Comput. Biol. Med. Pergamon Press 1971. Vol. 1. pp. 243–253. Printed in Great Britain.

The Correlation of Polluted Air with Tree Growth and Lung Disease in Humans

JAMES A. HEIMBACH, JR.
Department of Meteorology, The University of Oklahoma,
Norman, Oklahoma, U.S.A.

(*Received* 9 *July* 1970 *and in revised form* 23 *November* 1970)

Abstract—In Altoona, Pennsylvania, a study was done to correlate the incidence of lung cancer and pulmonary emphysema to air pollution. Spatial distributions of relative sulfur dioxide concentrations were calculated for Altoona through a pollution model based on the Pasquill–Gifford equation of diffusion using the city's climatological and source distributions. A statistical analysis was done comparing the incidence of lung disease with concentrations of sulfur dioxide. A positive correlation was indicated. Lateral growth rates of trees throughout the city were analyzed to compare with the concentration patterns obtained from the pollution model. Areas showing past stunting of tree growth coincided reasonably with those areas shown numerically to be polluted.

1. INTRODUCTION

THE NASHVILLE air pollution study,[1] run from 1957 to 1959, showed a relation between air pollution and bronchial asthma, anthracosis, mortality and morbidity. A study done by GILBERT[2] correlated the lack of certain types of lichens to air pollution in the Tyne Valley in England. THOMAS[3] listed some of the effects of specific types of air pollution to plant as well as animal health. To further investigate some possible harm that air pollution could have on Man, a correlation between lung disease and air pollution was sought. This endeavor was broken into three parts: investigating all cases of typical lung diseases based on hospital records over a period of one year, comparing these patterns with spatial distributions of relative sulfur dioxide concentrations predicted by a pollution model, and checking the results of the modelling by analyzing tree growth throughout a city.

It was thought best to use a small city which had at one time a severe pollution problem. The city was to be small enough to make a complete survey of lung diseases possible, as well as a complete tree growth analysis, but still large enough to give a meaningful trend. A changing pollution profile was sought to provide differing growth rates in trees, as will be explained later. Altoona, Pennsylvania, was chosen.

2. THE CITY

Altoona is about 100 miles to the east of Pittsburgh in the northern part of the Appalachian Mountains. The city was established about 1850 as a western maintenance center for the Pennsylvania Railroad (PRR), and grew into the main maintenance depot where all heavy repairs were done as well as the complete production of steam locomotives.

On either side of Logan Valley, where Altoona is located, are hills up to 1200 ft higher than the valley floor.[4] These southwest to northeast oriented hills would tend to influence the winds to prevail from the southwest, as verified when analyzing old railroad weather data. Because of the nature of the industry, there was heavy air pollution in Altoona, the peak being around 1945. Between 1950–1960, however, air pollution from coal burning has decreased to almost none because of the diesel engine replacing the steam locomotive and the fact that the railroad industry in Altoona had been drastically cut back.

3. DATA COLLECTION

Pollution sources

All the sources considered belonged to the Pennsylvania Railroad. There were three categories of sources: power generating stacks; line sources, i.e. railroad tracks; and area sources made up of switching yards and maintenance shops. During the era of the steam engine, coal burning accounted for practically all of the combustion products. Interviews with railroad men disclosed that anthracite coal was used and the sulfur content averaged from two to four per cent.

There were seven power stacks in Altoona which were considered (see Fig. 2). A main line of the PRR went through Altoona from Harrisburg to Pittsburgh with a small branch going to Hollidaysburg. Shops and switching yards were located in the northeast and central portions of the city.

Many records of the amount of coal burned by the various sources were destroyed by summer, 1969, when this investigation was begun. Estimates of the amounts burned by those sources for which there were no records available, were obtained from the memories of the older railroad workers. Interviews with railroad veterans also gave estimates of the plume rise. Plume rise is the lifting of the effluent above the source due to buoyancy and upward velocity. This added height varied from 15 per cent of the source height for power stacks to 50 per cent for the steam locomotives for normal weather conditions.

Tree data

Tree cores were taken throughout Altoona. An attempt was made to sample only maple trees because it is believed by some[5] that maples are more sensitive to sulfur dioxide pollution. In many areas, however, there were no maples to be found. In some areas, where there was previously very heavy pollution, the only older trees available were the *Ailanthus*, or there were no trees older than 1955 to be found.

Disease data

The types of lung disease used in the study were lung cancer and pulmonary emphysema. These were investigated because they can be a result of continued irritation.[6] The two public hospitals, the Altoona and Mercy Hospitals, had their records checked for all lung cancer and emphysema cases which were entered during the year 1965. This year was settled upon for several reasons. First, better records are available for recent years because of increased management efficiency at both hospitals. Second, medical science has developed more accurate diagnoses of lung diseases, and third, the smoking habits of the patients must be considered. Prior to 1960, little concern over smoking was evident, and even in 1965, only 70 per cent of the cases considered had anything entered on their hospital records about smoking.

The residence where each patient was living at the time of diagnosis was recorded and later converted to a grid address. This address was assumed to be where the patient lived from at least 1945, a safe assumption because the older Altoonians are not very nomadic. Only four of 171 emphysema and lung cancer cases were under 45—with ages 6, 26, 40 and 44—having emphysema. Also recorded were sex, extent of ailment, age of patient, smoking habits, source of data, income, place of employment, season of admission, and population density where the patient lived. Only those cases which resided within the city limits were considered.

4. COMPUTER ANALYSIS

The analysis done by computers was in three parts: sorting of disease and tree growth data, numerical air pollution modelling, and application of the results of this modelling to disease data for correlation purposes. Most of the computation was done on the CDC 6600 located at NCAR, Boulder, Colorado. The data organization was an elementary exercise of classifying and plotting grid locations of incidence in the case of disease data as shown in Figs. 5–7. With the tree data, this involved finding comparative growth rates and plotting these on a printed representation of the city (see Fig. 4). For this portion, as well as for the rest of the study, the city was sectioned into a 53 by 108 grid.

As there were no air quality control data for Altoona, a theoretical method of estimating air pollution values had to be used. The Pasquill–Gifford diffusion model, as explained in the next section, was applied to individual sources within Altoona. Each area of the 53 by 108 grid had the contributions from each source of three source and weather configurations summed. The model was run for an up valley wind using all sources mentioned in the previous section. Also, average relative pollution concentrations for two fields were estimated using a yearly sixteen point wind climatology with all sources and with only the seven main power stacks. In all three pollution fields the point source Pasquill–Gifford model was applied to a multi-source situation to obtain a dense grid of values.

The fields of pollution computed using the yearly climatology were applied to correlate the incidence of emphysema and lung cancer to relative values of air pollution. Because pollution values and disease locations were referenced to the same grid, statistical analyses simply involved searching for the respective pollution value of the address of each disease case. Later in the study, when a map giving population densities of Altoona was obtained, an analysis was done to check for randomness of disease incidence with respect to population.

5. POLLUTION MODELLING

The working equation used for computing pollution values is a special form of the Pasquill–Gifford diffusion treatment where pollution concentrations are found at the surface.[7]

$$\chi[x,y,0;H] = \frac{Q}{\pi\sigma_y\sigma_z u} \exp[-y^2/2\sigma_y^2 - H^2/2\sigma_z^2]. \tag{1}$$

In this approach, air pollution concentration is assumed to be bi-normally distributed with respect to vertical, z, and across wind, y, directions. The density of a normal distribution as a function of y only is denoted by $f(y)$

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} \exp[-y^2/2\sigma_y^2].$$

The density as a function of z only is $g(z)$

$$g(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z} \exp[-z^2/2\sigma_z^2].$$

The standard deviations of these distributions are σ_y, σ_z. The point source is assumed to be continuously emitting, i.e. steady state conditions are assumed, making stretching in the downwind direction a function of the inverse of average wind velocity, $1/u$. Multiplying these three contributions of diffusion together, and allowing for the plume to be reflected at the earth's surface from a source at an effective height H, gives:

$$\text{Joint density at surface} = \frac{1}{\pi\sigma_y\sigma_z u} \exp\left[-y^2/2\sigma_y{}^2 - H^2/2\sigma_z{}^2\right].$$

The concentration of pollution, χ, at a point $x, y, z = 0$; due to a source with effective height H, and emission rate Q, is $\chi[x,y,0;H]$. This is Q times joint density, giving equation (1).

The σ's are a function of downwind distance, x. As the plume travels downwind it widens vertically and horizontally having the σ's increase as x increases. During unstable atmospheric conditions, the increased turbulence would have the σ's increasing at a faster rate than for stable conditions. Graphed values of standard deviation as a function of downwind distance have been found empirically for six stability conditions.[7,8] These graphs were approximated by exponential functions which were used to give values of σ_y and σ_z for any x when the model was applied.

Although equation (1) can be applied to line and area sources with some reworking, it was more economical, in terms of computer time, to approximate the line and area sources by a series of point sources. Due to the roughness of the source data, and the uncertainty of mapping long-term averages of pollution, relative densities of pollution were mapped. These had the same order of magnitude as one could expect of actual long term averages.

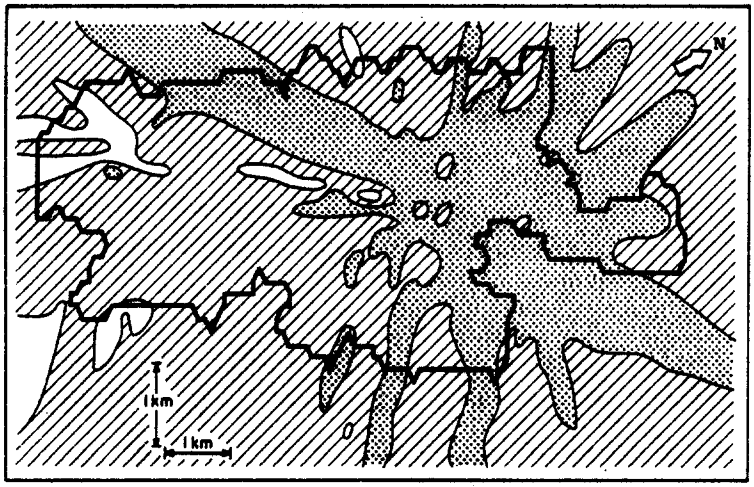

FIG. 1. A yearly average pollution field computed using the seven main power stacks of the Pennsylvania Railroad with the yearly wind climatology. The dotted areas represent values greater than 10^{-4} g/m^3, the hatched areas, greater than 10^{-5} g/m^3, and the other areas represent values of less than 10^{-5} g/m^3. This shading scheme is used for the other pollution figures. On this as well as all other figures, the map is stretched slightly along the northwest to southeast direction. This was done because output from line printing required some stretching in this direction to preserve resolution.

Three results of the modelling are presented. The first, Fig. 1, used only the seven power stacks as sources with winds and their directions averaged from five years of weather records. Each wind direction contributed to the yearly average pollution value with a weight equal to the amount of time the wind blew from that direction. The second, Fig. 2, used all pollution sources with the same climatology, and the third, Fig. 3, used all the sources with the prevailing up-valley wind only. The wind climatology has sixteen compass points and for each direction an average atmospheric stability was hypothesized. For example, winds up the valley (from the southwest) were associated with neutral stabilities. There are some similarities among the three results worth pointing out. All show heavy pollution values to the northeastern portions of the figures. These areas are known locally as Juniata, East Altoona, and Greenwood. Also, all figures show a heavy pollution ribbon from the center of the city to the west-southwest. The narrow bands of Figs. 1 and 2, which radiate from the center, are attributable to the sixteen discrete directions of the wind climatology.

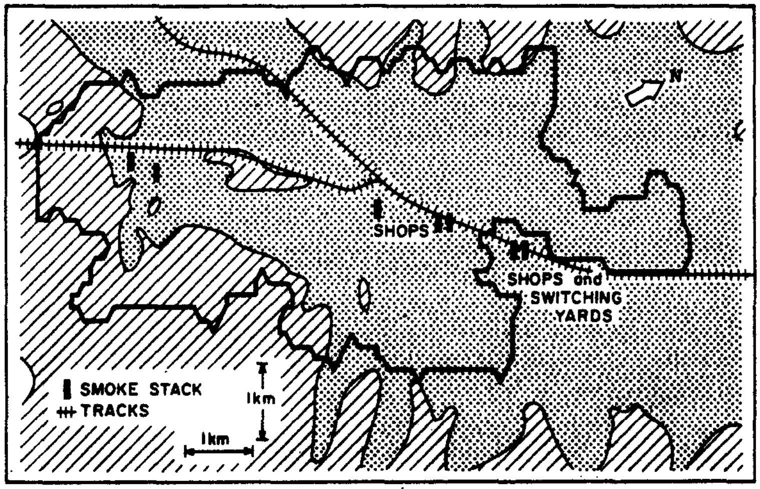

FIG. 2. The yearly average pollution field computed using all railroad sources with the yearly wind climatology. Shown also in this figure are the positions of the stacks, area sources, and line sources used in the study.

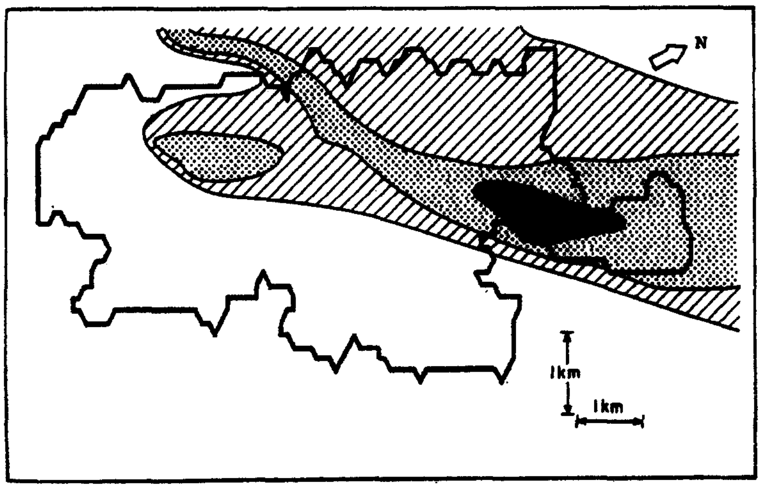

FIG. 3. The pollution field obtained using all railroad sources with prevailing upvalley wind. The black area represents relative values greater than 10^{-8} g/m^{3}.

6. TREE ANALYSIS

There are so many variables to consider when synthesizing a pollution field by a computer that some type of alternate method should be used to obtain a pollution profile. A combination of the two methods would make possible a better idea of the actual situation. As pollution measurements by instrument were nonexistent in Altoona, some type of indicator had to be used which was present in the city since 1945. The indicator used was the growth rate of trees.

It is believed by some [9] that air pollution stunts tree growth. If this is so, an area which had heavy air pollution in the past would have trees growing faster in the years of suppressed air pollution. There are, of course, other factors involved with tree growth such as rainfall, temperature, disease, damage by children, crowding by other trees, paving close to the tree, and the age of the tree. By measuring the amount of growth over several years and comparing this growth to an equal span of time, the short term effects of rainfall could be filtered out. If the other factors mentioned above, excepting age, were assumed constant and if no exceptionally old trees were analyzed, then the differing growth rates could be attributed to differing pollution. What was looked for was a rate of change of size with respect to time:

$$\frac{\delta\,(\text{size})}{\delta t}.$$

A gradient of this change of size with respect to time, or $\frac{\delta}{\delta x}\,\frac{\delta\,(\text{size})}{\delta t}$ would be meaningless because there were several types of trees considered under many growing conditions. To get around this, each tree was used as its own control and the following ratio called "FACTOR" was defined.

$$\text{FACTOR} = \frac{\text{Growth of tree in early period}}{\text{Growth of tree in later period}}$$

$$= \left[\frac{\delta\,(\text{size})}{\delta t}\right]_{\text{period } A} \Big/ \left[\frac{\delta\,(\text{size})}{\delta t}\right]_{\text{period } B}. \tag{2}$$

It is the gradient of this "factor" which provides an idea of the past pollution profile. If "factor" gives a result equal to 1, then, either there would be no pollution during the two periods, or the pollution rate has remained constant. If the result is less than 1, then there would be more pollution during the early period than the later. A "factor" greater than 1 would imply more pollution during the later period or something else which would cause the tree to grow slower in the later period. Significantly less rain for several years, damage by children, or insect damage during the later period are some examples of this.

Altoona had heavy air pollution before 1950. Between 1950–1960, the pollution decreased to a fraction of what it was, remaining at this low level through 1969. The "factors", as defined in equation (2), should be less than unity over areas of the city which were polluted, particularly when comparing the growth of 1945–1955 to later years.

Figure 4 shows the contours for the factors derived by comparing 1945–1955's growth to 1955–1965's growth. Areäs of large stunting occurred to the northeast of the city and along the Altoona to Pittsburgh railroad tracks. A patch of heavy stunting is to the southwest of the city, probably attributable to the two power stacks in that area. Areas of light and heavy stunting correspond closely to those areas shown by the theoretical method to be polluted.

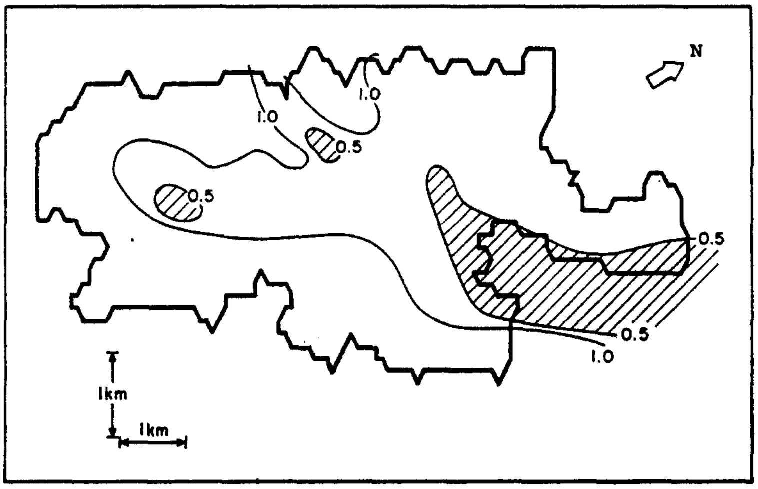

FIG. 4. A comparison of the 1945–1955 lateral growth of trees to 1955–1965 growth. Areas having no trees older than 1945 were included in the very stunted areas (hatched). Areas having stunted tree growth prior to 1955 correspond with areas shown numerically to be polluted.

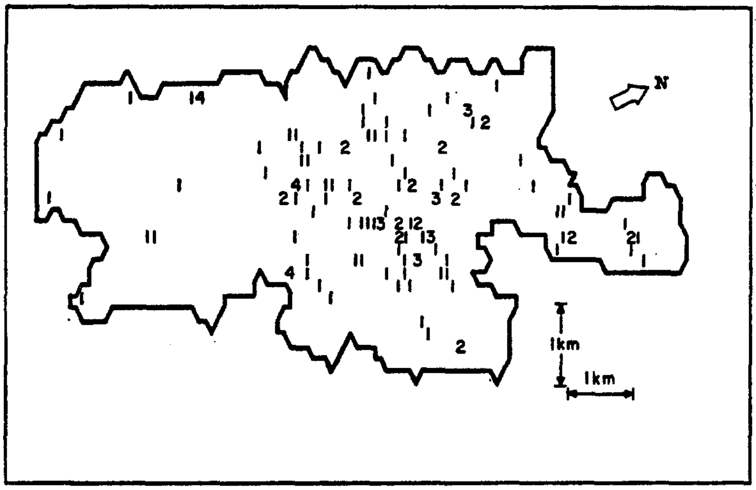

FIG. 5. The locations of all emphysema cases living within Altoona during 1965. All cases were diagnosed in the two local hospitals. Each integer represents the number of cases residing within each grid area.

7. DISEASE ANALYSIS

All locations of cases of lung cancer and emphysema were plotted on an outline of Altoona. (See Fig. 5 for emphysema). Both plots showed a grouping of cases near the center of the city and a lack of cases to the southwest. Some of this grouping was due to a difference of population density, but, as shown later, there was some other factor involved in this grouping. The definitely nonsmoking lung cancer (Fig. 6) and emphysema (Fig. 7) are also plotted. The plot of definitely nonsmoking lung cancer cases shows an interesting grouping around the Altoona to Pittsburgh railroad tracks. There was only one case not in this area and it turned out that this was the only nonsmoking lung cancer patient who did not work for the railroad.

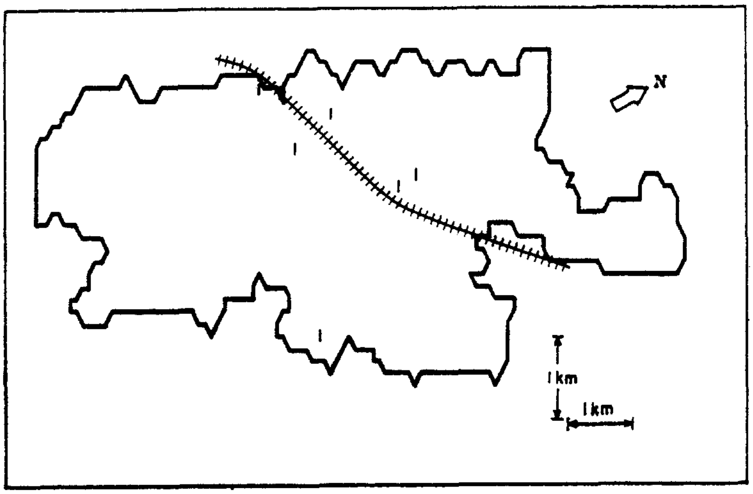

FIG. 6. The locations of definitely nonsmoking lung cancer patients reported in 1965. A section of the Altoona to Pittsburgh railroad track was drawn to show the grouping of five out of the six cases around these tracks. The case to the southeast, not in this grouping, was the only nonsmoking lung cancer patient who was not a railroad employee.

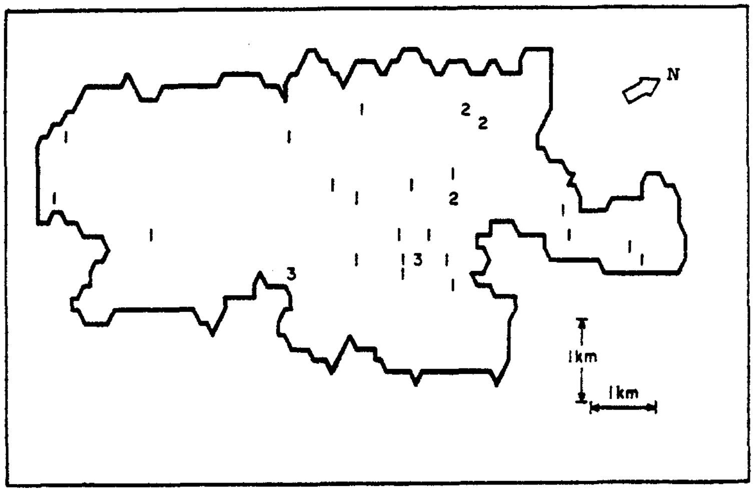

FIG. 7. The locations of definitely nonsmoking emphysema patients.

Comparing average pollution values

The mean pollution values of grid areas having an incidence of a type of lung disease were compared to the mean values of those points not having any lung disease reported. A version of Student's t-test was used to test for significance of difference.[10] This test was used because the parameter variances of the distributions of pollution values could not be assumed to be equal. The parameter is defined as follows:

$$t_s = \frac{[\bar{y}_1 - \bar{y}_2] - [\mu_1 - \mu_2]}{[s_1^2/n_1 + s_2^2/n_2]^{\frac{1}{2}}}. \tag{3}$$

The $\bar{y}$'s are the sample means and the null hypothesis has the parameter means, μ's, equal. The s^2's are the sample variances and the n's are the number of samples. The two pollution fields computed using the yearly climatology, were used to obtain the pollution values. All cases, as well as only nonsmoking cases of the two lung diseases, had their means compared with the means of those points not having any cases reported. No grid areas outside the city's boundaries were considered. The parameter is compared with a weighted value of t'_α to determine the significance of difference:[10]

$$t'_\alpha = \frac{t_{\alpha[\nu_1]}s_1^2/n_1 + t_{\alpha[\nu_2]}\, s_2^2/n_2}{s_1^2/n_1 + s_2^2/n_2}.$$

The subscript α is the critical probability, which is used to define a significant difference, and the $t_{\alpha[\nu]}$'s are found in statistical tables. The degrees of freedom are the ν's.

It turned out that all the means of the sets of points showing an incidence of lung disease were greater than those not having any incidence, whether considering all cases or nonsmoking cases only. Table 1 summarizes the results of the testing of these means.

Table 1
Summarized results of Student's t-test for significance of difference between mean pollution concentrations

	Pollution field computed using:	
	all sources	seven main sources
All emphysema	0·001	0·001
All lung cancer	0·05	0·4
Nonsmoking emphysema	0·05	0·1
Nonsmoking lung cancer	0·3	0·2

Figure 8 shows a plot of the ratios of the number of diseased grid areas over the nondiseased areas for classed pollution values. The pollution field of Fig. 2 was used for this plot. Lung cancer, and to a greater degree, emphysema, show an increase in this ratio toward higher pollution values. This illustrates that areas with higher pollution values had a greater chance of having a residing lung patient.

Analysis of variance

As an alternate method to test for significance of difference among the mean pollution values of diseased areas vs. nondiseased areas, an analysis of variance (Anova) was done.[11] The data were stratified into three groups: areas having no incidence of lung disease, areas having lung cancer, and areas having emphysema. Four anovas were done on the pollution values of these three groups. The pollution values were obtained from the pollution fields of Figs. 1 and 2. Two anovas used only the definite nonsmokers and two used all cases reported. Each of the anovas showed a significant variance among the mean pollution values of the groups. The significance levels obtained were 0·05 for both nonsmoker anovas, and 0·001 for both anovas which used all cases reported.

Population density

There was no correction for varying population densities in the above analyses. To test if the disease cases were randomly distributed with respect to population density, the city was divided into 65 areas each containing 1000 people. The areas were based on the 1960 census. The numbers of cases of lung cancer and emphysema occurring in each area were

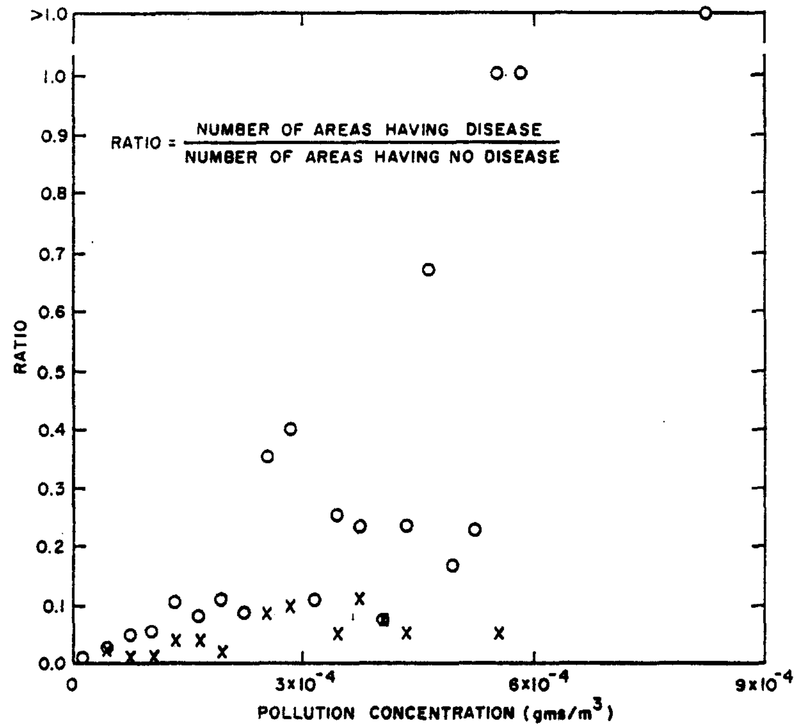

FIG. 8. A plot of the ratio of the frequency of grid areas having an incidence of lung cancer and emphysema to the frequency of nondiseased areas for classed pollution values. The pollution field of Fig. 2 was used for this figure. The circles indicate ratios for emphysema.

plotted vs. frequency of each number. These distributions were compared with the expected Poisson frequencies and both showed an excess of frequencies on the tails, demonstrating "clumping".[10] That is, there were more areas containing small and large numbers of cases than expected, whereas the frequencies in the middle of the distributions were less than expected. Clumping can be interpreted as meaning there was some influence in Altoona encouraging the incidence of these diseases. Whether this was air pollution or not cannot be decided solely by this analysis.

A chi-squared goodness of fit test was applied to determine the significance of difference between the actual frequencies and the expected frequencies of disease occurrence within these areas of equal population. The distribution of lung cancer was found to be different at a significance level of 0·2, and the emphysema at a level of 0·005. For emphysema, this can be interpreted as meaning that there was less than one chance in 200 that the emphysema cases could be distributed as they were, in a random fashion. This implies that there was some ordering imposed on the disease incidence.

8. SUMMARY AND CONCLUSIONS

Although the time of heavy pollution and time of disease occurrence were ten years apart, this project has found evidence that there was some residual effect of the past air pollution of Altoona on certain lung ailments. By fabricating several numerical pollution fields and checking their combined validity with tree growth analysis, long term average values of air pollution were obtained. Values of air pollution taken from two of these fields were used to see if grid points having an incidence of lung cancer or pulmonary emphysema had higher means than those not having any incidence. In all cases the diseased areas had higher pollution averages.

The effects of population density on the grouping of disease cases was shown to be offset by some factor, presumably air pollution. Analysis showed that areas tended to have either no cases of emphysema or lung cancer or many cases, i.e. a "clumping" of cases in certain areas. Through use of the *t*-test of means and analysis of variance, it was found that those areas where the disease cases clumped tended to be areas shown numerically to be polluted.

For the number of lung ailments obtained in 1965 (a total of 171 cases), the results have been most encouraging. Further work is planned along this line using a larger city. The pollution model will be updated to include topography, work schedules of factories, and the urban heat island effect. Future plans also include short term effects of several types of air pollution on lung ailments.

Acknowledgements—I would like to thank Y. Sasaki, G. A. Eddy and E. L. Rice for their help and encouragement while working on this project. Acknowledgment is also made to the National Center for Atmospheric Research, which is sponsored by the National Science Foundation, for use of its Control Data 6600 computer.

REFERENCES

1. J. D. Williams and N. G. Edmisten, *An Air Resource Management Plan for the Nashville Metropolitan Area.* U.S. Department of Health, Education, and Welfare, Cincinnati, Ohio, 1965, pp. 86–90.
2. O. C. Gilbert, Lichens as indicators of air pollution in the Tyne Valley, *Ecology and the Industrial Society*, A Symposium of the British Ecological Society (Symposium No. 15, 13–16 April 1964).
3. M. D. Thomas, Effects of air pollution, *Ecology and the Industrial Society*, A Symposium of the British Ecological Society (Symposium No. 15, 13–16 April, 1964).
4. Blair County Planning Commission, Historical background and physiography, Report No. 12 of the Blair County Planning Commission, Hollidaysburg, Pennsylvania, 1967, pp. 4–37.
5. W. E. Herman, From a personal interview, Department of Forests and Waters, Altoona, Pennsylvania, 1969.
6. S. W. Tromp, *Medical Biometeorology*, pp. 479–480 and 499–501. Elsevier, New York, 1963.
7. United States Atomic Energy Commission, *Meteorology and Atomic Energy*, p. 99. U.S. Department of Commerce, Springfield, Virginia, 1969.
8. D. B. Turner, *Workbook of Atmospheric Dispersion Estimates*, pp. 8–9. U.S. Department of Health, Education and Welfare, Cincinnati, 1969.
9. L. J. Battan, *The Unclean Sky*, pp. 74–77. Doubleday, Garden City, New York, 1966.
10. R. R. Sokal and F. J. Rohlf, *Biometry*, pp. 90–94, 221, 374. W. H. Freeman, San Francisco, 1969.
11. P. G. Hoel, *Introduction to Mathematical Statistics*, pp. 299–325. John Wiley, New York, 1962.