# WATER QUALITY OBSERVATION PROGRAM
# OTTY LAKE 1972-2002

Lorne Gold

## 1. Introduction

In 1971-72 a group of concerned property owners on Otty Lake began a water quality observation program. The fecal coliform count (hereafter fecal count or fc) was selected to be the measure of lake water quality. Trial measurements were made at 15 sites located along the Otty Lake southwest shore. It was decided from this experience to measure the fecal count each summer at 32 sites distributed reasonably uniformly around the lake. Sites were selected and the program initiated in 1973.

The GPS reading for the sites are given in Table 1. Figure 1 is a map showing their location. Nine measurements were made at each site in 1973, 74, 77 and 78. Seventeen measurements were made in 1975 by three students carrying out a study of the lake supported by the Federal Government Opportunities for Youth Program[1]. Six measurements were usually made in 1976, 1979 to 1983, 1985 to 1988, and 1999 to 2002, but only 3 in 1984. Three measurements per year were made in the ten-year period 1989 to 1998. The collection of samples began in late May or early June and ended late August or early September. Fecal count was reported as the number of counts in a 100-milliliter water sample.

Volunteers ran the program from 1973 to 1984. Water samples were obtained within fifty feet of the shore at a depth of six inches below the surface, following the instructions of the Ontario Department of the Environment. Fecal counts were determined, without cost, by the Laboratory Services Branch, Leeds, Grenville and Lanark Health Unit, Ontario Ministry of Health. This service was discontinued in 1985 and the Otty Lake Association paid for the analysis thereafter using private laboratories. In 1985 the Association initiated yearly contracts to have the samples taken by Dawson Girdwood. Following the advice of the Ontario Ministry of the Environment, he fixed the routine of taking the samples within six feet of the shore by dipping the sample bottle to the maximum distance that could be reached from the side of the boat.

For the first nine years of the program the testing laboratory measured the total coliform count and the fecal count. This was discontinued in 1982 and thereafter only the fecal count was determined. The ease with which fecal coliform can be counted depends on the level of the background total coliform, which tends to increase during the summer. Because of this, the accuracy to which the counts were determined and reported varied throughout the program. From 1973 to 1982 they were given in multiples of 2 with 0 being the lowest value. Odd values were given, in addition to values in multiple of 2, for 1979 and 1980. In 1983 this was changed to “less than 5” and in multiples of 5 thereafter. In 1986 counts were given as “less than 4” and in multiples of 2 thereafter. From 1987 to 1991 and in 1993 and the first half of 1994 they were reported as “less than 10” and in multiples of 10 thereafter. In 1992 and after 1994 the minimum count was given as “less than 1” or “less than 10” with even and odd values for low counts and multiples of 5 or 10 thereafter. If no fecal coliform could be seen, it was recorded as less than 1, less than 2 or absent. In addition to this large variation in the accuracy and way of reporting fecal counts, there was considerable random variation in the weather at the time of sampling and how and where in the test sites the samples were taken.

After thirty years of measurements it was decided to change from fecal coliform to e-coli as the measure of lake water quality. E-coli is a more specific measure of harmful bacteria and easier to count within the background of total coliform bacteria. Also, the time had come to determine if the large data set that had been compiled could provide information on the possible dependence of lake water quality on lakeshore development, one of the principle objectives of the program.

## 2. Characteristics of the fecal count data

Because of the variation in the number of samples obtained each year and in the ranges in which counts were reported, it was necessary to determine how to treat the measurements so that the results of an analysis of the data could be interpreted with a good level of confidence. To this end, all measurements for each station were compiled and listed in ascending order of value. Samples with a count of 0, less than 1, less than 2 or absent were grouped with those with a count of 1.

Coliform counts are often large and can overwhelm the contribution of small values when analyzing a set of measurements. It is found to be more meaningful for an analysis to use the logarithmic value of counts rather than the actual values. The form of the logarithm used throughout this report is the one that arises naturally in statistical theory. It is designated by “ln( )”. Minimum values of less than 4, less than 5 or less than 10 were replaced by the whole number closest to that corresponding to one-half their logarithm, i.e., 2 or 3 respectively.

Figure 2a shows, for station 1, a normal probability plot of the percentage of samples with a count less than or equal to a given value. The values of the fecal count are presented on a logarithmic scale. Looking at populations of fecal counts in this way, without regard as to when the samples were obtained, is less sensitive to errors of measurement and to recording in ranges of value. It allows samples with fecal count less than 1 to be included even when the reported value is 0, less than 1, less than 2 or absent. The line drawn through the sample points is the logarithmic best fit to the observations. It is described by the equation:

$$fc = A\exp[B(\mathrm{norm}(\%))] \qquad 1$$

or

$$\ln(fc) = \ln(A) + B(\mathrm{norm}(\%)) \qquad 2$$

A and B are constants for each station, norm(%) is obtained from a table of values for the standard normal distribution and ln( ) is the natural logarithm of the fecal count or of constant A.

“R”, the correlation coefficient, indicates how well the logarithmic relation describes the distribution in the fecal counts. If all the data values were on the best-fit line the correlation coefficient would be 1. A value of 0.95 shows that there is a very good correlation for station 1 between the logarithm of the fecal count and the percentage of counts equal to or less than that count, over the range of 30% to 99 %. Ln-normal probability plots, similar to fig. 2a, are given for all the stations in an addendum to this report[2]. Figure 2b is a bar chart showing the distribution in the correlation coefficients for all 32 stations. Twenty-seven have values of R equal to or greater than 0.95; the smallest value is 0.88. Figure 2c is a ln-normal probability plot for all the 5,493 counts obtained in the thirty years of sampling. The lower than expected percentage for fecal counts less than or equal to 1 may be due, in part, to missing values when the minimum was reported as less than 5 or 10. The steps in the data probably reflect the times when counts were reported in multiples of 10. The figure shows, however, that the distribution in the total number of counts is described to high degree of correlation (R = 0.99) by the ln-normal distribution function.

The high degree of correlation of the ln-normal distribution function with each of the thirty two station population of counts and with the total population of counts implies that the fecal counts are random variables and can be analyzed using statistical methods. It is also provides additional justification for the practice of working with the logarithm of values when reporting and analyzing coliform counts. For example, a measure of the average value for a collection of two or more sample counts can be calculated by dividing the sum of the logarithms of the counts by the total number of samples, including those with a count of 0, less than 1, less than 2 or absent.

The station annual average logarithm of the fecal count, ln(fc), was calculated for each of the thirty years of measurements. These values were used to obtain an estimate of the thirty-year average ln(fc) for each station and of the annual average ln(fc)s for the lake. The station annual average ln(fc)s, the station thirty-year average ln(fc)s and the lake annual average values, are given in tables in the Addendum.

Figure 3a gives the lake annual average ln(fc) for each year since 1972. It shows a general increase in the annual average for the first twenty-five years and a quite abrupt decrease over the last five. The best-fit line to the data values for the first twenty-five years was determined and is plotted in figure 3a. The difference between the observed values of the lake average ln(fc) for the first twenty-five years and the corresponding values calculated from the best-fit line, were determined. Figure 3b is a normal probability graph giving the percentage of years with a difference less than or equal to a given value. The population of difference values has a normal distribution to a high level of confidence (0.995) and a standard deviation (sd) of 0.25. Dashed lines spaced 1 sd above and below the best-fit line are plotted in fig.3a. About two-thirds of the station values are between these two lines; all values are within ± 3 sds. Figures 3a and 3b show that, although the lake annual average ln(fc) tends to increase linearly with years for the first twenty-five years, the yearly values, when considered as a population, are random variables with respect to the average rate of increase defined by the best-fit line.

The station thirty-year averages are plotted against station number in figure 4a. Stations 1, 2 and 3 (Mclaren Lake) are plotted as stations 33, 34 and 35 respectively. With this rearrangement the stations are ordered across the southwest end to the south shore, along the south shore to northeast end, across the northeast end to the north shore and along the north shore to the southwest end. The solid line in the figure is the best fit to the data points. It has a relatively good correlation factor, R, of 0.66. Figure 4b is a normal probability plot of the difference between the observed station values and the corresponding best-fit values given by the line in fig. 4a. The differences have a normal distribution to a high level of confidence (R = 0.98) and a standard deviation of 0.17.

Two lines, spaced 1 sd above and below the best-fit line, are plotted in fig. 4a. Again, about two-thirds of the data points are between these two lines and all are within ± 3 sds. The figure shows a marked trend for the station thirty-year average ln(fc) to decrease from the northeast end of the lake to the southwest end. This is consistent with development of Otty Lake beginning at the northeast end, the area closest to Perth, and gradually extending down the north and south shores. Figure 4b, however, shows that the difference between the station values of the 30-year average ln(fc) and their corresponding values from the best-fit line appear to be random variables when considered as a population without reference to the station number.

## 3. Characteristics of the stations

A plot of the dependence of the annual average ln(fc) on years since 1972 is shown for station 1 in figure 5a. There is a marked decrease in the values for years 26 to 30, similar to that observed for the lake annual average ln(fc) (fig. 3a). The solid line in fig. 5a is the linear best fit to the values for years 1 to 25. Similar plots for each station are given in the Addendum. The correlation coefficient for the best-fit line is 0.55. The correlation coefficient for the best-fit line for each of the stations is in the range of 0.33 to 0.79.

Figure 5b is a normal probability plot, for station 1, of the difference between the observed values of the annual average ln(fc) for years 1 to 25, and the corresponding values given by the best-fit line in fig. 5a. A correlation coefficient of 0.97 indicates that the differences are normally distributed to a high level of confidence. Similar plots for the other stations had correlation coefficients between 0.90 and 0.99, with 90% equal to or greater than 0.94, indicating that their difference populations are random variables to a good level of confidence. The standard deviation of the difference population for station 1 is 0.69. The dashed lines in fig. 5a are 1 sd above and below the best-fit line. About two thirds of the data values are between these two lines and all are within ± 3 sds. This was the case, also, for the other stations.

Figure 6a is a plot of the standard deviations of the difference populations against station number. It shows that the year-to-year variation in the annual average ln(fc) increases from the southwest end of the lake to about station 20 and is relatively constant for the northeast half of the lake (station 20 to station 2). Figure 6b is a normal probability plot of the percent of standard deviations less than or equal to a given value. It shows that the population of standard deviations can be described to a high level of confidence by the normal distribution function (R = 0.99). Figures 5a,b and 6a,b show that, although the twenty five year populations of station annual average ln(fc)s are random variables, the year to year variation in the station values depends in a significant way on their location along the lakeshore.

The correlation coefficient, R, for the linear best fit for the first twenty five years to the dependence of the annual average ln(fc) on the corresponding year since 1972, is plotted against station number in figure 7a. These are the coefficients given in fig 5a and similar figures in the Addendum. The correlation coefficients decrease from the southwest end (station 4) to a minimum in the region of station 22, and increase around the northeast end and down the north shore, i.e., stations 18 to 24 have the smallest correlation of the annual average ln(fc) with the corresponding year since 1972.

Figure 7b shows how the station annual rate of increase in the average ln(fc) for the first twenty-five years, depends on station number. The rate is given by the slope of the best-fit line to the dependence of the annual average ln(fc) on years for years 1 to 25, e.g., 0.062 for station 1 (fig. 5a). It has a dependence on station number similar to that found for the correlation coefficient, i.e., a decrease across the southwest end and down the south shore to a minimum in the range of stations 16 to 26, and an increase across the northeast end and down the north shore to station 2.

An estimate of the annual average ln(fc) at 0 years (1972) was obtained for each station from the best-fit line, presented in fig. 5a and the figures in the Addendum. Its value for station 1, given by the equation in fig. 5a, is 0.59. These estimated values of the annual average ln(fc) at the beginning of the program are plotted against station number in figure 7c. They increase from station 4 to a maximum at the northeast end of the lake and decrease thereafter along the north shore.

The normal probability plots for the correlation coefficients given in fig. 7a, the rates of increase given in fig. 7b and the station annual average ln(fc) for 1972 given in fig. 7c, had correlation coefficients of 0.99, indicating they are random variables to a high level of confidence. Table 2 gives the station values for years 1 to 25 for the standard deviation of the difference populations, fig. 6a, the correlation coefficients for the linear best-fit line to the dependence of the annual average ln(fc) on years, fig 7a, its annual rate of increase, fig. 7b and the estimated station values of the annual average ln(fc) for 1972, fig. 7c.

Figures 6a, 6b and 7a,b, and c, along with figures 2, 3, 4 and 5, indicate that there is a randomness and regularity in both the lake and station annual average values of ln(fc) and the 30-year station average values. The randomness appears when they are considered as populations without reference to time or station number; the regularity appears in their dependence on time and station number. The regularity seen in figures 6 and 7 suggests that environmental factors may affect the fecal counts.

## 4. Summary of the characteristics of the fecal count and station data

The following is a brief summary of the evidence presented in Sections 2 and 3 of a dependence of lake water quality on development on or near the lake. It was shown that the natural logarithm of the fecal counts are random variables if the number equal to or less than 1 is combined with the number equal to 1. This behavior provided a rational basis for calculating annual and 30-year average fecal counts for the stations and the lake. The lake annual average ln(fc) was found to be dependent on the number of years since 1972, shown in figure 3a, and the 30-year station averages on station number or location on the lake, shown in figure 4a.

The annual average ln(fc)s for each station increased with time during the first twenty-five years, then decreased abruptly during last five years in the same manner as the lake annual average ln(fc)s (fig. 5a and Addendum). The amount by which each varied from year to year with respect to the corresponding best-fit value increased from the southwest end to the northeast end (fig. 6a). The correlation between the station annual average ln(fc)s and corresponding year (fig. 7a) and the station rate of increase (fig. 7b) decreased from the southwest end to a minimum in the range of stations 16 to 25 and increased across the northeast end and down the north shore. Estimated values of the annual average ln(fc) for 1972 (year 0) increased with station number to a maximum in the region of station 22 and decreased across the northeast end of the lake and down the north shore (fig. 7c). These characteristics of the station annual average ln(fc)s are defined by populations of random variables that, like the lake annual average ln(fc)s, have a dependence on time and on station number, i.e., location on the lake. They suggest that the fecal counts, measured over a period of thirty years at thirty-two stations distributed approximately uniformly around the lake, are affected by factors in addition to lakeshore development.

## 5. Annual average fecal count and shoreline development

A survey, carried out in 1993 under the Environmental Youth Corp Program[3] and a follow-up study in 1994[4], provided information on the age of septic systems for 305 lakeshore properties. Unfortunately, information was not obtained on the age of structures on the lake or on whether a septic system was a replacement. The results of the survey provide, however, a basis for correlating the increase in the annual average lake fecal count with the corresponding increase in the number of septic systems.

Figure 8 presents the dependence on years since 1960 for the number of septic systems on the lake, as given by the 1994 study. The line is the best fit to the data values. Figure 9 gives the lake annual average ln(fc) plotted against the corresponding number of septic systems on the lake. The data points are for 0, 5, 10, 15, and 21 years since 1972. They were determined from the linear best fit to the dependence of the lake annual average ln(fc) on number of years, shown in figure 3a, and the corresponding number of septic systems given by the best-fit curve in figure 8. It was assumed that the number of septic systems on the lake increased by 10 per year from 1993 when plotting the lake annual average ln(fc) for years 26 to 30. It is very probable that this number of septic systems have not been installed, but presenting the lake annual average ln(fc)s in this way shows the remarkable general decrease in their values that occurred in the last five years of the program.

It was possible to get from the Environmental Youth Corp survey a rough estimate of the average date of installation of septic systems in the vicinity of each station. Figure 10 presents the station 30-year average ln(fc) plotted against the corresponding estimated average date of installation. There is considerable scatter in the data, but the selected values in the figure show a trend. The line in the figure is the linear best fit to the selected values. It is defined to a reasonably high confidence level with R equal to 0.89. Replacement of old septic systems may account for some of the stations being above the trend, for example, that near Camp Shomria. Two data points, identified by “x”, are for stations near wetlands. It is known that for one of the wetland stations there is considerable animal activity.

Figures 9 and 10 give evidence that there is a correlation between the annual average lake and station 30-year average ln(fc)s and the number and age of septic systems on the lake, respectively. The fecal count, however, is not a sensitive indicator of the performance of a septic system immediately adjacent to a measurement station. When a high fecal count was measured and re-measured immediately after receiving the result, which was usually a matter of days, it almost always had returned to a low value.

A study was carried out on the effectiveness of the dye method for detecting poorly performing septic systems[5]. Dye was introduced into three tanks with their tile beds 45, 72 and 84 feet from the lake, respectively. Water samples were taken near the shore during the forty-eight hours after the start of the test, but no change in color was seen. Dye was poured into two holes, one 15 feet from the lakeshore and the other about 1 foot. During the next twenty-four hours it was detected in the water only for the hole nearest to the lake. These observations indicate that the residence time of dye in the soil, or in a system that has not totally failed, is sufficiently long that it is effectively absorbed before it reaches the lake. It was concluded that periodic fecal count measurements and dye tests would not be satisfactory ways to identify poorly performing septic systems.

These tests, and the observed dependence of the annual average ln(fc) on time and station for the first twenty five years, suggest that active fecal matter must have a reasonably long residence time on shore and in the lake and increases in some manner from year to year. As there is a general movement of lake water from the southwest to the northeast, due to the flow out Jebb Creek, fecal counts could be affected by lake and shore conditions for some distance “upstream”. This suggests there may be a degree of station-to-station correlation between annual average fecal counts and this is explored in the next section.

## 6. Correlation analysis

A correlation analysis was carried out on the array of thirty values of the annual average ln(fc) for each of the thirty-two stations using the CORREL function of EXCEL. The calculated correlation coefficients between stations for their 30 yearly values and the correlation coefficients between years for the 32 station value, are given in tables in the Addendum.

Figures 11a, b and c provide an appreciation of the meaning of the correlation coefficient. Figure 11a is a plot of the annual average ln(fc)s for station 5 against the corresponding values of the annual average ln(fc) for station 8. The linear best fit to the observations has a correlation coefficient, R, of 0.81, which is the value given in the table in the Addendum. Figure 11b is a corresponding plot for stations 22 and 11 that have a correlation coefficient, R, of 0.0004. For these stations there is practically no correlation between their corresponding values of the annual average ln(fc). Figure 11c is a plot of the station annual average ln(fc)s for 1987 against the corresponding values for 1994. In this case the correlation coefficient is -0.46 and the best-fit line has a downward slope rather than an upward slope as in figure 11a. These figures show that the closer the value of the correlation coefficient is to 0, the smaller is the correlation between two sets of station or yearly values. A positive correlation indicates that corresponding values tend to increase or decrease together. If the correlation is negative, there is a tendency for one value to increase when the corresponding value decreases.

Figure 12 gives the percent of the correlation coefficients for both stations and years that have values smaller than a given value. It shows that there is a larger percentage of correlation coefficients for years below a given value of the coefficient than there is for stations. For example, about 82 % of the correlation coefficients between stations had values less than 0.6, whereas about 99.5 % of the correlation coefficients between years where less than that value. The average value of the correlation coefficient for stations (50% value) is about 0.45. The average value of the coefficient for years is about 0.15.

Figure 13 is a bar graph showing the correlation coefficients for the annual average ln(fc)s of station 1 with the corresponding values for the other stations. Figure 14 is a bar graph showing the correlation coefficients for the annual average ln(fc)s of year 1 with the corresponding values in subsequent years. The bar graphs for all the stations and years are given in the Addendum. Figures 13 and 14 have maxima and minima in the dependence of the correlation coefficients on stations or years, and this occurs also for the figures in the Addendum. This regularity may provide information on the nature of the affects of the environment, weather, and lake and terrain characteristics on the fecal count.

Figure 15a is an area display of stations with a correlation coefficient, R, less than 1 and greater or equal to 0.6. The correlation coefficient for a station with itself would be on the diagonal line and have a value of 1. The distribution of values above the diagonal line is the mirror image of that below because of the equality of the correlation coefficients for two stations, one with the other. It is of interest that stations 7, 21 and 29 do not correlate with any other station at a value of the correlation coefficient equal to or greater than 0.6

Figure 15b is a bar chart showing the number of times a station correlation coefficient with other stations equals or exceeds 0.6. There is a periodic maximum in the number of counts that are separated by a minimum at station 2, stations 9 to 12, stations 18 to 22 and stations 28 to 30. Stations 21 and 29 are in two of the minima, but station 7 falls in the range of a maximum.

There is a pattern in the distributions shown in figs. 13, 14 and 15a,b, and this pattern, along with knowledge of the characteristics of the sites, should give additional insight into what determines the fecal count at the stations. For example, the maximum in the number of correlations greater than 0.6 for stations 4 to 8 (excluding station 7), stations 13 to 17, stations 23 to 27 and stations 31 to 1 suggest that there may be factors common to these stations that affect the fecal counts to a greater extent than they do at stations associated with the minimum values. This implies that, to obtain a deeper understanding of the dependence of lake water quality on shoreline development, it will be necessary to consider several site and environmental factors. Obtaining a quantitative understanding of what determines water quality is truly a multivariate statistical problem.

There has not been an opportunity to obtain information on environmental factors that might be affecting the fecal count. Some of them can be measured while for others it may be possible to obtain only a qualitative measure, such as class of shoreline based on vegetation, drainage and shoreline configuration; exposure to wind and boat traffic; animal activity; etc. They are probably random at the scale of the lake and over a period of thirty years, and this could be the basis for the randomness in the fecal count populations. It may be possible to obtain information on some of them for the thirty years of the program, and to correlate it with the individual fecal count measurements and station and lake average values. The results presented in this initial analysis of the data provide a basis for such studies.

## 7. Conclusions

1. The fecal counts measured at stations on Otty Lake are populations of random variables that can be described by the ln-normal distribution function to a good level of confidence, fig. 2a,b,and c.
2. The lake annual average ln(fc) increased linearly with years for the first 25 years of the program, then decreased abruptly after that, fig. 3a.
3. The difference between the observed values for the lake annual average ln(fc) for the first 25 years and the corresponding values given by the best-fit line, is a random variable, fig. 3b.
4. The station 30-year average ln(fc) tends to decrease from the northeast end of the lake to the southwest end, fig. 4a.
5. The difference between the observed station 30-year average ln(fc) and the corresponding value given by the best-fit line to the observed values is a random variable, fig. 4b.
6. The dependence of the lake annual average ln(fc) on years (fig. 3a) and the dependence of the 30-year average ln(fc) on station number (fig.4a) are consistent with the history of development for the lake.
7. The station annual average ln(fc)s for the first 25 years has the same general dependence on years as the lake annual average ln(fc), fig. 5a.

8. The year-to-year variation with respect to the corresponding best-fit values for the station annual average ln(fc)s for the first 25 years increases with station number, i.e., location on the lake, fig. 6a.
9. The estimated annual average ln(fc) for 1972 (year 0) increases with station number from the southwest end of the lake along the south shore to the northeast end, and decreases thereafter to station 2, fig. 7c.
10. The correlation between the station annual average ln(fc)s and the corresponding 25 years since 1972, fig. 7a, and the rate of increase in their values, fig. 7b, decrease from the southwest end of the lake down the south shore to the region of station 22, and increase around the northeast end of the lake and down the north shore to station 2.
11. There is a positive correlation between the number of septic systems on the lake and the lake annual average ln(fc), fig. 9.
12. There is a positive correlation between the station 30-year average ln(fc) and the estimated average age of adjacent septic systems, fig. 10
13. The dependence of the lake annual average ln(fc)s on the corresponding number of septic systems (fig. 9), and the station 30-year average ln(fc) on the estimated average year of installation of adjacent septic systems (fig 10) show a dependence of the fecal count on lakeshore development.
14. The year to year variation in the amount the annual average ln(fc)s are above or below the corresponding expected value defined by the best-fit line suggests that factors such as terrain, vegetation, shoreline configuration, animal activity, weather, exposure to wind and boat traffic, may also affect fecal count determinations.
15. The correlation of corresponding values of the annual average ln(fc) for station 1 with all the other stations (fig. 13) shows that there may be a spatial regularity in the factors affecting the station annual average ln(fc)s.
16. The maxima and minima in the correlation coefficients for 1973 with all years to 2002 (fig. 14) suggests that the factors that determine the station annual average ln(fc)s vary in a relatively regular way over long periods of time.
17. The abrupt drop in the lake and station annual average ln(fc)s in year 26, including station 3 on McLaren Lake, suggests that Otty Lake and its watershed experienced something that lowered the lake annual average ln(fc)s during the last 5 years by an amount significantly greater and more persistent than the normal year-to-year variations found for the first 25 years (fig. 3a).
18. The water quality of a lake depends on many factors in addition to shoreline development and must be treated as a multivariate problem.
19. The initial analysis of this data set has shown that it is internally consistent and statistically valid. This provides some confidence that a greater understanding can be obtained of what affects lake water quality through a more detailed correlation with environmental factors for which there is a record for the period of the program.

## 8. Some speculations on the interpretation of the analysis

Fecal or e-coliform counts are effective for monitoring beaches when measured regularly during the day. Such a program of measurement is similar to daily weather observations. If samples are taken every two or more weeks, however, and it takes two or more days to get the results, they will give warning of an unacceptably high count only if it is persistent. This did not occur for any of the stations during the thirty years of measurement. The analysis of the fecal counts for Otty Lake has shown, however, that such a program can provide a coherent data set and yearly average values that could be used to a reasonable level of confidence to relate to factors that might be affecting lake water quality. In this case, calculating annual and longer-term average values of the fecal count is comparable to determining the yearly average of weather elements and the climate from daily weather observations.

Stations 7, 21 and 29 have interesting characteristics that may be significant. Station 7 is in a bay and relatively sheltered from the main body of the lake. Its estimated annual average ln(fc) for 1972 is higher than the values for the adjacent stations (fig. 7c). The station’s annual average ln(fc) has a low correlation coefficient for its dependence on years (fig. 7a) and the smallest rate of increase for the lake (fig. 7b). This behavior suggests that local factors have a great affect on the fecal count and that these factors vary significantly from year to year. It is of interest that station 8 tends to behave in a contrary manner. It is highly correlated with other stations and its estimated annual average ln(fc) for 1972 is consistent with stations 4 to 6 and station 9 and higher. Its annual average ln(fc) is well correlated with years since 1972 and has the highest rate of increase for the lake.

Figures 6a and 7a,b,c indicate that the stations in and near Parks Bay, which is just across the lake from the bay leading to Jebb Creek, experienced environmental conditions that caused their annual average ln(fc) and associated statistical characteristics to be in the range of either the maximum or the minimum for the lake. Station 21, at the bottom of the Bay and somewhat sheltered from the main body of the lake, has one of the highest values for the estimated annual average ln(fc) for 1972 (fig. 7c) and the highest 30-year average value (fig. 4a). It is adjacent to a wetland and birds and animals are active in the area where the samples were obtained. A creek from the wetland empties into the lake not far from the sampling site. The fecal counts did not have the marked drop in values for the last five years found for most of the other sites. Animal and bird activity associated with the wetland may be one of the important factors affecting the annual average ln(fc)s and their statistical characteristics for stations in and adjacent to the Bay.

The general movement of the surface water of the lake is from the southwest to the northeast. This, along with southwesterly winds down the length of the lake, are the probable cause of the drift of rafts of weeds and other material into the area associated with stations 23 to 27. It could be possible that the statistical characteristics of the annual average ln(fc)s for these stations, and the maximum in the number of station correlation coefficients greater than 0.6 associated with them, are due, in part, to their being in line with the main body of the lake. As can be seen in fig. 1, the entrance to Jebb Creek tends to the north of the main axis of the lake and there is little to deflect the movement of surface waters in that direction.

Station 29 is the one closest to Jebb Creek on the north shore. There is a marshland adjacent to it. It has the second highest estimated annual average ln(fc) for 1972. The rate of increase of its annual average ln(fc) and the correlation coefficient for its dependence on years since 1972 are well below the values for the neighboring stations and below those for station 21. Although it appears that local conditions might be an important factor affecting the fecal counts for the station, it, along with stations 28 and 30 define the minimum in the number of correlation coefficients greater than 0.6 (fig. 15b) that is associated with the entrance to the Creek. This minimum may be due to a preference for the surface water of the lake to move in the direction of stations 23 to 27 rather than into the bay leading to the creek.

No obvious reason has yet been found for the marked drop in the annual average ln(fc) in the last five years of the program. This drop has some interesting characteristics. It is more marked in the southwest end of the lake then in the northeast end. There is a greater spread in the values at the northeast end and in some cases they are not too far out of line with those for the first twenty five years. This is particularly true for station 21 in Parks Bay and to a lesser extent for station 7. The fecal counts for the last five years were included in the station probability plots presented in fig. 2a and the Addendum. These plots showed that the thirty-year station fecal count populations were described to a high level of confidence by the lognormal distribution function. This indicates that the counts for the last five years were statistically coherent with the counts for the first twenty-five years.

One possible explanation for this coherence may be that the random weather and weather related factors caused the fecal counts to be smaller in the last five years than would be expected from an extrapolation of the best-fit line for the first twenty-five years. Of interest in this regard is that the annual average ln(fc) for station 1 in 1973, which is on the best-fit line in fig. 5a, has a positive correlation with the values in 1998 to 2002, but a negative or zero correlation with the values in the preceding three years. Analysis of such behaviour could help determine the possible affects of weather on fecal counts.

The Otty Lake Association changed the measure of lake water quality from fecal coliform to e-coli in 2003. In preparation for this change, the e-coli count was determined for each station at the same time as the fecal count for three of the six sampling periods in 2001. The ratio between the lake annual average ln(fc) and the lake annual average ln(e-coli) determined from these measurements was 2.30. In 2003 and 2004 e-coli samples were taken on six days. The sets of samples were alternated between even and odd number stations, resulting in six estimates of the lake annual average ln(e-coli). Samples were taken at ten stations in 2005 (stations 3, 5, 7, 8, 15, 20, 21, 25, 29 and 31). One sample was obtained for station 3, two for stations 7, 20, 21 and 29, four for stations 5, 8, 15 and 31, and five for station 25.

The lake annual average ln(e-coli) was determined from the measurements at stations 5, 8, 15, 25 and 31 for 2001, 2003, 2004 and 2005. They are plotted in fig. 3a. It is of interest that the annual rate of increase given by this small number of years (0.052/year) is essentially the same as that for the rate of increase in the annual average ln(fc) for the first twenty-five years (0.053/year). The ratio of lake annual average ln(fc) to the lake annual average ln(e-coli) calculated for 2001 from the five station values is about 2.40. Applying this factor to the values for the annual average ln(e-coli) places the corresponding estimated annual average ln(fc) for 2003 - 2005 within 3 sds of the extrapolated best-fit line for the first 25 years of measurement. It is statistically possible, therefore, that the low values for the last 5 years are associated with persistent environmental factors that cause the annual average ln(fc) to be below the linear best-fit line.

## References

1. Gold, P. E. and S. D. Rogers. 1975. Background Study for a Remedial Lake Management Plan for Otty Lake. Opportunities for Youth Program. Otty Lake Association, Perth, ON.
2. Addendum to the Water Quality Observation Program, Otty Lake, Otty Lake Association, Perth, ON.
3. Peace, E., D. Nuttall, J. Tufts, A. Gainforth and T. Skillen. 1993. Otty Lake Sewage System and Shoreline Survey. Environmental Youth Corp Project. Otty Lake Association, Perth, ON.

4. Peden, C., J. Leitch and L. Buchner. 1994. Otty Lake Follow up Study. Environmental Youth Corp Project. Otty Lake Association, Perth ON.
5. Schriever, W. R. 1997. How to Save Your Septic System. Captain Otty's Log, October issue. Otty Lake Association, Perth, ON.

## Acknowledgements

The Otty Lake Association's water quality testing program is an example of what can be achieved by dedicated volunteers. It was initiated in 1971 as a trial program under the leadership of Fred Green. Fifteen sites were selected along the shore from Mud Lake to the bay in which station 9 of the 30 year program is located (see fig. 1). It was decided to proceed with the program covering the whole lake after discussions with J. J. McNeely and D. Monteif of the Cottage Pollution Control Program under the Ontario Ministry of the Environment.

A complete record was not kept, unfortunately, of all the volunteers that collected water samples and delivered them to the testing laboratories. The following names for the first twelve years of the program were found in notes written by Fred Green and some reports from the testing laboratories.

J. K. Bisson, Ken Daines, Dave Eastham, Al Hoffman, Garth Johnson, Miles Johnson, Laurie and Carol Liberty, Bob Mears, Dave Mears, S. Roy, Bill Schriever, P. Tweedie, J. Wilson.

Dawson Girdwood who, over the years, has accumulated a rich store of knowledge concerning the lake, the people and animals that use it and the fish that live in it, started collecting the samples in 1985 under a contract with the Association. Lorne Gold and Bill Schriever delivered the samples to the laboratories from 1986 to 2002 and filled in for Dawson when required. Murray Hunt helped with this near the end of the program, determined the geographic coordinates for the stations and took responsibility for sample taking and delivery when the change was made from fecal count to e-coli in 2003.

Through the vision of its members and the support of its Board, the Otty Lake Association has established a valuable, comprehensive, thirty-year record of development on the Lake and changes in water quality that is truly unique in Canada. It could not have undertaken this program without the dedicated work of volunteers, the encouragement of the Cottage Pollution Control Program of the Ontario Ministry of the Environment and the support of the Laboratory Services Branch, Leeds, Grenville and Lanark Health Unit, Ontario Ministry of Health, that analyzed the samples without cost until 1986. The members of the Otty Lake Association and its Board can be justly proud of this achievement.

I wish to thank, in particular, Karen Hunt, Pat Larson and my daughters, Judi Brouse and Pat Martin, for their comments on the report. I am deeply grateful for the opportunity given to me by the Board of the Otty Lake Association to analyze the fecal count measurements made in it's water quality monitoring program in the period 1973 to 2002, and for making this report available to all who have an interest in the affects of human activity on the water quality of lakes.

It is with deep appreciation that I acknowledge the help I received from colleagues at the Institute for Research in Construction, National Research Council, concerning the statistical analysis and application of computer programs and procedures.

It is very possible that some volunteers for the program have been missed in the acknowledgements.  If you know of any would you please give their name to a member of the Executive of the Board of the Association.

Lorne Gold
May, 2007

| Stn # | Latitude | Longitude | | Stn # | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 1 | N44$^{O}$ 50.535’ | W76$^{O}$ 13.956’ | | 17 | N44$^{O}$ 50.250’ | W76$^{O}$ 13.151’ |
| 2 | “ 50.166’ | “ 14.344’ | | 18 | “ 50.420’ | “ 13.047’ |
| 3 | “ 49.318’ | “ 15.542’ | | 19 | “ 50.603’ | “ 12.620’ |
| 4 | “ 49.380’ | “ 14.737’ | | 20 | “ 50.793’ | “ 12.495’ |
| 5 | “ 49.557’ | “ 14.461’ | | 21 | “ 51.002’ | “ 12.224’ |
| 6 | “ 49.408’ | “ 14.280’ | | 22 | “ 51.200’ | “ 12.202’ |
| 7 | “ 49.192’ | “ 14.235’ | | 23 | “ 51.426’ | “ 12.220’ |
| 8 | “ 49.082’ | “ 14.190’ | | 24 | “ 51.593’ | “ 11.920’ |
| 9 | “ 48.868’ | “ 14.433’ | | 25 | “ 51.887’ | “ 12.044’ |
| 10 | “ 48.909’ | “ 14.070’ | | 26 | “ 51.779’ | “ 12.449’ |
| 11 | “ 49.079’ | “ 13.857’ | | 27 | “ 51.373’ | “ 12.693’ |
| 12 | “ 49.188’ | “ 13.845’ | | 28 | “ 51.580’ | “ 12.917’ |
| 13 | “ 49.504’ | “ 13.654’ | | 29 | “ 51.301’ | “ 13.251’ |
| 14 | “ 49.883’ | “ 13.771’ | | 30 | “ 50.960’ | “ 13.431’ |
| 15 | “ 50.298’ | “ 13.610’ | | 31 | “ 50.776’ | “ 13.590’ |
| 16 | “ 50.211’ | “ 13.447’ | | 32 | “ 50.675’ | “ 13.793’ |

**Table 1:** GPS coordinates for the fecal count sampling sites on Otty Lake

| Stn # | ln(fc) 1972 | Sd diff | Slope ln(fc)/yr | R | | Stn # | ln(fc) 1972 | Sd diff | Slope ln(fc)/yr | R |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.59 | 0.69 | 0.062 | 0.55 | | 17 | 0.84 | 0.54 | 0.047 | 0.54 |
| 2 | 0.59 | 0.57 | 0.056 | 0.59 | | 18 | 0.72 | 0.50 | 0.025 | 0.35 |
| 3 | 0.82 | 0.52 | 0.067 | 0.69 | | 19 | 0.88 | 0.62 | 0.046 | 0.48 |
| 4 | 0.25 | 0.40 | 0.069 | 0.79 | | 20 | 0.60 | 0.66 | 0.051 | 0.50 |
| 5 | 0.40 | 0.32 | 0.052 | 0.77 | | 21 | 1.02 | 0.66 | 0.053 | 0.51 |
| 6 | 0.20 | 0.38 | 0.060 | 0.76 | | 22 | 1014 | 0.54 | 0.035 | 0.43 |
| 7 | 0.76 | 0.40 | 0.021 | 0.36 | | 23 | 0.71 | 0.65 | 0.041 | 0.42 |
| 8 | 0.39 | 0.63 | 0.089 | 0.72 | | 24 | 0.90 | 0.64 | 0.045 | 0.46 |
| 9 | 0.67 | 0.48 | 0.063 | 0.70 | | 25 | 0.74 | 0.50 | 0.048 | 0.58 |
| 10 | 0.77 | 0.41 | 0.031 | 0.48 | | 26 | 0.77 | 0.55 | 0.047 | 0.53 |
| 11 | 0.56 | 0.44 | 0.041 | 0.56 | | 27 | 0.39 | 0.51 | 0.077 | 0.74 |
| 12 | 0.76 | 0.54 | 0.053 | 0.58 | | 28 | 0.77 | 0.60 | 0.059 | 0.58 |
| 13 | 0.59 | 0.48 | 0.060 | 0.68 | | 29 | 1.05 | 0.50 | 0.031 | 0.42 |
| 14 | 0.63 | 0.36 | 0.037 | 0.60 | | 30 | 0.63 | 0.60 | 0.069 | 0.65 |
| 15 | 0.51 | 0.47 | 0.053 | 0.64 | | 31 | 0.65 | 0.54 | 0.068 | 0.68 |
| 16 | 0.65 | 0.50 | 0.042 | 0.52 | | 32 | 0.41 | 0.78 | 0.088 | 0.64 |

**Table 2:** Characteristics of the station annual average ln(fc)

R is the correlation coefficient for the linear best-fit line to the dependence of the annual average ln(fc) on years for the first 25 years; ln(fc) 1972 is the estimated value of the annual average ln(fc) for 1972; slope is the rate at which annual average ln(fc) increases, ln(fc)/year; and sd is the standard deviation for the population of difference values between the observed annual average ln(fc) and the value given by the linear best-fit line for the corresponding year.

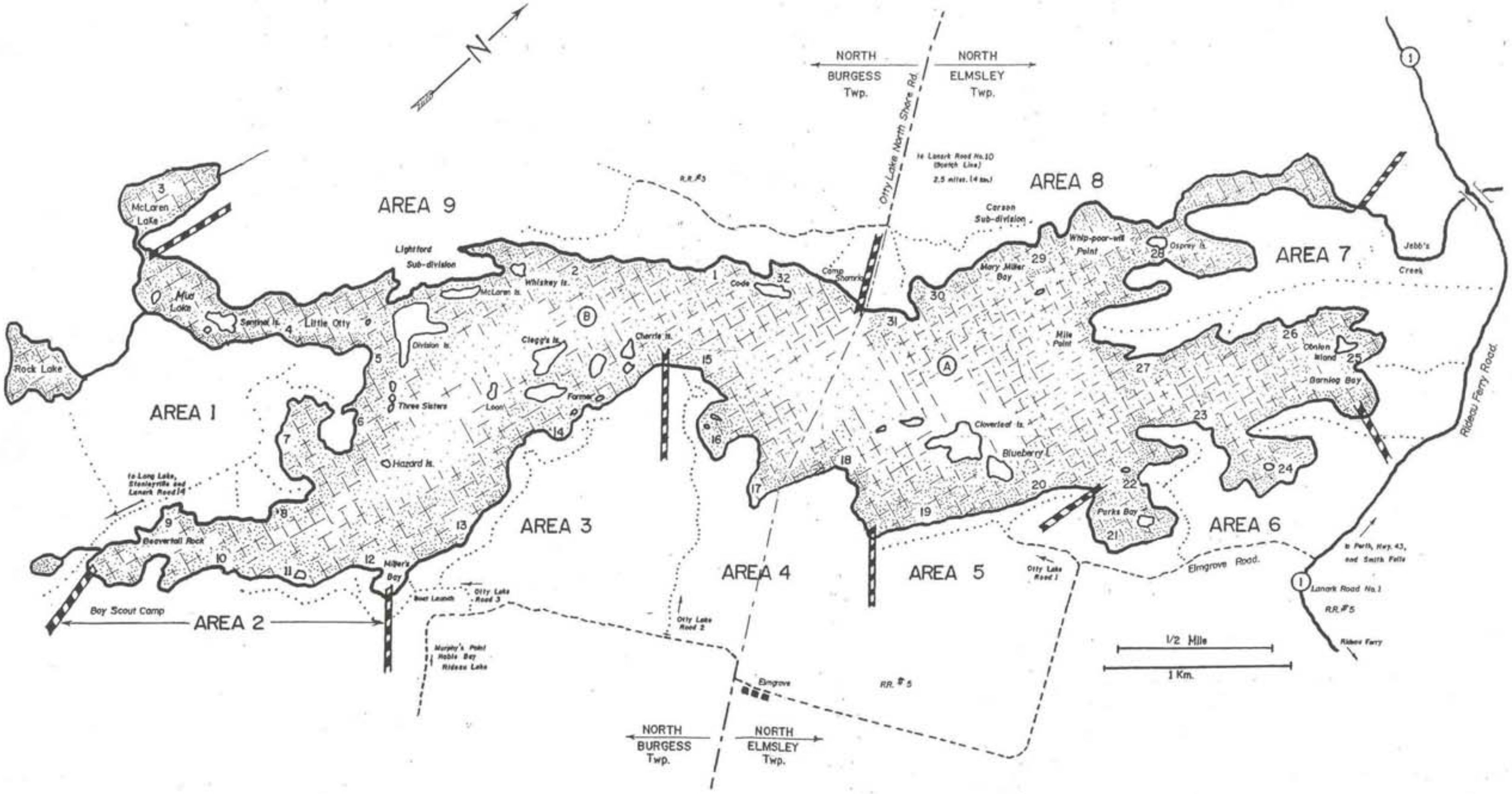


**Figure 1:** Map of Otty Lake

Elevation, 430 ft (133m); length, 4.5 mi (7.2 km); depth, av., 30 ft (9 m), max, 90 ft.
Area, 1545 acres (625 hectares); volume, 46,000 acre ft.

Numbers show bacterial testing sites. A & B are sampling sites for chemical analysis and water clarity.

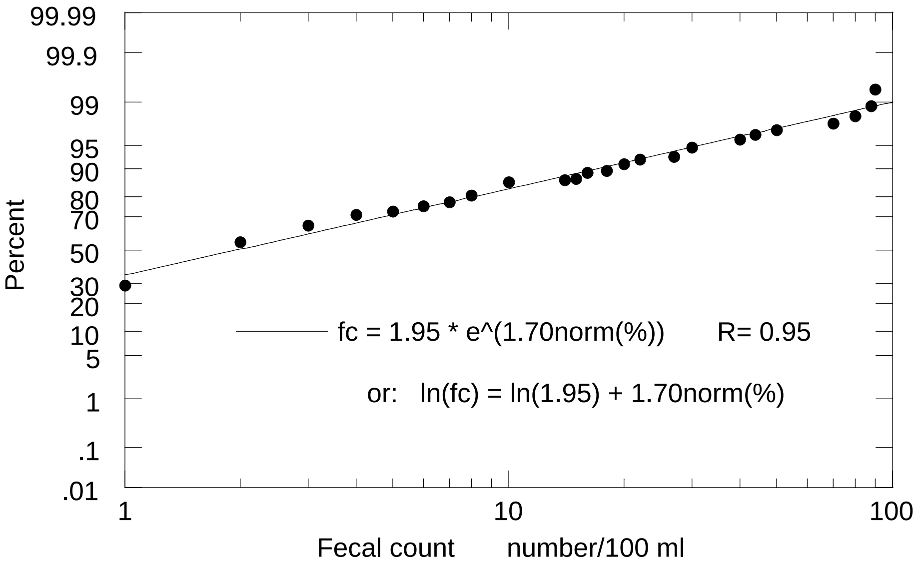


**Figure 2a:** Normal probability plot of fecal count for station 1, on a logarithmic scale. The number of fecal counts less than or equal to 1 is 49; total number of counts is 169.

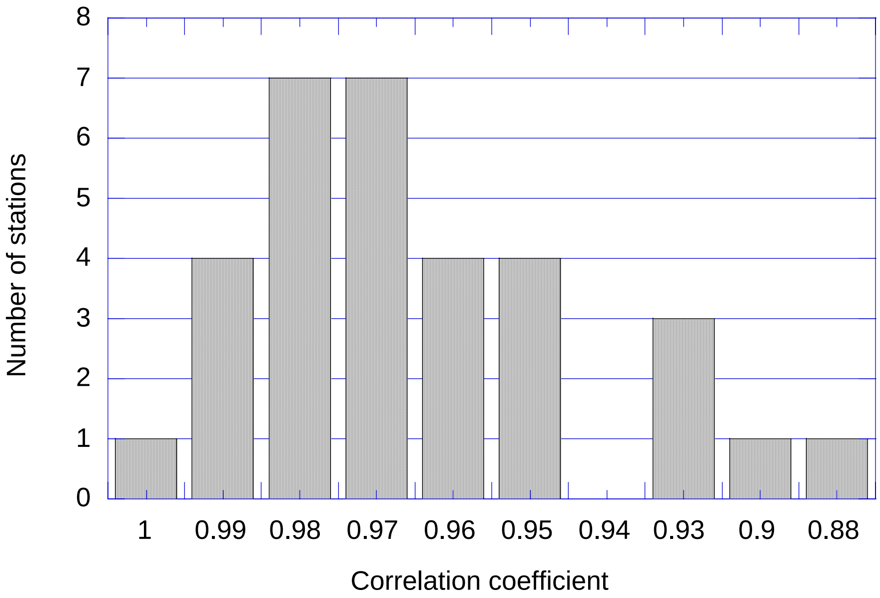


**Figure 2b:** Bar chart showing the distribution in the correlation coefficients for the 32 station normal probability plots of fecal count given in the Addendum.

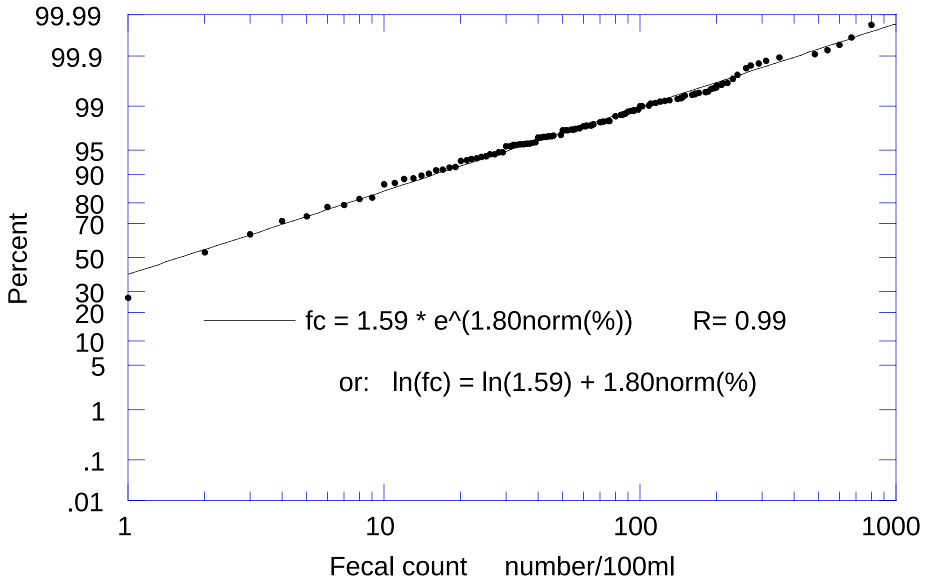
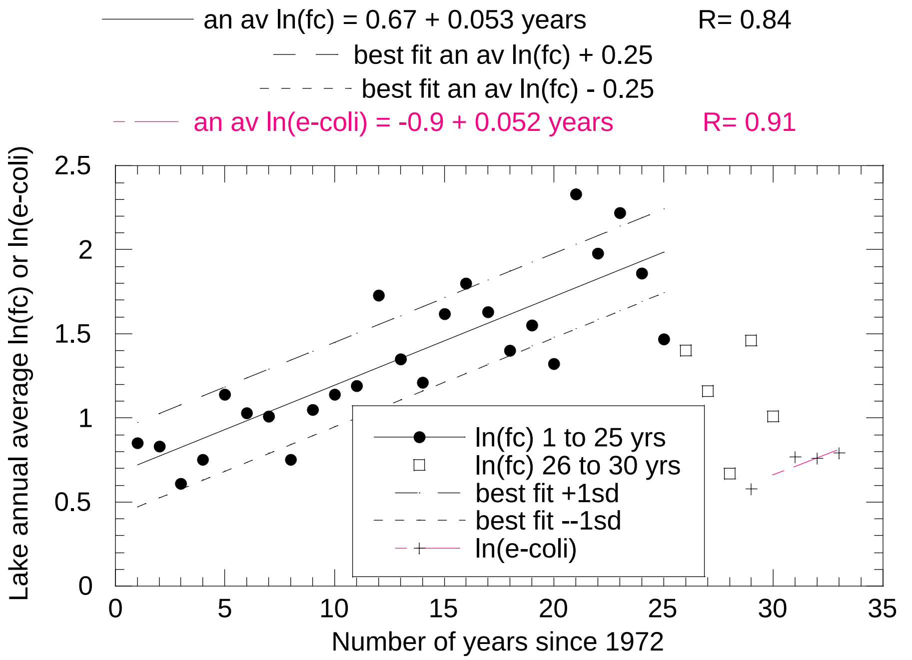


**Figure 2c:** Normal probability plot for all the fecal count measurements for thirty years. The number of counts less than or equal to 1 is 1,472; equal to 0 is 1,160. The total number of counts is 5,493.

**Figure 3a:** Dependence of the lake annual average ln(fc) on years since 1972. The solid line and the dashed lines are linear best fits as noted in the legend.



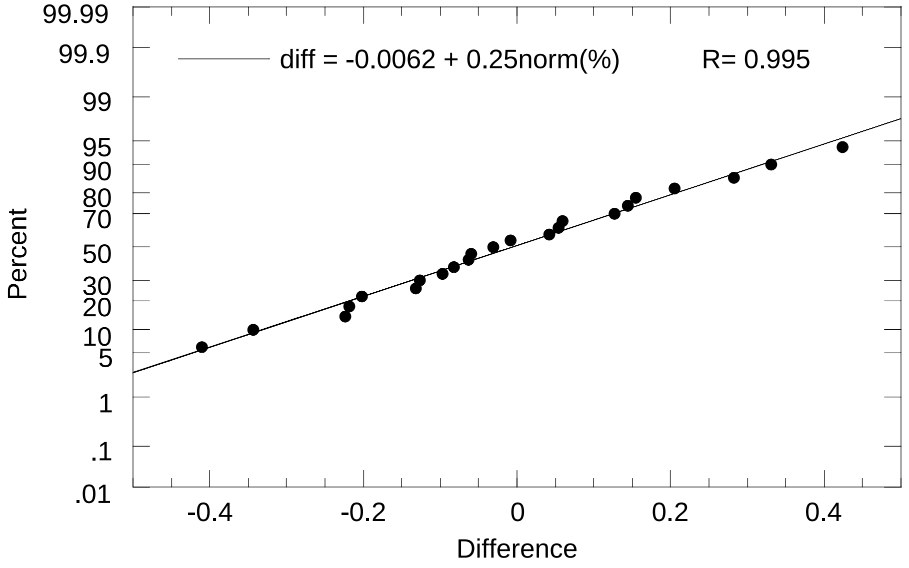
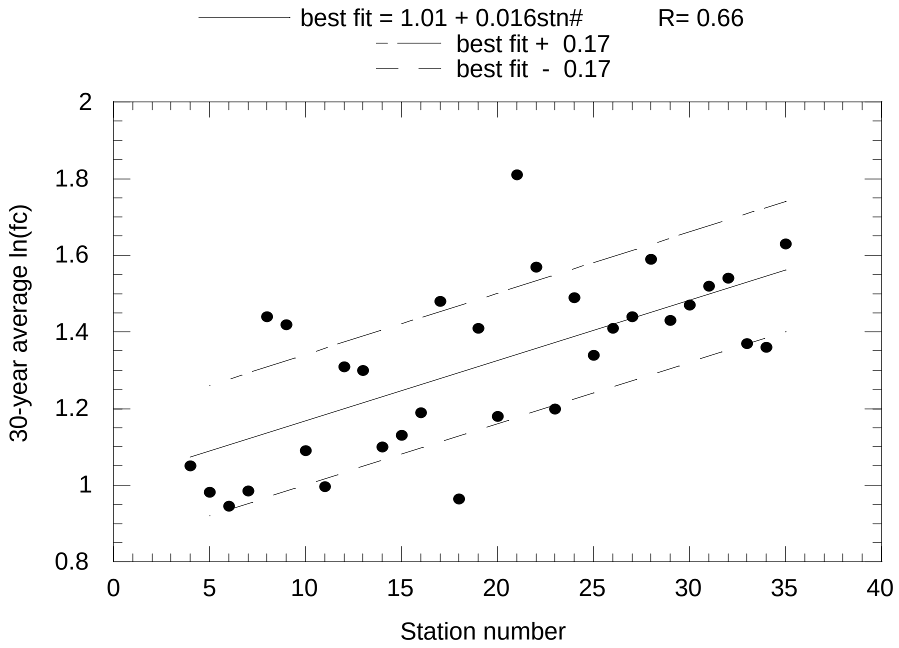


**Figure 3b:** Normal probability plot of the difference between the observed lake annual average ln(fc)s for the first 25 years of the program and the corresponding yearly average given by the best-fit line.

**Figure 4a:** Dependence of the station 30-year average ln(fc) on station number. Stations 1, 2 and 3 have been plotted as stations 33, 34 and 35 respectively. The solid line is the best fit to the data values.



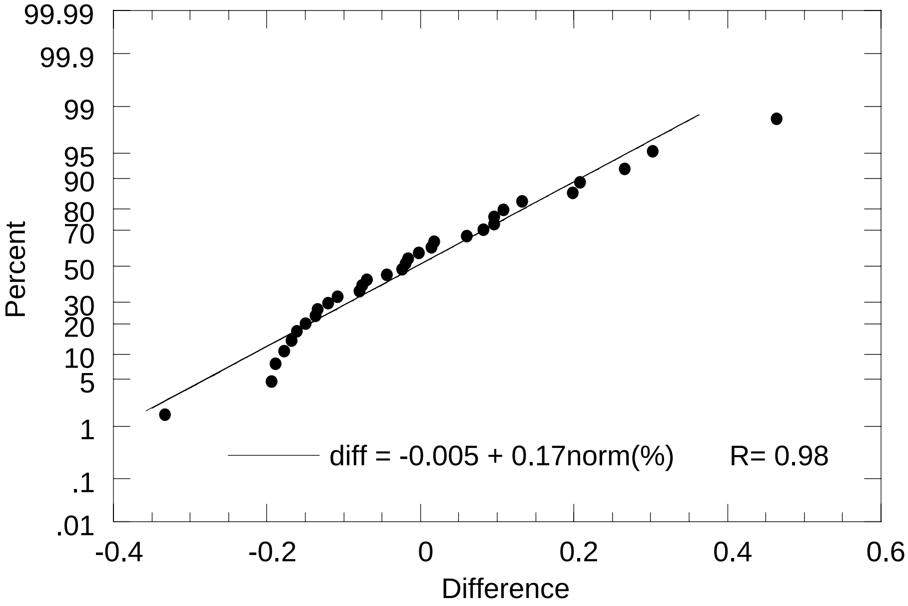


**Figure 4b:** Normal probability plot of the difference between the observed station values for the 30-year average ln(fc) and the corresponding average values calculated from the best-fit line in fig. 4a.

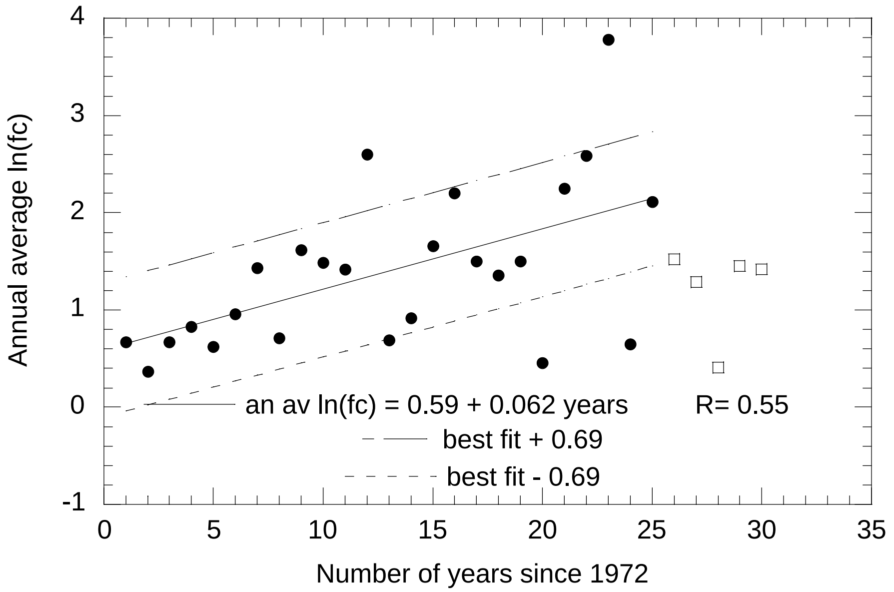


**Figure 5a:** Dependence of the annual average ln(fc) on number of years since 1972 for station 1.

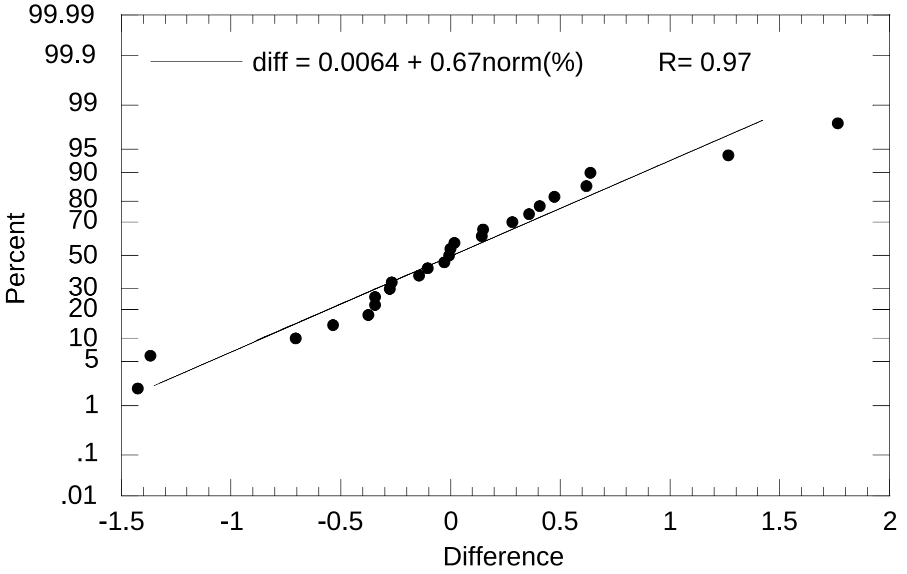


**Figure 5b:** Normal probability plot of the difference between the observed values of the annual average ln(fc) for station 1 and the corresponding values given by the best-fit line in fig. 5a.

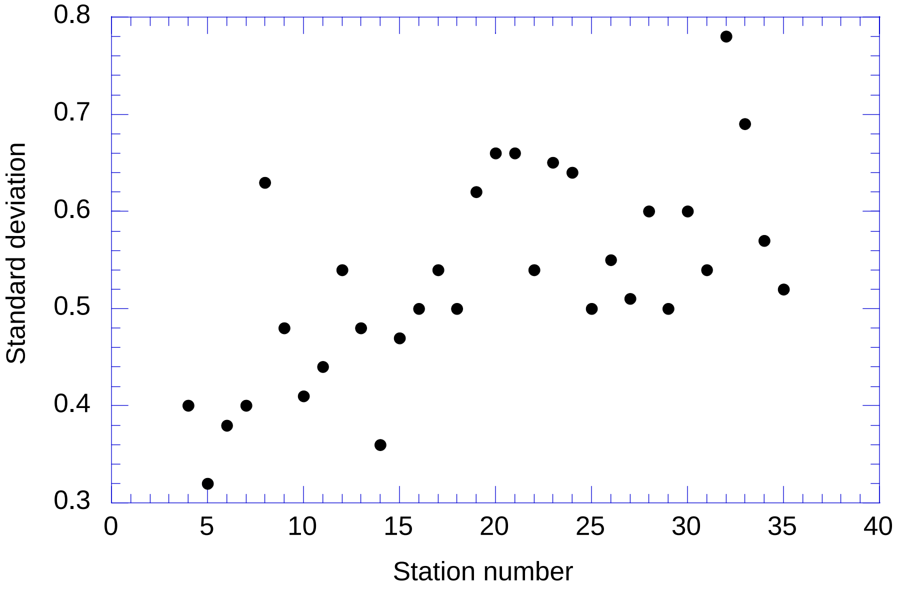


**Figure 6a:** Dependence on station number for the standard deviation of the populations of differences. Stations 1, 2 and 3 moved to 33, 34 and 35.

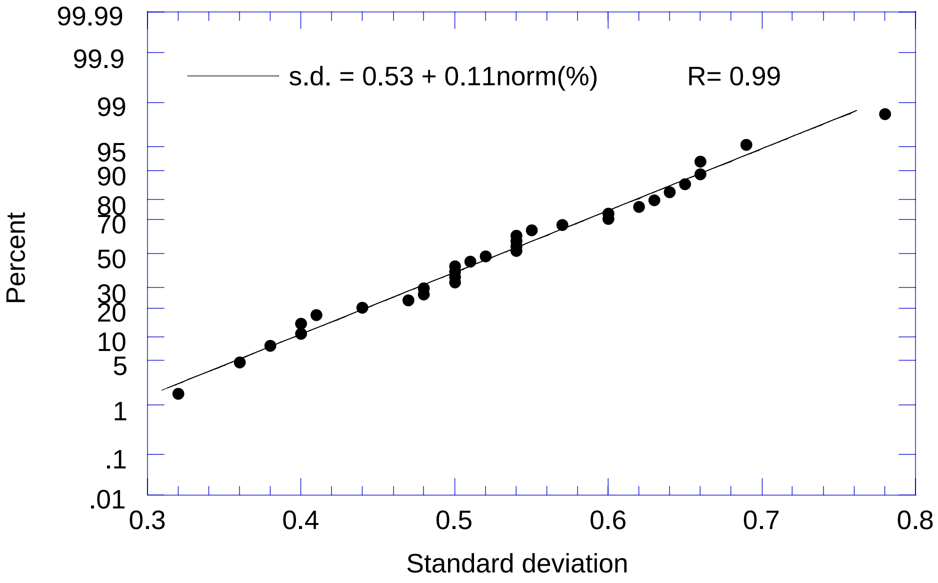


**Figure 6b:** Normal probability plot of the percent of standard deviations less than or equal to a given value for the station difference populations.

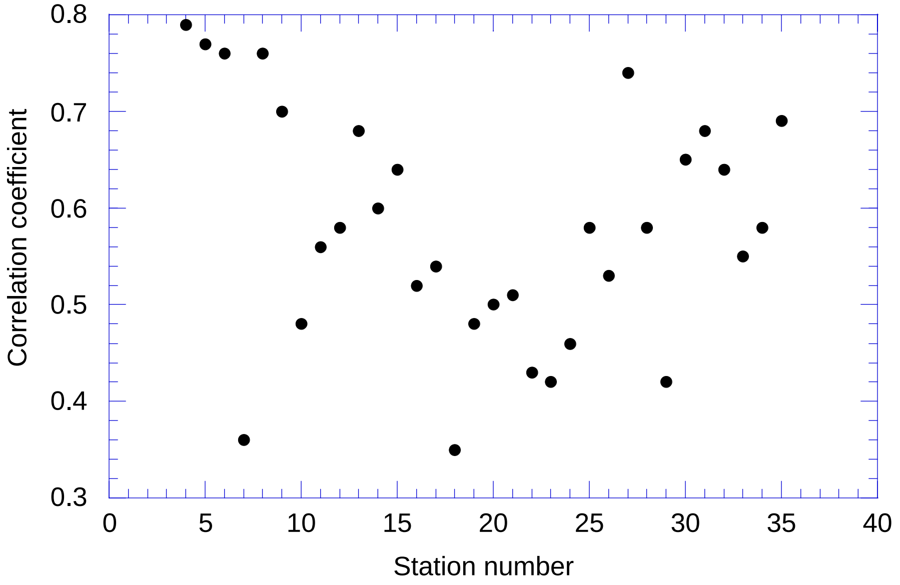


**Figure 7a:** Plot of the correlation coefficient for the dependence of the station annual average ln(fc) on years since 1972 against station number. Stations 1, 2 and 3 shifted to 33, 34 and 35.

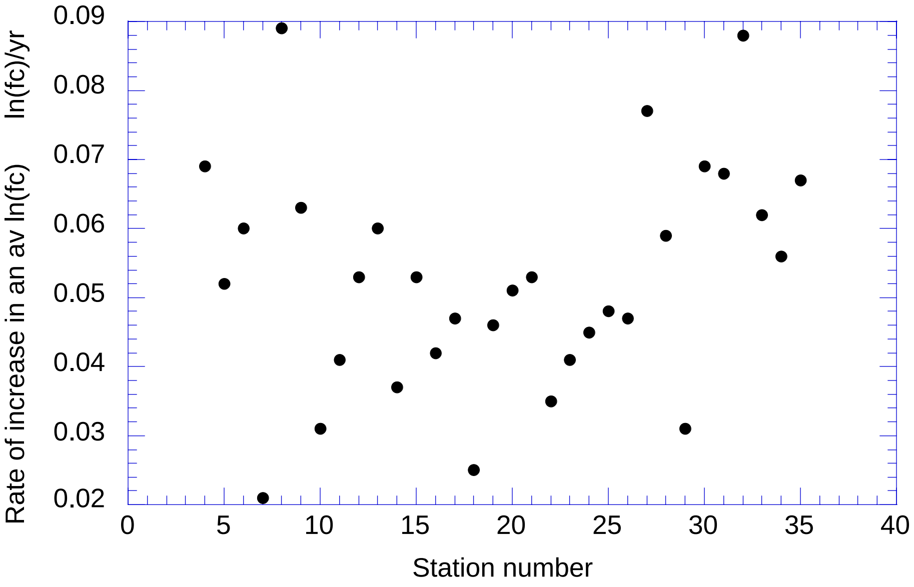


**Figure 7b:** Dependence of the station rate of increase in the annual average ln(fc) on station number for years 1 to 25. Stations 1, 2 and 3 shifted to 33, 34 and 35.

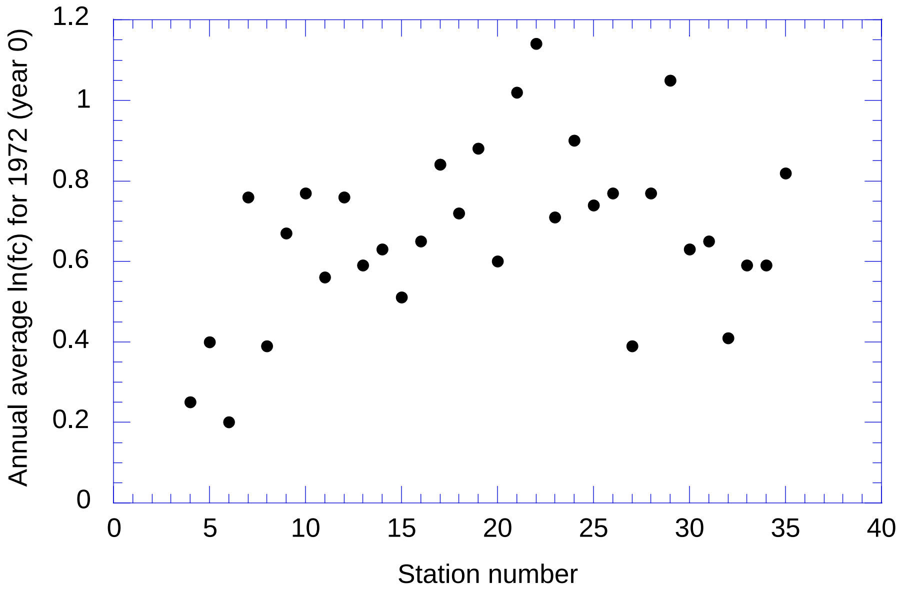


**Figure 7c:** Dependence of the annual average ln(fc) for 1972 (year 0) on station number, determined by the best-fit line to the dependence of the station annual average l(fc) on years for years 1 to 25. Stations 1, 2 and 3 shifted to 33, 34 and 35.

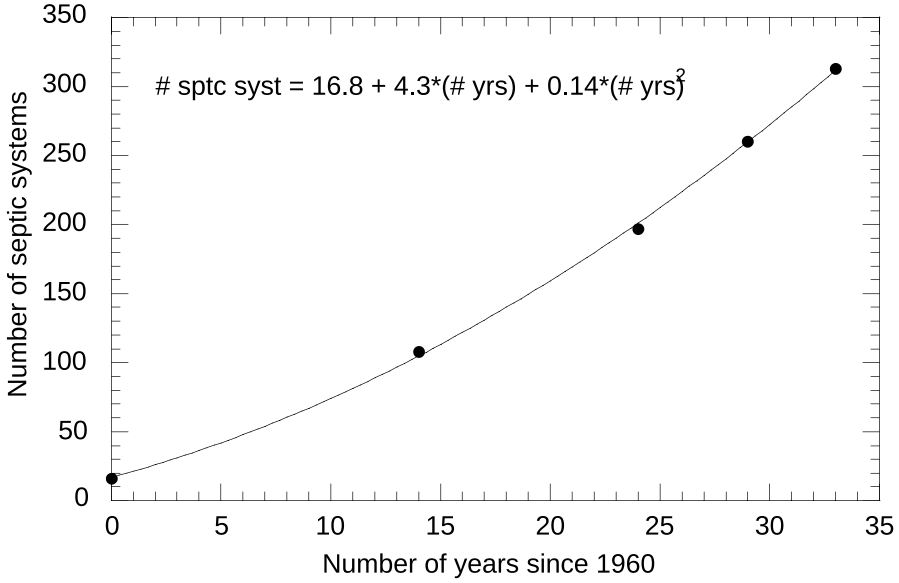


**Figure 8:** Dependence of the number of septic systems on the lake on number of years since 1960.

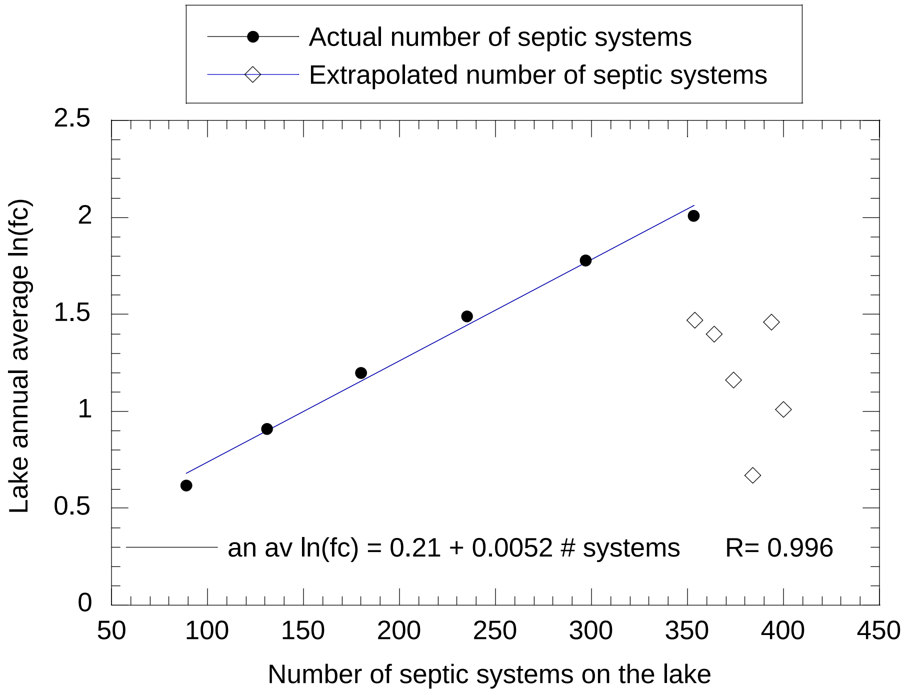


**Figure 9:** Dependence of the lake annual average ln(fc) on the corresponding actual or extrapolated number of septic systems.

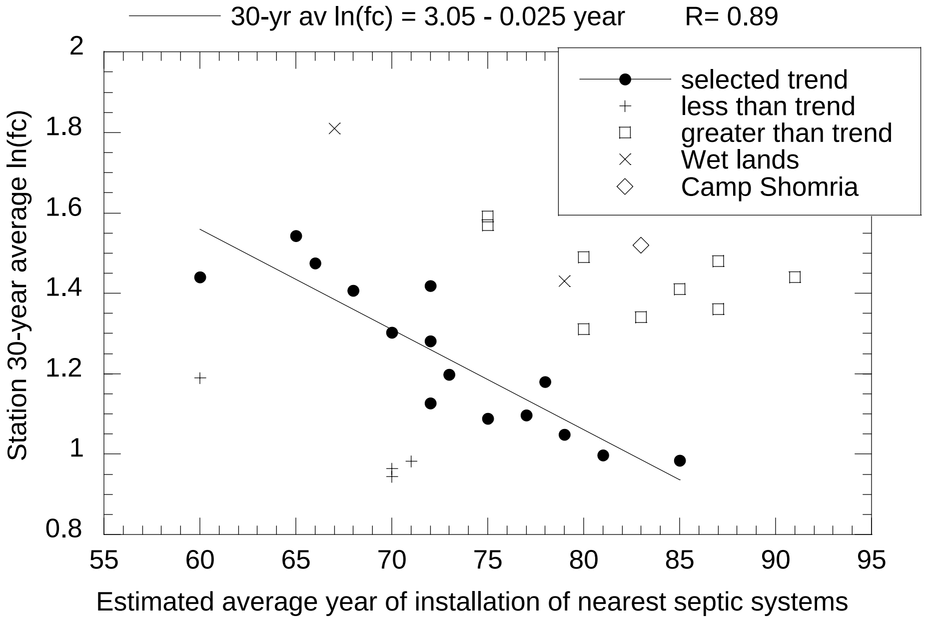


**Figure 10:** Dependence of the station 30-year average ln(fc) on the estimated average year of installation of the nearest septic systems.

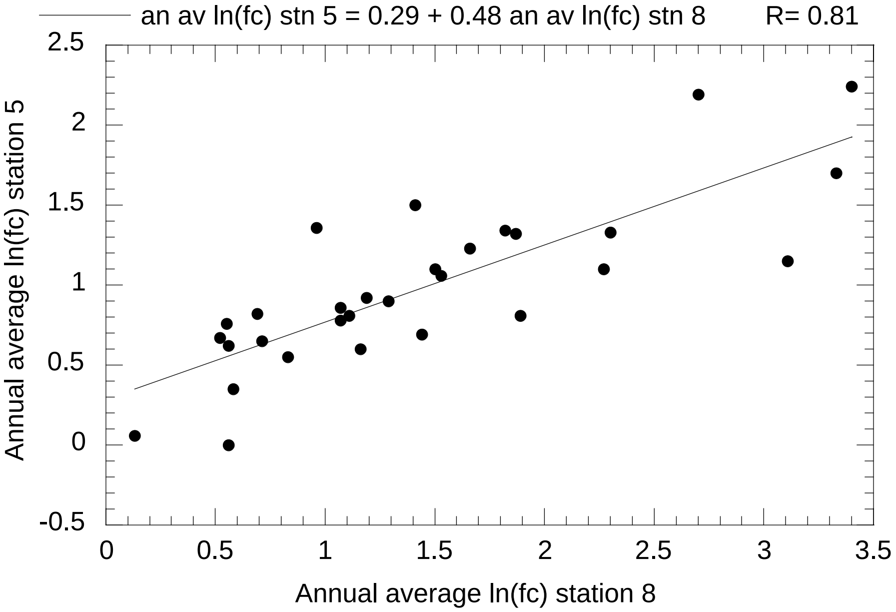


**Figure 11a:** Plot of the annual average ln(fc) for station 5 against the corresponding annual average ln(fc) for station 8.

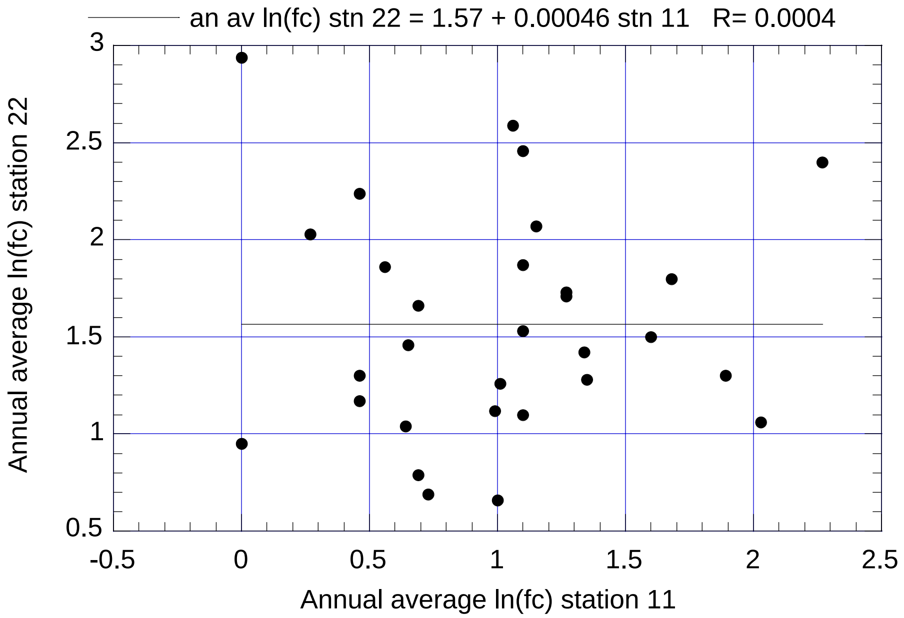


**Figure 11b:** Plot of the annual average ln(fc) for station 22 vs the corresponding annual average ln(fc) for station 11, all years.

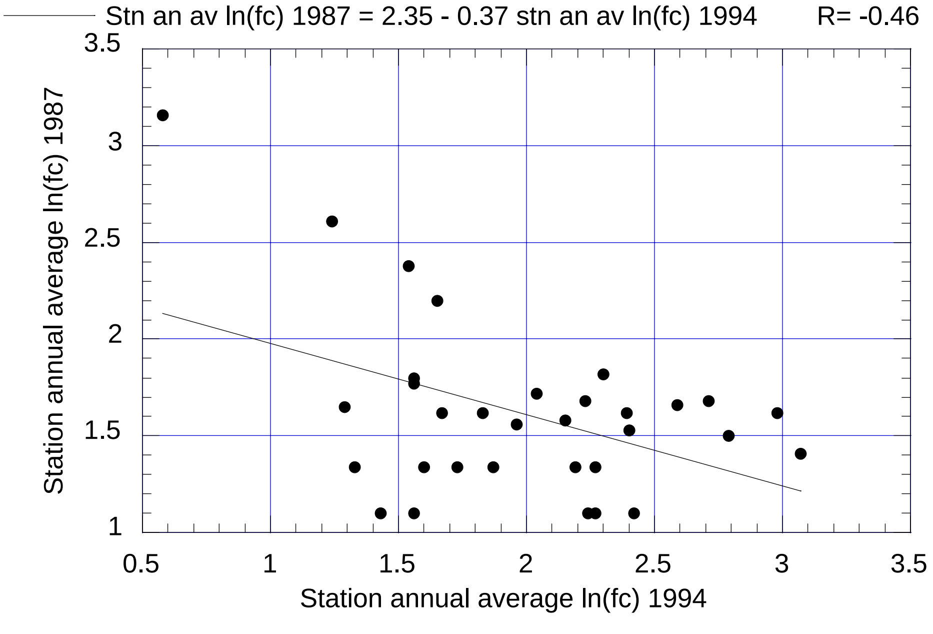


**Figure 11c:** Plot of the station average ln(fc) for 1987 vs the corresponding station annual average ln(fc) for 1994, for all stations.

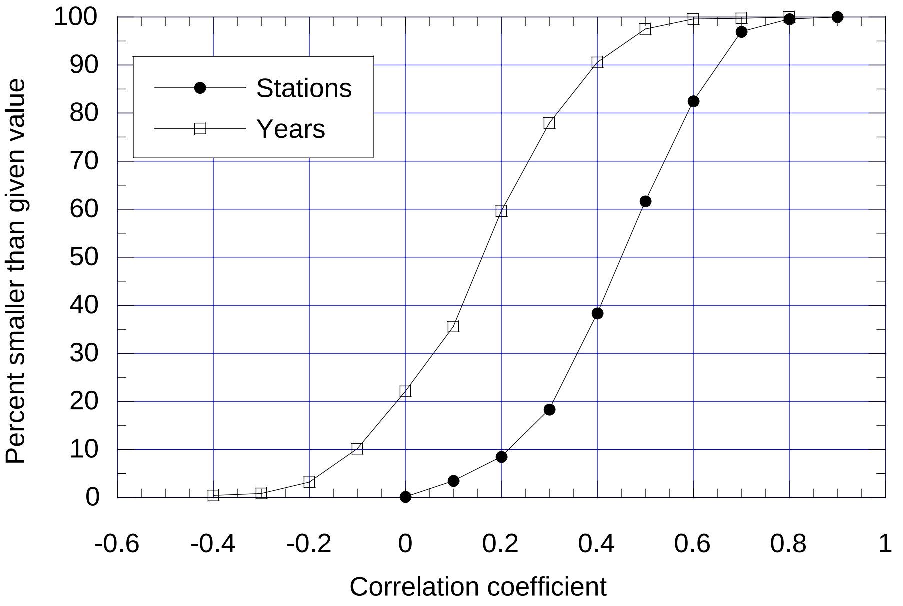


**Figure 12:** Plot of the percent of correlation coefficients smaller than a given value for corresponding values of annual average ln(fc) for stations and years.

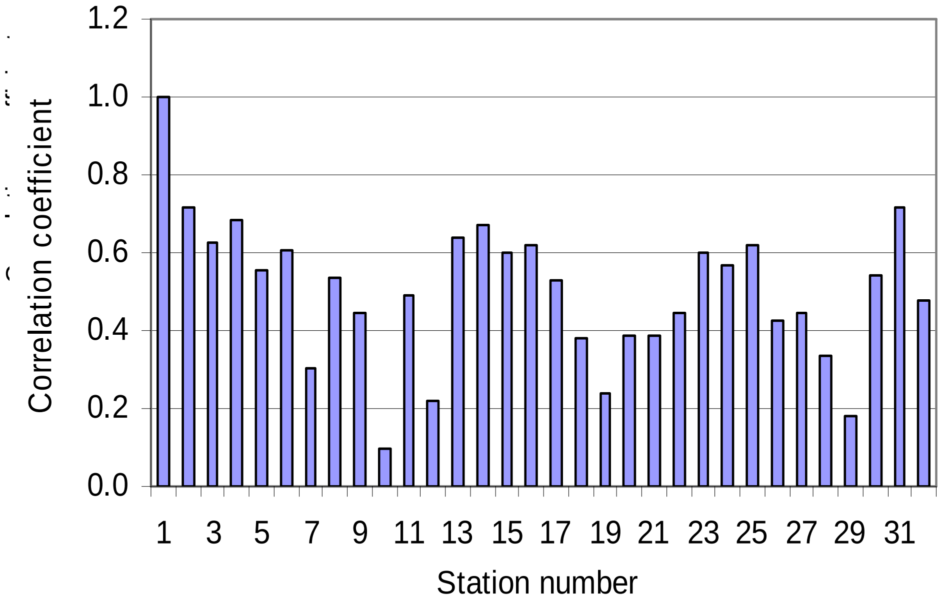


**Figure 13:** Correlation coefficients for station 1 with all stations

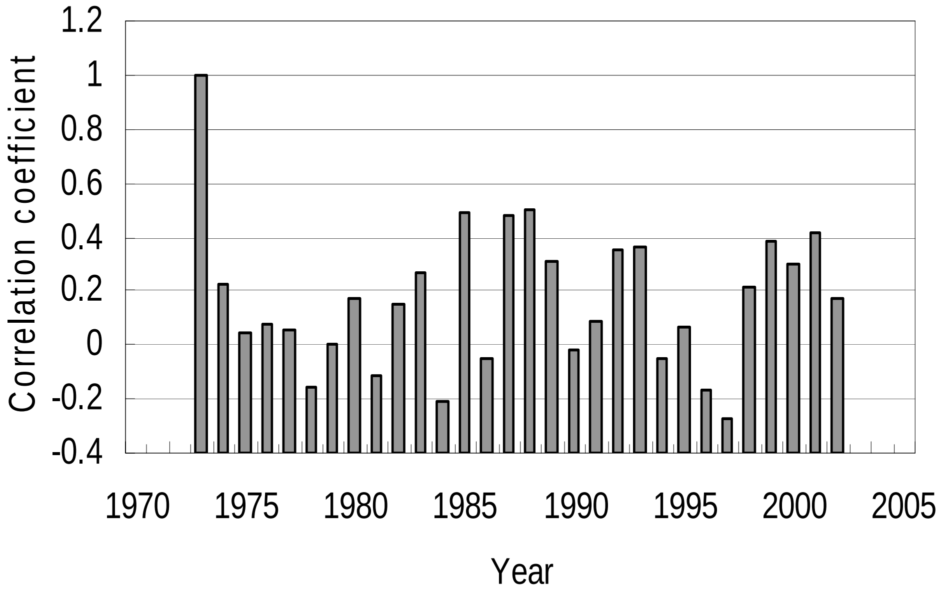


**Figure 14:** Correlation coefficients for 1973 with all years to 2002

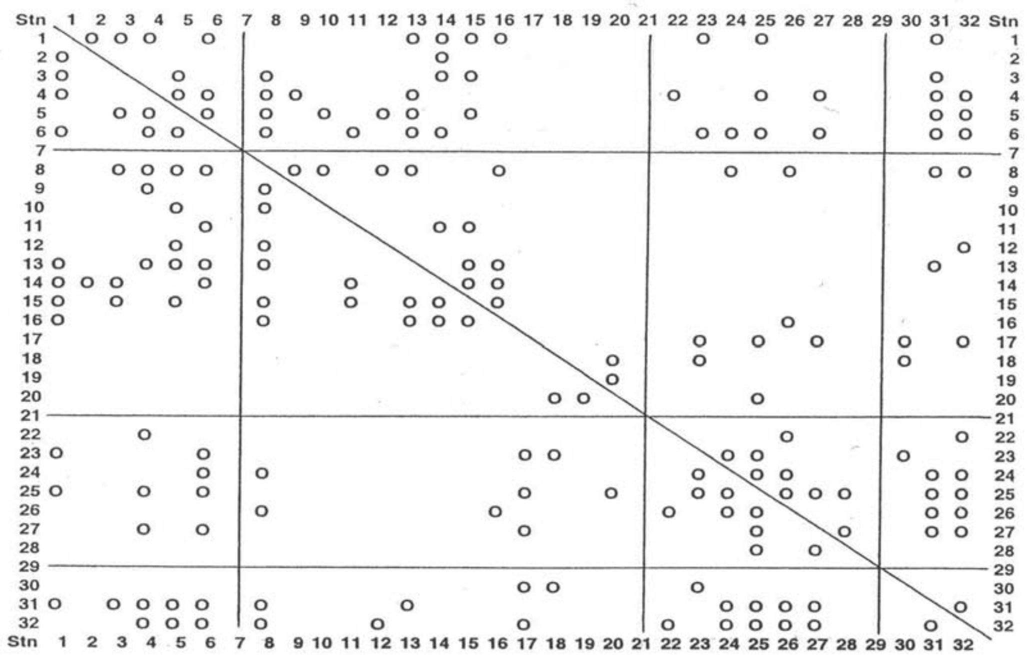


**Figure 15a:** Plot showing the stations with a correlation coefficient equal to or greater than 0.6, but less than 1. The distribution of values above the diagonal is the mirror image of that below. Stations with no correlation coefficients in the range of 0.6 to 1 are shown by the horizontal and vertical lines.

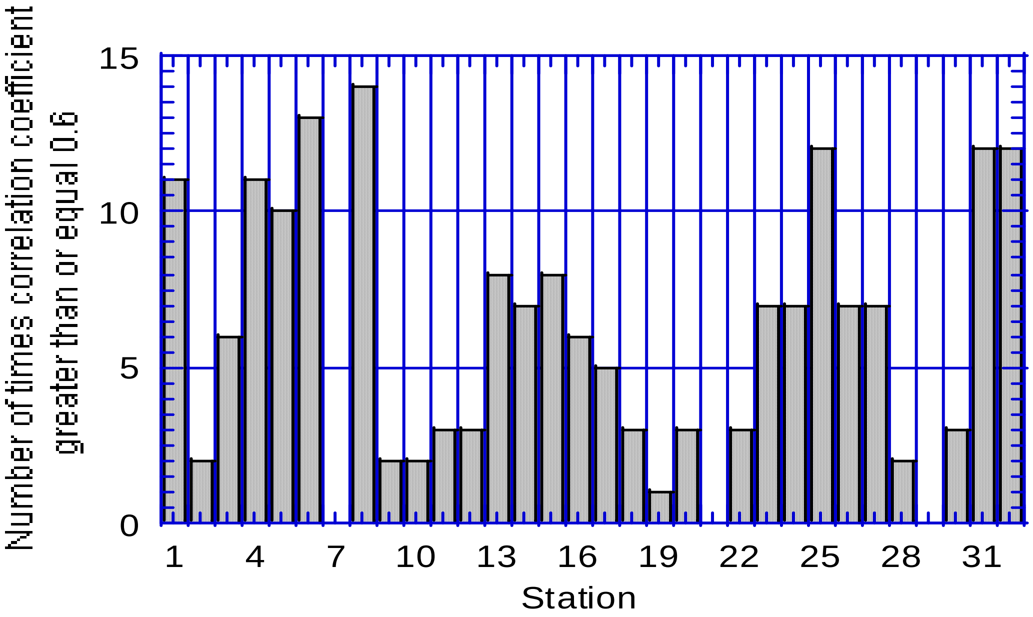


**Figure 15b:** Bar graph showing the number of times a station's correlation coefficient with other stations is greater or equal to 0.6.