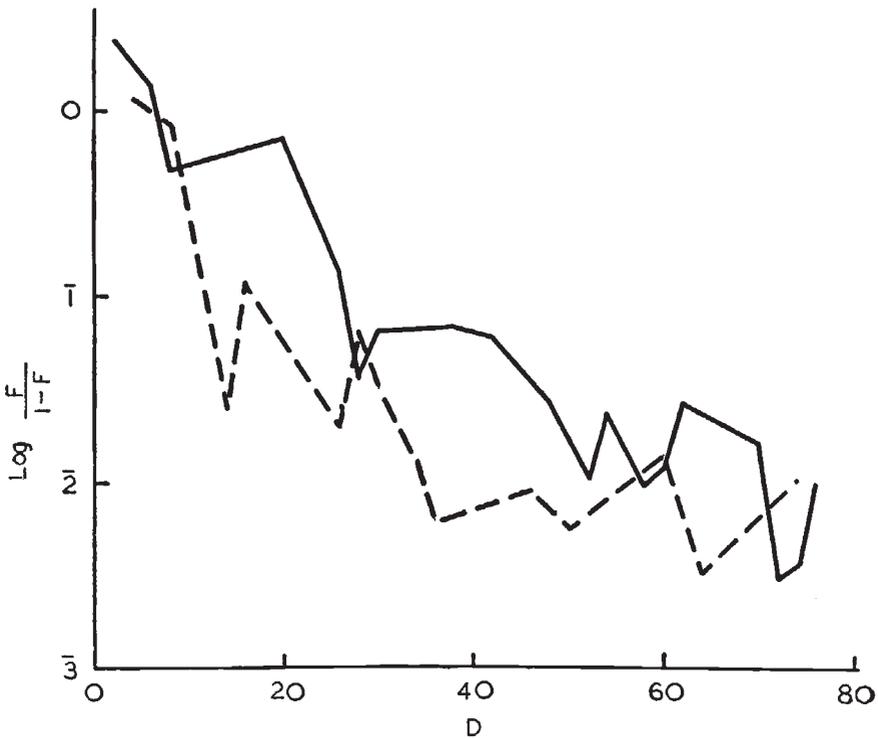


FIG. 7.—Contamination in maize. $\text{Log } \frac{F}{1-F}$ plotted against D in feet. Data in Bateman (1947b). The entire and broken lines represent replications.

where in the simplest case $m = 0$. There is no need to make any detailed examination of the contamination results in the two wind-

pollinated crops beet and maize. Since the error variance of the contamination is much higher than that of pollen distribution a comparison between various regressions will be much less sensitive. It has been seen (fig. 5*d*) that the beet curve for $\log F$ against D shows upward curvature. Since F is small the curve for $\log \frac{F}{1-F}$ would be very similar. The results are thus compatible with either insect or wind pollination in this crop.

Fig. 7 shows $\log \frac{F}{1-F}$ plotted against D for two sets of maize data from the North and South arms of the same experiment. Only those ears which silked before 11th August (which proved to be the data at which contamination almost ceased) are represented. The curves are very erratic in comparison with curves for pollen concentrations, because whilst in the former each ear represented a sample of pollen taken at its own time of silking and with local variation in amounts of non-contaminant pollen, in the latter all samples were taken simultaneously with only one source of pollen. The erratic nature of the results makes the calculation of a regression superfluous. By appearance the N curve is almost straight, whereas the S curve shows slight upward curvature. The results are at any rate not incompatible with the expectations based on pollen studies.

CONCLUSIONS

It is now possible to compare the formulæ derived for insect- and wind-pollination. In insect-pollinated crops two formulæ appear equally suitable :

$$F = ye^{-kD^{\frac{1}{2}}} \quad \text{and} \quad F = \frac{ye^{-kD}}{D}$$

In wind-pollinated crops the corresponding formula is

$$\frac{F}{1-F} = \frac{ye^{-kD}}{D}$$

In spite of the very different ways in which these formulæ have been derived, they all give similar results in practice. If F is small so that $F \simeq \frac{F}{1-F}$, one formula can be used in all cases, viz. :

$$F = \frac{ye^{-kD}}{D}$$

If this formula is expressed as a graph in which $\log F$ is plotted against D (see fig. 4) a curve of negative slope is obtained, its steepness decreasing as D increases. If the curve relating $\log F$ to D were a *straight line* of negative slope the *proportionate* decrease in contamination with unit increase in distance would be constant. The proposed

formula means, therefore, that even the proportionate decrease in contamination becomes less with each increase in distance. In practical terms successive increases in isolation distance become less and less effective in improving the isolation. If, for example, contamination is complete for no isolation distance, and 100 yards reduces it to one in a hundred, the contamination at 200 yards will be greater than one in ten thousand.

The practical importance of the proposed formula to seed growing would be in the prediction of the contamination at any isolation distance provided the contamination at two distances were known with sufficient accuracy and that other conditions were constant. Under given conditions it would be possible to predict the distance at which contamination of one in a hundred or one in a thousand would be obtained.

The formula also has a bearing on population genetics. It gives the distribution in a continuous population of the pollen parents of the progeny of an individual seed parent. The nature of this distribution suggests that if the seed were not widely dispersed, there would be a correlation between the proximity of two plants and the closeness of their relationship. Even in an infinite and continuous population, therefore, gene combinations would spread slowly and there would be slight inbreeding. In a population of finite size more genetic variability and greater inbreeding would occur than would be expected on the basis of population size alone. In the past, the occurrence of greater variability than was anticipated from the population size has been explained by the assumption of an internal discontinuity in the population (Sewall Wright, 1939). A limited dispersal of pollen or seed would effectively explain the same situation.

It is worth noting that in a seed crop which is harvested in bulk, mixed mechanically and resown, there is no correlation between the proximity of plants in the field and their relationship. Crop plants are therefore probably the only examples in which the total number of plants in a population really corresponds to Sewall Wright's N statistic.

SUMMARY

Insects and air produce pollination by very different means which can be contrasted as follows :—

1. Insects move independently of one another, but the experiments show that they do so in a statistically predictable manner. Air, on the other hand, moves in large masses broken up by turbulence, the effects of which are also predictable; but the movement is modified by variation in wind direction and velocity which is unpredictable.

2. Insects carry pollen systematically from flower to flower of the same species: under normal wind conditions, if the species is evenly dispersed, the pollen is distributed equally in all directions. Air-borne

pollen, on the other hand, is distributed down-wind and alights on a stigma without regard to its species.

3. An insect can carry only a limited amount of pollen available for pollination. Consequently the amount of pollen of one variety on an insect can increase only at the expense of other varieties. In the air, on the other hand, the amount is almost unlimited. The atmospheric concentration of pollen of one variety, therefore, has no direct influence on the concentration of another.

Bearing in mind these distinctions, formulæ are derived for the effect of distance on contamination in insect- and wind-pollinated crops. These agree well with the experimental results now described. The formulæ for the two classes of pollination are unexpectedly similar; so similar indeed that one can derive a common formula for the two.

With this formula one can use the contamination observed at two distances to predict what will be found at a third. The formula also throws some light on the breeding behaviour of natural populations.

Acknowledgments.—This work has been carried out under the auspices and financial assistance of the Agricultural Research Council. The author also wishes to express his gratitude to Dr K. Mather for his guidance during the work.

REFERENCES

- BATEMAN, A. J. 1947*a*.
Contamination in Seed Crops. I. Insect Pollination.
J. Genet. 48, 257.
- BATEMAN, A. J. 1947*b*.
Contamination in Seed Crops. II. Wind Pollination.
Heredity 1, 235.
- BOSANQUET, C. H., and PEARSON, J. L. 1936.
The spread of smoke and gases from chimneys.
Trans. Faraday Soc. 32, 1249.
- BROWNLEE, J. 1911.
The mathematical theory of random migration and epidemic distribution.
Proc. roy. Soc. Edinb. 31, 262.
- CRANE, M. B., and MATHER, K. 1943.
The natural cross-pollination of crop plants with particular reference to the radish.
Ann. appl. Biol. 30, 301.
- CURENCE, T. M., and JENKINS, J. M. 1942.
Natural crossing in tomatoes in relation to distance and direction.
Proc. Amer. Soc. hort. Sci. 41, 273.
- FRAMPTON, V. L., LINN, M. B., and HANSING, E. D. 1942.
The spread of virus diseases of the yellow type under field conditions.
Phytopathology 32, 799.
- GREGORY, P. H. 1945.
The dispersion of air-borne spores. *Trans. Brit. mycol. Soc.* 28, 26.

JENSEN, I., and BØGH, H. 1941.

On conditions influencing the danger of crossing in the case of wind-pollinated cultivated plants.

Tidsskr. Planteavl. 46, 238.

PEARSON, K., and BLAKEMAN, J. 1906.

Mathematical contributions to the theory of evolution. XV. A mathematical theory of random migration.

Drap. Co. Mem. biom. Ser. 3, 1.

SCHMIDT, W. 1925.

Der Massenaustausch in freier Luft und verwandte Erscheinungen.

Probl. kosm. Phys. 7, 1.

SUTTON, O. G. 1932.

A theory of eddy diffusion in the atmosphere.

Proc. roy. Soc. A, 135, 143.

WADLEY, F. M., and WOLFENBARGER, D. O. 1944.

Regression of insect density on distance from centre of dispersion as shown by a study of the Smaller European Bark Beetle.

J. agric. Res. 69, 299.

WRIGHT, S. 1939.

The distribution of self-sterility alleles in populations.

Genetics 24, 538.