WORKING PAPER

A STATISTICAL MODEL OF BACKGROUND AIR POLLUTION FREQUENCY DISTRIBUTIONS

M. Ya. Antonovski
V.M. Bukhshtaber
E.A. Zaleniuk

November 1988
WP-88-102

IIASA
International Institute
for Applied Systems Analysis

A STATISTICAL MODEL OF BACKGROUND AIR POLLUTION FREQUENCY DISTRIBUTIONS

M. Ya. Antonovski
V.M. Bukhshtaber
E.A. Zaleniuk

November 1988
WP-88-102

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

The authors of this paper describe an approach for identifying *statistically stable* central tendencies in the frequency distributions of time series of observations of background atmospheric pollutants. The data were collected as daily mean values of concentrations of sulfur dioxide and suspended particulate matter at five monitoring stations - three in the USSR, one in Norway, and one in Sweden.

In their approach, the authors use well-developed statistical techniques and the usual method of constructing multimodal distributions. The problem is subdivided into two parts: first, a decomposition of the observations in order to obtain a description of each season separately and second, an investigation of this description in order to derive statistically stable characteristics of the entire data set. The main hypothesis of the investigation is that dispersion processes interact in such a way that in the zone of influence of one process (near its mode) the "tails" of the other process are not observed. This permits illumination of interrelations between the physics and the chemistry of the atmosphere.

During the last 15-20 years, a wide range of monitoring programs has been initiated at national and international levels including, for example, the European Monitoring and Evaluation Program (EMEP) under the auspices of the ECE, and the Background Air Pollution Monitoring Network (BAPMoN) under the auspices of the WMO.

The flow of data from the system of monitoring stations has led to national and international projects for the development of extensive environmental data bases such as NOAANET (NDAA), GRID/GEMS/UNEP/NASA, etc. The degree of information obtained should be sufficient for the goals of the analysis but often there is an overabundance of such data. The methods discussed in this paper therefore help in air pollution assessments, particularly with respect to distinguishing the baseline components, and their trends over decades.

R.E. Munn
Leader, Environment Program

CONTENTS

A STATISTICAL MODEL OF BACKGROUND AIR POLLUTION FREQUENCY DISTRIBUTIONS

M. Ya. Antonovski, V.M. Bukhshtaber and E.A. Zaleniuk***

1. INTRODUCTION

1. Problems of Background Air Pollution Monitoring

The natural environment experiences ever-increasing anthropogenic effects. In order to estimate the magnitude of these effects and to prevent dire consequences, it is necessary, first of all, to have unbiased information concerning the actual state of the natural environment.

The results of investigations on the distribution of pollutants from various sources are present in Izrael and Novikov (1985). According to outerspace exploration data, air pollution occurs in the form of "aerosol fields" and entire zones of anthropogenic effects can be distinguished at distances of hundreds to thousands of kilometers from the source of pollution. The degree of anthropogenic pollution in such areas can be determined only by using estimates of normal, or background pollutant concentrations in the atmosphere of the respective regions. The National System for monitoring background values, adopted in the USSR (Izrael, 1984) calls for investigations and observations on the composition, transformation and migration of pollutants. Among the pollutants of priority are ozone, dust, sulfur and nitrogen compounds, lead, mercury, and some other substances. At present special attention is centered on the system of observations on background air pollution. In the USSR the first phase of establishing a network of monitoring stations has been completed (Rovinskii and Buyanova, 1982). A special global system of stations for background air pollution monitoring is being realized within the framework of the program of the World Meteorological Organization (Izrael, 1984). The projects and recommendations in regard to national and global background monitoring systems are widely discussed in the literature (see, for example, Wiersma, 1985; Lynn, 1976).

The major problems to be solved by global background monitoring have been formulated in Wiersma (1985) as follows:

1. Establishment of the relative concentration levels for pollutants, capable of estimating global distributions.
2. Early warning of trends in global pollutant distributions.
3. Establishment of normal concentration levels for parameters of ecosystems and their comparison with concentration levels of impact zones.

The problems bearing on the detection of the continental and global behavior of pollutants are treated in Rovinskii and Buyanova (1982), where it is pointed out that it is necessary to analyze regional background air pollution processes. At

* All-Union Research Institute of Physicotechnical and Radiotechnical Measurements, USSR.
** Natural Environment and Climate Monitoring Laboratory, USSR.

present it is considered that the most satisfactory sites for the location of background monitoring stations are biosphere reserves and other natural reserves. Among the other criteria recognized in Rovinskii and Cherkhanov (1982) for the selection of station sites are geographical zonality, distance from the source of pollution, and the degree of representativeness of the derived data. Representativeness has a special meaning - the absence of any obvious anthropogenic effects on the measured normal pollutant concentrations, and the comparability of the aerometric data with data derived from other stations. Such a comparison in certain cases is fraught with difficulties, on account of the high variability of the aerometric data, resulting from measurement errors as well as from the influence of physical and geographical factors. The latter involve, first of all, the location of stations in regions that differ according to the degree and nature of anthropogenic effects, and according to the processes determining the pollutant concentration variations. An idea of the degree of variability of the estimate of background concentration levels can be conceived from the data presented in Izrael (1984) on the lead concentrations in the lower atmosphere: for the lowland areas of Western Europe - over 100 ng/m^3; for normal regions of the USSR - 2-40 ng/m^3; for mountainous areas of the USSR - 2-6 ng/m^3, for mountainous areas of North America - 4.6-21ng/roman m^3. In order to reduce the variability of the data and of the derived estimates of background concentration levels, it is commonly suggested (see, for example, Rovinskii and Buyanova, 1982) that observational data, averaged in time and space, should be used. In terms of spatial averaging, the global, semi-global, continental and regional types of backgrounds can be distinguished. The concept of regional background allows one to attribute part of the variability of the background level estimates to the specific features of the region and of the locality of the observing station. In this case the background value is determined (Rovinskii and Buyanova, 1982) as the mean of the minimal content values of the given substance during a definite time-interval. Such an approach makes it possible to attribute the effects of high "abnormal" concentrations to local sources, or to associate them with anomalous meteorological conditions.

Hence, the problem of the extraction of "background" information from a series of monitoring data is associated with the development and application of statistical assessment of aerometric data, and can be formulated as the problem of determining statistically stable characteristics of the derived data.

1.2. Probabilistic Approach to Investigations of Background Air Pollution.

The present study is devoted to the statistical analysis of background air pollution monitoring data, having as its objective the design of a statistical model of background air pollution and its application for the determination of statistical characteristics describing the probability laws governing the behavior of impurities in the atmosphere.

Statistical models of air pollution distribution have been widely discussed in the literature (see, for example, Augustinyak and Sventz (1982); Berlyand (1975); Berlyand (1984); Benarie (1982); Mage (1981). However, background monitoring data possess certain specific features, creating difficulties in the use of traditional models (such as, for example, the two-parameter lognormal distribution LN2 (Harris and Tabor, 1956; Larsen, 1961). Measurements of background air pollution levels are conducted in areas where the direct effects of strong pollution sources are practically excluded. This implies that the observed data variability is to a considerable degree due to the effects of large-scale atmospheric processes, that determine the mode of occurrence of different concentration levels in the area, rather than to the effects resulting from point sources of pollution. Most of the air pollution models employed are designed for use under the assumption of the

existence of point sources. Studies of the probability concentration distribution laws for the atmosphere of normal regions allow one to get an idea of the qualitative mechanisms governing the formation of different concentration levels. Statistics, describing these laws, reflect certain regularities in the formation mechanisms and can be used for assessment of background air pollution. Such an approach enables one to validate statistically the intuitively derived concepts of the normal (background) level as the mean of the minimal measurements for a given time-interval (Rovinskii and Buyanova, 1982), or as the minimal but most distinctly expressed concentration level, typical of the region (Izrael, 1984). The derived statistics represent an informative description of the time series of background air pollution monitoring data and, in turn, can be used to obtain explicit inferences bearing on the nature of the measurements and their behavior.

The major stages in designing, analysis and application of the statistical model of background air pollution are as follows:

1. Statistical analysis of background air pollution monitoring data. Studies of the logarithmic concentration distribution functions for data series of different time-intervals.
2. Investigation of the possibilities of describing the logarithmic concentration series by multimodal distributions, and the physical prerequisites for the origin of multimodality.
3. Simulation of data series in terms of composite distributions of a specific type, and development of graphical methods for estimation of performance parameters.
4. Description of seasonal observational data series by central tendencies of multimodal frequency distributions. Development of techniques for identification of statistically stable grouping intervals.
5. Analysis of statistically stable grouping intervals and their manifestations in seasonal and multiyear data series.
6. Analysis of the air pollution components described by statistically stable grouping intervals; comparative analysis of the components and their manifestations at different background monitoring stations; development of recommendations for the assessment of background concentration levels.

2. PRESENT STATUS OF STATISTICAL ANALYSIS OF AIR POLLUTION

2.1. Review of the Application of Probabilistic Methods to Descriptions of Air Pollution.

Probabilistic models are often used for the description of aerometric data, and provide the basis for obtaining estimates and approximate descriptions of the distribution of air-pollutants. The models are used in many aspects of air quality planning, when prediction is one of the main aims of the study. The application of theoretical-probability and statistical methods to solving such problems has been discussed in a number of publications. For instance, one review (Hunter, 1981) treats many characteristic aspects of the application of statistical methods to problems of environmental control. The major problem on which these authors center their attention concerns the existence of the gap between the demands for a good descriptive model, connecting observed processes with the environmental parameters introduced into the model, and the real possibilities for assessment and measurement of such parameters. A typical systematization of the applied models can be found in Benarie (1982). Dividing air pollution models into descriptive, computational and predictive, Benarie assigns time-series analysis to the first

method. In the second and third cases, including regression and simulation, the methods demand information on the state of independent parameters that have a direct bearing on pollution dispersion. Examining the possibility of application of the first method, the author has presented characteristics examples illustrating their limited applicability. For instance, experiments with the Box-Jenkins model (analysis of time-series) for the extrapolation of air pollution data from 100 days of observations, showed that estimates of model concentrations for the 101st, 102nd, etc., days were no better than any random predictions. To improve the forecast, it is necessary to introduce into the model some assumptions concerning meteorological or other conditions affecting the pollutant distribution and to assume continuity of these conditions, both for initial and extrapolated data. A similar, if not even greater appeal to the development of precise concepts on processes occurring in the atmosphere is to be found by Benarie (1982) in various computational models. In citing examples of the parameters employed in these models, Benarie (1982) expresses his doubts as to the possibility of predetermining many of the matching parameters in the context of a logical description of natural pollution conditions. In order to define the bounds within which these methods can be applied, Benarie (1982) proceeds on the basis of two considerations. The first concerns the objective of the study. If for deriving mean estimates of certain pollution characteristics, or for a general description of pollutant distributions, statistical methods are to be useful, they should reflect certain general or typical characteristics of atmospheric processes. The second consideration directly concerns those characteristics of atmospheric processes, that render impossible the application of certain detailed analytical schemes, and the prediction of the behavior of impurities in the atmosphere. The design of such models commonly proceeds under the assumption of monotypic behavior of the parameters and mode of pollution distribution within an area that should be large enough to preserve certain common properties in the course of a period long enough for investigations, but should be small enough that it might be attributed to some common properties reflected in the parameters introduced into the model. In Benarie (1982) certain characteristic time-periods are presented for the existence of such areas, within which adequate functioning of most of the proposed models is ensured. For an area with a side of 300 km, this time-period is, according to different estimates, between 12 and 75 hours with a pronounced mode in the histogram at about 45 hours.

Quite a large number of examples can be offered of the successful application of mathematical models to the description of pollution in different environments (see, for example, Anokhin and Ostromogil'skii, 1978; Ostromogil'skii, 1982). Berlyand (1975,1984) demonstrates the possibility of using mathematical models, based on equation-solving of turbulent diffusion in the atmosphere, for the prediction of a possible sharp increase in concentrations during a period lasting from several hours up to several days, under unfavorable climatic conditions. Examples illustrating the successful application of regression models for the prediction of air-pollutant concentrations are presented in Singpurwalla (1972); the model parameters are chosen not on the basis of physical considerations, but so as to derive the best forecasts, and Singpurwalla (1972) finds it necessary to produce evidence justifying the use of such "non-physical" models. An example of a model designed in terms of probability considerations on the behavior of pollutants over long time intervals, and that serves to estimate the dynamics of pollutant distributions in space and time, is presented in Augustinyak and Sventz (1982) based on data derived from several stations.

Thus we see that, notwithstanding the serious difficulties encountered in the application of air-quality models due to the complicated nature and rapid occurrence of atmospheric processes, interesting results can nevertheless be

derived. In each case this can be achieved by clearly defining the class of problems that should be solved by the model, and the choice of adequate mathematical or statistical methods for their realization, taking into account the difficulties cited at the beginning of this chapter. Concentrating their attention on analytical treatment and comprising data derived from several background monitoring stations, the present authors recommend the use of a body of statistics, reflecting certain general characteristics in the behavior of atmospheric pollutants, imposing a minimum of assumptions on the usage of the data, and not offering any direct meaningful conclusions; such a design philosophy is feasible for the description of the data. Such a description in itself often furnishes the basis for the development of new hypotheses concerning the data and leads to important results from testing these hypotheses. The use of applied statistical techniques and analytical treatment of the data for solving problems bearing on the assessment and description of air pollution, is exemplified in the construction of statistical models describing the behavior of pollutants in the atmosphere according to their frequency distributions of concentrations.

2.2. Descriptive Air-Pollution Models

Descriptive air-quality models have been employed in routine investigations since the 1950s. A review of existing models can be found, for example in Mage (1981).

One of the earliest models to be used is the two-parameter lognormal distribution LN2, with a density function:

$$f(x) = \frac{1}{2\pi\sigma \cdot x}\, x \exp\left[\frac{-(\ln x - \ln a)^2}{\sigma^2}\right].$$

Particle sizes formed during crushing, also dust particles, are well-described by the LN2, and it was assumed that use of this function could be expanded to describe particulate matter in the atmosphere, not only according to size, but also according to concentration distributions (Zimmer and Tabor, 1959). The major conclusion drawn in Zimmer and Tabor (1959) on the basis of the results of suspended particulate measurements performed over cities and beyond urban areas, is that, notwithstanding certain deviations, the concentrations reveal a lognormal distribution. It should be mentioned that the widespread applicability of lognormal distributions was illuminated by Aitchison and Brown (1957). In 1961 the LN2 distribution was also used for the description of gaseous air-pollutant concentrations (R.J. Larsen, 1969a). It was established that the CO concentrations in the area of Los Angeles "... reveal a tendency towards a lognormal distribution" (Larsen, 1969a). Later the LN2 model was widely used by the same author to describe all types of air pollution (Larsen, 1969b). The wide application of the LN2 model, that renders possible its use for estimation of the air-quality under practically any measurement conditions, furnishes the basis for designing techniques for the assessment of air pollution characteristics as statistical parameters of the proposed model. The use of these parameters in setting national standards is described in Larsen (1969b).

Let the concentrations of pollutants measured during successful time-intervals be denoted as $C_0, C_1, C_2, \ldots, C_n$. These values result from many meteorological, geophysical and other factors, and Khan (1973) suggests using the assumption that changes in concentrations from one time interval to another can be described in the form:

$$C_j - C_{j-1} = p_j \cdot C_{j-1} \qquad (2.1)$$

where C_j and C_{j-1} are the concentrations measured during the time intervals j and $j-1$. The random variable p_j represents the impact from many effects that form the random realization of the concentration value during the time $j-1$. Equation (2.1) is commonly known as the law of proportional effect. For concentrations to obey this law, it is postulated that the change in the concentration during any time interval is proportional to the concentration that has been attained up to this moment.

Equation (2.1) can be rewritten as

$$(C_j - C_{j-1}) / C_{j-1} = p_j \ .$$

Then

$$\sum_{j=1}^{n} \frac{C_j - C_{j-1}}{C_{j-1}} = \sum_{j=1}^{n} p_j \ .$$

Assuming that the changes at any time instant are small, we get

$$\sum_{j=1}^{n} \frac{C_j - C_{j-1}}{C_{j-1}} = \int_{C_0}^{C_n} \frac{dx}{x} = \ln C_n - \ln C_0 \ ,$$

from which it follows that

$$\ln C_n = \ln C_0 + p_1 + p_2 + \ldots + p_n \ .$$

The central limit theorem permits one to state than $\ln C_n$ is of asymptotic normal distribution, regardless of the p_j distribution and, consequently, the random value C_n is of lognormal distribution. This result is also given by Aitchison and Brown (1957).

As is indicated in Aivazyan et al. (1983), C_0 can be regarded as a "true" value C in an idealized scheme, when the effects of all random factors have been eliminated, and the $p_1, p_2, \ldots, p_n$ quantities are the numerical expression of the effects of the above-mentioned random factors. In this connection, it is noted in Aivazyan et al. (1983) that although the values of the logarithmic distribution of the random quantity are formed as random errors of a certain "true" value C, the latter emerges, in the long run, not in the role of a mean value, but as the median. This serves to define the role of the median as the best estimate of central tendency air-pollutant concentrations.

It is interesting to note also that the direct application of the central limit theorem presumes independence of the random variables p_j. Khan (1973) does not claim that such independence can be proved, although he considers that indirect proof can be found in the results of empirical investigations.

Real observational data seldom show precise correspondence to the LN2 law, even in cases when its application can be strictly proved. Data on pollutant concentrations in the atmosphere also include deviations from the "pure" LN2. This has served and still serves as the basis for the critical analysis of the LN2 model and for the use of alternative models. Examples of this can be found in Khan (1973), Mage (1980, 1981) and Mage and Ott (1975); related problems are discussed in Horowitz and Baracat (1979), Roberts (1979) and Soeda and Sawaragi (1979).

Lynn (1976) was among the first to study the applicability of several probabilistic models to air pollution data. The analysis involved the normal law, LN2, the three-parameter lognormal distribution LN3, the I and IV types of the Pearson distribution and the Gamma-distribution. The conclusion was drawn that the LN2 was the best of all the above-cited distributions. Here a situation occurring frequently in statistical analysis was observed. Namely, in many cases a distribution can be selected (even among those cited above) that most closely approximates the distribution of the sampled data. However, not one of these distributions can be applied to the description of all types of samples of aerometric data. For their description, several distributions should be employed. However, the LN2 distribution is of greatest value.

During the 1960s-1970s many case studies were accumulated concerning the application of LN2 to the description of air pollution data. The observed deviations from the LN2 and the regularities perceived in them were used by several authors to design models that could ensure a high degree of applicability for the description of the available data, as good as that of the LN2 model. Such an approach is exemplified in Mage (1980, 1981) and Mage and Ott (1975), where several types of distributions suitable for this purpose are proposed, and in de Nevers et al. (1979), where the possibility of describing the data by employing combined distributions is discussed. This method of describing the data is characterized by the search for the best statistical design for the description of event-data, secured at the expense of general model applicability.

For instance, in Mage and Ott (1975) the authors conclude that all air pollution data studied by them reveal a common behavior in their deviations from the LN2 - their distribution functions plotted on lognormal probability paper demonstrate characteristic "curving". In order to take account of this effect, they suggest using the LN3 model - a three-parameter lognormal distribution with a density

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma(x-\lambda)} \exp\left[-\frac{1}{2}\cdot\frac{(ln(x-\lambda)-ln\alpha)^2}{\sigma^2}\right].$$

The fact that this is not the only means for describing such deviations from LN2 is apparent from Mage (1981). In this work concerning the best description of the data, it is proposed to use the limited distribution models, with the introduction of nonstatistical prerequisites concerning the probable origin of such distributions in the problems under study.

In de Nevers et al. (1979), after analytical treatment of a large number of event-data on atmospheric particulate matter, the authors distinguished not one (as in the former example) but four types of deviations from the straight line, typical of distribution functions plotted on lognormal probability paper. These four types are depicted in Figure 2.1. The authors analyzed in detail the reasons for such deviations and proposed to describe them by a combination of two LN2 distributions. In the same work, an example is given illustrating how in reality such a meteorological situation leading to a "composite" distribution can arise, and an analytical treatment is presented of real data corresponding to such a situation. It is obvious that, from the point of view of increasing model applicability, the last line of attack on the problem is best. By retaining the well-studied and convenient LN2 distribution as the base-distribution, one may perform a uniform description of practically all observed deviations from LN2 by postulating that several different types of meteorological processes affect the concentrations.

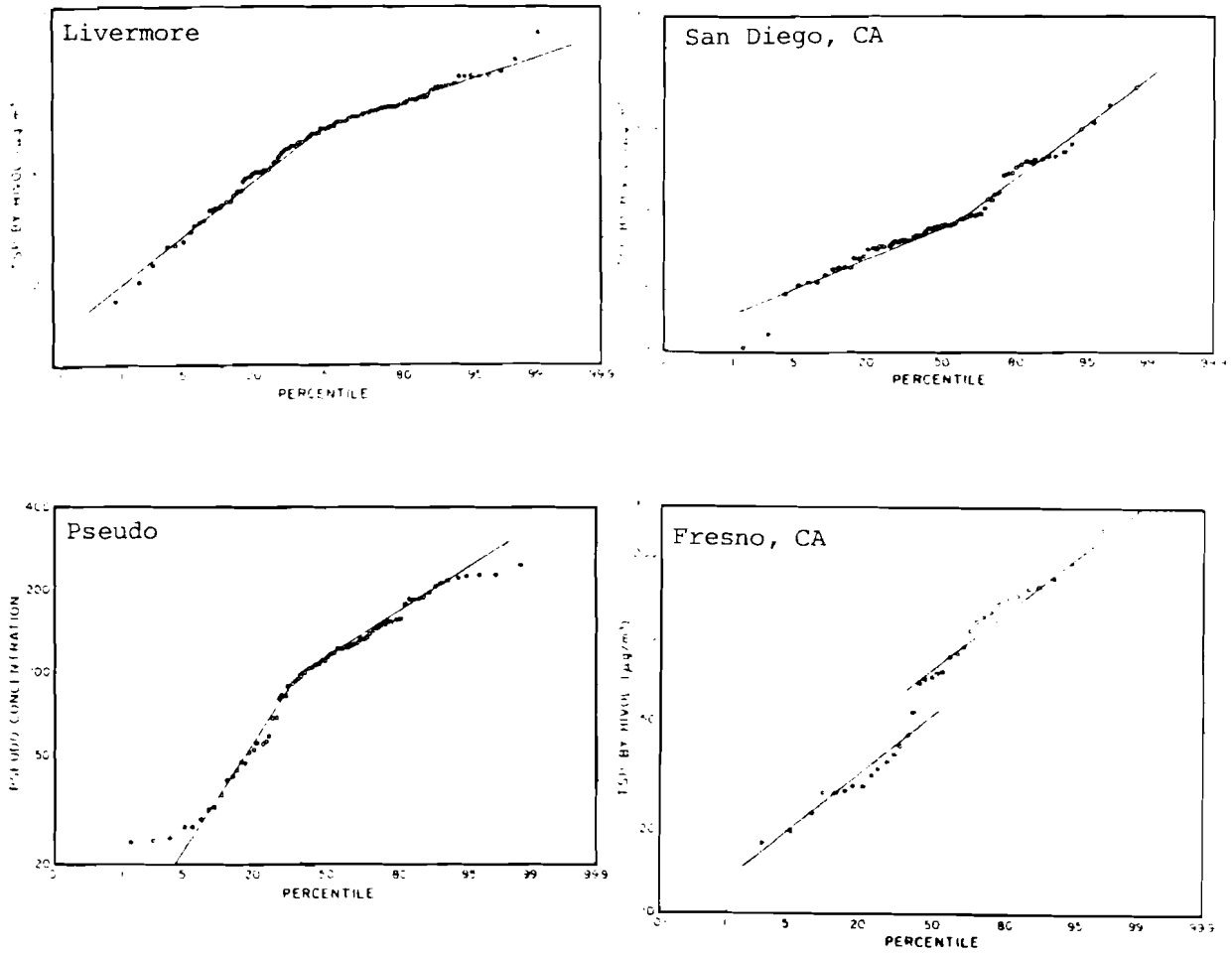


Figure 2.1 Four types of deviations from the LN2 distribution for data on concentrations of aerosol dust (de Nevers et al. 1979).

2.3. The Use of Descriptive Models for Studies of Background Air Pollution Monitoring Data.

Studies of different air pollution probability models gives ever-more convincing evidence that analytical treatment of air pollution data should be performed by means of thorough analysis of the chosen statistical model, and by the use of that model that from the point of view of the statistical criteria provides the best descriptions of the event-data. For this purpose a set of automatic facilities is proposed in Bencala and Seinfeld (1976) which performs the choice of the best distribution from the point of view of the maximal similitude principle. Such an approach which, probably, is applicable to the analysis of air pollution data at the impact level, can hardly be used for the description of background monitoring data. The construction of such statistical air pollution models leads to a loss of generality in the physical presentation of pollutant concentrations since, on the basis of statistics proposed by different models, it becomes impossible to establish any common factors controlling the formation of pollution concentration frequency distributions.

In the majority of problems using statistical air pollution models, the authors are interested, first of all, in the possibility of the application of the model to obtain extreme value statistics. For instance, most of the studies mentioned in the present chapter relate to the air quality standards adopted in the USA, and the formulation is in terms of the frequency of exceeding maximum permissible concentrations during a given period of time (week, month, year) and for a given averaging time. For example, the CO concentrations averaged for one hour might be permitted to exceed 35 ppm only once a year, which is equivalent to the statement that the 1-hour time-averaged concentrations of CO may exceed 35 ppm in no more than 0.011% of the observations. It is obvious that if the hourly concentration distributions of CO are known, and a probability model of this distribution exists, then from a number of observations with specified occurrence, it is possible to define the distribution parameters and to evaluate whether they conform to the standard distribution under the specific conditions.

The formulation of air pollution background monitoring problems in such a context has not been encountered (Rovinskii and Buyanova, 1982). Our attention is centered mainly on the determination of the mechanisms governing the formation of different pollution concentration distributions; and the statistical design employed should be sufficiently general that it could be applied to different air pollution background monitoring time-series. The choice of the model from among numerous statistical models should be prompted by the problems to be solved; and the degree of generality of the model should correspond to the degree of generality of the results.

We have chosen the two-parameter lognormal law of pollutant concentration distributions - the LN2. Of all the laws studied, this is most widely used, owing to the fact that it performs well for all pollutants within any observational area, and for various time averages and, most likely, reflects certain general conditions in the formation of different air pollutant concentration levels. Taking into account the fact that we are often confronted with the necessity of studying distributions that deviate from LN2, we adopt here the hypothesis postulating an increase of model applicability by the use of combined LN2 distributions.

3. CONSTRUCTION OF A STATISTICAL MODEL SIMULATING BACKGROUND AIR POLLUTION FREQUENCY DISTRIBUTIONS

3.1. Estimates of Background Concentration Levels

Measurements of background atmospheric pollution concentrations have been obtained over a long period of time merely for the general evaluation of air quality, and have been episodic in nature. The areas chosen for such measurements were usually located far from industrial pollution sources, outside of urbanized districts. They included at time mountainous areas, located at great heights above sea level. In Burtseva et al. (1982) some data are given on lead concentrations in Western Europe and North America: in non-urbanized areas of Norway in 1971-1972, at heights of 3600 m above sea level in Switzerland, and during a four-year (1968-1971) observational cycle in California at heights of 3800 and 1860 m above sea level. These data provided the basis for determining the mean concentration values of lead for the USA and Central Europe. For the USA, the mean has been taken equal to 8 ng/m^3; for Central Europe - 4 ng/m^3. At the same time certain specific features were noted in the behavior of lead in different physical-geographical areas; for instance, the Californian data revealed a seasonal trend in concentration, with summer maximum and winter minimum, and the absence of a correlation between lead concentrations and suspended particulate matter, while in England a maximum was apparent in the winter concentrations of lead and other heavy metals.

In Rovinskii, Burtseva et al. (1982) and Rovinskii, Egorov et al. (1982), an attempt is made to analyze and summarize the data available in the world literature on the distribution of the major pollutants in nonindustrial areas. Due to the geographical position of the areas under study, it is assumed that these are background data. It is pointed out that the existing data are related to episodic observations performed at different time-intervals and in different localities. The reason for this is that only at the end of the 1960s and the beginning of the 1970s did background air pollutant concentrations begin to attract attention, when it was realized that anthropogenic effects are of large-scale global importance. Because of this, answers to many questions concerning the long-term state of the atmosphere cannot be obtained. For instance, it is impossible to answer the question raised in Rovinskii, Burtseva et al. (1982) as to whether there are upward trends in the concentrations of air pollutants. In general, it can be stated that the world data studied in Rovinskii, Burtseva et al. (1982); Rovinskii, Egorov et al. (1982) reveal very great variability in space and time. In Rovinskii and Buyanova (1982) it is suggested that different types of backgrounds should be distinguished: global, hemisphere, continental and regional. It is suggested that the minimal mean values for different time intervals should be used for estimates of normal (background) concentrations. This idea conveys implicitly the concept of the nature of background air pollution. As a matter of fact, the proposed types of data-averaging, allowing for data smoothing over given time-periods and given spatial areas, and eliminating the effects of concentration increases in local zones and during short-time intervals, enable one to derive an integrated picture of the background setting. The construction of such a picture requires local measurements continuously conducted over a long time. In Rovinskii and Buyanova (1982) stress is laid on the particular importance of regional background investigations for different regions taken together. The regional background regularities are, seemingly, the only predictors of regional, continental and global long-term behavior of pollution concentrations.

Air-pollution background monitoring stations have been established in the USSR and in many other countries within biosphere reserves, also in localities not subjected to the influence of any apparent sources of pollution. These programs involve measurements of air-pollutant concentrations. Since 1976 such aerometric data have been accumulated in the USSR which makes it possible to estimate background concentration levels for particular regions, to analyse the data for different regions and for the world as a whole, to study the principles governing the formation of different concentration levels, and to obtain estimates of normal air pollution concentrations over continents (Burtseva, Lapenko et al. 1982; Burtseva, Volonseva et al. 1982; Pastukhov et al. 1982). Annual data publications have begun (see for example, Bulletin of background pollution of the natural environment in the region of East-European Members-Countries of CMEA, 1982, 1983).

The data on heavy metal concentrations in the area of the "Borovoe" station are discussed in Burtseva, Lapenko et al. (1982). In the case of lead, the lower limit of measurement error was found to be 0.5 ng/m^3, the coefficient of variation not exceeding 20%. According to the data presented in Burtseva, Volosnea et al. (1982), lead concentration measurements at background monitoring stations are performed within an accuracy of about 10%. The data represent daily mean concentrations in the lower atmosphere. Analysis of the histograms of daily mean values for lead concentrations measured over a four-year period, 1977-1980, shows a strong asymmetry in the frequency distribution, with a pronounced concentration maximum in the left lower quartile and a long "tail" in the right upper quartile. Burtseva, Lapenko et al. (1982) used the histograms for simple statistical inferences on the possibility of obtaining relatively stable estimates of lead concentration levels, the major maxima in the frequency distribution being chosen. For the samples in Burtseva, Lapenko et al. (1982), such an interval included 65-85% of the observations. The upper limit of the interval was taken as the upper estimate of the background concentration level; thus, according to the authors' estimates, the background concentration level in the atmosphere for lead in the area of the "Borovoe" station is between 0.5 to 30 ng/m^3. For the four years studied, no clearly evident time changes in the concentration distributions occurred; during 230-310 days per year, the concentrations varied within the limits typical of normally pure continental areas.

The proposed method for estimation of the background concentration level has a number of shortcomings. One of these is that the method does not explain the behavior of the concentrations in the frequency distribution. For instance, in Burtseva, Lapenko et al. (1982), the authors could not offer a plausible explanation for the increase in the frequency of lead concentrations in the interval of 30-60 ng/m^3 in 1979, or the presence of arsenic concentrations in the interval of 3-6 ng/m^3) for 30% of the observations in 1980 (the arsenic background level being defined at 1-3 ng/m^3). Analysis of the possible various types of effects of meteorological and other conditions on concentration variations fails to explain the observed events (Burtseva, Lapenko et al. 1982). Analysis of background monitoring data for sulfur dioxide was performed in Pastukhov et al. (1982), the average monthly concentrations varying between 0.3 to 18.9 μg/m^3 during the period of investigations - from 1977 to 1981. The highest values were recorded during the winter, the lowest - during the summer, which is a general result found also in data from the Repetek and Berezin B.Z, background monitoring stations. The annual cycle is associated with two factors - the considerable increase in anthropogenic emissions from fuel-burning during the cold periods of the year, on the one hand, and the drop in the rate of oxidation of sulfur dioxide, on the other hand. Analysis of the monthly concentrations of sulfur dioxine, separately performed for the warm and cold seasons, made it possible for the authors (Pastukhov, 1982) to estimate the sulfur dioxide concentration level in the area of the "Borovoe" station at

0.5-1.0 $\mu g/m^3$ - for the warm period and at 3.2-13.7 $\mu g/m^3$ for the cold period. Similar analysis of the average monthly values at the "Berezin B.Z." and "Repetek B.Z." background monitoring stations gives the values 1.0-2.4, 10 $\mu g/m^3$ - for the first and 0.3, 1.0 - for the second. Analysis of meteorological conditions and trajectories indicated that the extreme concentration values cannot be unambiguously correlated with the vector wind directions in the "Borovoe" station area. The derived estimates for different observational areas are incommensurate and doubt arises concerning their possible use in estimating characteristics of continental and global background concentration levels.

In Szepesi (1982) estimates are presented on air pollution characteristics, plotted on different scales. It is suggested that the horizontal extent of the districts should be determined from two meteorological considerations: the lower measure is specified by the distance, within which the background level is determined by the mixing processes in the atmospheric boundary layer, whereas the upper one - by the relative extent of the fetch over which the meteorological parameters remain constant. By such estimates, boundaries were defined (Szepesi, 1982) that delimit the area of action of the estimates of the regional air pollution level; under the assumption of regional uniformity, they should operate over a radius of between 20 to 300 km. Notwithstanding the rough nature of these estimates, it is possible to formulate the problem of determining the magnitudes of background concentration levels by means of comparison of data from several stations. In Szepesi and Fakete (1987) it is assumed that continental and global air pollution background concentration levels are subject to the influence of processes occurring over thousands and tens of thousands of kilometers. It is obvious that the mutual influence of such processes leads to an intricate picture of formation of pollutant concentration levels. If the station network covering the continent is sufficiently dense, it might be possible to define differences in background air pollution levels, to distinguish zones where the factors affecting the formation of different concentration levels are uniform, and to determine a certain integral characteristic describing the mean background level of pollutants for the entire continent. It might be interesting to compare such mean levels, derived daily at many stations for different time-periods, in order to check the hypothesis postulating that lengthy periods occur when the background level does not undergo changes across the continent as a whole, although daily variations are registered in the aerometric data for each station. This hypothesis underlies the assumption of the existence of a continental and a global background value. In a somewhat different formulation, this hypothesis can be found in Izrael (1984).

In Augustinyak and Sventz (1982) approximate estimates are given for the number of observing stations that permit one to plot the area describing the behavior of the pollutants through time within a certain territory. When the linear law is used, the minimal number of measurement points is 9, for the square law - 18, the cubic law - 30, etc. At present, the density of background monitoring is not sufficient to apply such models.

3.2. Statistical Analysis of Background Monitoring Data.

The data to be used are from three background monitoring stations - Borovoe, Berezin biosphere reserve, and Repetek biosphere reserve in the USSR. Descriptions of the data are given in bulletins (Bulletin of background pollution of the natural environment in the region of East-European Members-Countries of CMEA, 1982 and 1983). The techniques used to derive the data and a discussion of their reliability can be found in Burseva, Lapenko et al. (1982), Burseva, Volosneva et al. (1982) and Pastukhov et al. (1982).

In the present study, three pollutants have been selected - sulfur dioxide, lead, and total suspended particulates, for which daily observations were available during 1976-83 at the Borovoe station and 1980-83 at the Berezin and Repetek stations. The three pollutants differ according to their physical-chemical behavior, and the stations are located in different physical-geographical areas. A joint analysis of the sampled data with a view to finding common statistical characteristics can enable one to define some common principles governing the behavior of air pollutants, and can provide a basis for designing techniques for evaluation of background pollutant concentration levels on a wide scale - both in space and time.

The first stage of statistical data analysis should be the construction of the statistical data model. Then, the statistical characteristics describing the data series can be investigated, and their applicability for obtaining non-statistical conclusions can be explored. Techniques for designing statistical models and the use of the statistical information in hydrometeorological and geophysical applications are described in Aivazyan et al. (1983), Gruza and Reitenbakh (1982) and Kleiner and Gradel (1980). In Aivazya et al. (1983), some general techniques used: in designing statistical models are presented. In practice two different methods of analysis are used: mathematical, relying on theoretical-probabilistic considerations, and computational - by way of direct reproduction of the model function on a PC. The first method calls for hypotheses and *a priori* assumptions concerning the data that should serve to validate the choice of model; the second requires some preliminary formalized knowledge of the data, that could be reflected in algorithmic form, and could be used to develop or refine the theoretical-probabilistic method. In the present study, both of these mutually complementary methods are employed: the first stage, presumably, should involve the development of certain general theoretical-probabilistic concepts of the model.

In Burtseva, Lapenski (1982) several histograms were examined that describe the heavy metal frequency distribution at the Borovoe background monitoring station. These histograms exhibit a lognormal distribution, with the mode shifted to the left and a long "tail" at the right. Histograms of this type can be perceived in the distribution of all three pollutants, sampled for statistical analysis at all stations and for any period. It is therefore possible already to utilize the logarithmic form in the analysis and for checking the hypothesis of a lognormal distribution.

In Figures 3.1, 3.2, 3.3, plots are shown that characterize the lead concentration distributions at the "Borovoe" station during the four-year period of observations. Because much of the subsequent analysis is based on studies of these plots, we shall dwell upon them. These plots portray graphically the empirical density and cumulative distribution functions (3.1 and 3.2), and depict the deviation of the empirical density function from the theoretical one (3.3). Methods for graphical assessment of the distribution parameters are discussed in Murzewski and Sowa (1978-1979) and problems bearing on graphical estimates are treated also in Rubin (1976), Aivazyan et al. (1983) and other publications. Kleiner and Gradel (1980), note that the use of graphical methods is generally typical of statistical analysis of geographical data. Those authors consider that the reason is that geophysical data usually involve daily, seasonal, annual and inter-annual variations, apart from other more pronounced effects, characteristic of short-time intervals, and, inasmuch as the major objective of these methods is to illuminate these relationships and structures, representation of the data in the most recognizable form becomes particularly important. For evaluation of the degree of agreement of the data with the chosen LN2 distribution, various methods can be used. Methods of evaluation, in particular for the lognormal distribution are discussed in Rovinskii and Cherkhanov (1982) while in Selvin (1976) and Gnanadesican and Kettering (1972), several methods are examined for numerical estimation of the model distribution under conditions of different types of data errors. Many of the methods

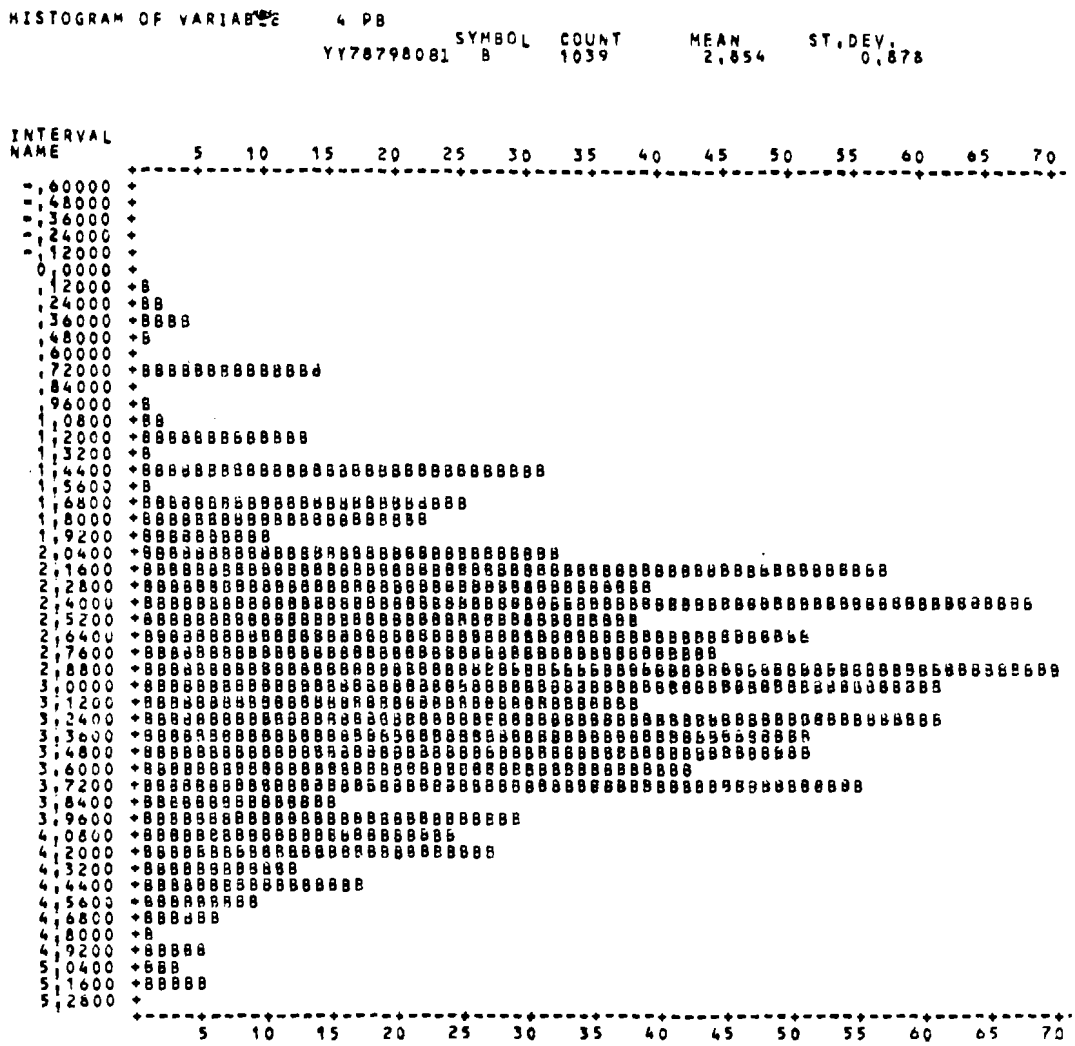


Figure 3.1 Histogram of logarithmic concentrations of lead. Borovoe station, 1978-1981.

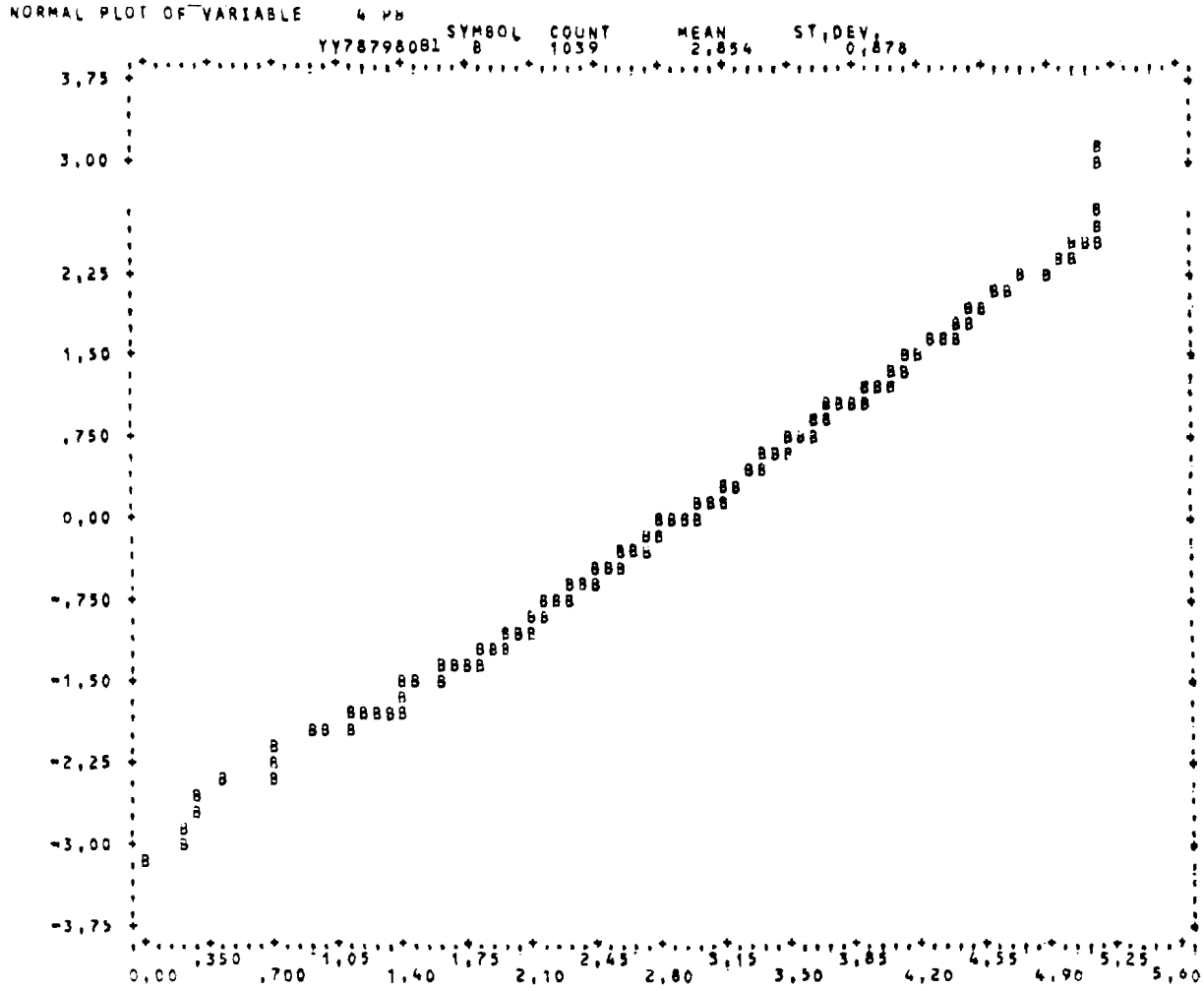


Figure 3.2 Normal plot of cumulative logarithmic concentrations of lead. Borovoe station, 1978-81.

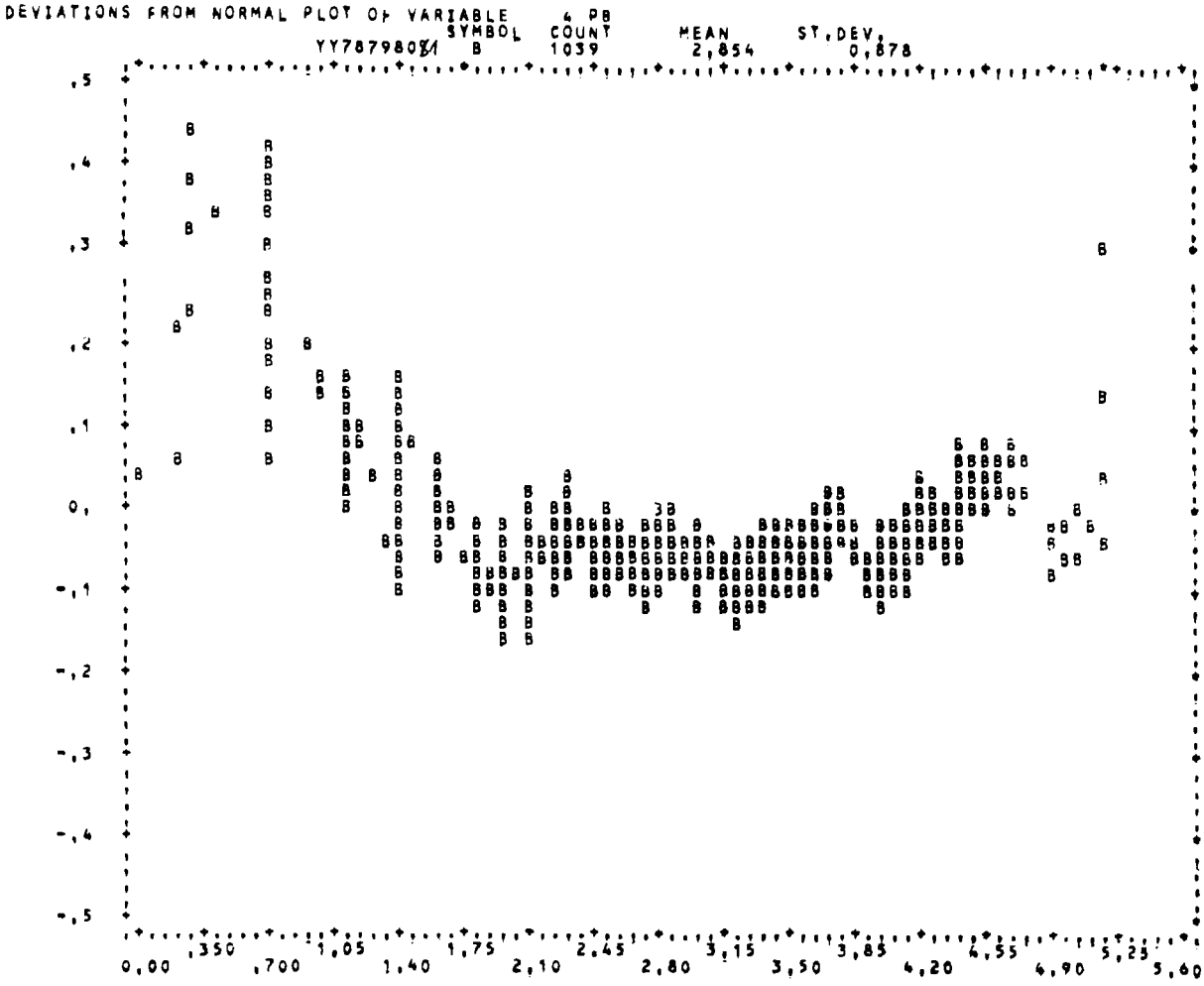


Figure 3.3 Deviations from normal plot of logarithmic concentrations of lead, Borovoe station, 1978-1981.

discussed in these works are designed to derive numerical statistics that best describe the empirical distributions. Graphical qualitative evaluations of the distribution pattern are also used. In order to determine how much the observed distribution differs from a given theoretical distribution, various criteria of goodness of fit can be used. However, according to Kleiner and Gradel (1980), the numerical result derived from their use does not indicate in what places and for what reasons the observed distribution deviates from the model one. In the case of a normal distribution, there would be an exactly symmetrical bell-shaped curve in Figure 3.1, a straight line in Figure 3.2, and a very narrow spread in Figure 3.3.

Histograms are often constructed when the number of observations becomes large. The length of the interval is taken equal to

$$h = \frac{x_{max} - x_{min}}{10 \cdot lg(N)+5} , \tag{3.1}$$

where x_{max} and x_{min} are the maximal and minimal points on the logarithmic concentration scale for the given sample, N - the number of observations in the sample.

The distribution function is plotted on normal probability paper as distribution quantiles against the observed variable,

$$F(x_n) = \Phi^{-1}\left[\frac{(3n-1}{(3N+1)}\right] \tag{3.2}$$

where n is the number of the variable x_n in the variational series, arranged in ascending order. The value of the $F(x_n)$ function corresponds to the probability

$$(3n-1) \;/\; (3N+1)$$

of the centered and normalized normal distribution

$$\Phi(t) = \int_{-\infty}^{t} N(x;0,1)dx ,$$

where

$$N(x;0,1) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left[-\frac{1}{2}\cdot\frac{x^2}{\sigma^2}\right]. \tag{3.3}$$

Equation 3.4 represents Equation 3.2 with linear trend removed:

$$F'(x_n) = \Phi^{-1}\left[\frac{3n-1}{3n+1}\right] - \frac{\bar{a}-x_n}{\bar{\sigma}} \tag{3.4}$$

where $\bar{a}$ and $\bar{\sigma}$ denote the sample average and variance, respectively. This equation shows the deviation from the straight line, specified by estimates of parameters $\bar{a}$ and $\bar{\sigma}$, and thereby gives a qualitative display of the degree of agreement between the event-data and a LN2 distribution, graphically revealing the nature of inconsistencies with the theoretical distribution.

Discussions of the problems concerned with plotting and evaluation of the distributions by employing graphs of this type can be found in Aivazyan (1983) and Kleiner and Gradel (1980).

As can be seen from Figures 3.1, 3.2, 3.3, the empirical density and distribution functions, as expected, differ from the theoretical ones. The question as to how to proceed in the case of such deviations is discussed at length in Aivazyan

(1983). It is obvious that if we have available a sufficiently large class of model densities, for example the Pearson curves, we can find a density function that best approximates the behavior of the empirical density under study, and, in the long run by expanding the number of hypothetical model densities, we can attain a very high degree of approximation, even in cases of "crevices" in the model density frequency curves. However, the result has an essential shortcoming, which can be easily perceived when we attempt to apply the model law to the description of model density for any other sample from the same statistical population. In most cases the attempt is a failure. As a consequence, this approach cannot be used to solve the major modeling problem - expansion of the regularities perceived in the behavior of the sampling data over the general population. Thus, in analytical treatment of the data with the purpose of defining common statistical characteristics of air pollution, we shall use model laws and statistics that, perhaps, are less than optimal in terms of formal criteria, but have characteristics that are of much greater importance in our investigations, namely, degree of stability and invariance of the derived results with respect to methods of sample organization, different types of pollutants and geographic areas. Let us consider from this point of view the characteristics common to the distributions of the data plotted in (3.1), (3.2), (3.3).

Figures 3.4 - 3.6, also Figures A.3.1 - A.3.18 of the Appendix, show the empirical density distribution functions, the empirical distribution functions on normal probability paper, and the deviation from the normal distribution function (termed hereafter the histogram, normal graph and deviation from the normal graph) for logarithmic concentrations. For comparison with the model law, we can use the series generated by a random-number-generator; the respective graphs for this series are presented in Figures A.4.1 to A.4.3.

As is evident from these graphs, the distributions are quite similar to the normal ones. However, when these multi-year data-series are divided into seasonal data series, i.e. from May to September, and from November to March, then the departure from normal becomes apparent. For comparison we show similar graphs for suspended particulate concentrations, represented in Figures A.3.10 - A.3.18 of the Appendix, where the distribution pattern does not change and is preserved in all three samples under study. In the general case, the data exhibit a lognormal distribution. This is due to the fact that the deviations from the straight line on the respective graphs, although causing distortions in the form of the line, are not so great as to obscure the normal distribution. It is obvious also that this lognormal distribution is formed under the effects of a large number of diverse factors, among which are yearly and seasonal variations. Probably a plausible explanation is offered also by the hypothesis of a similar influence of the factors reflecting the effects produced by the background constituents, anthropogenic local sources, As a matter of fact, if we compare graphs 3.3 with 3.6, A.3.3 with A.3.6, and A.3.9, A.3.12 with A.3.15 and A.3.18, a number of common characteristics can be distinguished.

From comparison of Figures 3.3 and 3.6, some similarities and differences can be seen, from which we can get an idea of how the multi-year lognormal distributions are formed. These plots differ greatly in their form. We could hardly have expected it to be otherwise, since the second sample is a non-random sample taken from the first, and comprises less than 10% of its population (91 out of 1039 observations). However, a common feature is apparent in these graphs, reflecting the concentration distributions of lead in different areas. The type of deviation from the straight line clearly changes beyond the value 2.1 in both plots, which serves to indicate some common formation process, where the operating factors strongly affect the low concentration range, and their manifestations are common to all seasons and years of observations. The second consideration is that the logarithm of

HISTOGRAM OF VARIABLE 13 PB

		SYMBOL	COUNT	MEAN	ST.DEV.
	BZT81	B	91	2.521	0.624

INTERVAL NAME	5 10 15 20 25 30 35 40	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
.1200	+	0	0	0.0	0.0
.2400	+	0	0	0.0	0.0
.3600	+	0	0	0.0	0.0
.4800	+	0	0	0.0	0.0
.6000	+	0	0	0.0	0.0
.7200	+B	1	1	1.1	1.1
.8400	+	0	1	0.0	1.1
.9600	+	0	1	0.0	1.1
1.080	+	0	1	0.0	1.1
1.200	+	0	1	0.0	1.1
1.320	+	0	1	0.0	1.1
1.440	+BBB	3	4	3.3	4.4
1.560	+BB	2	6	2.2	6.6
1.680	+B	1	7	1.1	7.7
1.800	+BBB	3	10	3.3	11.0
1.920	+BBBBBB	6	16	6.6	17.6
2.040	+BBBBBBBBBB	10	26	11.0	28.6
2.160	+BBBBB	5	31	5.5	34.1
2.280	+BBB	3	34	3.3	37.4
2.400	+BBBBB	5	39	5.5	42.9
2.520	+BBBB	4	43	4.4	47.3
2.640	+BBBBBB	6	49	6.6	53.8
2.760	+BBBB	4	53	4.4	58.2
2.880	+BBBBBBBB	8	61	8.8	67.0
3.000	+BBBBBB	6	67	6.6	73.6
3.120	+BBBBBBBBB	9	76	9.9	83.5
3.240	+BBBB	4	80	4.4	87.9
3.360	+BBBBB	5	85	5.5	93.4
3.480	+BB	2	87	2.2	95.6
3.600	+BB	2	89	2.2	97.8
3.720	+B	1	90	1.1	98.9
3.840	+B	1	91	1.1	100.0
3.960	+	0	91	0.0	100.0
4.080	+	0	91	0.0	100.0
4.200	+	0	91	0.0	100.0
4.320	+	0	91	0.0	100.0
4.440	+	0	91	0.0	100.0
4.560	+	0	91	0.0	100.0
4.680	+	0	91	0.0	100.0
4.800	+	0	91	0.0	100.0
4.920	+	0	91	0.0	100.0
5.040	+	0	91	0.0	100.0
5.160	+	0	91	0.0	100.0
5.280	+	0	91	0.0	100.0
5.400	+	0	91	0.0	100.0
5.520	+	0	91	0.0	100.0
5.640	+	0	91	0.0	100.0
5.760	+	0	91	0.0	100.0
5.880	+	0	91	0.0	100.0
6.000	+	0	91	0.0	100.0

5 10 15 20 25 30 35 40 45 50 55

Figure 3.4: Histogram of logarithmic concentrations of lead. Berezina, warm season, 1981.
1 - first mode; 2 - second mode.

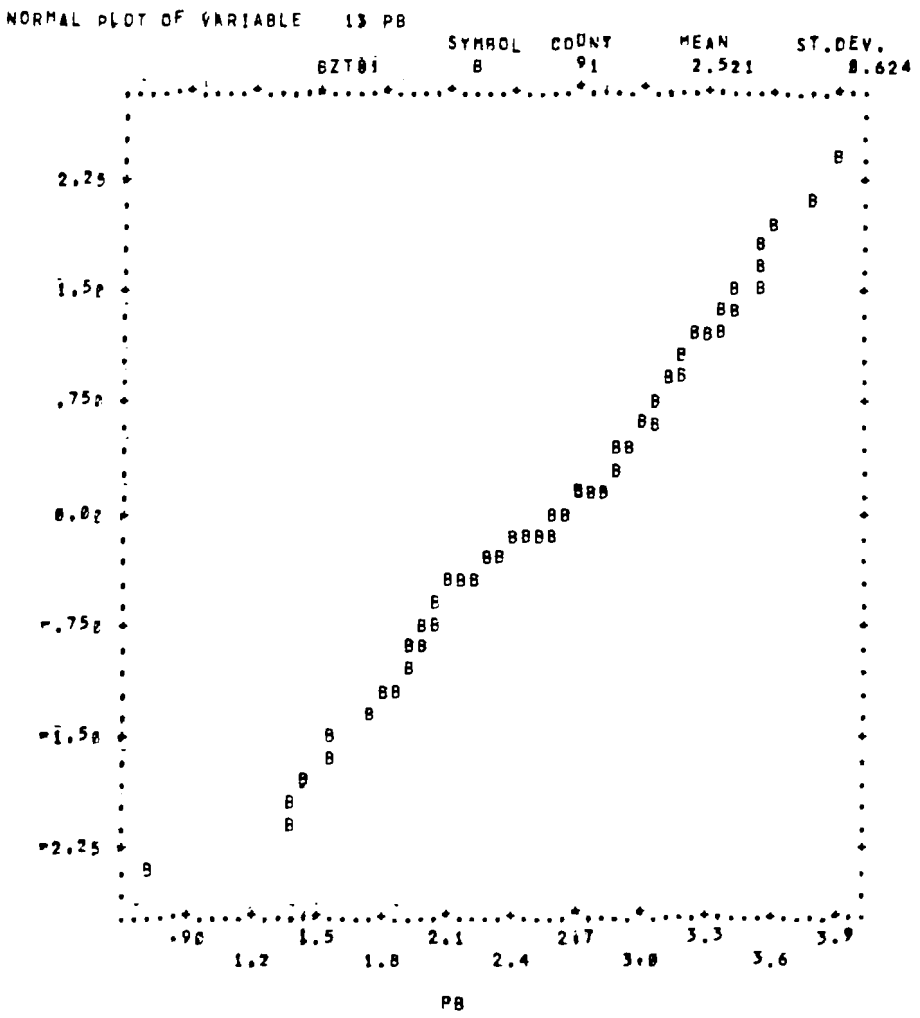


Figure 3.5: Normal plot of logarithmic concentrations of lead. Berezina, warm season, 1981.

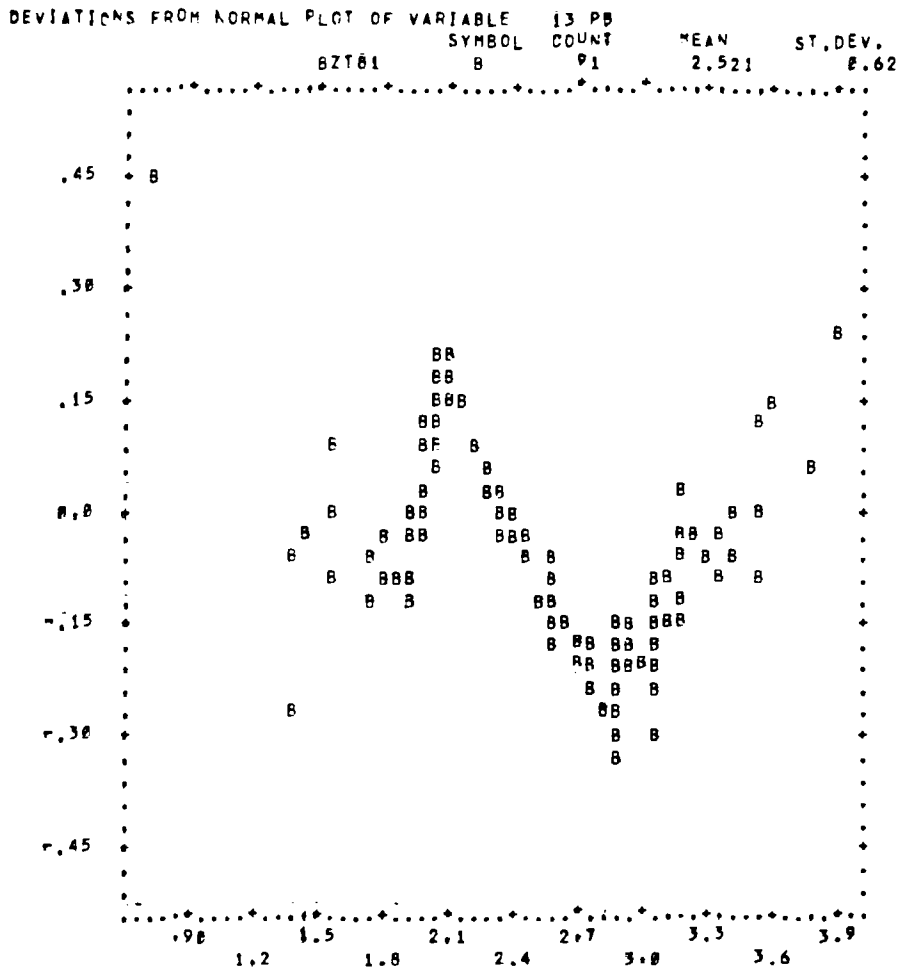


Figure 3.6: Deviations from normal plot of logarithmic concentrations of lead. Berezina, warm season, 1981.

the upper concentration limit varies during the warm seasons around 3.9-4.0; this means that all the visible points in Figure 2.3 that lie beyond the boundary 3.9 reflect the influence of specific "winter" factors, and from the form of the plot, it can be established that these effects do not coincide with the ordinary effects of the formation factors that we observe on the line-segment (2.1, 3.9). But then it becomes apparent that the lognormal distribution along this line-segment is a reflection of the effects of a very large number of formation factors, and the elimination of the effects of these factors results in manifestations of the operating mechanisms of other factors that are reflected in graphs of the (3.4) type as deviations from the straight line. That is, the subject-matter under study should be concerned not so much with the search for agreement between the observed distribution and the LN2, as with the search for deviations from the normal and plausible explanations of their cause.

If we compare similar plots for sulfur dioxide concentrations (A.3.1 - A.3.9), we perceive very great differences in the mechanisms governing the formation of the concentrations during the warm and cold seasons. The average values for the logarithms of the concentrations differ considerably - -0.29 for the warm seasons and 1.76 for the cold seasons. Differences are likewise reflected in the respective plots - most of the winter concentrations are located above a very small interval (-0.5, 0.0) and about half of the concentrations for the warm season lie beneath it. The "warm" concentrations terminate at about the logarithmic value 1.8, whereas most of the "winter" concentrations lie within this range. That is, on the multi-year plot A.2.1., zones can be distinguished that reflect the effects of cold and warm seasons. Even such a cursory examination makes it apparent that in order to determine natural background concentrations, it is necessary at least to get rid of the effects associated with the cold seasons, that are clearly contingent upon anthropogenic effects of the heating season.

The data series derived from observations on particulate matter display a similar distribution pattern of the constituents, with a breakpoint on the curve, that depends upon the season.

Thus we conclude that it is necessary to design a statistical model that would enable the observed effects of various groups of factors to be taken into account, and to obtain quantitative estimates on the basis of statistical characteristics. Since the desired statistical model should describe the effects of different groups of factors, we are confronted with the problem of how to distinguish some typical samples from among the data. These samples should reflect quite fully the effects of different groups of factors and, at the same time, the regularities derived on their basis should be typical of a specific pollutant and area of observation. In order to determine such sampling characteristics, several hundreds of plots were analyzed, which show the logarithmic concentration distributions for periods between a decade to eight years. As a typical example, a data series of five-month duration was chosen, that characterized the warm or cold seasons. The period from May to September, inclusive, is regarded as the warm season; the period from November to March of the next year is regarded as the cold season. Such a time-interval is, on the one hand, sufficiently long to show the effects of the major groups of factors and, on the other hand, sufficiently distinct from other observational series. Evidence that the seasonal observational series are actually the major carriers of information on the effects of different groups of factors is found in the fact that in contrast to all data series, these data series include the highest percentage (over 80%) of deviations from the "pure" lognormal distribution. Examination of Figure 3.6 makes it immediately apparent that the fluctuations are "organized" into three line-segments, where each can be interpreted as manifestations of the effects of a group of factors controlling the formation of pollutant concentrations. The respective histogram (Figure 3.4) clearly displays a bimodal

density function. These modes can be considered as central tendencies for each group of controlling factors. This implies that the model simulating the characteristics sample, which we have adopted for the seasonal observational series, should reflect the effects of different groups of factors, treated in the form of "composite" distributions.

These results were derived only on the basis of graphical analysis and data presentation. For such an analysis, the authors used the package of applied statistical programs BMDP. The techniques used for analyzing and processing the meteorological data have been described by Zelenyuk (1984) and Zelenyuk, Zubenko et al. (1984) and a model has been proposed that describes the event-data series derived from background air pollution observations.

3.3. *Construction of a Statistical Model for Background Air Pollution Monitoring Data.*

Studies of seasonal observational series enable one to establish the specific multimodality of the density distributions, and the presence of characteristic deviations from the theoretical distribution function of the probability of logarithmic concentrations related to manifestations of different groups of causative factors. As a natural consequence, three problems arise. The first concerns the design of the statistical model for characteristic samples, with model parameters selected to reflect sample specifics (type of pollutant, area of observation, season, and mainly, the nature of the effects of the causative factors). The second problem refers to the method of analysis of the model with a view to determining the causative factors and the development of techniques for estimating model parameters. The third problem involves the derivation of statistical inferences concerning the entire data population from the seasonal sample population.

The next two chapters are devoted to the second and third problems. Here we shall describe the statistical model simulating background air pollution concentrations.

The following composite multimodel distribution model is used:

$$f(x) = p_1(x)\, f_1(x) + p_2(x)\, f_2(x) \tag{3.5}$$

where f_1 and f_2 are the density distributions, and p_1 and p_2 are the frequencies of realization, respectively. In contrast to classical composites (see, for example, Aivazyan et al. 1983) the frequencies are treated as a function of x, in order to distinguish the effects from the separate operation of the models within different intervals of the logarithmic concentration axis.

Let us take p_1 and p_2 as convolute functions of the "switching" action of the laws, and let us impose normal noise with a zero mean value and variance σ. If we take

$$H(x;a) = \begin{cases} 1, & x \le a \\ 0, & x > a \end{cases}, \quad H_1(x;a) = H(x;a), \quad H_2(x;a) = 1 - H(x:a) \tag{3.6}$$

then we can write

$$p_{tH,N}(x) = H_t(x:a) * N(x;0,\sigma) .$$

It will be recalled that

$$N(x;a,\sigma) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left[-\frac{1}{2}\frac{(a-x)^2}{2\sigma^2}\right], \quad \Phi(x;a,\sigma) = \int_{-\infty}^{x} N(t;a,\sigma)dt .$$

Then

$$p_{H,N}(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot H(y;a) \cdot e^{\frac{-1}{2}\frac{(y-x)^2}{\sigma^2}} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-x}{\sigma}} \exp -\frac{z^2}{2} \cdot dz = \Phi(\frac{a-x}{\sigma}) . \tag{3.7}$$

By introducing the constants π_1 and π_2 characterizing the frequency of $x \leq a$ and $x > a$, respectively, we get

$$f(x) = \pi_1 \Phi\left(\frac{a-x}{\sigma}\right) f_1(x) + \pi_2 \Phi\left(\frac{x-a}{\sigma}\right) f_2(x) . \tag{3.8}$$

It is obvious that π_1 and π_2 should be connected through normalization,

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

In the case under discussion, when $f_k = N(x;b_k,s_k)$ then for $k = 1.2$ we get

$$\int_{-\infty}^{\infty} \left[\pi_1 \Phi(\frac{a-x}{\sigma})\, N(x;b_1,s_1) + \pi_2\, \Phi(\frac{x-a}{\sigma})\, N(x;b_2,s_2)\right] \cdot dx = 1 .$$

Using the identity

$$\int_{-\infty}^{\infty} \Phi\,(\frac{a-x}{\sigma}) N(x;b,s) = \Phi\left(\frac{a-b}{\sqrt{\sigma^2+s_1^2}}\right),$$

we derive the normalization conditions for two distributions

$$\pi_1 \Phi\left(\frac{a-b_1}{\sqrt{\sigma^2+s_1^2}}\right) + \pi_2 \Phi\left(\frac{b_2-a}{\sigma^2+s_2^2}\right) = 1 .$$

The proposed method of compositing is derived from a qualitative analysis of the mechanisms governing the separate distribution models. The "switching" mechanism is presumed to operate in such a manner that the zone of concentration grouping, defined by one of the formation mechanisms, should include more of its "own" concentrations than those of a neighbor, which is possibly an even stronger mechanism (i.e., including a larger number of observations). In practice, upon examination of the plots of density and distribution functions, the major zones of central tendencies are distinguished under the assumption that each one reflects certain regularities in the behavior of formation factors. In those cases, when the

mutual influence of groups of factors is very great, their action can be regarded as a single formation mechanism. Such an approach to the analysis of the distributions, as was shown in the preceding section, is justified above all by the fact that the composite action of different groups of seasonal factors forms a distribution closely similar to the LN2. Thus, each component of the distribution $f_1(x)$ may be regarded as the result of the operation of relatively independent causative factors.

Then, the parameters a and σ acquire a meaningful interpretation. From the attributes of the function it follows that

$$f(x) = \begin{cases} f_1(x), & x \leq a - 3 \\ f_2(x), & x > a + 3 \end{cases} \tag{3.9}$$

That is, the parameter "a" plays the role of a "switching point" for each of the models f_1. Parameter σ defines the extent of the "switching zone" where composite factors do not reveal sufficient manifestations to form individual concentration groups. When $\sigma \rightarrow \infty$, the composition under study shows a tendency towards the well-known type of compositing.

This statistical model can represent data series of air-pollutant concentrations. These samples can be described by a set of quantities

$$\left\{a_{t-1}, \sigma_{t-1}, \left\{b_t, S_t, \pi_t\right\}, a_t, \sigma_t\right\}_{t=1}^{k}. \tag{3.10}$$

Each set enables one to reproduce the information that we derive from examination of the plots, representing empirical density and distribution functions. The value π can be substituted by the percent of observations within the respective distribution, as related to the total number of observations.

4. ASSESSMENT OF BACKGROUND AIR-POLLUTION MODEL PARAMETERS

4.1. *Discussion of the Possibilities of a Credible Interpretation of the Model Parameters.*

The problem of assessing model parameters discussed in the former chapter, is closely associated with the possibilities of non-statistical meaningful interpretation of the model. Such an interpretation, defining the possible use of model parameter estimates, may allow one to formulate certain requirements for the estimation itself. At any rate, it should be useful in ascertaining what types of statistics proposed by the model are consistent with our knowledge of the processes studied, leading to the solution of practical problems of evaluation of background air-pollution.

In Marchuk (1982), a descriptive air-pollution model was considered for the following case. During a season or a year within a given region, several types of air mass flows occur, each of which, for a characteristic period of its existence, can be regarded as invariant and stationary. After each of these periods, the air-mass flow changes and a new steady state commences. Since the change in circulation occurs during a period of time much shorter than that of the existence of this specific type of motion, it may be assumed that the change in types is instantaneous. For the description of such a situation, in the case of n types of synoptic patterns, the following system of independent equations has been suggested in Marchuk (1982).

$$div\ U_i C_i + \sigma C_i = f \tag{4.1}$$

$$C_i = C_{is} \quad \text{by } S \text{ when } U_{in} < 0, \quad i = 1,n \tag{4.2}$$

where $i = 1,\overline{n}$, $C(x,y,z,t)$ denotes the rate of transport of the pollutant by the wind, $U = Ui + vj + wk$, where i,j,k are vectors along the x,y,z axes, respectively, U is the velocity vector of the pollutant as a function of x,y,z,t, accordingly, and $\sigma \geq 0$ is a value inversely proportional to time; $C_{i,s}$ denote values of the C_i function at the boundary S and $U_{i,n}$ the projection of the wind velocity vector on to the normal. The problem (4.1), (4.2), can be solved for each of the intervals $t_i < t < t_{i+1}$, the length of which equals Δt_i. Now the problem of estimation of the air-pollution characteristics can be formulated as follows: assuming that all the equations (4.1), (4.2) are solved, how can we find the mean pollutant concentration for a certain time period? In Marchuk (1982), it is suggested that the average over the concentration distributions for the time period T should be considered as a linear combination

$$T = \sum_{i=1}^{n} \Delta t_i, \quad \bar{c} = \frac{1}{T} \sum_{i=1} C_i\ \Delta t_i \tag{4.3}$$

Marchuk (1982) suggests that (4.1, 4.2, 4.3) should be termed the statistical model of pollutant distributions in the atmosphere. It can easily be seen that the proposed statistical model is nothing but a simple composite of ordinary type of distributions. This example serves to indicate that for the use of compositing distributions as a model describing air-pollutant distributions, of paramount importance are not so much the concepts on the mechanisms governing different types of pollutant distributions, as are some more general principles underlying the model, namely, the assumptions of the role of the nature of the air mass flow, and of the presence of stationary periods in such a flow.

The statistical model, proposed herein, also relies on the assumptions of the presence of certain "stationary" periods in the mechanisms of concentration formation, with the periods replacing each other, and the changes being described by the "switch". However, the composite distributions proposed in paragraph 3.3 for the description of the resultant concentration distribution exhibit a greater generality, insomuch as (4.3) is included as a particular case.

Thus, for determining the applicability of the proposed model, it is not so important to define the specific mechanisms governing the pollution distributions and formation of different concentration levels, as to show that relationships exist between concentration levels and different types of meteorological conditions within the area of observations. The relationship between the air-pollutant concentration levels and the meteorological processes in the region is a special subject-area under study in a number of publications (Smirnov, 1982; de Nevers et al., 1979; Schmidt and Velds, 1969). For instance, in Smirnov (1982) the data available on the mean values and the daily trend in aerosol concentrations are systematized and discussed. This work had as its objective the formation of concepts on the normal composition of particulate matter in the lower atmosphere. The author used measurements taken under conditions typical of the formation of the concentrations and pollutant distribution mechanisms, i.e., during summer and winter anticyclones. At the same time, in Smirnov (1982) the author calls attention to the fact that systematized data on average values and variability of factors for typical weather situations in different geophysical regions, are practically lacking. The

relationship between the variations in the meteorological conditions and sulfur dioxide concentrations is treated in Schmidt and Velds (1969). In de Nevers et al. (1979), the suggestion is made that for the description of air-pollutant distributions, "composite" distributions should be used, ones that on lognormal paper look like a graphical combination of two LN2 distributions. de Nevers et al. (1979) offer an example illustrating how two quite different types of meteorological conditions in the area of observation (with air-mass transfer across a canyon to the area of observation during particular time periods) lead to the formation of the distribution, depicted in Figure 2.1 under number 1.

Analytical treatment of plots showing pollution concentration distributions led to the generation of concepts postulating the existence of several different mechanisms governing the formation of different pollutant concentration distributions. This concept is reflected in the respective statistical model in terms of a set of model parameters describing characteristic samples. In evaluation of the model and application of the statistical techniques to derive inferences, errors coming from different sources may occur. A review of the most frequent errors can be found in McKay and Bornstein (1981), which treats models simulating air quality and their application. However, most of the sources of errors occur when the model is used for prediction purposes and are of little significance in descriptions of model behavior. de Nevers et al. (1979) examine, as one source of error that can cause deviation from the LN2 model, samples of the general population that reveal, as a whole, LN2 distributions. However, in the previous chapter we have mentioned that deviations from the LN2 distributions observed in the multi-year data series are associated with manifestations of different concentration formation mechanisms. de Nevers et al. (1979) claim that deviations of one of the four types noted by them were perceived in about 25% of the studied cases. The specific features of the background monitoring data and of the characteristic samples selected for investigations, are such that deviations from the LN2 distribution are apparent in more than 80% of the seasonal samples. The principal relationships between the observed deviations and the characteristics of the meteorological processes have been discussed in Zelenyuk and Cherkhanov (1985) and Izrael et al. (1985).

An analytical treatment of data series in relation to meteorological processes is presented in Zelenyuk and Cherkhanov (1985) and Izrael et al. (1985). Taking into account the fact that difficulties are encountered in establishing weak relationships with meteorological processes, and the complex nature of the different atmospheric factors that control the concentration patterns, Zelenyuk and Cherkhanov (1985) and Izrael et al. (1985) restricted their studies to the identification of two major components of the seasonal data-series, viz., components that accommodate the lowest and highest concentration regions, and that cover not less than half of the entire population. Such a data-series is represented in Table 4.1. In the upper section of each row is shown the dates of observations for one of the components of the distribution, in the lower section - the other component. The pattern exhibited in the table is typical of practically all observed distributions of seasonal concentration data-series; the duration of individual spells is long and, accordingly, changes, concomitant with the formation of concentrations in "transitional periods" which are represented in the table by the interval 2.8 to 3.2, are not long. Such a separation leads to various hypotheses on the relationship between formation mechanisms of different concentration levels and types of atmospheric states. It is obvious that examination of only two components is to a large extent a teaching example, distinguishing those components that can be easily interpreted. A more detailed analytical treatment, incorporating the regularities in atmospheric processes, demands more precise concepts of how to distinguish among the diversity of meteorological parameters and situations that bear on

the formation of different concentration levels in the areas of observation. From the analysis presented in Zelenyuk and Cherkhanov (1985) and Izrael et al. (1985), certain common features in the behavior of different substances in different areas can be perceived. There is a considerable coincidence - between 60 to 90% - on days when the sulfur dioxide and lead contents exceed the "average" level (concentrations exceeding the "average" level were graphically determined). These pollutants differ considerably, according to their origin, input rates, transport and dispersion, from which the inference can be drawn that the factors that control the formation of these two major components are characteristic of large-scale meteorological processes. These are associated with two major groups of factors. The first, which causes high pollution concentration levels, is related to the stable anticyclonic state of the atmosphere, when the concentrations of pollutants are determined mainly by diffusion. The second group of factors is associated with the unstable cyclonic state that is favorable for the dispersion of pollutants by turbulent processes, and, accordingly, with low concentration levels. Analysis of these relationships revealed that the coefficients of cross-correlation amount to 65-75%. Thus graphical estimation of the components of composite distributions, specified by equations (3.2) and (3.4), enables one to interpret components of the background air-pollution process. In order to show this, cluster analysis of the composite distribution was performed with a view to ascertaining whether the two major components comprise "natural" clusters in the event-data. As was demonstrated in Perone et al. (1975), the application of such techniques to solving problems bearing on air-pollutant concentrations, is permissible even when there is a weak assumption of homogeneity and uniformity of measurement scales. As in Perone et al. (1975), the present authors employed cluster-analysis to distinguish the compositing components. The results of the discrimination are presented in Table 4.2. The last three lines of the table show the distances between centers of gravity of common ground point data, their relative coordinates and numbers. It can be seen that the last distance is much greater than the others, which serves as evidence of the presence of two natural data "clusters". Notwithstanding the relatively low accuracy of discrimination, derived in this manner, nevertheless, it provides evidence justifying the distinction between the two major components.

Thus, analysis of the sampled data series made it possible to establish the essential relationship between the graphically defined compositing components and the general characteristics of the meteorological processes.

4.2. *Theoretical Principles Underlying the Statistical Background Air-Pollution Model*

Let us discuss the physical prerequisites that provide the basic premise for the two most essential characteristics of the proposed model, the probable reasons for its good approximation to the LN2, and the physical prerequisites for the origin of multimodal distributions.

The first problem is treated in Karasev (1980; 1982). In Karasev (1980), it is shown that the lognormal distribution can be derived from the description of energy fluctuations in equilibrium systems and the maximal entropy principle. In Karasev (1980) the author takes advantage of the observation that the lognormal law is the law of greatest entropy, as compared to all laws with specified logarithmic dispersion. The density probability distribution function can be written in terms of energy as

$$f[E] = [E(2\Pi)^{1/2} \cdot \sigma_{\ln}]^{-1} \times \exp\{-(lnE - \overline{lnE})\}^2/_2 \sigma_{\ln}^2$$

Table 4.1: Periods of formation of two major components in the observational series of logarithmic concentrations of dust. Berezina B.Z., warm season, 1982. See text for explanation.

LN C > 3.2 - DATE			01.05	02.05			04.05	05.05	06.05	
LN C ≤ 2.8 - DATE						03.05				
07.05	08.05	09.05	10.05	11.05	12.05	13.05				
									14.05	
		18.05	19.05	20.05			24.05	25.05		
15.05	17.05					23.05				
	27.05	28.05	29.05	30.05	01.06	02.06	03.06			
26.05										
04.06	05.06	06.06					10.06	11.06		
			07.06	08.06	09.06					
12.06	14.06									
		15.06	16.06	17.06	19.06	20.06	21.06			
23.06	24.06	25.06	26.06	27.06						
					28.06	29.06	30.06			

Table 4.2: Hierarchical cluster analysis of logarithmic concentrations of sulfur dioxide. Repetek, V-VII, 1981.

AMALGAMATION ORDER	DIST.		
1	0,001	0,401	2,000
2	0,002	-1,650	2,000
3	0,002	0,818	2,000
4	0,002	0,528	2,000
5	0,004	0,830	2,000
6	0,004	0,969	2,000
7	0,005	0,586	2,000
8	0,005	0,719	2,000
9	0,005	0,653	2,000
10	0,006	0,497	2,000
11	0,006	0,223	2,000
12	0,010	0,436	2,000
13	0,011	0,219	3,000
14	0,013	0,824	4,000
15	0,013	0,648	3,000
16	0,013	-1,021	2,000
17	0,014	-1,074	2,000
18	0,014	0,492	3,000
19	0,014	-1,107	2,000
20	0,017	0,533	3,000
21	0,019	0,820	5,000
22	0,020	-1,718	8,000
23	0,022	0,295	2,000
24	0,022	0,463	3,000
25	0,024	0,824	6,000
26	0,026	0,977	3,000
27	0,027	-1,012	3,000
28	0,029	0,926	2,000
29	0,030	0,641	4,000
30	0,033	-1,785	21,000
31	0,033	-1,091	4,000
32	0,038	0,649	5,000
33	0,041	0,513	6,000
34	0,042	0,426	5,000
35	0,043	0,309	3,000
36	0,049	0,831	7,000
37	0,051	-1,101	5,000
38	0,052	0,957	5,000
39	0,053	0,206	4,000
40	0,062	0,631	7,000
41	0,067	-1,752	6,000
42	0,071	1,241	2,000
43	0,076	-1,535	2,000
44	0,086	0,474	11,000
45	0,088	0,650	9,000
46	0,089	-1,067	8,000
47	0,102	-1,718	6,000
48	0,103	0,250	7,000
49	0,119	-1,495	3,000
50	0,126	0,883	12,000
51	0,167	-1,086	9,000
52	0,177	0,553	20,000
53	0,188	-1,745	7,000
54	0,189	-1,067	10,000
55	0,231	1,318	3,000
56	0,250	-1,670	10,000
57	0,303	0,475	27,000
58	0,409	0,600	39,000
59	0,603	-1,368	20,000
60	0,718	0,652	42,000
61	2,020	0,000	62,000

where $\overline{\ln E}$ is the mean logarithmic energy value, $\sigma_{\ln}$ the logarithmic dispersion. The author's reasoning in terms of the energy distribution of the system can be applied to the distribution of the studied system according to the concentration levels. The general physical considerations on which the reasoning is based, are valid for our case. Karasev (1980) assumes that the preservation of logarithmic dispersion when the average value changes, is a general attribute common to systems with fluctuations and that in many cases this attribute determines the system distribution in terms of energy distribution (in our case, concentration distribution). Let us consider several examples illustrating the preservation of logarithmic concentration dispersion.

The numerical values -0.29, -0.52, 0.73, 1.76, are a series of mean logarithmic concentrations of sulfur dioxide for each of the four seasons based on samples taken over 3-years. As can be seen, there are strong fluctuations about the mean value. At the same time, the respective logarithmic dispersion values are 1.010, 1.008, 1.154, 1.019. A similar relationship has been observed for two other pollutants under study.

The possibility of applying the lognormal law to the description of diffusion phenomena is discussed in Karasev (1982) and it is demonstrated that the spatial distribution of Brownian particles can be well-described by formulae, connecting the distribution spread with the average value of the logarithmic dispersion, based on the lognormal law.

Thus, it becomes apparent that from very general prerequisites, by employing statistical information available from the event-data series, we can justify the model distribution adopted by us, viz., the two-parameter lognormal distribution.

Turning our attention to the possible factors causing multimodal distributions, it should be noted that empirical distribution functions with more than one mode, originate from time to time in various investigations. Unfortunately, in routine practice, the occurrence of multimodal distributions commonly makes it necessary to abandon the ordinarily employed unimodal distributions, such as the normal, exponential, Weibull, etc., and to use a much more complicated statistical form such as, for example, the Pearson curves. The use of this form usually requires a large number of assumptions concerning the nature of the data, which seldom can be provided in real conditions. The method of attacking the problem, proposed by the author of the present study, makes it possible to use specifically multimodal distributions, frequently observed in routine practice, and furnishes the investigator with a relatively simple and available body of mathematics for obtaining informative statistics in the multimodal data series, under quite general assumptions concerning the cause of the multimodality.

As another promising alternative, let us discuss publications by Cobb (1978) and Hangos (1983), where a multimodal distribution is employed as a natural consequence of non-linear stochastic processes generating it. In Cobb (1978), by "nonlinear" is meant the presence of more than one stable state in the system. A stochastic system is one that constantly undergoes perturbation by random effects. Fluctuating around a mean value, it undergoes a step-wise transition to another mean value. A random sample composed of multiple steady states of such a system, should exhibit a multimodal distribution.

Systems, in which the steady states are described by a differential equation, where one of the variables is random, are of wide use for describing diffusion processes. Let the system be described by the differential equation

$$\frac{\partial x}{\partial t} = -\frac{\partial P}{\partial x} \tag{4.4}$$

where $P(x;a,b,...)$ is the real-valued function of x and parameters $a,b,c,...$ (in our case x refers to the logarithmic concentration of the pollutant). Multiple steady states of such a system can be described by a set of x, such that $\partial x / \partial a = 0$, i.e., by a set of solutions of equation $\partial P / \partial x = 0$. Let us assume that the change in the variable x is not strictly defined by equation (4.3) but that a probability density exists that determines the possible changes in x. Then for $\partial x / \partial t$ we can substitute the expression

$$m(x) = \lim_{h \to 0} h^{-1} E\{x(t+h) - x(t) \mid x(t)\}$$

where E refers to the expected value of the random variable in the braces. This expression will be denoted as $m(x)$, understood as the average measure of variation of x. Cobb (1978) calls $m(x)$ the "drift function" of the diffusion process. It is obvious that

$$m(x) = -\frac{\partial P}{\partial x} . \tag{4.5}$$

Also, a natural assumption is introduced, that the trajectory $x(t)$ is smooth, i.e., that instantaneous, abrupt changes in x are impossible. The range of variation of x is defined as

$$v(x) = \lim_{h \to 0} h^{-1}\{1/2[x(t+h) - x(t)]^2 \mid x(t)\} . \tag{4.6}$$

The arbitrary probability density of the diffusive process varies in accordance with the differential equation

$$\frac{\partial f}{\partial t} = -\frac{\partial F}{\partial x} , \tag{4.7}$$

where

$$F(x,t) = mf - \partial(vf)/\partial x . \tag{4.8}$$

The second term in (4.8) defines the stochastic diffusion process. The probability density function $f^*(x)$ can be defined as any solution of $\partial f / \partial t = 0$, i.e., any time-invariant density function. This is possible when the process is constant for all x values. That is, the problem of finding the density is reduced to solving $F = 0$, for a certain f^*, which is a simpler problem than the solution of (4.7). Then equation (4.8) can be rearranged into an ordinary differential equation

$$f^{-1} \cdot \partial f / \partial x = (m - v')/v . \tag{4.9}$$

If we consider m a linear function and v a quadratic function, then the solution of equation (4.9) conforms with the Pearson-class curve. However, since in problems of the theory of catastrophes, m is always regarded as a polynomial, we derive an expansion of the Pearson-class curve. From (4.9) it follows that the integral

$$\varphi(x) = \int v^{-1}(m - v') , \tag{4.10}$$

if it exists, gives a solution for (4.9). This solution is

$$f^*(x) = k \cdot \exp(-\Phi(x)) , \tag{4.11}$$

k being chosen so that $\int_{-\infty}^{\infty} f^*(x)dx = 1$, while Φ is defined by equation (4.10). That is, the stochastic densities are accommodated in the family of canonical exponential densities.

The relationship between the density f^* and the potential function P for the stochastic model under discussion, is clearly apparent when $v(x) = \varepsilon$, in other words, when an infinitesimal increment is determined by a small constant. In this case, $\Phi = P/\varepsilon$, and a solution of equation (4.10) is simplified to

$$f^*(x) = k \cdot \exp[-P(x)/\varepsilon]. \tag{4.12}$$

In this density class, the system potential is proportional to the logarithm of the probability density. For instance, if P is a linear function, then f^* is an exponential. In problems of practical concern, the potential P is a polynomial of power higher than 2 which, naturally, leads to the general multimodal density functions. In Cobb (1978), techniques for estimating the parameters of multimodal distributions are designed on this basis. In accordance with the proposed theory, the estimates are designed as estimates of the proportional factor and coefficients of the polynomial P.

The foregoing enables one to gain an insight into possible physical mechanisms governing the origin of multimodal densities in problems of statistical analysis of air-pollution data, and to obtain a qualitative depiction of the resulting concentration distribution. However, the construction of a mathematical model is fraught with difficulties, since the density class generated by the model is very wide, and thus the verification of this model is practically impossible. Besides, the proposed techniques, although suitable for estimates of the polynomial coefficients and for construction of the best of these in terms of the proposed theory, make it impossible to compare distributions other than according to the proportionality factor of the polynomial, due to the fact that the meaning of the comparison of the coefficients is not clear. In order to design a mathematical model of the described phenomena on the basis of the techniques proposed by Cobb (1978), a large amount of additional information is required; in fact, it is necessary to develop cogent hypotheses concerning the factors that control the concentration distributions.

Thus, analysis of the theoretical prerequisites for the application of the statistical model of background air-pollution, proposed in the previous chapter, served to demonstrate the possibility of its theoretical validation, from the point of view of its use as a particular case of the LN2 distribution, and revealed that the time series accommodate statistical information that shows the possibility of this kind of its application. The theory of catastrophes also gives theoretical support for the possible occurrence of multimodal distributions in the class of problems under study herein. However, the associated methods are unsatisfactory for our purposes. The following section is devoted to a discussion of this problem.

4.3 *Assessment of Model Parameters in Terms of Simulation Data.*

The problem of obtaining estimates for the parameters of the statistical model of background air-pollution frequency distributions incurs difficulties due to the impossibility of verifying the estimates. As a matter of fact, it is not possible to evaluate any of the model parameters, using values obtained from direct measurements or reproduced from physical considerations; from the data series description proposed by model (3.11), we cannot choose the most informative set of parameters. In order to test the model function, to develop methods for graphical evaluation of the parameters, to elaborate concepts on the precision of such

estimates and to use them, models simulating the observational data series were therefore designed.

To perform a simulation, it is necessary to establish what attributes of the model should be simulated and investigated. Taking into account the main problem of our investigation, the determination of statistics that provide a description of the time-series, ensuring comparison with observational data and the possibility of distinguishing several components in the simulated composite distribution, we shall restrict our attention to the composite performance of two distributions, for which we shall attempt to find an answer to the questions: what statistics fit the description of different composite distributions, and what is the accuracy of such a description?

Figures 4.1, 4.2 and 4.3 respectively depict a histogram, a distribution function plotted on normal probability paper, and the same function with the linear trend removed, for a composite distribution, conforming to normal laws, with parameters $N(x;25,1)$ and $N(x;27,1)$ respectively. The number of observations in the first case is 35 and in the is -85. (Henceforth, we shall speak of composite distributions 35N/25,1/ and 85N/27,1/). For evaluation and graphical analysis, the same data-presentation is used as for ordinary observation series. Simulation is performed employing the normal distribution, represented in Figures A.4.3, A.4.4 and A.4.5.

The plots presented in Figures 4.1, 4.2 and 4.3 enable us to demonstrate the techniques for assessment of the composite distribution parameters. At first, the major components are singled out in Figure 4.3, they lie, obviously, on opposite sides of value 26.5. Then the intervals for the grouping of both composite components can be pointed out: intervals [21.9, 26.5] and [26.5, 28.95]. Within the intervals thus distinguished, we can find the respective central tendencies. In determining these, the influence of the "transitional process" should be considered, which consists in the existence of observations of different components near the "switching point" 26.5, i.e., it is necessary to exclude from consideration the region of the "switching point"; for practical purposes it is sufficient to exclude three intervals to the right and to the left of it. In the remaining intervals, the points located at the centers of the intervals give the maximum total contribution to the set of observations. This contribution is estimated from the histrogram in Figure 4.1, where the respective intervals are denoted by stars. In practical estimates, three successive intervals are considered, choosing that triplet that provides the maximum contribution; the right boundary of the middle interval of the triplet is taken as the estimate of the central tendency. The estimates for the composite distribution, derived in this manner, amount to 24.9 and 27.45, the true values being 25.0 and 27.0, respectively. This method of estimation underwent multiple evaluation in regard to different kinds of composites, and the estimates for the central tendencies, derived from them, differ by no more than 10% from the known centers. Estimates of the variance were not performed: as such an estimate for the switching points, we adopted the length of $\pm 3\ \sigma$. Using the histogram, it is possible also to evaluate the relative weight-percent of the composite components. This is determined as the percent of observation points lying on opposite sides of the "switching point". This estimate is less accurate, revealing in our case proportions of 44 and 56% for the left and right sides of the distribution, respectively. Nevertheless, it enables both sides of the distribution to be evaluated qualitatively.

Thus, returning to equation (3.11)

$$a_{i-1},\ \sigma_{i-1},\ (b_i, S_i, \Pi_i),\ a_i, \sigma_i\ {}_{i=1}^{k}$$

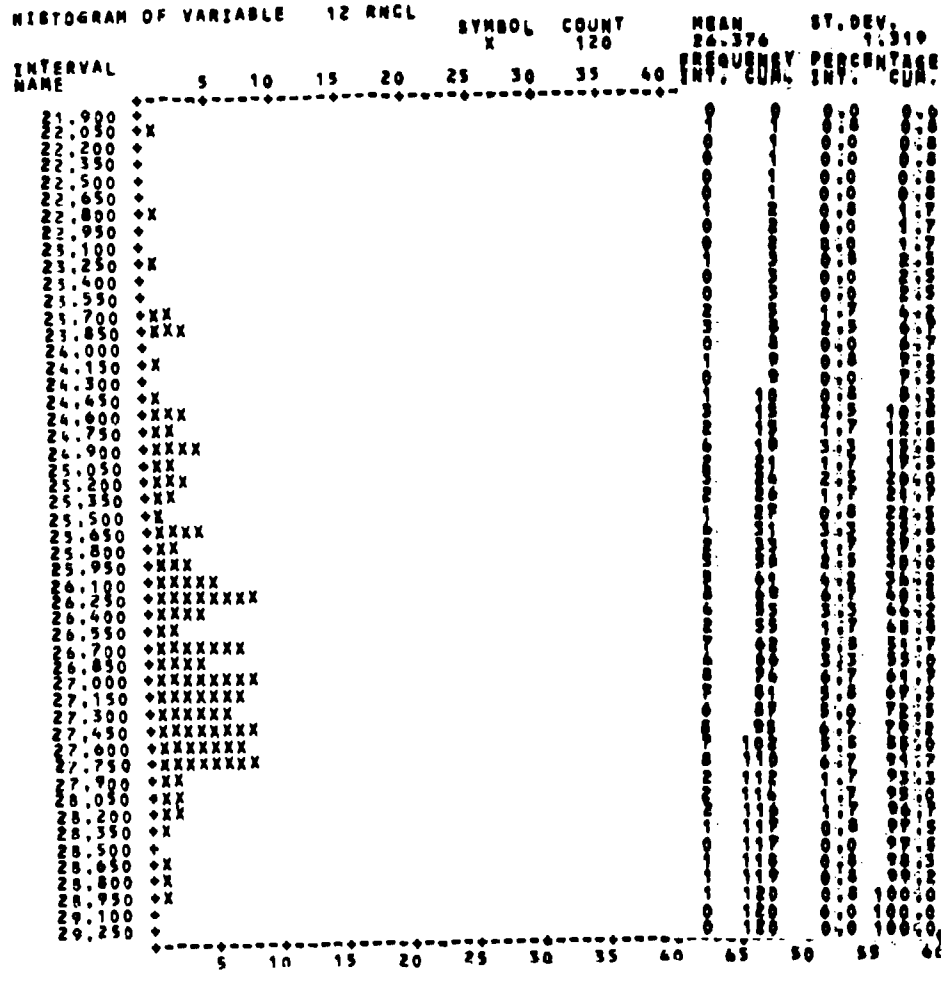


Figure 4.1: Histogram of composite distributions 85N/27,1/ and 35N/25,1/ (Simulation).

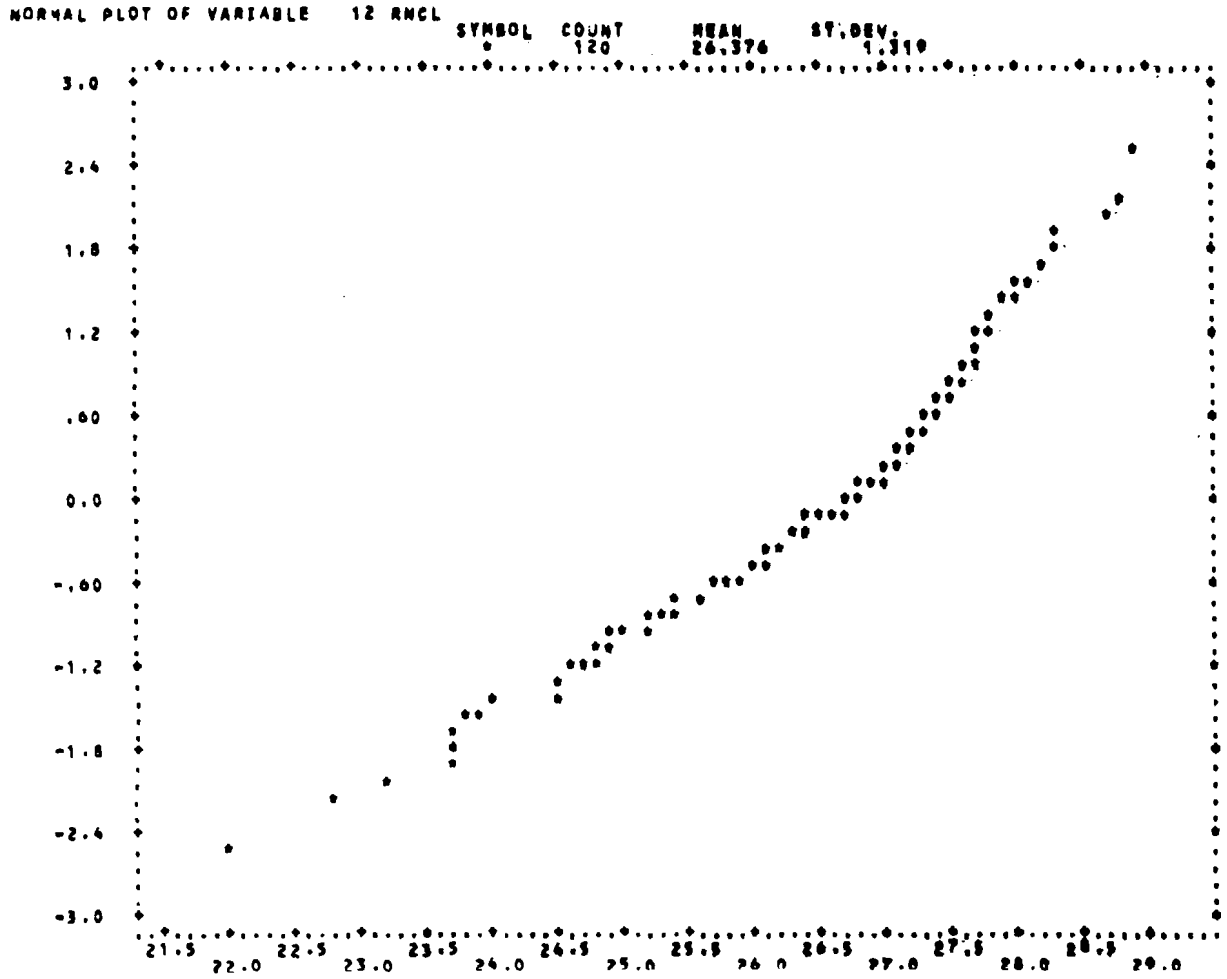


Figure 4.2: Normal plot of composite distributions 85N/27.1/ and 35N/25,1/ (simulation).

Figure 4.3: Deviations from normal plot of composite distributions 85N/27,1/ and 35N/25,1/ (simulation).

we can represent the distribution under study as

23.0, 0.15, (24.9,44%), 26.55, 0.15, (27.45, 56%), 28.95, 0.15 .

As is evident from comparison of Figures 4.1, 4.2, 4.3, this description quite precisely reflects the observed distribution pattern. The present author analyzed over one hundred plots of model functions describing the composite distributions. In general, the estimation techniques and results are quite satisfactory and can be used for estimates of the model parameters and for the logical design of the model. However, a question of great interest arises: which of the parameters are of the greatest statistical stability in relation to changes in the formation of the composition? In order to answer this question, a further analysis was performed, varying the compositing parameters. Let us examine some of the results.

In the Appendix, A.4.6, A.4.7, A.4.8 present plots describing the distribution of the same components, i.e., of the composite distribution $N(x;25.1)$ and $N(x;27.1)$. These distributions differ only in the amount of implemented components - 35-85 in the first case already discussed, and 55-70 in the second. This example is typical from the point of view of the extent to which it is possible to rely on the configuration of the distribution pattern in any hypotheses offered on the nature of the data. It is obvious that with real concentration distributions, accepting the hypothesis validated in Section 4.1 that two major types of meteorological factors influence the formation of the composites, situations may occur when days with different meteorological conditions exhibit ratios of 35/85, as well as 55/70. Notwithstanding the differences in these plots, they obviously represent similar processes, which should be reflected in the respective estimates. What estimates for the second group of plots can serve to disclose the similarity with the first group?

In Figure A.4.8 three parts of the composite distribution can be clearly distinguished. These are represented by the intervals [22.50, 25.05], [25.20, 27.15], [27.15, 28.95]. As a matter of fact, we are dealing with three distributions, defining the composite components; the middle distribution characterizes the combined influence of factors under which the unmixed effects of two extreme components of the distribution were formed. It is obvious that the newly derived estimates for the central tendencies, and for the weights of the components, bear no resemblance to the respective parameters of the first distribution depicted in Figures 3.3, 3.5. The fact that they describe processes that have some similarities can be established only upon comparison of the intervals. The similarity is obvious between intervals [21.9, 26.55] and [22.50, 25.05], also between intervals [26.55, 28.95] and [27.15, 28.95]. This example is a reflection of the regularity repeatedly observed by the authors: when the parameters are subject to variations, then of greatest stability are the interval estimates of the composite components. Taking into account the fact that for our purposes they quite satisfactorily describe the distribution, they may be adopted as the major statistics characterizing the component distributions.

The statistics that we have selected in the form of estimates of the grouping intervals can, most likely, reveal the major regularities. However, the question remains: how does one obtain estimates of such grouping intervals. The application of computer-programmed discrimination analysis is formulated in Bukhshtaber et al. (1983) and the possibility of introducing *a priori* information into the classification algorithms is discussed. The present authors, participating in the development of the algorithms and programs for the work (Bukhshtaber et al. 1983) performed a large number of experimental computations on the application of different algorithms. Experience gained from this work (Bukhshtaber et al. 1983),

tends to indicate that each of the studied problems is independent and intricate, to which the application of monotypic discrimination methods is impossible.

This calls for visual, graphical, discrimination of the compositing components. Such a method, by the way, is widely used in routine practice, when the data can be represented in the form of two-dimensional patterns.

Figures A.4.6-A.4.9 present plots of composite distributions that cannot be unambiguously classified. For instance, the distribution can be represented as the sum of three components, specified by the intervals [5, 15], [17, 30], [27, 37], and specified by two intervals [5, 17] and [17, 37]. Notwithstanding the outward dissimilarity of these classifications, not less than 50% of the respective intervals reveal a trajectory cross-over, which enables one to judge the similarity in the studied distributions represented by such intervals.

Thus, it can be considered an established fact that the most statistically stable distribution, in the sense of the preservation of the major characteristics of different components of the composite distribution, represented by the statistical model of background air-pollution herein proposed, are the intervals of the observational central tendencies. Evaluation of the intervals is performed graphically, which ensures quite a reliable distinguishing of the effects caused by the components, this process being supplemented with the percentage of the number of observations falling within the band. That is, analysis of the simulation data and of the possibilities of graphical evaluation and interpretation of the parameters, makes it possible to reduce the informative description of the time-series from (3.10) to

$$\{\alpha_{t-1}, (\Pi_t \chi), \alpha_t\}_{t=1}^{k} . \tag{4.13}$$

5. DISCRIMINATION OF THE COMPONENTS OF BACKGROUND AIR POLLUTION

5.1. *Estimation of the Central Tendencies of Multi-modal Frequency Distributions of Seasonal Data Series*

The proposed statistical model of background air pollution was used in Izrael et al. (1985) to estimate the components of an observational series, sampled at the "Borovoe" background monitoring station. Discrimination of the components was performed by using a simplified model that assumes that the data series is the result of two major processes that affect the concentration frequency distribution. It is clear that estimates of the central tendencies in the lower part of the distributions can provide meaningful inferences concerning the model and the background concentration levels themselves. With respect to the model, it can be assumed that in the case of agreement between the graphically defined distribution characteristics and the processes occurring in the atmosphere, on the one hand, and the established concepts on the nature of these processes, on the other hand, it would be possible to predict the behavior of the estimates of the central tendencies in the lower parts of the distribution. The hypothesis can be offered that such estimates should possess essentially greater statistical stability, as compared to estimates for the upper parts of the distributions, considering that the physical processes forming the lower part lead to averaging of the concentration levels over the entire lower troposphere, thus "smoothing" the pollution concentrations, as a consequence of which the processes are characteristically defined as being of a greater scale and, naturally, of greater statistical stability. Evidence that lends support to this hypothesis is found in Izrael et al. (1985). The plots in Figure 5.1 demonstrate the difference in the behavior of two major components. This implies

that the application of the proposed model to background air pollution concentrations, and of the techniques used to derive estimates for the model parameters is fully justified.

The statistical stability of the estimates of the central tendency of low concentration frequently suggests its possible use for assessment of background concentration levels of air-pollutants. As a matter of fact, upon comparison with ordinarily employed mean values, it becomes apparent that the derived estimates have a number of advantages. They enable one to validate statistically, and to calculate the characteristics of, regional background levels as an average over minimal concentration values for a given time period (Rovinskii and Buyanova, 1982).

Zelenyuk and Cherkhanov (1985) discuss the possible use of the estimates for the central tendencies, derived graphically, in solving problems related to the evaluation of background pollution concentration levels in the atmosphere. For instance, pollutant deposition should be calculated taking account of the entire set of factors controlling the formation of concentration levels, which can be feasibly performed by using the estimates of the central tendencies that are common to multiyear seasonal data series. It may be found necessary to make such estimates for harmful pollutants that cause deleterious effects on vegetation. Analysis of the upper parts of the distribution may be necessary for estimates of the probable pollution level from anthropogenic effects in impact areas. In problems connected with short-time variations of background concentration levels or, to be more exact, of the concentration level of pollutants in impact areas, estimates can occur that characterize data series of one-two decades or one-two months duration. From practice (Zelenyuk and Cherhanov, 1985), it is apparent that use of such data series for discrimination of components is difficult due to the fact that the formation mechanisms are not sufficiently understood. A further serious handicap is the inadequacy of the results of averaging meteorological characteristics over such short time-periods, which leads to strong scattering of the derived estimates and makes it impossible to compare estimates for different periods. The use of observational series for a period of more than a year does not allow one to obtain the necessary estimates, namely due to the mixing of several distributions. This is in good agreement with the statistical design of the investigations, from which it can be inferred (see paragraph 3.2) that data-series of less than one season and more than one-year duration demonstrate much less deviation from the lognormal distribution than seasonal data-series. In the first and second case, it can be assumed that the distributions are unimodal, if this is required for deriving estimates of the central tendencies. In the case of short-time series, these centers describe the average effects of different types of pollution distributions, typical of this short period and small areas of observation. For data series of one-year duration and more, such centers describe the average effects of all the probable pollution formation mechanisms operating within a sufficiently large region under different synoptic conditions.

5.2. *Estimates of Selective Grouping Intervals.*

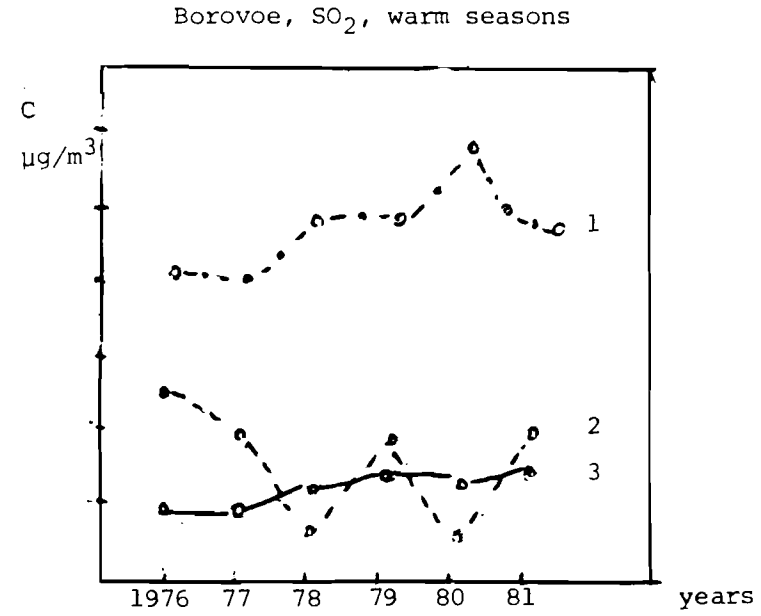
A method for estimating the grouping intervals in time-series was discussed in the previous chapter. Now let us see whether this method can be used to discriminate the components in seasonal data series.

Figures 5.2 - 5.10 present several characteristic data series, derived for different pollutants during different seasons and at different stations. As is apparent from the graphs, the components of background air pollution can be reflected in various types of deviations from the normal pattern. Analysis of such deviations does not enable one to judge which types of deviations are more typical for a particular station, pollutant or season. At the same time, the presence of distinctly

pronounced components in most of the graphs, again confirms the proposed model. In order to obtain evidence demonstrating the general nature of the regularities governing the formation of concentrations, we included data available on measurements of sulfur dioxide concentrations, that were performed in accordance with the international program on long-range transport of air pollutants. The data were taken from observational series obtained at stations located in the impact areas of Jergul, Norway and Abisko, Sweden. These data, differing according to their sampling and analytical techniques (which in practical work led to a need to recalculate the values; for comparison with the rest of the data array, the logarithmic concentrations should be reduced by 1.6), demonstrate the same type of distribution pattern as the data from the "Borovoe", "Berezina B.Z.", and "Repetek B.Z." stations. From this, the inference can be drawn that pollution concentration distributions are controlled by processes having common features, at least on the scale of continents. This implies that the statistical model enables one to describe observational data from background monitoring stations over entire continents. Such a description for each separately taken data series represents a set of intervals and corresponding weighted quantities, specified in the form of Equation 4.13.

The graphical estimate allows for informal interpretation of the compositing components, using two classes, "well" and "poorly" defined. As was shown in Section 3.3., such deviations do not influence essentially the principal "recognition" of the existing components. Studies in which experts analyzed the plots and distinguished the grouping intervals, served to indicate that the discrepancies in the estimates are small and, at any rate, much less than the probable variations in the estimates associated with the redistribution of the relative weighted quantities of the components, the case to be considered in Section 5.3. Estimates performed for each interval enable one to identify some rather general and statistically stable characteristics, so that we can judge the effects caused by different concentration formation mechanisms. In the Appendix, several plots are presented that are typical for estimates of the components, this refers to graphs A.5.1 - A.5.20. These graphs, together with the plots depicted in Figures 5.2 - 5.10 give a rather fair idea of the techniques used to distinguish the central tendencies on the basis of graphical discrimination.

Each of the series analyzed specified its own series of grouping intervals, that henceforth are termed the series of central tendencies. These intervals and series, tabulated in Tables 5.1 - 5.11, include practically all the statistical information available for analysis. In order to derive general conclusions, characterizing seasonal data series that are large in time and space, certain new statistics should be designed that generalize the random central tendencies. The following section is devoted to the design philosophy of such statistics.



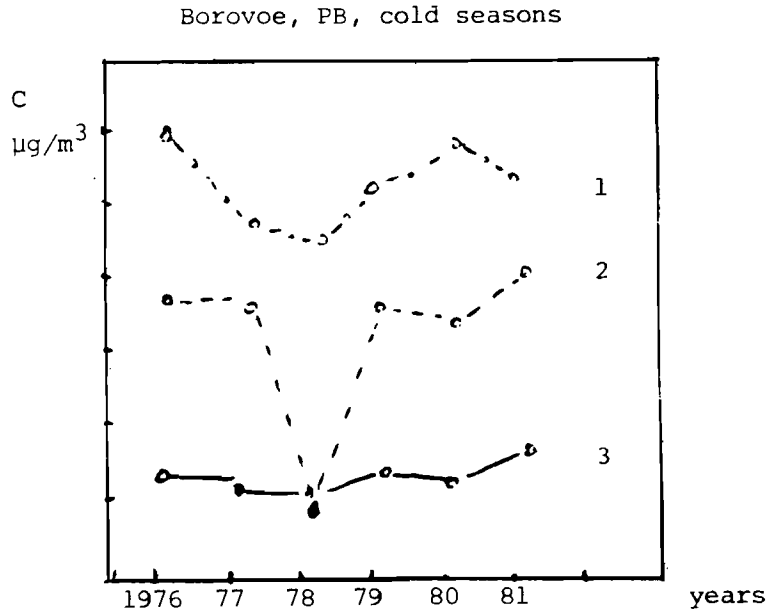


Figure 5.1: Results of estimation of two concentration central tendencies.
1 - lower component; 2 - upper component; 3 - composite series.

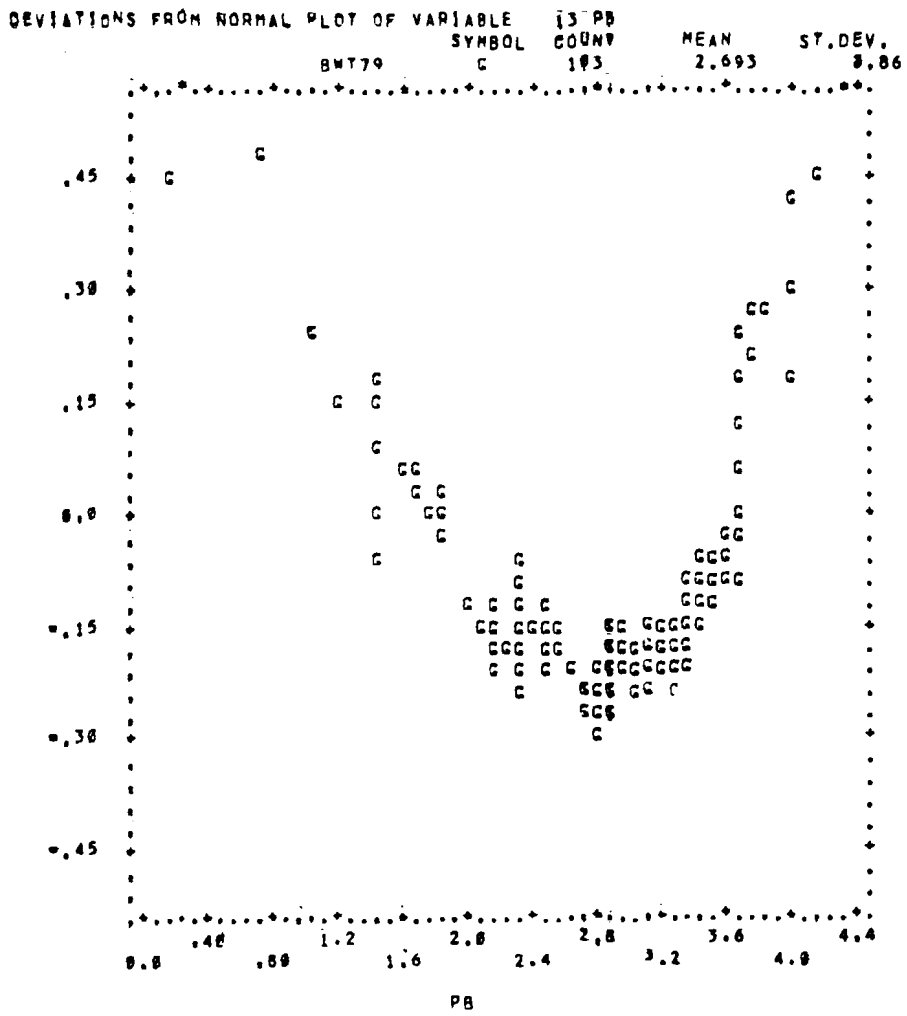


Figure 5.2: Deviations from normal plot of logarithmic concentrations of lead. Borovoe, warm season, 1979.

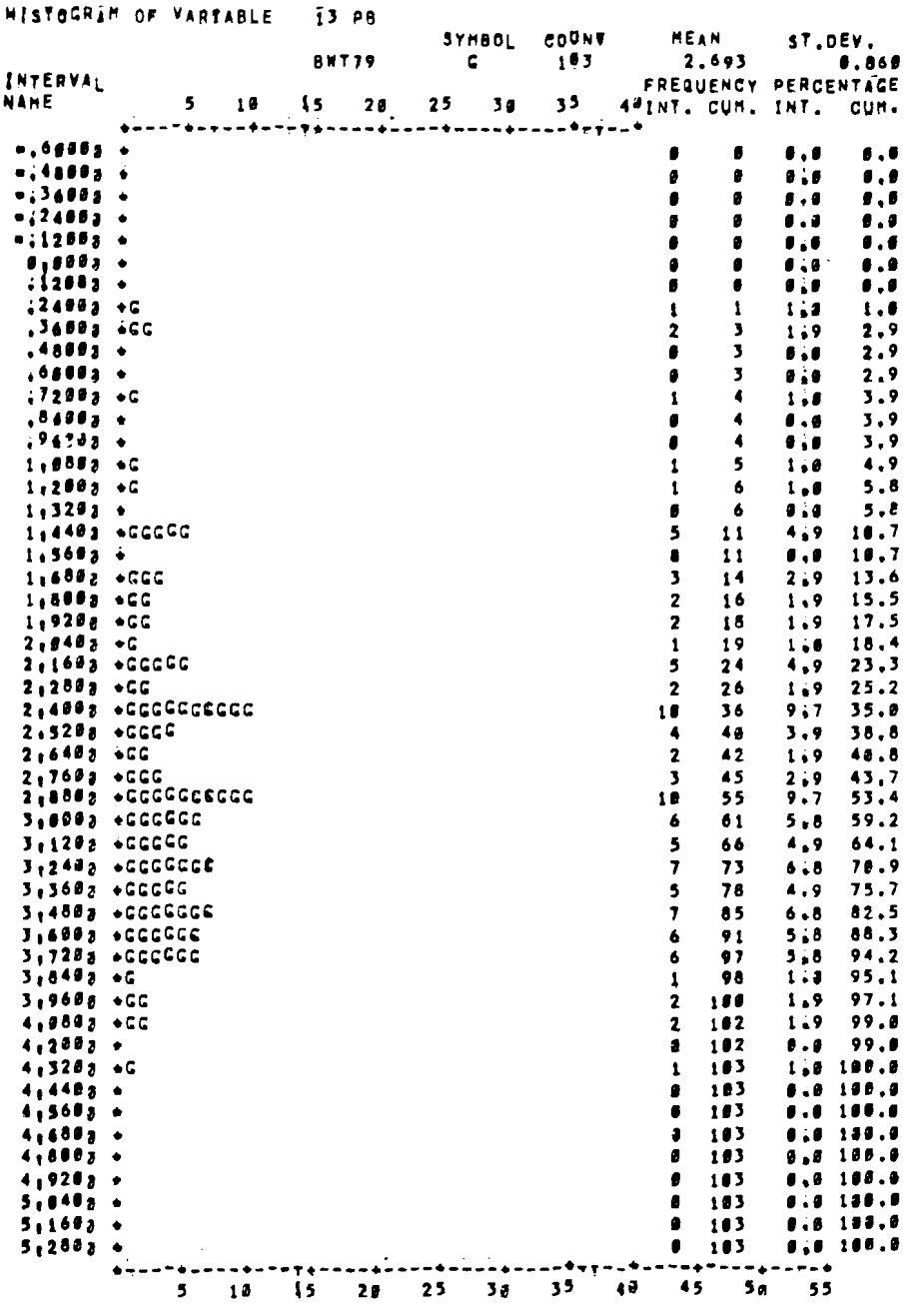


Figure 5.3: Histogram of logarithmic concentrations of lead. Borovoe, warm season, 1979.

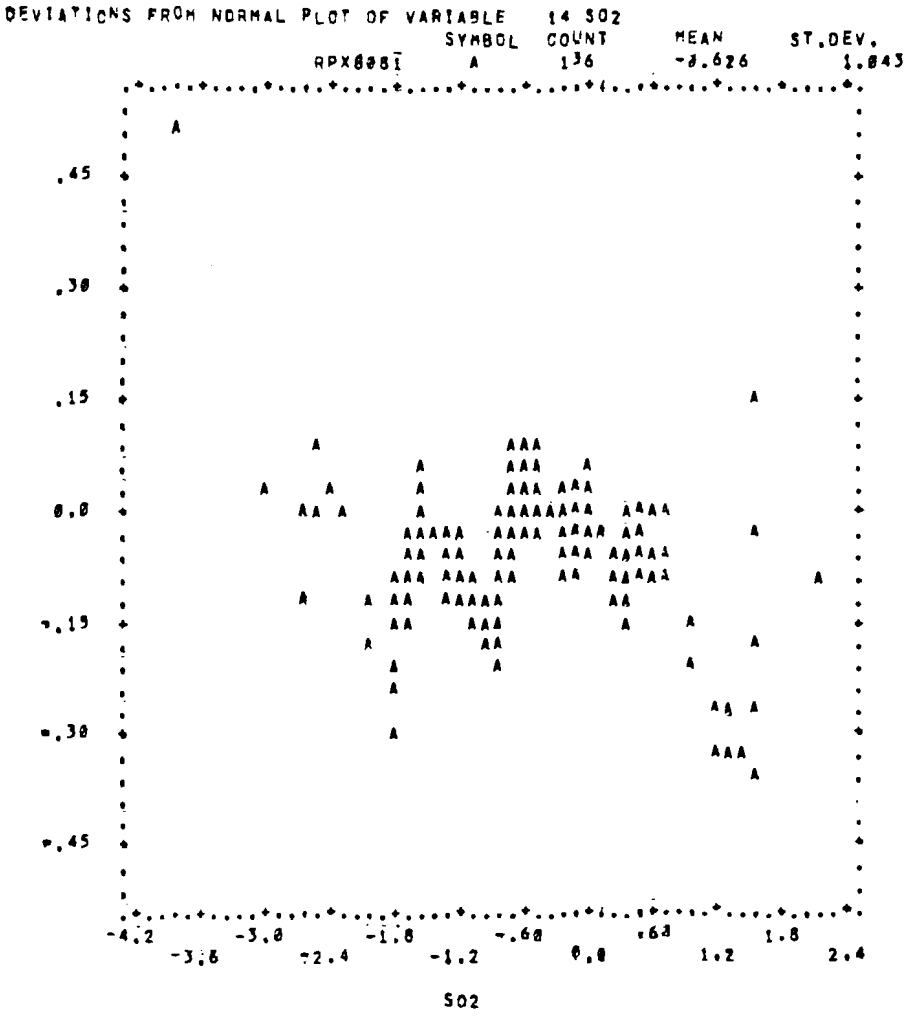


Figure 5.4: Deviations from normal plot of logarithmic concentrations of sulfur dioxide. Repetek, cold season, 1980-1981.

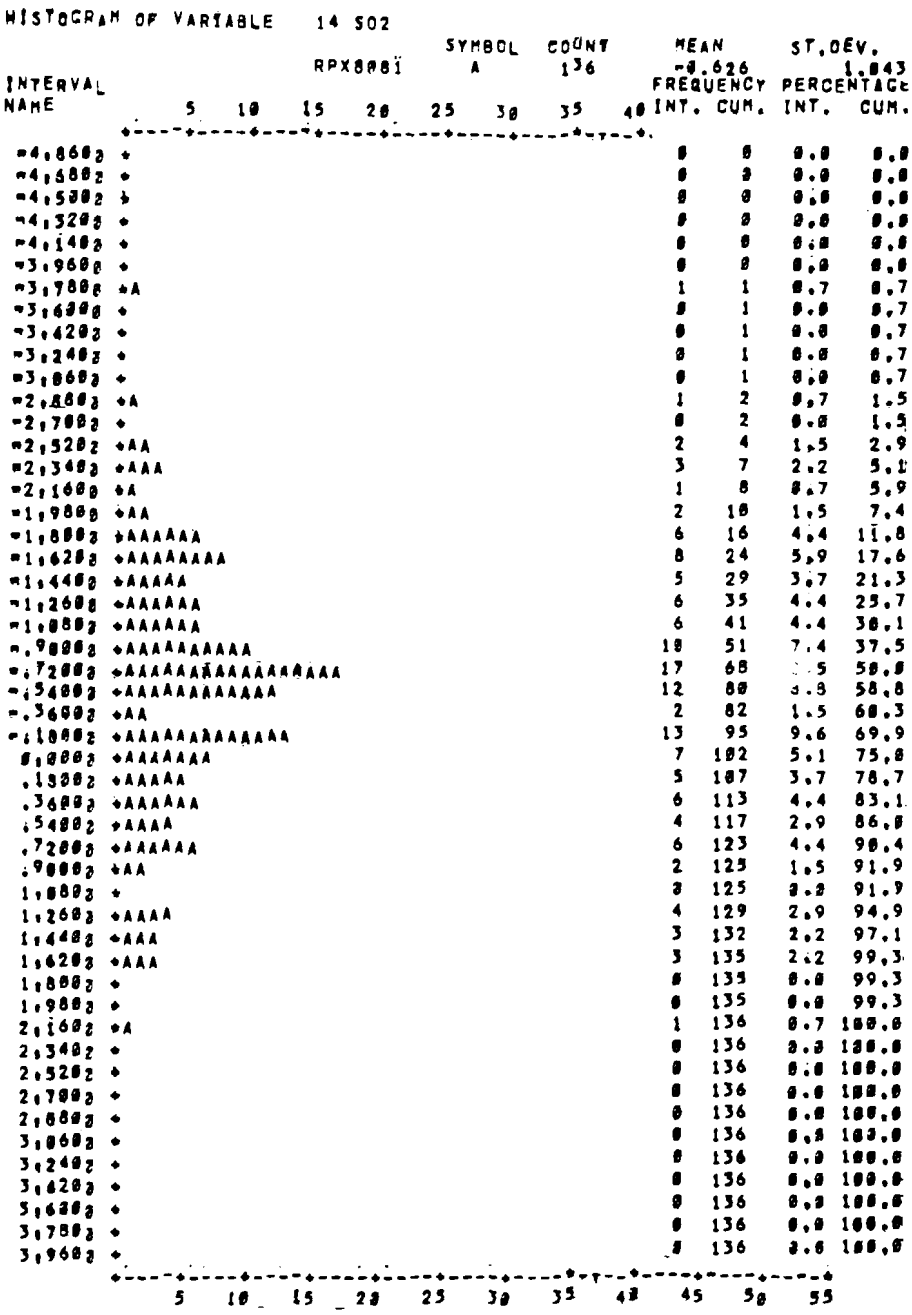


Figure 5.5: Histogram of logarithmic concentrations of sulfur dioxide. Repetek, cold season, 1980-1981.

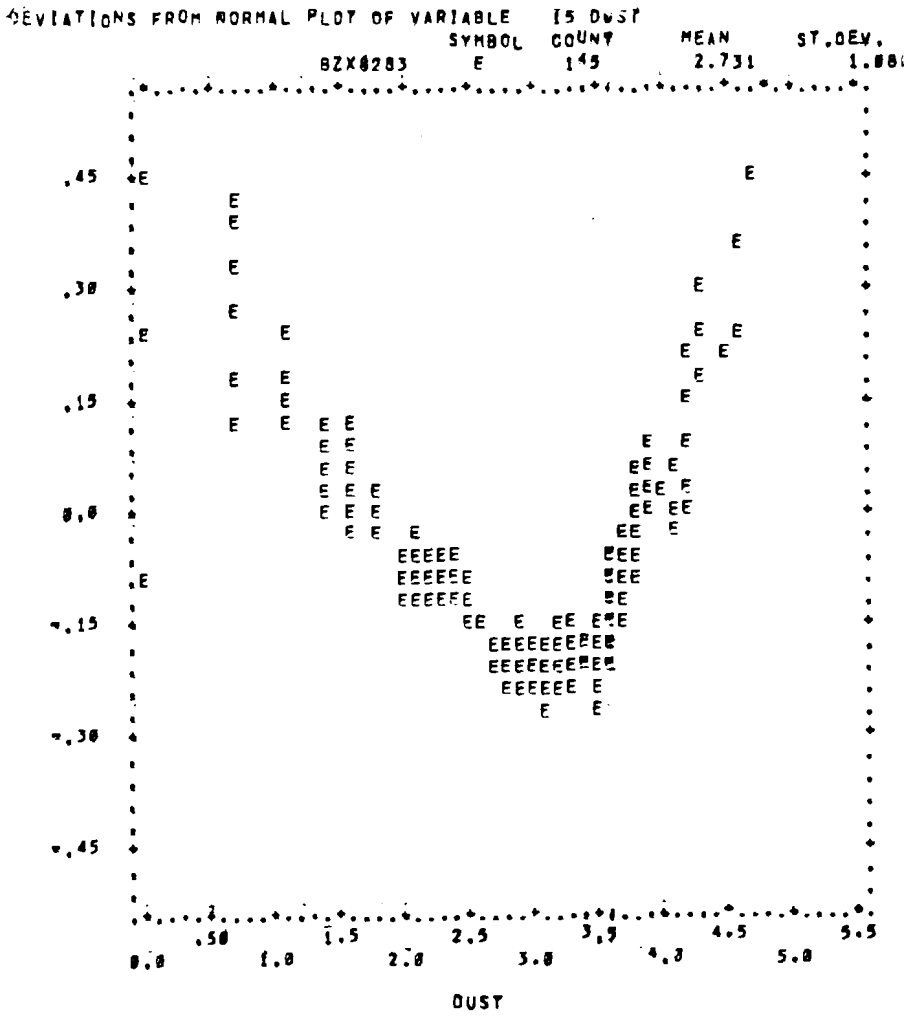


Figure 5.6: Deviations from normal plot of logarithmic concentrations of suspended particulate matter. Berezina, cold season, 1982-1983.

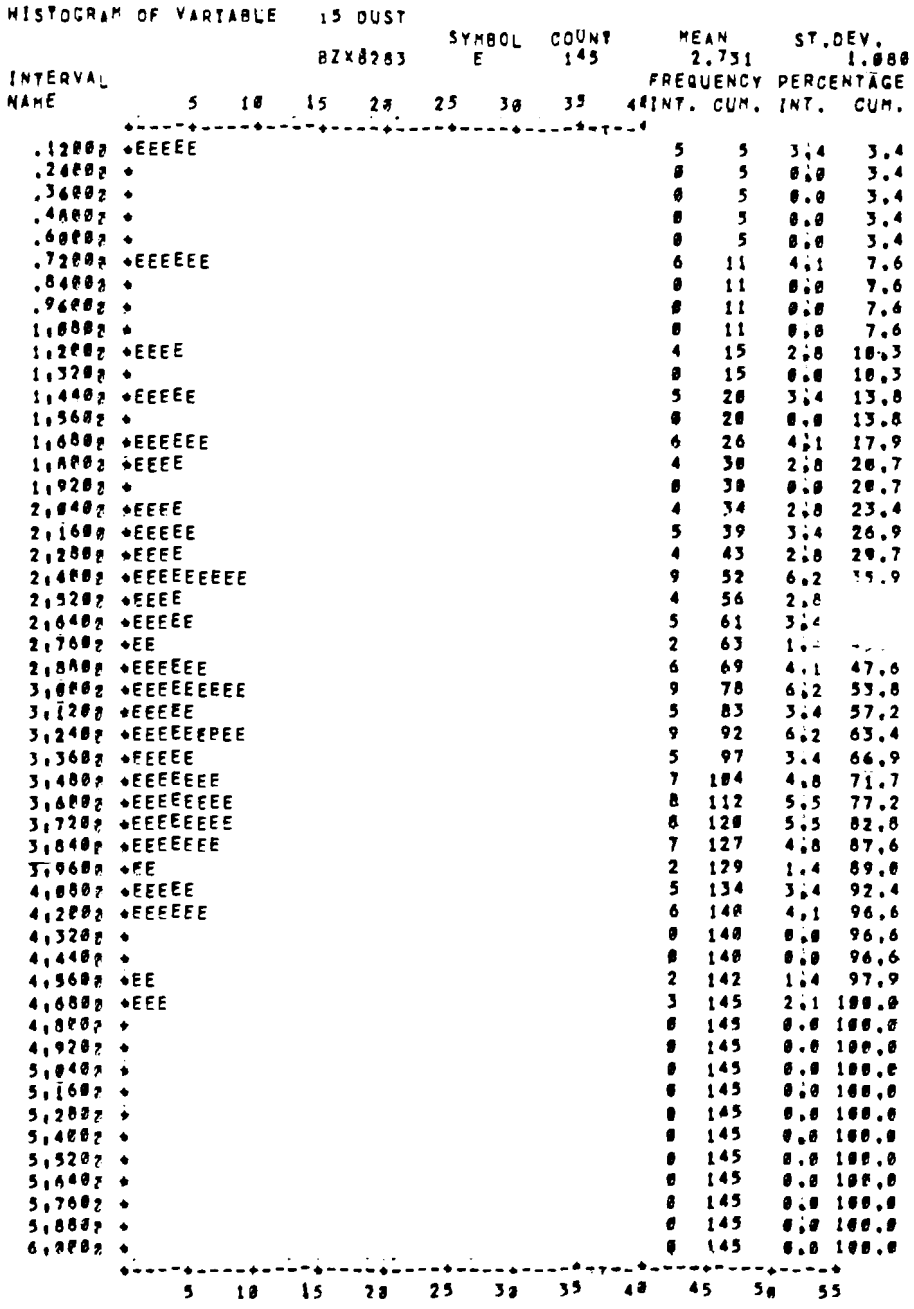


Figure 5.7: Histogram of logarithmic concentrations of suspended particulate matter. Berezina, cold season, 1982-1983.

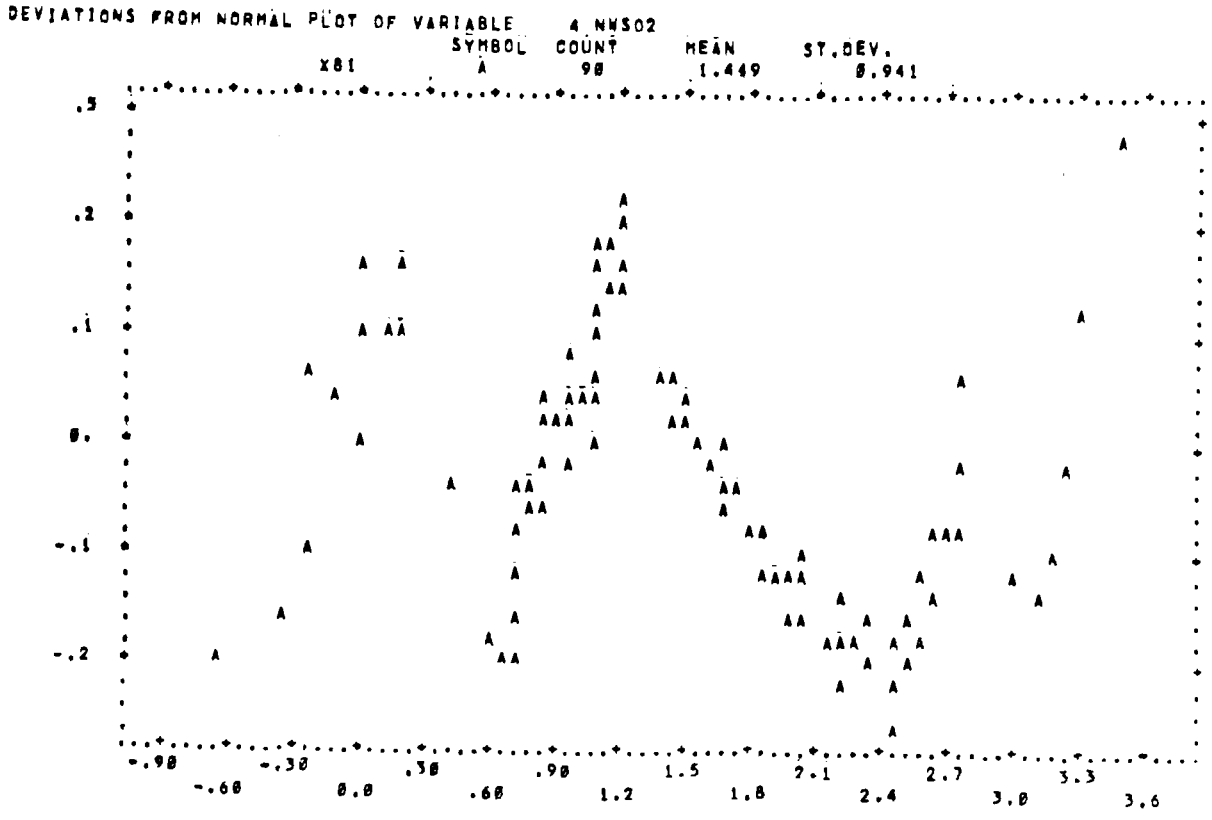


Figure 5.8: Deviations from normal plot of logarithmic concentrations of sulfur dioxide. Jergul, Norway, cold season, 1981.

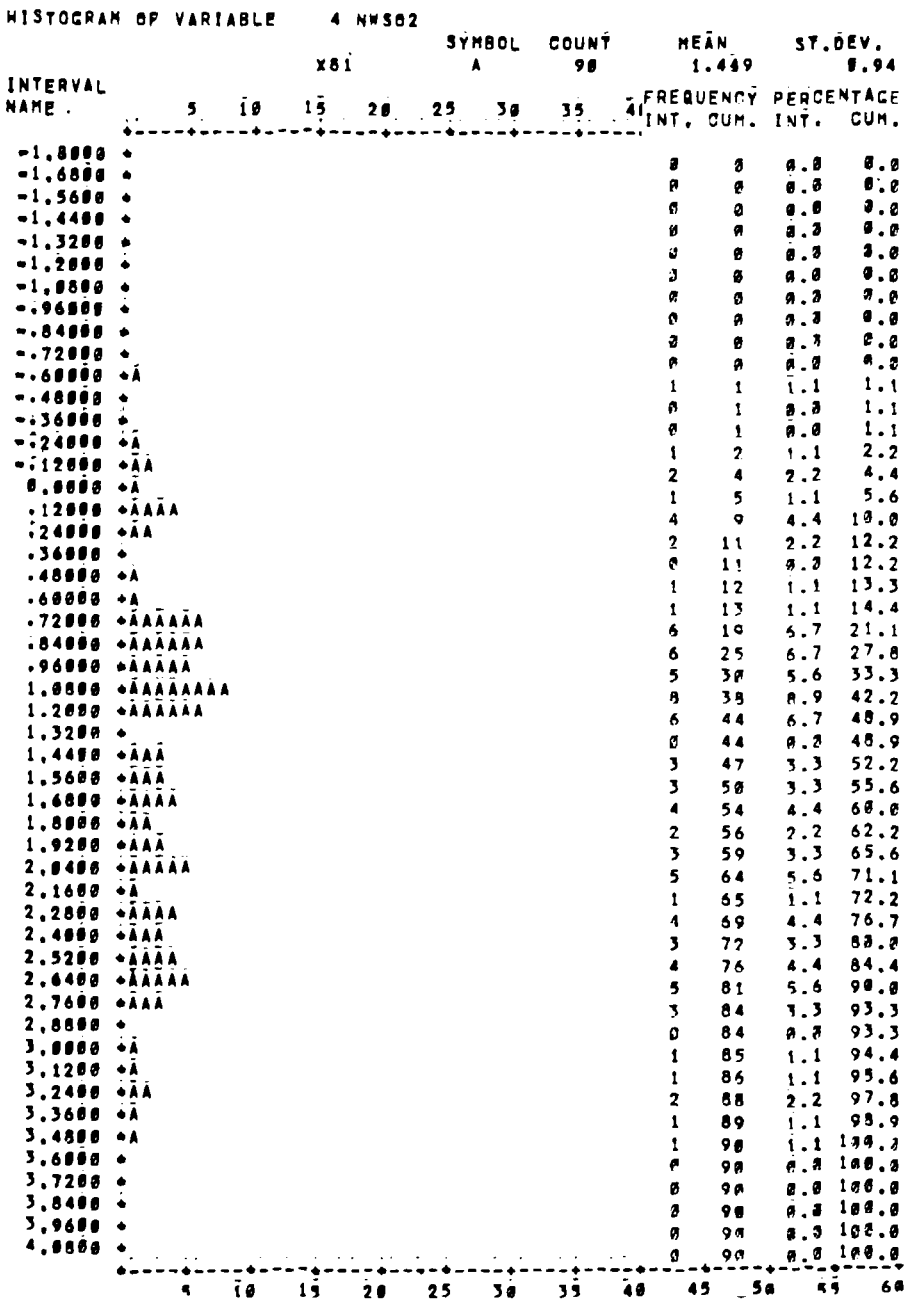


Figure 5.9: Histogram of logarithmic concentrations of sulfur dioxide. Jergul, Norway, cold season, 1981.

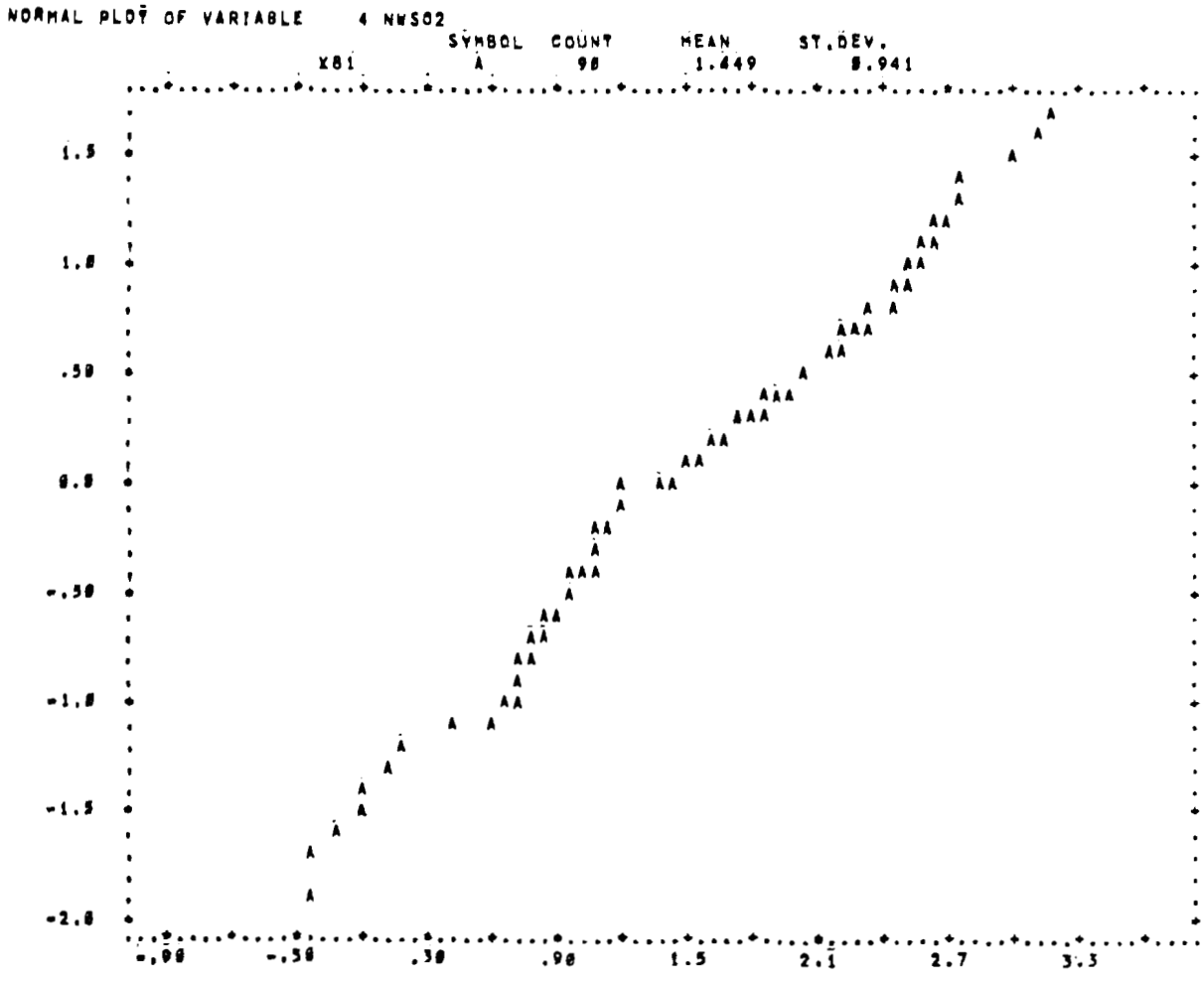


Figure 5.10 Normal plot of logarithmic concentrations of sulfur dioxide. Jergul, Norway, cold season, 1981.

Table 5.1: Random grouping intervals of logarithmic concentrations of sulfur dioxide. Borovoe, 1976-1983. T - warm season, 1978; X - cold season, 1979-1980; % - percentage of points falling within the interval.

T76	%	X7677	%	T77	%	X7778	%	T78	%
-2.0, -0.7	18			-3.0, -1.8	13	-2.4, -0.4	16		
				-1.8, -1.2	30				
-0.7, +0.1	17	-1.6, 0.2	26	-1.2, 0.2	57			-1.4, -0.1	40
-0.1, 0.8	60	0.2, 1.0	41			-0.4, 1.0	32	-0.1, 0.7	52
0.8, 1.5	5							0.7, 1.2	8
		1.0, 2.5	33			1.0, 1.6	22		
						1.6 , 2.6	30		

X7879	%	T79	%	X7980	%	T80	%	X8182	%
		-2.0, -1.2	10			-2.2, -1.2	30		
		-1.2, -0.8	30			-1.2, -0.2	20		
		-0.8, -0.4	37						
0.0, 1.1	32	-0.4, 0.0	13	-0.9, 0.5	14	-0.2, 0.5	16	-1.0, 0.5	17
		0.0, 0.7	10	0.5, 1.2	16	0.5 , 1.5	34	0.5, 1.9	42
1.1, 1.4	23			1.2, 3.3	70			1.9, 2.7	18
1.4, 2.2	29								
2.2, 2.9	15								
								2.7, 4.5	13

T81	%	X8182	%	T82	%	X8283	%	T83	%
								-2.0, -0.4	14
-1.5, -0.6	12			-1.4, -0.2	22				
-0.6, 0.2	25			-0.2, 1.0	72	-0.5, 1.1	12	-0.4 , 1.6	86
0.2, 1.2	46			1.0, 1.8	6				
1.2, 2.2	17	1.5, 2.1	15			1.1, 2.6	70		
		2.1, 3.1	46						
		3.1, 3.9	39			2.6, 4.0	18		

	X8384	%
	-0.5, 1.7	52
	1.7, 3.2	48

Table 5.2: Random grouping intervals of logarithmic concentrations of sulfur dioxide. Berezina B.Z., 1980-1983. T - warm season, 1981; X - cold season, 1981-1982. % - percentage of points falling within the interval.

X8081	%	T81	%	X8182	%	T82	%	X8283	%
		-1.5, -0.5	8			-1.4, -0.3	14		
-0.5, 0.6 0.7, 1.5	10 20	-0.5, 0.3 0.3, 1.5	42 37	0.3, 1.2	5	0.1, 1.0 1.0, 1.6	63 15	0.0, 0.8 0.8, 1.5	7 17
1.5, 2.1 2.1, 2.7	14 37	1.5, 2.5	13	1.2, 2.8	50	1.6, 2.0	5	1.5, 2.5	53
2.8, 3.3 3.3, 4.5	10 10			3.0, 4.0	45			2.5, 3.8	23

T83	%	X8384	%
-1.7, -0.2	20		
-0.2, 1.0	68	0.2, 1.3	16
1.0, 2.2	12	1.4, 1.8 1.8, 2.8	32 25
		2.9, 4.0	27

Table 5.3: Random grouping intervals of logarithmic concentrations of sulfur dioxide. Repetek B.Z., 1980-1983. T - warm season, 1981; X - cold season, 1981-1982. % - percentage of points falling within the interval.

X8081	%	T81	%	X8182	%	T82	%	X8283	%
-3.0, -2.4	6	-3.0, -2.3	18	-3.0, -2.5	2				
-2.0, -1.5	15	-2.3, -1.2	62	-2.3, -1.2	48	-2.3, -1.2	13	-2.2, -1.3	19
-1.5, -1.0	15	-1.2, -0.6	10	-1.2, 0.0	30	-1.2, 0.5	75	-1.2, 1.0	69
-1.0, -0.4	30								
-0.4, 1.0	25	-0.66, 0.9	10	0.2, 1.5	20	0.5, 1.0	12		
1.0, 2.0	10							1.0, 2.3	12

T83	%	X8384	%
-2.0, -1.3	4	-1.8, -0.8	23
-1.2, -0.5	24	-0.8, 0.1	43
-0.5, 0.4 0.4, 1.0	47 10		
1.0, 2.3	10	0.1, 2.3	32

Table 5.4: Random grouping intervals of logarithmic concentrations of sulfur dioxide. Repetek B.Z., 1980-1983. T - warm season, 1981; X - cold season, 1981-1982. % - percentage of points falling within the interval.

X8I	%	T8I	%
-2.5 , -I.5	I2	-2.6 , -I.6	8
-I.5 , -0.4	36	-I.6 , -I.I	I4
		-I.I , -0.2	23
-0.4 , 0.8	32	-0.2 , 0.8	38
0.8 , I.2	I3	0.8 , I.3	I4
I.3 , 2.0	7	I.3 , I.9	3

Table 5.5: Random grouping intervals of logarithmic concentrations of sulfur dioxide. Abisko, Sweden, 1981. T, X - warm and cold seasons, respectively. % - percentage of points falling within the interval.

X8I	%	T8I	%
-2.2, -I.6	7	-3.3, -2.5	5
		-2.5, -I.6	I3
-I.6, -0.7	26	-I.6, -0.8	23
-0.7, 0.0	27	-0.8, 0.7	40
0.0, I.6	35	0.7, I.9	I9
I.6, 2.3	5		

Table 5.6: Random grouping intervals of logarithmic concentrations of lead. Borovoe, 1976-1983. T - warm season, 1979; X - cold season, 1979-1980. % - percentage of points falling within the interval.

T76	%	X7677	%	T77	%	X7778	%	T78	%
0.3, I.5	26								
I.5. 2.3	34	I.2. 2.6	20	I.0, 2.0 2.0, 2.7	37 33	I.2. 2.5	I7	0.7, 2.5	45
2.3, 3.7	40	2.6, 3.2	36	2.7, 3.7	30	2.5, 2.9 2.9, 4.2	I5 68	2.5. 3.0 3.0. 4.2	4I I6
		3.2. 5.0	44						

X7879	%	T79	%	X7980	%	T80	%	X808I	%
		0.2. I.6	II						
I.5. 2.5	40	I.7. 2.6	30			I.3. 2.4	27	I.2. 2.4	2I
2.5, 3.I 3.I. 3.8	I6 24	2.6, 3.7	53	2.5, 3.7	40	2.4, 3.5	66	2.4. 3.I 3.I, 4.0	2I 49
3.8. 4.7	20	3.7. 4.3	6	3.7, 4.5	35	3.5. 4.I	7		
				4.5. 5.2	25			4.0. 5.2	9

T8I	%	X8I82	%	T82	%	X8283	%	T83	%
0.7, I.5	7					0.7, I.6	I0	0.0, I.5	25
I.5. 2.7	50	I.4. 2.8	32	I.0. 2.8	67	I.7. 2.3	24	I.5, 2.7	52
2.7. 3.5	43	2.8, 3.4	3I	2.8. 3.6	27	2.4, 3.2 3.2. 3.6	44 8	2.7, 3.7	23
		3.4, .4.6	37	3.6, 4.3	6	3.7, 4.7	I2		

Table 5.7: Random grouping intervals of logarithmic concentrations of lead. Berezina B.Z., 1980-1983. T - warm season, 1980; X - cold season, 1981-1982. % - percentage of points falling within the interval.

X808I	%	T8I	%	X8I82	%	T82	%	X8283	%
		I.I, I.7	6	0.6, I.5	8	0,6, I.I	I5	I.0, I.7	20
I.5, 3.I	3I	I.7, 2.I	22	I.6, 2.0	I2	I.I, 2.I	40	I.7, 2.9	43
		2.I, 2.8	36	2.I, 3.2	65	2.I, 2.9	30		
3.2, 3.5	22	2.8, 3.8	49	3.2, 3.8	I5	3.0, 3.9	I2	3.0, 3.7	27
3.6, 4.I	20							3.9, 4.3	I0
4.2, 5.2	I0								

T83	%	X8384	%
I.0, 2.I	29	I.0.I.7	I2
2.I, 2.5	I6	I.8, 2.6	42
2.5, 3.3	39		
3.3, 3.8	7	2.5, 3.5	29
		3.5, 3.7	I7
3.8, 4.3	7		

Table 5.8: Random grouping intervals of logarithmic concentrations of lead. Repetek B.Z., 1980-1983. T - warm season, 1981; X - cold season, 1981-1982. % - percentage of points falling within the interval.

X808I	%	T8I	%	X8I82	%	T82	%	X8283	%
				0.7, 2.2	42	0.6..I.6	2I	0.6, I.4	25
I.0, I.8	I0	I.3, 2.0	32					I.5, 2.0	I8
2.0, 2.7	45	2.0, 2.3	26			I.7, 2.3	36		
		2.3, 2.6	24						
2.8, 3.5	35	2.6, 3.2	I8	2.2, 3.0	I5	2.4, 3.0	49	2.0, 3.2	49
						3.0, 3.5	5		
3.5, 4.0	8			3.0, 4.2	I3			3.3, 4.3	8

T83	%	X8384	%
0.6, I.3	7	0.8, I.7	I2
I.3, I.7	27		
I.8, 2.4	50	I.7, 2.2	88
		2.3, 2.7	I2
2.4, 3.2	I5		
		3.2, 4.3	8

Table 5.9: Random grouping intervals of logarithmic concentrations of dust. Borovoe, 1976-1983. T - warm season, 1978; X - cold season, 1979-1980. % - percentage of points falling within the interval.

T76	%	X7677	%	T77	%	X7778	%
0.3, I.6	I6	0.8, 2.0	20			0.8, I.8	28
I.6, 3.8	60	2.0, 2.4 2.4, 3.I	I0 20	I.5, 3.3	39	I.8, 2.7 2.7, 3.3	23 39
3.8, 4.5	24	3.I, 4.0	50	3.3, 4.8	6I	3.3, 4.2	I0

T78	%	X7879	%	T79	%	X7980	%
		0.5, 2.0	34	0.9, 2.2	I7		
I.3, 2.8	23	2.0, 2.7	30	2.2, 3.2	23	I.0, 2.5 2.5, 3.2	20 42
2.8, 4.3	77	2.7, 4.0	36	3.2, 4.8	60	3.2, 4.7	35

T80	%	X808I	%	T8I	%	X8I82	%
2.2, 3.I	I5	I.6, 2.7 2.7, 3.7	2I 42	I.5, 3.3	26	I.6, 3.3	85
3.I, 3.8 3.8, 4.6	59 26	3.7, 4.3	37	3.3, 5.0	74	3.3, 4.0 4.0, 4.4	I0 5

T82	%	X8283	%	T83	%
		0.7, I.6	I2		
I.2, 3.2	20	I.6, 3.I	57	I.3, 3.2	53
3.2, 4.7	80	3.I, 3.8	3I	3.2, 4.2	47

Table 5.10: Random grouping intervals of logarithmic concentrations of dust. Berezina B.Z., 1980-1983. T - warm season, 1981; X - cold season, 1981-1982. % - percentage of points falling within the interval.

X808I	%	T81	%	X8I82	%	T82	%	X8283	%
				0.I, I.5	I3			0.I.I.0	8
		I.5, 2.5	II			I.0. I.8	8	I.0, 2.6	35
2.0, 2.6 2.8. 3.6	I5 42	2.6 .3.3	I9	I.8. 3.I	29	2.0. 2.8 2.8. 3.3	I7 20	2.6, 3.3	22
3.6 .4.0	30	3.6, 4.2	50	3.2. 4.3	67	3.3. 4.0	53	3.3. 4.0	25
4.2, 4.5	I2	4.2. 4.8	20					4.0, 4.8	II

T83	%	X8384	%
I.3. I.9	3		
2.0, 3.3	33	I.8, 2.7 2.8, 3.5	27 29
3.3, 4.8	60	3.5, 3.9	24
		4.0, 4.5	I0

Table 5.11: Random grouping intervals of logarithmic concentrations of dust. Repetek B.Z., 1980-1983. T - warm season, 1981; X - cold season, 1981-1982. % - percentage of points falling within the interval.

X808I	%	T8I	%	X8I82	%	T82	%	X8283	%
2.4. 2.9 2.9. 3.6	8 20	2.6 .3.8	53	2.7. 3.3	II	2.9, 3.6	I6	2.6 ,3.2	20
3.6 .4.7	53	3.8. 4.2	32	3.3, 4.3	50	3.6, 4.0	23	3.2. 3.9	5I
		4.2, 5.4	I5	4.4, 5.0	3I	4.0. 4.3 4.3. 5.0	23 38	3.9. 5.0	29
4.7, 6.4	I9			5.0. 6.0	8				

T83	%	X8384	%
2.9. 3.7	27	2.5. 3.5	20
3.7, 4.3	37	3.5, 4.3	49
4.3, 5.I	34	4.3. 5.0	3I

5.3. ***Methods of Construction of Statistically Stable Estimates of Pollution Components.***

Each series specifies, in fact, the subdivision of the logarithmic concentration axis into intervals. Thereby, the search for statistically stable characteristics of the series proceeds within the region over which they are defined, i.e., over the range of all intervals of the logarithmic concentration axis, henceforth denoted as I. Each of Tables 5.1 to 5.11 exemplifies a sample in space I. Let us denote the sample specified by the array of central tendencies as W. In order to compare the intervals and the subdivisions of the axis with each other, a measure of similitude is needed. Let us consider the intervals $I_1 = [\alpha_1, \beta_1]$, $I_2 = [\alpha_2, \beta_2]$. As a measure of their similitude, it is natural to take the quantity characterizing the cross-over of the intervals in relation to their sizes, i.e., the ratio of their cross-over to their unification.

$$\mu^1 = \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|}, \tag{5.1}$$

where $|I|$ is the length of interval I.

Such a definition of the measure of similitude, although convenient, has two shortcomings: firstly, it is inconvenient for computations and, secondly, it performs well only for central tendencies of common ground points. The next proximity measure eliminates such shortcomings:

$$\mu(I_1,I_2) = \frac{\min(\beta_1,\beta_2) - \max(\alpha_1,\alpha_2)}{\min(\beta_1,\beta_2) - \min(\alpha_1,\alpha_2)} \tag{5.2}$$

The measure μ assumes the value $-1 < \mu \leq 1$. It can be easily seen that if $\mu > 0$, $\mu'(I_1,I_2) = \mu(I_1,I_2)$.

Now, if we take a certain interval, for example I_0, we immediately obtain a set of numbers where each number corresponds to one of the intervals of the sample W, and characterizes the measure of its similitude to I_0. Having chosen a certain scalar measure for this set, for example, the average value for the measure of proximity, we get the functional number

$$F : I \rightarrow R^1 \tag{5.3}$$

estimated over the sample. The fact that such a representation is possible reflects the objective of the investigation, the search for stable statistics to describe the sample.

From examination of the Tables, where random grouping intervals are represented, it is apparent that many intervals are similar from one season to another. The postulate can therefore be offered that within the entire data series, a certain relationship exists which is differently manifested during different seasons. For a description of this relationship, we shall use the method of per-interval estimates: if the relationship can be defined by the intervals, then it itself can be described by the interval that represents the integrated effects of the grouping factors. Such an integral we shall term the statistical grouping interval. Now we can state the problem concerning the algorithm for discrimination of a statistically stable grouping interval.

Using μ, let us synthesize an algorithm for estimation over intervals of sample W of the measure as to how much the given interval I_0 can be regarded as a statistical concentration grouping interval:

1. Let us specify I_0 and $\mu_0 \in (0,1)$.
2. Let us design the section $S(I_0)$ in the form of a set of intervals, ordered in accordance with the series to which they belong, and for which the measure of similarity with I_0 exceeds μ_0.
3. For the section $S(I_0) = \{I_l\}_1^L$, we shall design a matrix M of $L \times L$ sizes, where the element (k,l) equals $\mu(I_k, I_l)$.
4. For the matrix M, we calculate the value

$$F_{W,\mu_0}(I_0) = \frac{1}{L^2} \sum \gamma_{k,l} \cdot \mu(I_k, I_l) . \qquad (5.4)$$

(In this case, the weighted quantities $\gamma_{k,l}$ are taken equal to 1).

Now the functional F can serve for description of the statistical characteristics of sample W, that can be expressed as extremal statistics.

For instance, by using the method of exhaustive search, we can find an interval I^*, on which F_{W,μ_0} attains its maximum. This interval should be naturally regarded as the statistically stable grouping interval μ_0. The threshold value for μ_0 is statistically stable for the defined interval I^*, if

$$F_{W,\mu_0}(I^*) = F_{W,\mu_0 \pm \sigma}(I^*) \qquad (5.5)$$

i.e., small variations of μ_0 do not cause changes in the functional number.

Let us assume that a statistically stable grouping interval has been found for the given sample W, if for μ^{*-} stable, a statistically stable interval I^* is found such that

$$F_{W,\mu^*}(I^*) > -\varepsilon .$$

As is evident from examination of the Tables of random central tendencies, the intervals are already ordered in accordance with the concentration values and their position within the "chain" of intervals, which ultimately designates a single statistical grouping interval. We shall demonstrate how such an interval can be found from the data tabulated in Table 5.12, including random central tendencies of logarithmic concentrations of sulfur dioxide for the "Borovoe" background monitoring station.

Let us examine the last four series from the Table corresponding to seasons T 82, X 8283, T 83, X 8384. Let us assume that we are interested in identifying the interval containing point 0.0. As a trial value of I_0, let us take $I_0 = [-1.0,1.0]$. The measures of proximity of the trial intervals to all other intervals of the series can be calculated from the sequence of intervals with maximal μ values. These intervals are [-0.2,1.0], [-0.5,1.1], [-0.4,1.6], [-0.5,1.7]. Besides, the interval [-0.2,1.0] is joined to the following interval [1.0, 1.8] - this does not follow from the requirements for a maximal value of the proximity measure, but from the fact that this interval includes only 6% of the total number of observations and cannot be used for independent analysis, since in this case it represents some weakly expressed process. The minimal value for the proximity measure is estimated in this case over the unified interval [-0.2, 1.8], and is 0.42. That is, it can be said

that the interval [-0.1,1.0] is of 0.4 statistical stability in relation to the intervals of the sample under study describing the four seasons. It is obvious that for the four intervals taken from the sequence, an interval can be specified that ensures maximal μ^* value and functional F. This interval, derived by averaging the limits, is [−0.7,1.5]. Now, considering the interval as I^*,we shall obtain the following μ-values for four of the sampled intervals: 0.77, 0.75, 0.95, 0.86. The chosen statistical grouping interval is of 0.75 statistical stability. The corresponding value for the functional, calculated from the matrix of mutual proximity of the intervals, is 0.8. This procedure represents a method for practical realization of the proposed algorithm, that enables the major conditions to be fulfilled, and ensures attainment of the functional extremum. It is obvious that such an exhaustive search is possible only because of the small sizes of the samples used. When the available intervals amount to several hundreds, then it becomes necessary to optimize the procedures.

This procedure for the search for statistically stable grouping intervals was applied to all of the tables including series of central tendencies. The derived estimates are the optimal ones, from the point of view of μ^*, i.e., maximal μ^*-stability is attained. Estimates for the central tendencies are statistics to be used to derive estimates of the next level - of statistical grouping intervals. With the aid of these estimates, statistical inferences can be drawn concerning multi-year processes and processes of large scales. The possibility of drawing such inferences is discussed in the following section.

5.4. *Analysis of Components of Background Air Pollution Components.*

The statistical grouping intervals for different pollutants and stations are listed in Tables 5.12 - 5.14. In addition to information on the limits of the intervals, the Tables include the mean weighted values of the components, averaged over the intervals involved in the formation of central tendencies, also the number of warm and cold seasons, during which the components, distinguished by the respective interval, were evident.

The selection of statistical grouping intervals is quite a statistically stable procedure in relation to seasonal changes or variations in the interval-estimations. This is supported by the fact that the estimate derived in the previous chapter of the interval [-0.7, 1.5] for the four seasons is close to the estimate for the entire data series [-0.2, 1.4) - their proximity measure is 0.7. Concerning variations in estimates of the intervals, it should be noted that the applied method of evaluation (averaging) strongly reduces their influence, and the possibility of unification of the central tendencies in the formation of a statistically stable grouping interval ensures stability of a number of intervals. In order to obtain series of central tendencies, the help of several experts was enlisted. Although differences in the estimates reached 50% at times, variations in the resultant estimates for the statistical grouping intervals did not exceed 10%. Of importance here is the choice of the μ_0 value which to a great extent determines the selection of the intervals. In our case, $\mu_0 = 3$ was used. For practical purposes this is quite sufficient, geometrically it is a measure of proximity, for example, of two equal intervals that overlap each other by half their length.

Thus, statistically stable characteristics have been obtained that reflect the relationships in background air pollution monitoring data series. These relationships are caused, presumably, by the presence of several mechanisms governing the concentration formation processes that differ in their effects, on account of which the statistical grouping intervals are interpreted as estimates of the areas of action of different components of background air pollution. The formulation of the problem of component discrimination implies detection of components inherent

Table 5.12: Statistically stable grouping intervals of sulfur dioxide concentrations, $\mu g/m^3$. Column 1 shows intervals; columns 2 and 3 - number of seasons during which this interval was observed, and the average percentage of observation falling within the given intervals (2 - warm, 3 - cold seasons, respectively), 4 - seasonal effects. A - Berezina B.Z.; B - Borovoe; C - Repetek B.Z., D - Jergul, Norway; E - Abisko, Sweden.

A

	1	2	3
	0.22, 0.8	3(14%)	0
	0.9, 4.0	3(79%)	4(21%)
4	4.0, 12.0	3(10%)	4(53%)
	16.0, 54.0	0	4(29%)

B

1	2	3
0.1, 0.33	5(23%)	1(16%)
0.25, 0.8	7(27%)	1(20%)
0.8, 4.0	7(68%)	7(38%)
4.0, 16.0	1(17%)	8(54%)
16.0, 56.0	0	3(23%)

C

	1	2	3
	0.06, 0.1	1(18%)	2(4%)
	0.12, 0.3	3(26%)	4(21%)
	0.3, 1.0	3(36%)	4(44%)
	1.0, 2.7	3(27%)	2(22%)
4	2.2, 9.0	1(10%)	3(16%)

D

	1	2	3
	0.08, 0.22	1(8%)	1(12%)
	0.22, 0.8	1(37%)	1(36%)
	0.8, 2.2	1(38%)	1(32%)
	2.2, 3.3	1(14%)	1(13%)
4	3.3, 6.7	1(3%)	1(7%)

E

1	2	3
0.04, 0.08	1(15%)	0
0.08, 0.2	1(13%)	1(7%)
0.2, 0.5	1(23%)	1(26%)
0.5, 5.0	1(59%)	1(63%)
5.0, 10.0		1(5%)

Table 5.13: Statistical grouping intervals of lead, ng/m^3. Columns 1, 2, 3 - see Table 5.12 for legends. 4 - seasonal effects. A - Berezina B.Z.: B - Borovoe; C - Repetek B.Z.

A

	I	2	3
	2.7, 5.5	3(16%)	3(13%)
	5.5, 18.0	3(60%)	4(50%)
	18.0, 45.0	3(23%)	4(27%)
4	45.0, 59.0	1(7%)	2(15%)
	57.0,100.0	0	1(10%)

B

I	2	3
1.5, 4.5	4(17%)	1(10%)
4.5, 12.0	8(47%)	6(23%)
12.0, 40.0	8(39%)	7(59%)
40.0, 56.0	3(6%)	5(30%)
56.0,100.0	0	2(17%)

C

	I	2	3
	1.8, 4.5	2(14%)	3(26%)
	4.5, 6.7	3(30%)	2(14%)
	6.7, 12.0	3(45%)	2(62%)
	12.0, 24.0	3(29%)	3(33%)
4	24.0, 57.0		4(9%)

Table 5.14: Statistically stable grouping intervals of concentrations of suspended particulate matter, $\mu g/m^3$. Columns 1, 2, 3 - see Table 5.12 for legends. 4 - seasonal effects. A - Berezina B.Z.; B - Borovoe; C - Repetek B.Z.

A

	1	2	3
4	1.1, 4.0	0	3(9%)
	3.3, 9.0	3(7%)	1(35%)
	7.0, 27.0	3(30%)	4(41%)
	27.0, 57.0	3(54%)	4(45%)
	54.0, 75.0	1(20%)	3(11%)

B

1	2	3
2.2, 7.0	2(16%)	4(21%)
4.5, 24.0	8.32%)	7(55%)
24.0, 56.0	8(67%)	7(29%)

C

	1	2	3
4			
	13.0, 36.0	3(32%)	4(15%)
	36.0, 59.0	3(31%)	4(50%)
	59.0, 100.0	3(37%)	3(30%)
	90.0, 160.0		2(14%)

to the given region, e.g., components associated with long-range transport, local and global anthropogenic effects, etc. It should be mentioned that such estimates can be derived only if additional information and parameters are introduced into the statistical model. The areas of action distinguished for the components by the proposed method, reflect certain general processes occurring in the impact regions and represent statistically derived relationships that, consequently, may be used for knowledgeable interpretation of the data. Such an analysis is outside the scope of the present study.

Let us examine the results of estimates of the effects of compositing of concentration formation factors, and the statistical relationships reflected in these estimates. From analysis of Table 5.12, it is apparent that the two lower intervals, which include the highest concentrations for all stations, result from the effects of factors that occur mainly during cold seasons. Thus, it become possible to distinguish these intervals for sulfur dioxide as effects of the heating season. In this case, the levels characterized by lower concentrations should be considered as background values; namely, below 4.0 $\mu g/m^3$ for Berezin B.Z. and Borovoe, and below 2.7 $\mu g/m^3$ for Repetek B.Z. For the impact regions of Norway and Sweden, these background levels are 3.3 $\mu g/m^3$ and 5.0 $\mu g/m^3$. 70% of the observations for all stations, including those in Norway and Sweden, fall within these limits. As can be seen, for all areas of observation, the derived estimates are close to each other, that is, they may be used as estimates of background concentration level for continents.

From analysis of the lead component, based on the data listed in Table 5.13, the following estimates were derived for the upper level of background values: 18 ng/m^3 for Berezin B.Z. and 12 ng/m^3 for Repetek B.Z. and Borovoe. Over 60% of the observations are below this limit.

On the basis of the data presented in Table 5.14, similar estimates can be made for total suspended particulates. According to the available data, the background concentration levels lie below 57 $\mu g/m^3$ for Berezin B.Z., 56 $\mu g/m^3$ for Borovoe, and 59 $\mu g/m^3$ for Repetek B.Z. Over 50% of the observations lie below these limits.

The estimates derived in this manner for background concentration levels of atmospheric pollutants are in agreement with concepts on the background value as a statistically stable concentration level typical of the area. However, this is not the only consideration on which the determination of the background concentration level should be based. In Section 5.1 we demonstrated the practical use of another intuitive method for determination of the background value (Rovinskii and Buyanova, 1982).

The intervals represented in Table 5.12 can serve as examples illustrating the foregoing. It is apparent that the interval 0.2 - 0.8 $\mu g/m^3$ is a statistically stable interval in relation to the series of data from the other stations. The same refers to the interval 5 - 30 ng/m^3 for lead shown in Table 5.13. These examples illustrate that statistical grouping intervals can be derived on these intervals themselves. It is obvious that the series of statistical grouping intervals, determining the subdivision of the concentration axis, can be used for the identification of intervals of the "second order of statistical stability". In this case, statistical stability should be understood in relation to the influence of specific regional factors, which leads to concepts of processes developing over continents and, accordingly, to the concept of a background air pollution level for continents. At least it can be stated that the estimates derived for sulfur dioxide and suspended particulate matter are better than those available from the literature. The use of lower concentration intervals as background estimates incurs difficulties due to their weak expression in terms of the weighted quantities of the components and their frequency of occurrence during different seasons. In respect to the estimates

presented herein, it can be said that they are manifested at Borovoe, Repetek and at the stations in Jergul, Norway and Abisko, Sweden with a frequency of about 30% for sulfur dioxide, and with a similar frequency for suspended particulate matter at Borovoe, Berezin and Repetek.

The components so identified may be used not only for the development of concepts on the background levels of pollution, but also to analyze the dynamics of the effects of air pollution at the background level. An example can be offered illustrating such effects for sulfur dioxide at the Borovoe station. Consider the concentration interval 4.0 - 16.0 $\mu g/m^3$. This interval is present in all cold seasons, the frequency distribution revealing that 50% of the data fall within this band. There is reason to believe that the chosen interval reflects the winter effects of anthropogenic factors, particularly of the heating season. If we examine this interval during different seasons, and estimate the corresponding central tendencies, an interesting fact emerges - the centres, appearing relatively stable, exhibit a distinct trend. They can be represented by the following row of numbers: 5.5, 6.0, 7.3, 9.0, 10.0, 10.0, 6.0, 12.0, showing a more than two-fold increase over an eight-year period. Thus our method of analysis may provide estimates of trends in the background air pollution components. However, such an investigation demands greater knowledge of the process of formation of background air pollution components, which requires, first of all, a feasible analytical treatment of the estimates derived with the aid of the statistical model of background air pollution.

In concluding this chapter, let us discuss another statistical characteristic, relying on the use of the components of pollution distinguished in this study. Since, as has been shown above, different components make unequal contributions to the general level of pollution, it is natural to ask - what is the share of one or another component? An answer cannot be obtained from analysis of the data presented in the Tables, reflecting only the level of pollution typical of the components, and the frequency of their occurrence in the data array. The "weight" of the component can be characterized by the total concentration of the components during the entire period of their occurrence, as related to the sum of all concentrations for the period of time under study. In practice, such a summation is performed separately for the observations that fall within different grouping intervals, i.e., the sum C_i can be represented as:

$$\sum_{i=1}^{N} C_i = \sum_{C_i > a_0}^{C_i < a_1} C_i + \sum_{C_i > a_1}^{C_i < a_1} + \ldots + \sum_{C_i > a_{k-1}}^{C_i < a_k} C_i .$$

The weights of the components, λ_i are defined as

$$\sum_{C_j \in [a_{i-1}, a_i]} C_i \; / \sum_{1}^{N} C_i , \quad \sum_{1}^{k} \lambda_i = 1 .$$

Let us now present some results derived from analysis of the weights of pollution components.

From an analytical treatment of the data series of sulfur dioxide, the total weight of the components, clustered above the 1 $\mu g/m^3$ level, was found to be: at Berezin B.Z. for the warm periods - over 95%, for the cold periods - 100%; at Borovoe for the cold period - over 90%, for the warm period - over 85%; at Repetek B.Z. for the cold periods - over 70%, for the warm periods - 70%. From the data series on lead concentrations, the total weight of the components exceeding 5 ng/m^3 were over 95%, both for the warm and cold seasons, for the data from the Berezin, Borovoe and Repetek B.Z. stations. At the same time, in nearly all cases, concentration

intervals can be distinguished that are responsible for the major part of the pollution. For instance, for sulfur dioxide such intervals are: at the Berezin B.Z. station for the warm periods - [0.9, 4.0] (73%); at Borovoe for the warm periods - [0.8, 4.0] (80%); at Repetek B.Z. for the warm period - [1.0, 2.7] (53%).

Analysis of the weights and dynamics of the components is a subject area of special interest.

It can be seen that the two lower components from intervals [0.1, 0.3] and [0.25, 0.8] make an insignificant contribution. Statistically most stable is the contribution of the component [4.0, 16.0]. The highest variability is typical of the components from intervals [0.8, 4.0] and [16.0, 56.0]: during the period under discussion they changed 9 - 11 times, the changes being mutually interrelated. This pattern shows that in the area of the Borovoe station, the effects of the heating season have increased and shows the specific component responsible for this increase.

On the basis of the foregoing, the weighted values of the pollution components can be recommended as statistical characteristics, reflecting the nature of the background air pollution in impact areas, and the suggestion is advanced that they should be used as criteria for tracking the dynamics of background pollutants, which is an essential monitoring problem.

6. CONCLUSIONS

The design philosophy employed in this study of a statistical model of background air pollution had the following objectives:

1. Evaluation of the information content of background monitoring data for the description of the behavior, in space and time, of atmospheric pollutants arriving from impact areas, and in order to distinguish background air-pollution characteristics common to all time-periods and different monitoring stations.

2. Elaboration of methods for the derivation of background air-pollution characteristics of temporal and spatial statistical stability.

The major results are as follows:

1. It has been demonstrated that the data obtained from measurements of different pollutants at different background monitoring stations, serve to define the subdivisions of the concentration scale into zones of action of several major pollution components. The problem of distinguishing the pollution characteristics common to different time periods and stations is therefore reduced to the problem of comparing the subdivisions of different concentration scales. This result has been derived on the basis of the design, analysis and interpretation of the statistical model, herein proposed.

2. Mathematical techniques have been designed for the discrimination of concentration grouping intervals in the event-data series, that are statistically stable in time and space, and can be interpreted as manifestations of background air pollution. This result was derived by employing methods specially developed for estimation of statistically stable grouping intervals, and for their interpretation as characteristics of different components of background air pollution.

A number of inferences have been drawn concerning the nature of the data and analytical methods. They can be formulated as follows:

- Notwithstanding the high variability of the event-data, the body of available information and its accuracy allow one to distinguish in the data-series the effects of the same probabilistic processes.

- The information embodied in the observational series can be retrieved with the aid of the proposed statistical model, and represents different subdivisions of concentration scales, typical of the monitoring station and period of observation.
- Background air pollution in each area is controlled by several groups of factors that are manifested in the forms of different levels of pollution, i.e., the physical effects of different components of pollution are a statistical manifestation of the different zones of the concentration scales.
- A typical example, representing a typical time interval, that best reflects the action of different pollution components, is a seasonal data series.
- The subdivisions of the concentration scales, characterizing the effects of pollution by different pollutants measured at one observing station during several seasons, reveal common features that enable one, with the use of specially designed techniques, to distinguish the components of pollution that are statistically stable and typical of the given area of observation.
- On the basis of the statistically stable pollution components, inferences can be drawn concerning the processes of air pollution in impact areas, and normal pollutant concentration levels in these areas.
- Analysis of the frequency of occurrence of the components during different seasons, and of the share of different components to the general pollution of the atmosphere, enables one to describe the seasonal concentration variations; to identify the components that experience distinct anthropogenic effects, and to distinguish them from the components that define the background air-pollution level proper and to draw conclusions concerning the time variations for different components. Employing these techniques, estimates were derived for background air-pollution concentration levels for different pollutants and different monitoring stations, and the inference was drawn that sulfur dioxide has undergone a considerable increase in the area of the Borovoe station during winter periods, due to anthropogenic effects.

Hence, the statistical model herein proposed is a device designed to derive statistical information that can be interpreted explicitly. The problem concerning the comparison of the statistics, reflecting local and regional effects, and incidents of global effects (Dege, 1982; Zelenyuk, 1984) can be solved within the framework of the proposed model by comparison of different statistical characteristics, using them as source-material for designing "statistics from statistics" that in terms of the model can be used for the description of the effects of events of large magnitudes. For instance, the use of the method of statistically stable interval estimation in application to intervals describing manifestations of pollution components, typical of different stations, enables one to pinpoint the intervals that are not only statistically stable in space, but also in time, i.e., to distinguish concentration intervals revealing certain common features within the continents.

These techniques require computational facilities and advanced computer programs; see Vipke (1985) and Dlikman and Katz (1982). Further development of the background monitoring network, and expansion of the proposed methods over a broad class of background monitoring problems, call for the creation of effective man-machine systems in order to obtain relevant inferences from the accumulated data.

Use of the proposed model and application of related statistical characteristics for estimation of background air pollution, enable valid and statistically stable estimates to be derived concerning the actual state of the natural environment. This is regarded as one of the most essential problem areas to be solved using the background monitoring system.

REFERENCES

Aitchison, J. and Brown, J.A.C. (1957) *The Lognormal Distribution.* Cambridge University Press, London.

Aivazyan, S.A., Enyukov, I.S. and Meshalkin L.D. (1983) *Prikladnaya statistika (Applied statistics),* Moscow: Nauka (in Russian).

Anokhin, Yu.A. and Ostromogil'skii, A.Kh. (1978) Matematicheskoe modelirovanie i monitoring okruzhayushchei sredy (Mathematical modeling and environmental monitoring, Obninsk:VNIIGMI-MTsD (in Russian).

Antonovskii, M.Ya., Bukhshtaber, V.M., and Zelenyuk, E.A. (1985) Background atmospheric pollution analysis on the basis of multi-mode distributions. Proceedings of the Internal Symposium on Integrated Global Monitoring of the State of the Biosphere, II Tashkent, USSR, 14-19 October, 1985. Tech. Doc. WMO/TD No. 151, Feb. 1987. *Tezisy dokl. III Mezhdunarodnovo simpoziuma "Kompleksnyi global'nyi monitoring sostoyaniya biosfery"* (Abstract of Report to III International Symposium "Complex Global Monitoring of the Biosphere State") Leningrad: Gidrometeoizdat, p.53-54 (in Russian).

Augustinyak, S. and Sventz, S. (1982) Determination of Environmental Changes on the Basis of the Generalized Signal Theory, in: *Problemy fonovov monitoringa sostoyaniya prirodnoi sredy* (Problems of natural environment background monitoring). Vip.2, Leningrad: Gidrometeoizdat, pp.205-213 (in Russian).

Benarie, M.M. (1982) Air-Pollution Modeling Operations and Their Limits. In: *Mathematical Models for Planning and Controlling Air Quality.* IIASA Proceedings Series, v.17, pp.109-117.

Bencala, K.E. and J.H. Seinfeld (1976) On Frequency Distribution of Air Pollution Concentrations. *Atmospheric Environment,* **10,** 941 p.

Bencala, K.E. and Seinfeld, J.H. (1979) An Air Quality Performance Assessment Package. In: *Atmospheric Environment,* **13,** pp. 1181-1185.

Berlyand, M.E. (1975) Sovremennye problemy atmosfernoi diffizii i zagryanzneniya atmosfery (Modern problems of atmospheric diffusion and air-pollution) Leningrad: Gidrometeoizdat (in Russian).

Berlyand, M.E. (1984) Bearing on the Fundamental Principles for Air-Pollution Prediction. In: *Sb. dokl. na Mezhdunarodnom soveshchanii VMO PA VI* (Coll. of Reports to the International Conference VMO PA VI), Leningrad: Gidrometeoizdat, pp.9-15 (in Russian).

Berlyand, M.E., Volberg, N.S., Lavrinenko, N.F. and Rusina, E.N. (1982) Problems of Correlation Between Local and Global Monitoring of Air Pollution. *Environmental Monitoring and Assessment.* v.2, No.4, pp.393-402.

Burtseva, L.V., Lapensko, L.A., Volosneva, T.A., and Vas'kovskii, A.T. (1982) Lead, Cadmium, Arsenic and Mercury Concentrations in the Atmosphere According to the Results Derived at the Borovoe Background Monitoring Station During the Period 1977-80. In: *Monitoring fonovovo zagryazheniya prirodnoi sredy (Background pollution monitoring of the natural environment)* Vip.1, Leningrad: Gidrometeoizdat, pp.101-111 (in Russian).

Burtseva, L.V., Volosneva, T.A., Lapenki, L.A., and Pastukhov, B.V. (1982) Comparison of Methods for Monitoring Heavy Metals, Sulfir Dioxide and Sulfates. In: ***Monitoring fonovovozagryazneniya prirodnoi sredy*** (Background pollution monitoring of the natural environment) Vip.1, Leningrad: Gidrometeoizdat, p.212-224 (in Russian).

Bukhshtaber, V.M., Zelenyuk, E.A., and Maslov, V.K. (1983) Methods of Analysis and Development of Automatic Classification Algorithms on the Basis of Mathematical Models. In: ***Prikladnaya statistika. Uchenye zapiski po statistike.*** (Applied Statistics. Scientific notes on statistics). T.45, Moscow: Nauka, p. 126-144 (in Russian).

Byulleten' fonovovo zagryazneniya okruzhayuschchei prirodnoi sredy v regione vostochno-evropeiskikh stran-chlenov SEV (1983) (Bulletin of background pollution of the natural environment in the region of East-European Members Countries of CMEA) Vip.1, Moscow: Gidrometeoizdat (in Russian).

Byulleten' fonovovo zagryazneniya okruzhayushchei prirodnoi sredy v regione vostochno-evropeiskikh stran-chlenov SEV (1984) (Bulletin of background pollution of the natural environment in the region of East-European Members Countries of the CMEA) Vip.2, Leningrad: Gidrometeoizdat (in Russian).

Cobb, L. (1978) Stochastic Catastrophe Models and Multimodal Distributions. *Behavioral Sciences*, No.23, pp.360-374.

Dege, S. (1982) Simulation of Intraregional and Interregional Transfer of Pollutants and Evaluation of the State of the Natural Environment. In: ***Proplemy fonovo monitoringa sostoyaniya okruzhayushchei prirodnoi sredy (Problems of natural environment background monitoring)*** Vip.1, Leningrad: Gidrometeoizdat, pp.141-147 (in Russian). Warsaw, v.82, No.3-5, pp.23-37.

de Nevers, N., Lee, K.W. and Franc, N.H. (1979) Patterns in TSP Distribution Functions. *Journal of APCA,*

Dlikman, F., and Katz, B. (1982) Certain Specific Features of Informational Provision for Problems of Background Monitoring of the Natural Environment. In: ***Problemy fonovovomonitoringa sostoyaniya okruzhayushchei prirodnoi sredy*** (Problems of natural environment background monitoring). Vip.1. Leningrad: Gidrometeoizdat, pp.141-147 (in Russian).

Gibbons, D.T. (1978) An Evaluation of Two Mode Specification Tchniques for a Lognormal Distribution. *IEEE Transactions*, v. R27, No.1, pp.60-63.

Gnadanesican, P. and Kettering, J.R. (1972) Robust Estimates, Residuals and Outlier Detection with Multiresponse Data. *Biometrics*, No.28, pp.81-124.

Gruza, R.V. and Reitenbakh, R.G. (1982) ***Statistika i analiz gidrometeorologischeskikh dannykk*** (Statistics and analysis of hydrometeorological data), Leningrad: Gidrometeoizdat (in Russian).

Hangos, K.M. (1983) Application of Catastrophe Theory Models in Simulation of Industrial Noise Sources. ***Computer and Automation,*** Hungarian Academy of Sciences, Budapest, pp.53-59.

Harris, E.D., and Tabor, E.D. (1956) Statistical Considerations Related to the Planning and Operation of National Air Sampling Network. *Proceedings of the 49th Annual Meeting of APCA, Buffalo, pp.7-9.*

Horowitz, J. and Baracat, S. (1979) Statistical Analysis of the Maximum Concentrations of Air Pollutants, Effects of Autocorrelation and Non-Stationarity. *Atmospheric Environment*, v. 13, pp.811-818.

Hunter, J.S. (1981) Environmetrics: Mathematics and Statistics in the Service of the Environment. *Environmetrics-81, Selected Papers*, SIAM, Philadelphia, pp.3-11.

Izrael, Yu.A. (1984) ***Ekologiya i kontrol' sostoyaniya prirodnoi sredy*** (Ecology and Environmental Control) Moscow: Gidrometeoizdat (in Russian).

Izrael, Yu.A. and Novikov, Yu.V. (1985) Cosmic Geological Monitoring of the Anthropogenic State of the Natural Environment. In: ***Tez. dokl. III Mezhdunarodnovo simpoziuma "Kompleksnyi global'nyi monitoring sostoyaniya biosfery"*** (Abstr. Reports III International Symposium "Complex Global Monitoring of the Biosphere State") Moscow: Gidrometeoizdat, pp.15-16 (in Russian).

Izrael Yu.A., Rovinskii, F.A., Antonovskii, M.Ya., Bukhshtaber, V.M., Zelenyuk, E.A., and Cherkhanov, Yu.P. (1985) K statisticheskomu obosnovaniyu komponent zagryazneniya atmosfery v fonovom raione. (To the Statistical Validation of Air-pollution Components in Normative Areas). *DAN, 276, No.2*, Moscow, pp.334-337.

Kalpasanov, W. and Kurchatova, G. (1976) A Study of the Statistical Distributions of Chemical Pollutants in the Air. *Journal of APCA*, No.26, p.981.

Karasev, B.V. (1980) Lognormal'nyi zakon raspredeleniya i sokhranenie logarifmicheskoi dispersii (Lognormal distribution law and preservation of the logarithmic variance) *Zh.F.Kh.T. IV, No.2*, pp.3032-3037 (in Russian).

Karasev, B.V. (1982) Vozmozhnosti primeneniya lognormal'novo zakona k opisaniyu diffuzionikh yavlenii (On the application of the lognormal law to the description of diffusive phenomena) *Zh.F.Kh., T. VI, No.2*, p.357-365.

Khan, D.T. (1973) Note on the Distribution of Air Pollutants. *Journal of APCA*, No.2, p.973.

Kleiner, B. and Gradel, T.E. (1980) Exploratory Data Analysis in Geophysical Sciences. *Reviews of Geophysics and Space Physics*, V.18, No.3, pp.699-717.

Lamb, R.J. (1981) Air Pollution Modeling as a Problem in Statistics. *Environmetrics-81, Selected Papers*, SIAM, Philadelphia, pp.13-28.

Larsen, R.J. (1961) A method for Determining Source Reduction Required to Meet Air Quality Standards. *Journal of APCA*, No.11, p.71.

R.J. Larsen (1969a) Determining Reduced-Emission Goals Needed to Achieve Air Quality Goals: a Hypothetical Case. *Journal of APCA*, No.19, pp.24.

Larsen, R.J. (1969b) A New Mathematical Model of Air Pollutant Concentrations Averaging Time and Frequency (1969b), *Journal of APCA*, No.19, pp.24.

Larsen, R.J. (1979) An Air Quality Performance and Data Analysis System for Interrelating Effects, Standards and Needed Sources Reductions. *Atmospheric Environment*, v.15, pp.372-368.

Lodge, G.B. and West, P.N. (1971) Discussion. *Journal of APCA*, No.7, p.979.

Lynn, D.A. (1976) Air Pollution, *Treat and Response*, N.Y., pp.179-196.

Mage, M.D. (1980) An Explicit Solution for S_B Parameters Using Four Percentile Points. *Technometrics*, No.22, pp.247.

Mage, D.T. (1981) A Review of Applications of Probability Models for Describing Aerometric Data. *Environmetrics-81: Selected Papers*, SIAM, Philadelphia, pp.42-52.

Mage, D.T. and Ott, W.R. (1975) An Improved Model for Analysis of Air and Water Pollution Demands. *International Conference on Environmental Sensing and Assignment*, IEEE-ICESA, v.1., pp.20-5.

Mage, D.T. and Ott, W.R. (1978) Refinements of the Lognormal Probability Model for Analyzing Aerometric Data. *Journal of APCA*, No.29, pp.286-295.

Marchuk, G.I. (1982) ***Matematicheskoe modelirovanie v probleme okruzhayushchei sredy*** (Mathematical modeling in appliction to environmental problems). Moscow: Nauka (in Russian).

McKay, K.P. and Bornstein, R.D. (1981) Statistical Evaluation of Air Quality Simulation Models. *Environmetrics-81: Selected Papers*, SIAM, Philadelphia, pp.28-42.

Murzewski, J. and Sowa, A. (1978-1979) Graficzne metody estymacji parametrow i weryficacji typy rozkladu prawdopodobienstwa, ***Czasopismo technicze***, No.1, pp.32-37.

Ostromogil'skii, A.Kh. (1982) Mathematical Models Simulating Pollutant Circulation in Natural Environments. In: ***Problemy fonovo monitoringa sostoyaniya prirodnoi sredy*** (Problems of natural environment background monitoring) Vip.1, Leningrad: Gidrometeoizdat, pp.185-192 (in Russian).

Pastukhov, B.V., Popova, E.V. and Syroegina, O.A. (1982) Background Monitoring of Sulfur Compounds. In: ***Monitoring fonovovo zagryazneniya prirodnoi sredy***. (Background pollution monitoring of the natural environment) Vip.1, Leningrad: Gidrometeoizdat, pp.83-95 (in Russian).

Perone, S.P., Pichler, M, Gaarenstrom, P. and Mayers, G.H. (1975) The Application of Pattern Recognition Techniques to the Characterization of Atmospheric Aerosols. *International Conference on Environmental Sensing and Assignment*, IEEE-ICESA, v.1, p.5-4.

Roberts, E.M. (1979) Review of Statistics of Extreme Values with Application to Air Quality Data. *Journal of APCA*, No.29, p. 632.

Rovinskii, F.Ya. and Buyanova, L.D. (1982) Background Monitoring of the Natural Environment. In: *Problemy fonovovo monitoringa sostoyaniya prirodnoi sredy* Problems of natural environment background monitoring) Vip.1, Leningrad: Gidrometeoizdat, pp.5-11 (in Russian).

Rovinskii, F.Ya., Burtseva, L.V., Petrukhin, V.A., Cherkhanov, Yu.P. and Chicheva, T.B. (1982) Background Contents of Lead, Mercury, Arsenic and Cadmium in Natural Environments (World Data). In: ***Monitoring fonovovo sostoyaniya prirodnoi sredy*** (Background pollution monitoring of the natural environment) Vip.1, Leningrad: Gidrometeoizdat, pp.3-14 (in Russian).

Rovinskii, F.Ya., Egorov, V.I., Pastukhov, B.V., and Cherkhanov, Yu.P. (1982) Background Contents of Ozone, Dust, Nitrogenous and Sulfuric Compounds in the Atmosphere. In ***Monitoring fonovovo zagryazneniya prirodnoi sredy*** (Background pollution monitoring of the natural environment) Vip.1, Leningrad: Gidrometeoizdat, pp.23-24 (in Russian).

Rovinskii, F.Ya., and Cherkhanov, Yu.P. (1982) Recommendations for the Organization of Observations at Complex Background Monitoring Stations. In: ***Monitoring fonovovo zagryazneniya prirodnoi sredy*** (Background pollution monitoring of the natural environment). Vip.1, Moscow:Gidrometeoizdat, pp.19-26 (in Russian).

Rovinskii, F.Ya. and Wiersma, G.B. (1987) Procedures and Methods for Integrated Global Background Monitoring of Environmental Pollution, *WMO Tech. Doc. No.178, GEMS Info. Series No.5.*

Rubin, D.B. (1976) Inference and Missing Data. *Biometrics*, v.63, No.3, pp.581-592.

Schmidt, F.N. and Velds, C.A. (1969) On the Relation Between Changing Meteorological Sequences and the Decrease of Sulfur Dioxide Concentration Around Rotterdam. *Atmospheric Environment*, No.3, pp.455-460.

Selvin, S. (1976) A Graphical Estimation of the Population Mean from Censored Normal Data. *Applied Statistics*, v.25, No.1, pp.8-11.

Singpurwalla, N.D. (1972) Extreme Values from a Lognormal Law with Application to Air Pollution Problems. *Technometrics*, No.14, pp.703-711.

Smirnov, V.V. (1982) Variations in the Aerosol and Ionic Background Composition of the Lower Atmosphere. In: ***Monitoring fonovovo zagryazneniya prirodnoi sredy***, Vip.1, Leningrad: Gidrometeoizdat, pp.83-95 (in Russian).

Soeda, T. and Sawaragi, Y. (1979) Arima and GMDH Forecasts of Air Quality. ***Mathematical Models for Planning and Controlling Air Quality.*** IIASA Proceedings Series, v.17, pp.196-215.

Szepesi, D.J. (1982) Generalizing the Concept and Factors of Air Quality Management. ***Mathematical Models for Planning and Controlling Air Quality.*** IIASA Proceedings Series, v.17, pp.125-137.

Szepesi, D.J. and Fekete (1987) Background levels of air and precipitation quality for Europe Atmospheric Environment, **21**: 7:1623-1630.

Vipke, B. (1985) Data Collection, Transmission and Storage in the Framework of Background Monitoring: In *Tez. dokl. III Mezhdunarodnovo simpoziuma "Kompleksnyi global'nyi monitoring sostoyaniya biosfery"* (Abstr. Reports to III International Symposium "Complex global monitoring of the biosphere state) Moscow: Gidrometeoizdat, pp. 172-177 (in Russian).

Wiersma, J.B. (1985) Complex Global Monitoring Network. In: *Tez. dokl. III Mezhdunarodnovo simpoziumz "Kompleksnyi global'nyi monitoring sostouaniya biosfery"* (Abstr. Reports III International Symposium "Complex global monitoring of the biosphere state) Moscow: Gidrometeoizdat, pp. 41-42 (in Russian)

Zelenyuk, E.A. (1984) Statistical Analysis of Metrological Data in the System of Background Air-Pollution Monitoring. In: *Tez. dokl. pyatoi Vsesoyuznoi konferentsii "Problemy metrologicheskovo obespecjeniya sistem obrabotki izmeritel'noi informatsii - SO11 - 5"* (Abstr. Reports 5th All-Union Conference "Problems of metrological provision for aerometric data processing systems - S011 - 5") Moscow, p.38-41 (in Russian).

Zelenyuk, E.A. (1985) Statistical Data Analysis in Problems of Air-Quality Control. In: *Tez. dokl. tret'ei Vsesoyuznoi konferentsii'Primenenie mnogomernovo statisticheskovo analiza v ekonomike i otsenke kachestva produktsii"* (Abstr. Reports 3rd All-Union Conference "Application of multivariate statistical analysis in economics and estimates of the product quality). Tartu, p228-230 (in Russian). Leningrad: Gidrometeoizdat, p.49-55 (in Russian).

Zelenyuk, E.A., Zubenko, A.A., and Cherkhanov, Yu.P. (1984) Sampling and Data Analysis in the Background Monitoring System. In: *Tez. dokl. pyatoi Vsesoyuznoi kunferentsii "Problemy metrologicheskovo obespecheniya sistem obrabotki izmeritel'noi informatsii"* (Abstr. Reports 5th All-Union Conference "Problems of meteorological provision for aerometric data processing systems- S011-5"), Moscow, p.41-44 (in Russian).

Zelenyuk, E.A. and Cherkhanov, Yu.P. (1985) Investigations of Background Air-Pollution in Terms of the Statistical Model. In: *Monitoring fonovo zagryazneniya prirodnoi sredy* (Background pollution monitoring of the natural environment) Vip.2,

Zimmer, C.E., Tabor, E.C., and Stern, A.C. (1959) Particulate Pollutants in the Air of the United States. *Journal of APCA*, No.9, p.136.

ACKNOWLEDGEMENT

The authors wish to thank Professor R.E. Munn who assisted greatly in clarifying some of the technical details and in editing.

APPENDIX TO CHAPTER 3.

Distributions of concentrations of pollutants in long time series of observations.

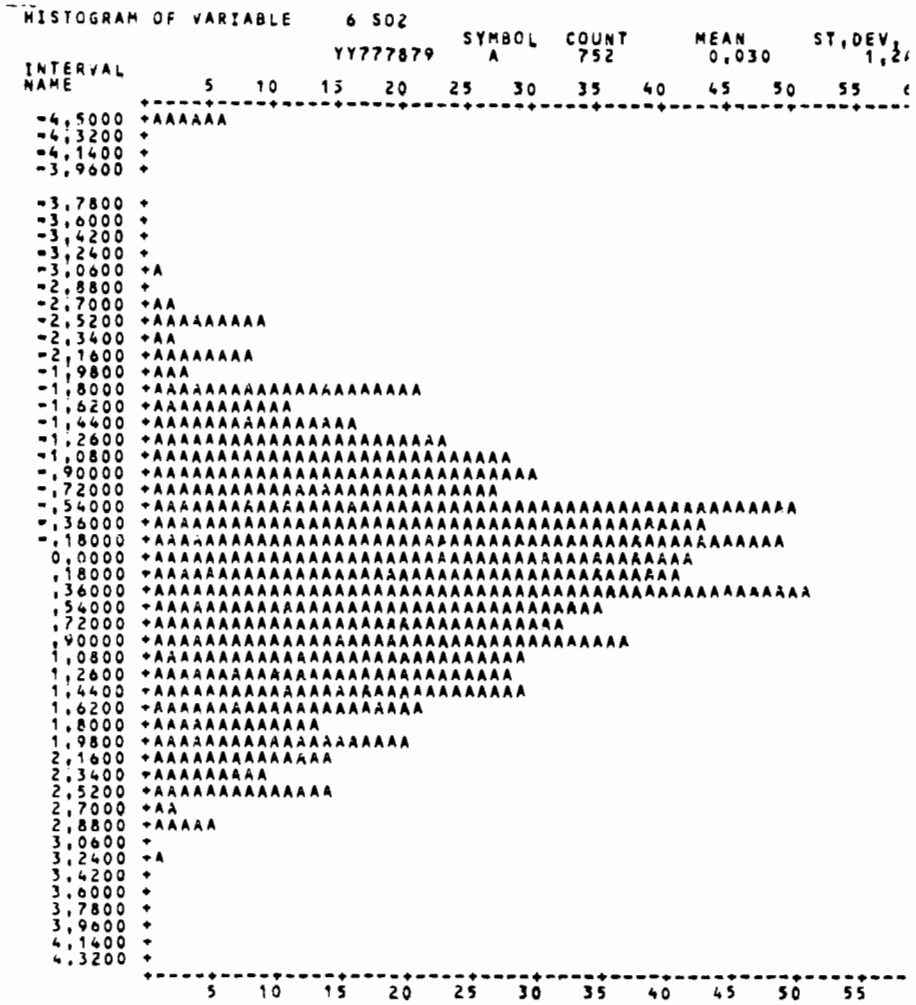


Figure A.3.1.:Histogram of logarithmic concentrations of sulfur dioxide. Borovoe station, 1977-1979.

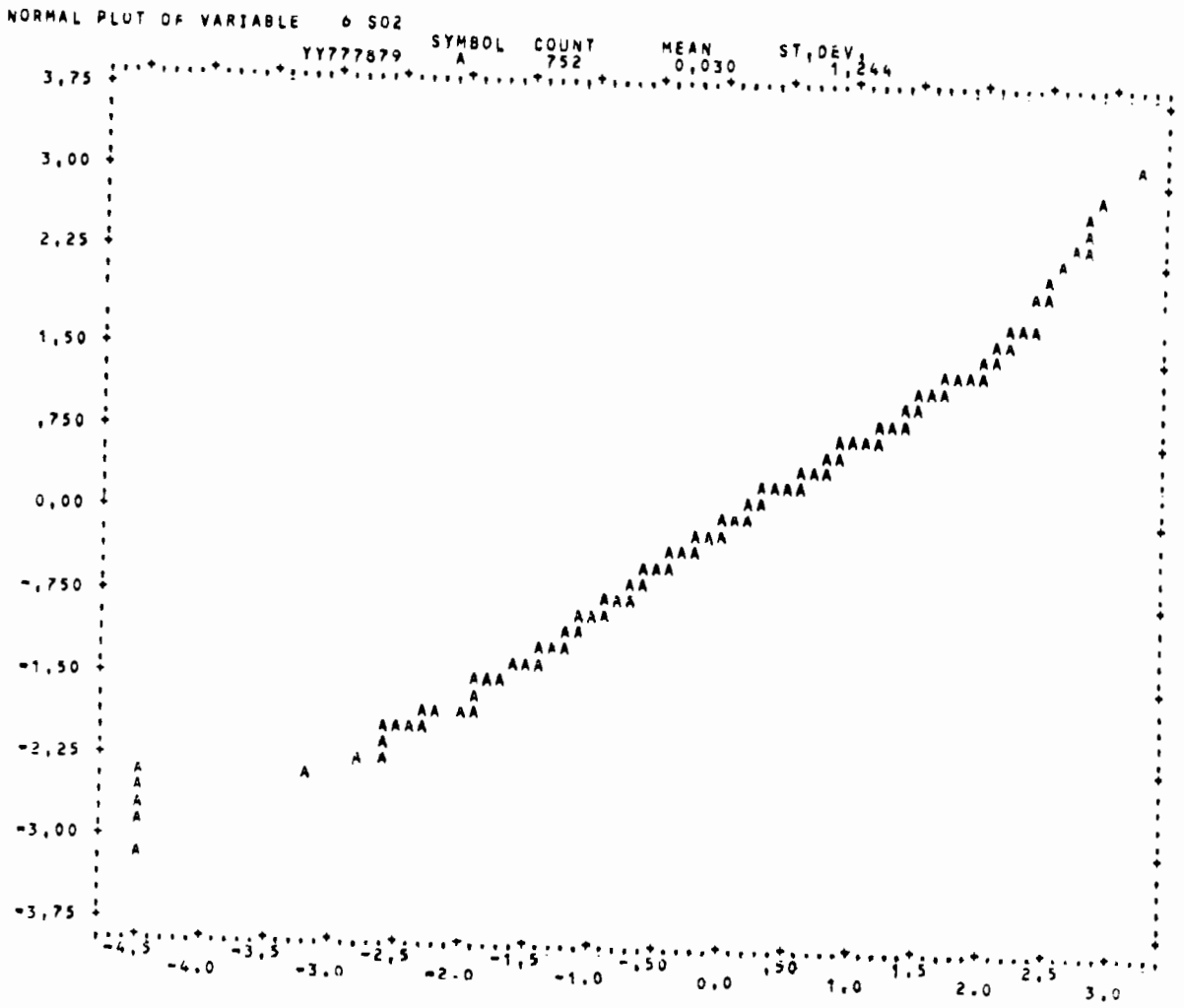


Figure A.3.2.:Normal plot of cumulative logarithmic concentrations of sulfur dioxide. Borovoe station, 1977-79.

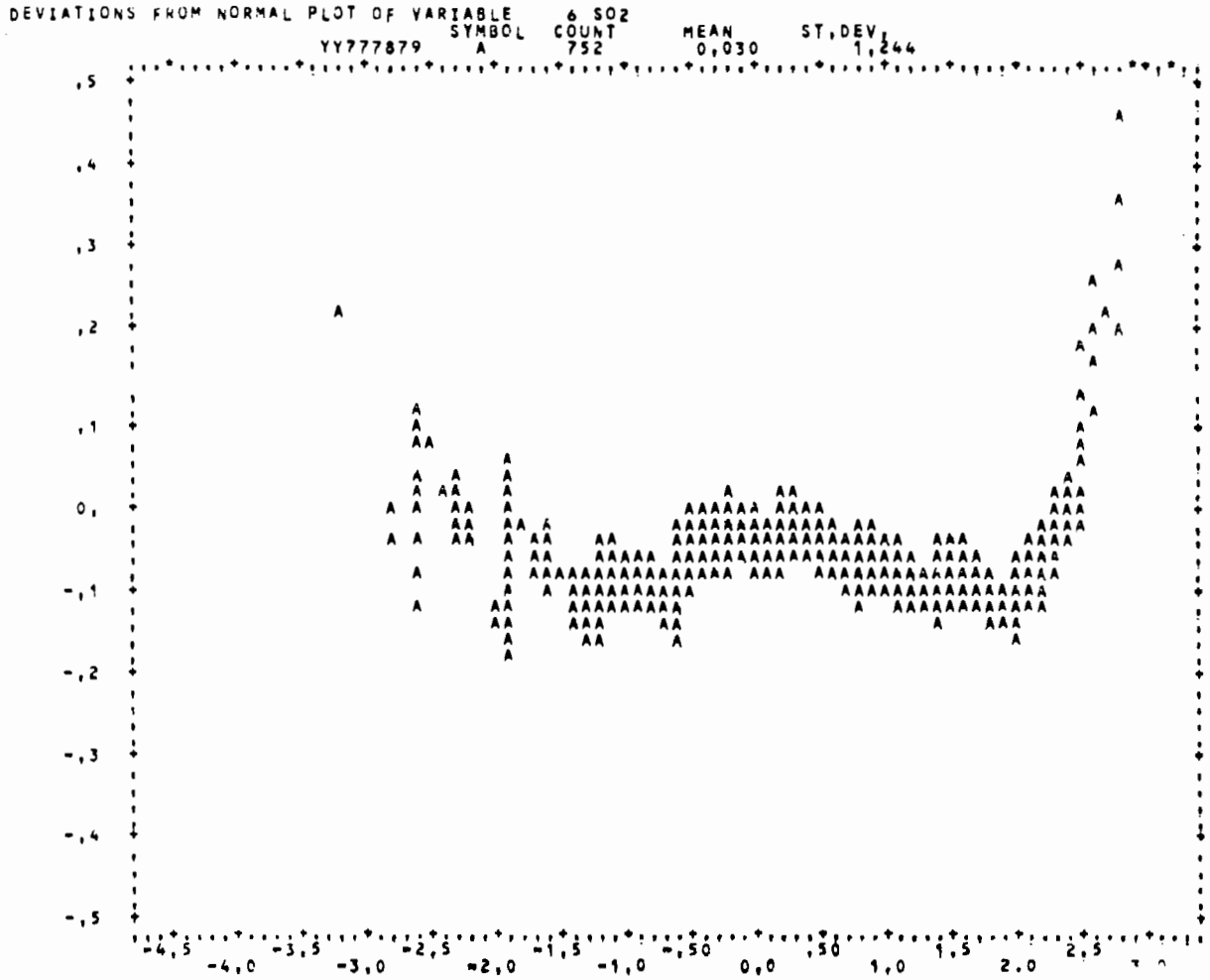


Figure A.3.3.: Deviations from normal plot of logarithmic concentrations of sulfur dioxide. Borovoe station, 1977-79.

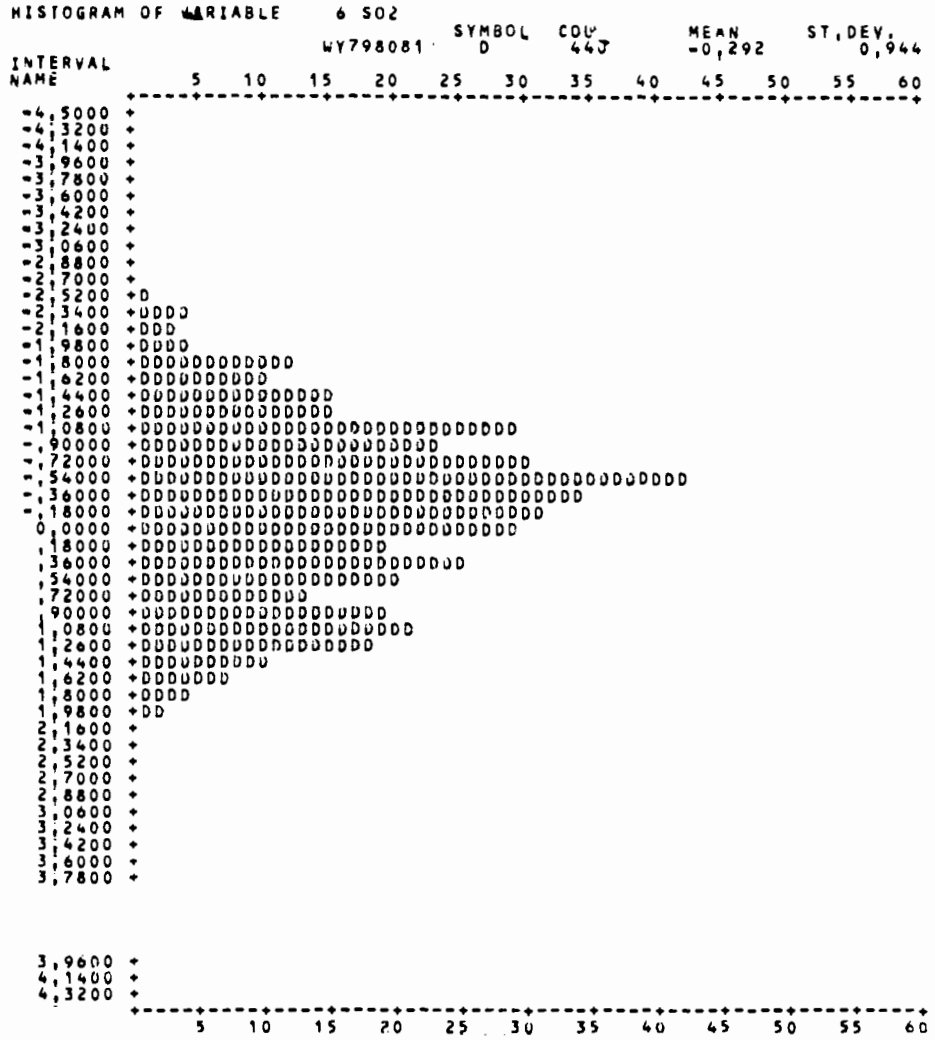


Figure A.3.4.:Histogram of logarithmic concentrations of sulfur dioxide, Borove station, warm seasons, 1979-1981.

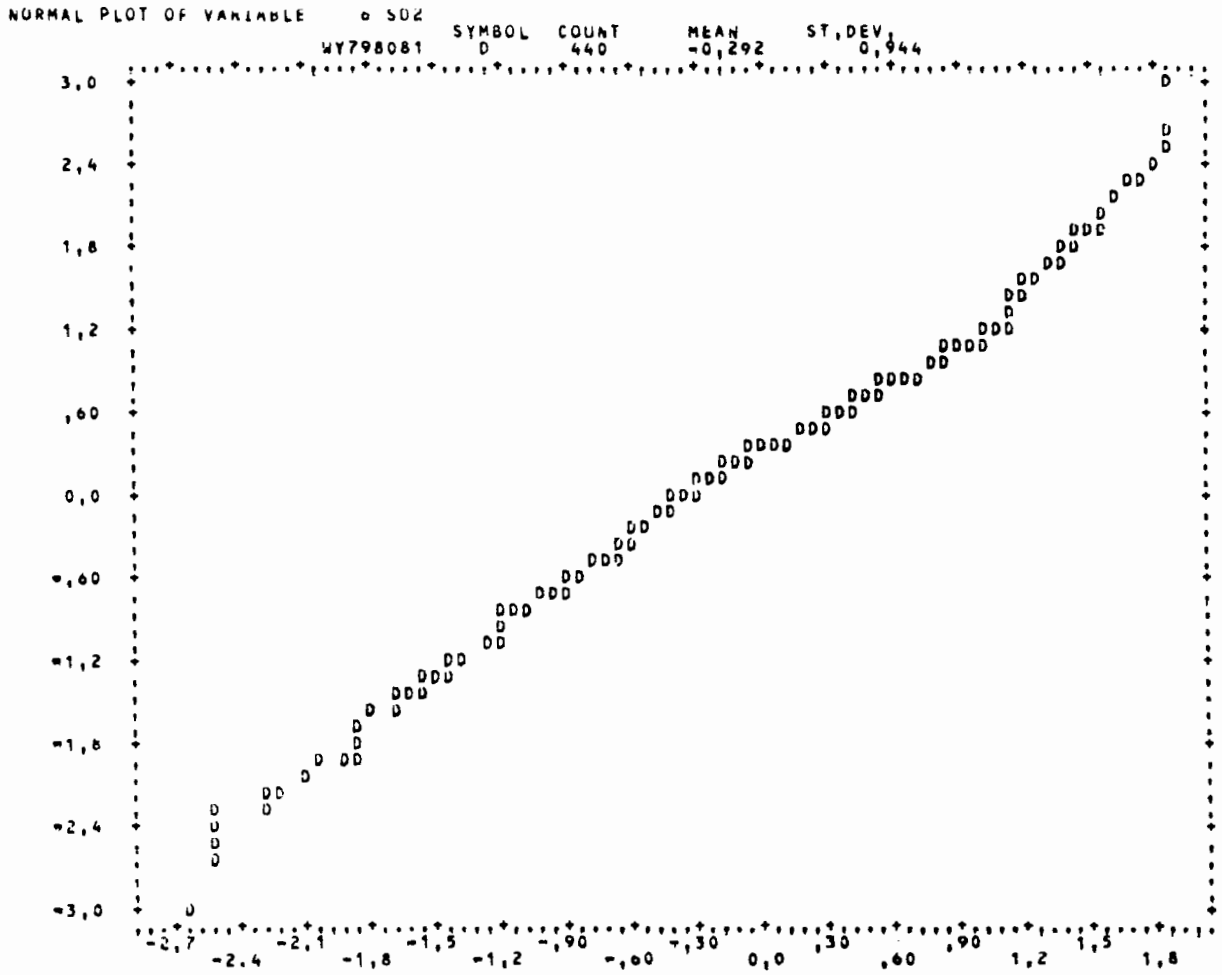


Figure A.3.5.: Normal plot of cumulative concentrations of sulfur dioxide. Borovoe station, warm seasons, 1979-1981.

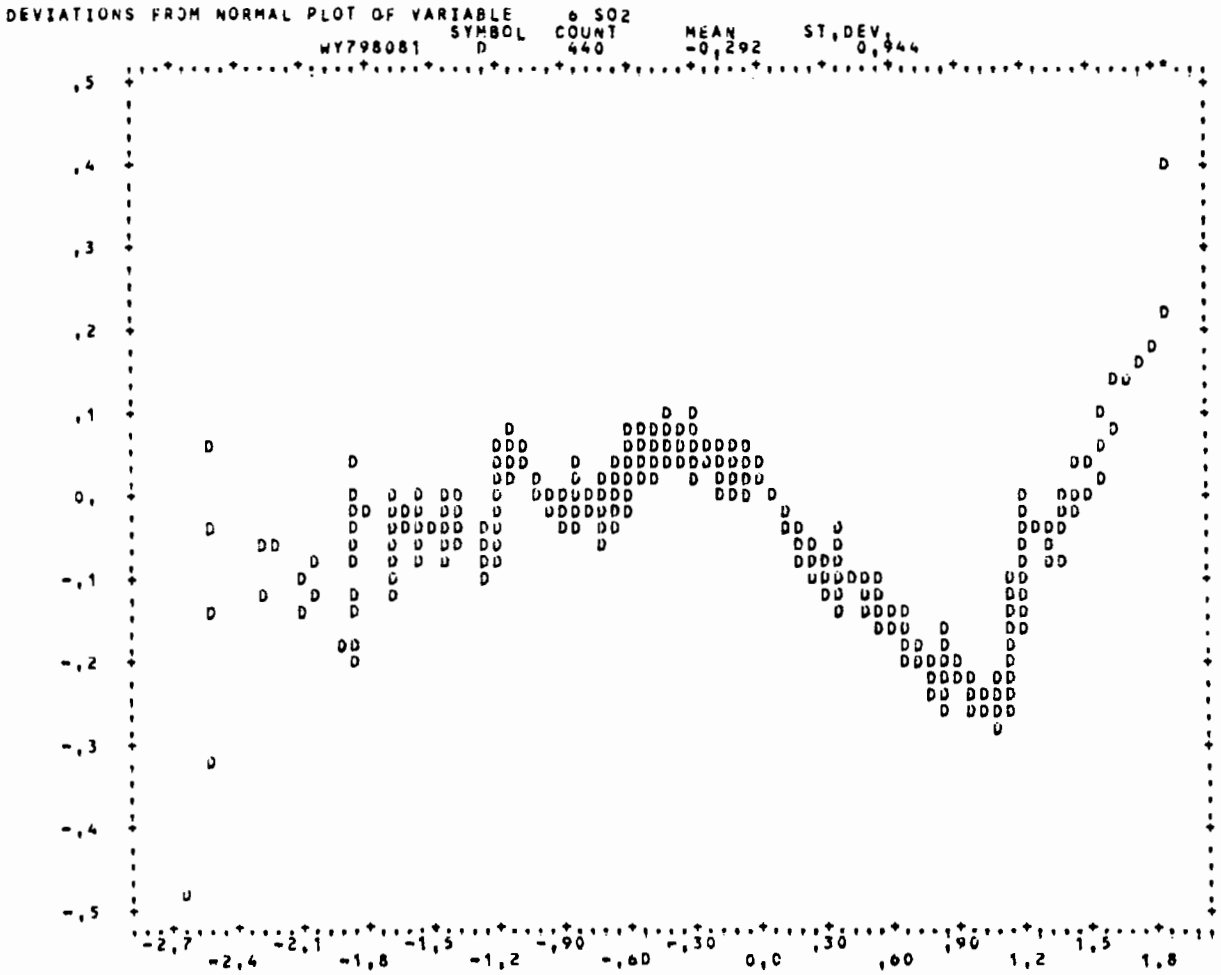


Figure A.3.6.: Deviations from normal plot of logarithmic concentrations of sulfur dioxide. Borovoe station, warm seasons, 1979-81.

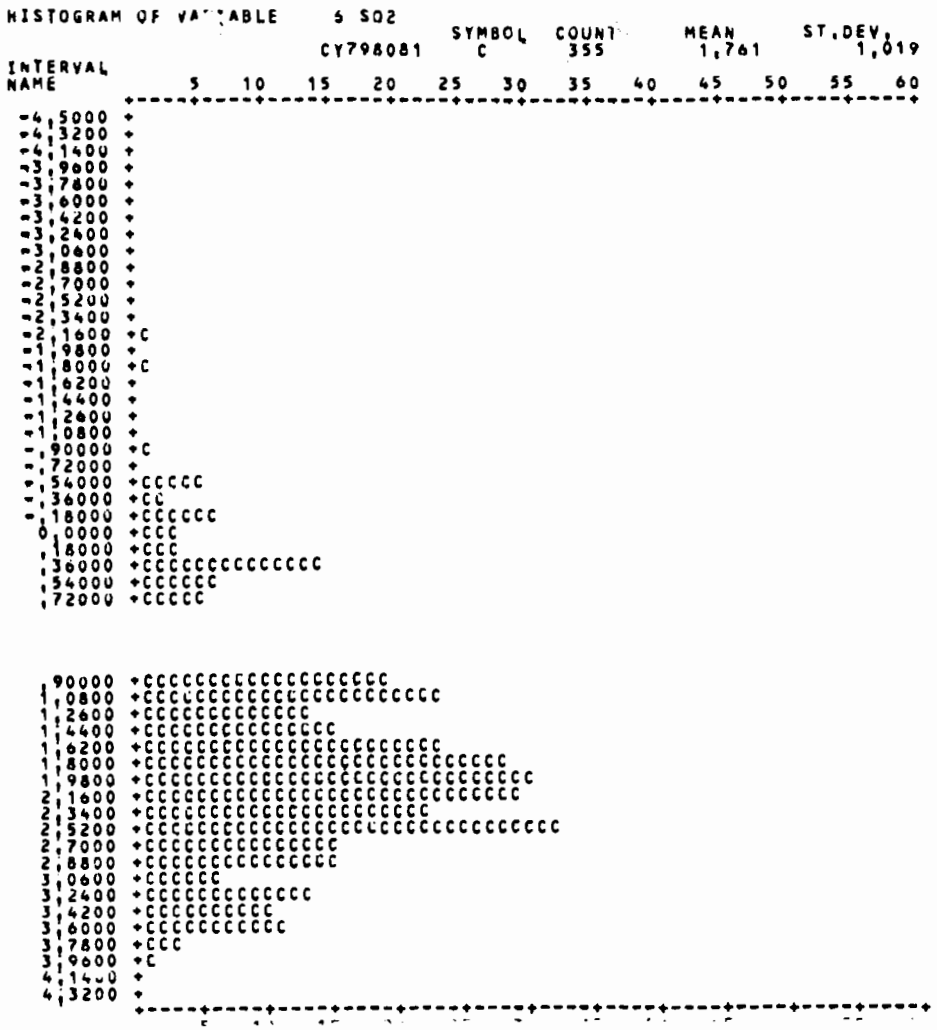


Figure A.3.7.:Histogram of logarithmic concentrations of sulfur dioxide. Borovoe station, cold seasons, 1979-81.

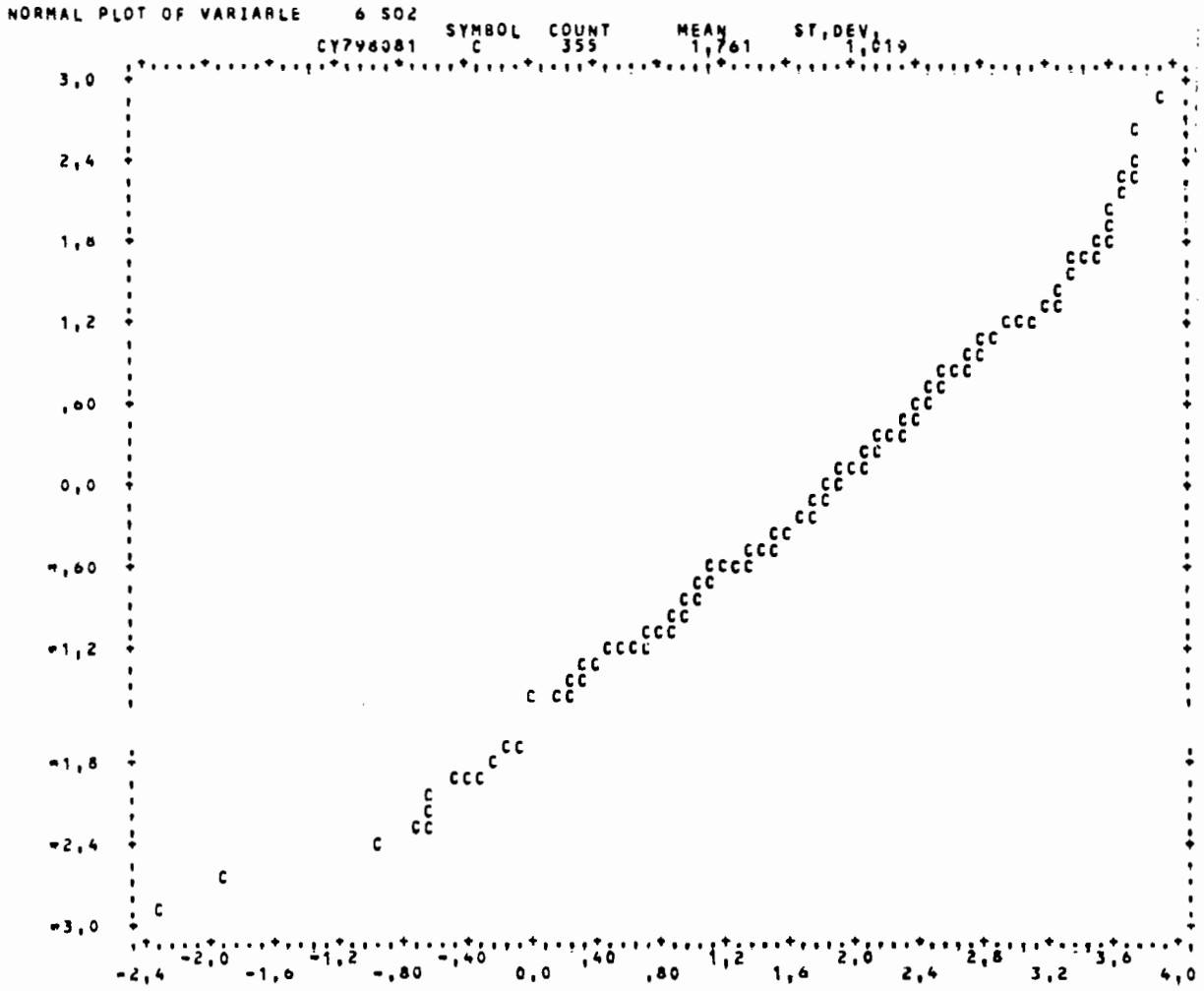


Figure A.3.8.:Normal plot of cumulative logarithmic concentrations of sulfur dioxide. Borovoe station, cold seasons, 1979-81.

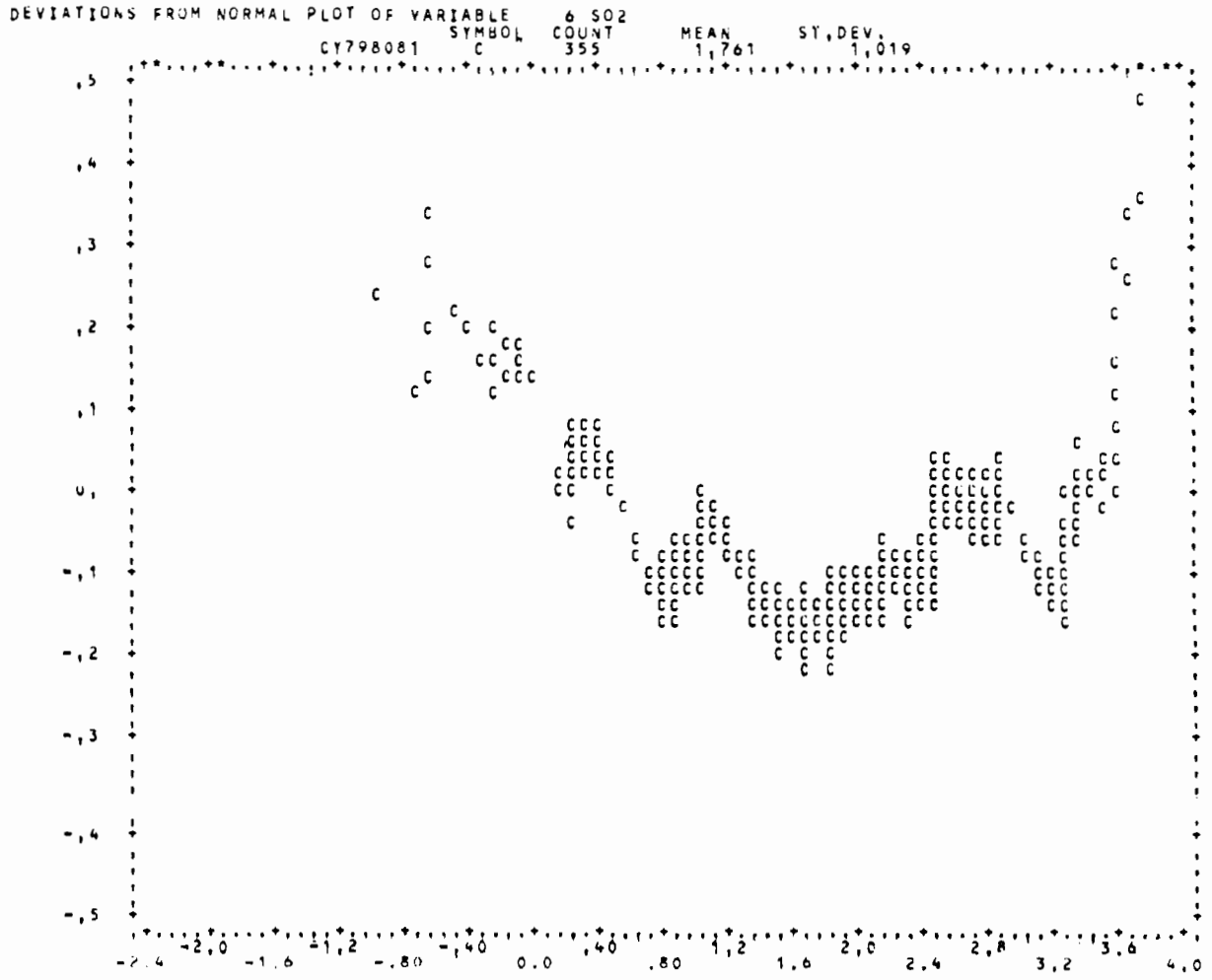


Figure A.3.9.: Deviations from normal plot of logarithmic concentrations of sulfur dioxide. Borovoe station, cold seasons, 1979-81.

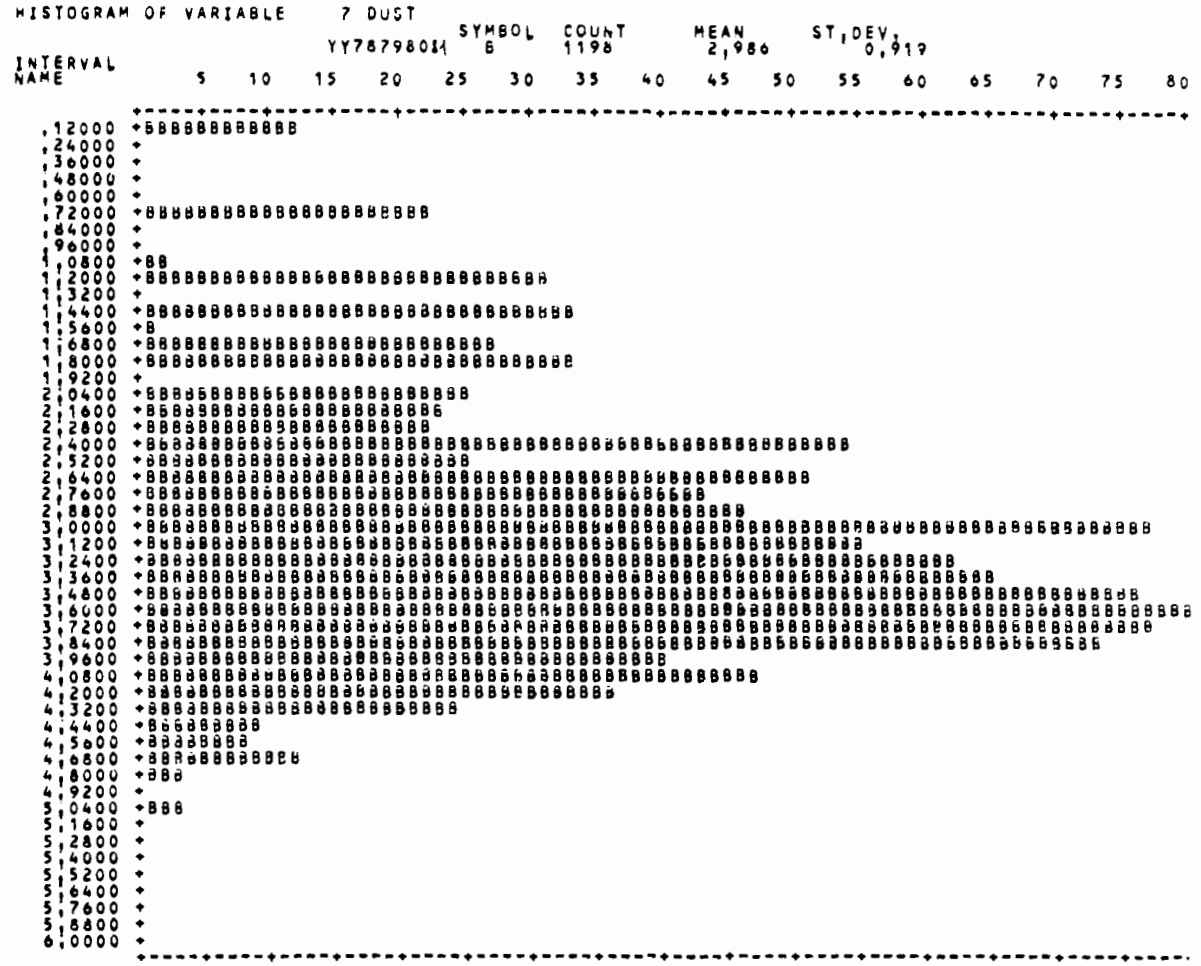


Figure A.3.10.:Histogram of logarithmic concentrations of suspended particulate matter. Borovoe station, 1978-81.

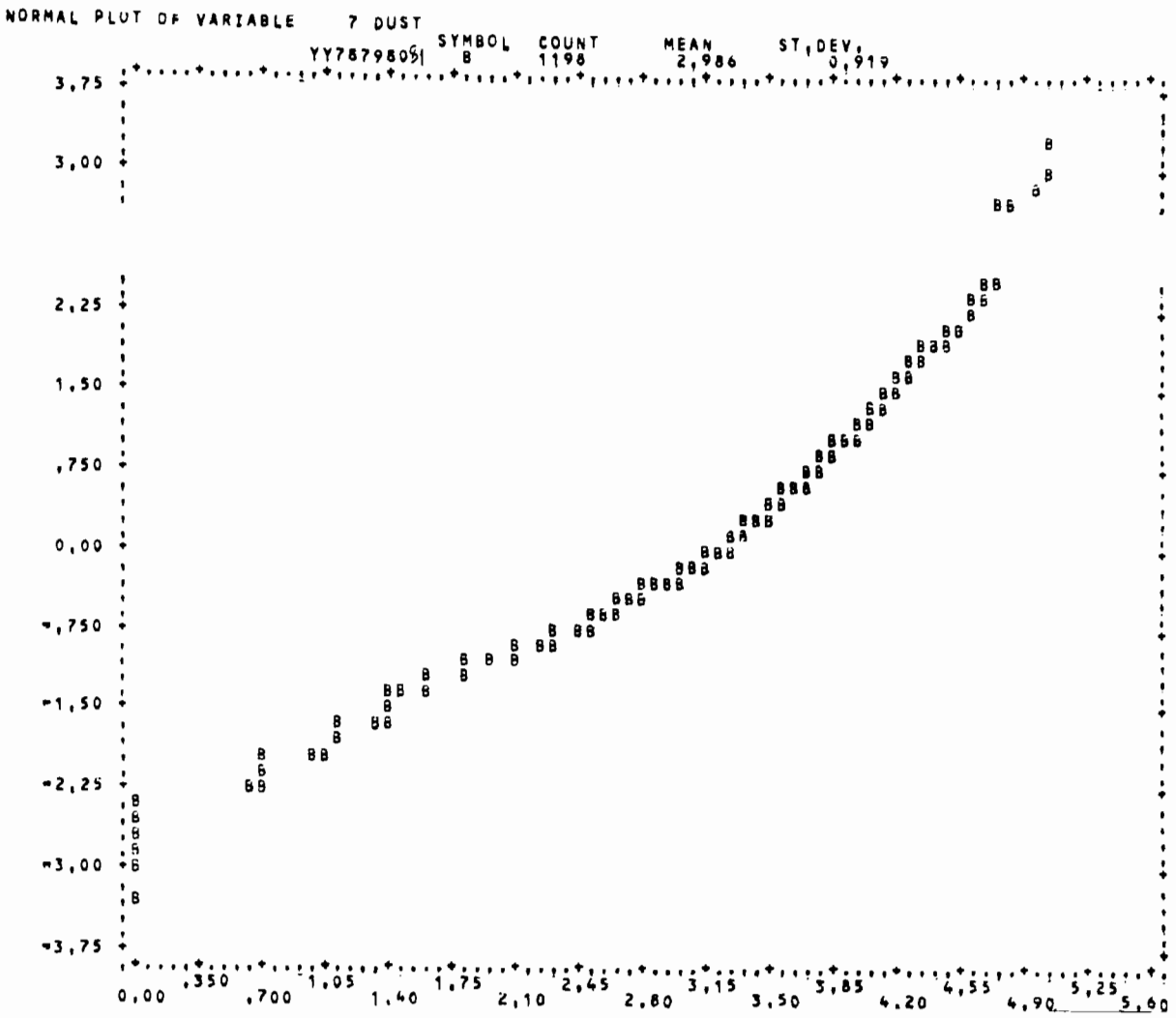


Figure A.3.11.:Normal plot of cumulative concentrations of suspended particulate matter. Borovoe station, 1978-81.

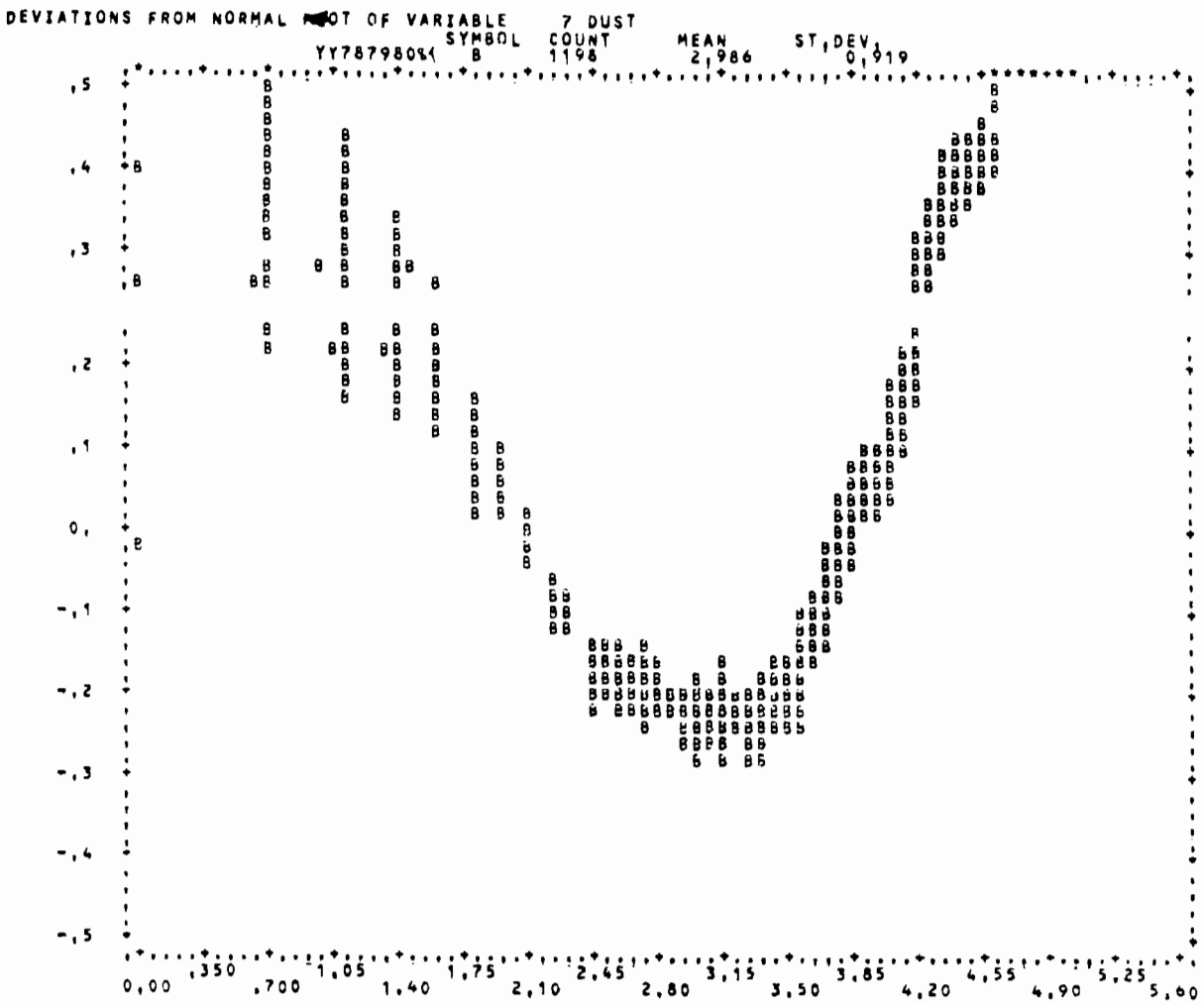


Figure A.3.12.:Deviations from normal plot of logarithmic concentrations of suspended particulate matter. Borovoe station, 1978-81.

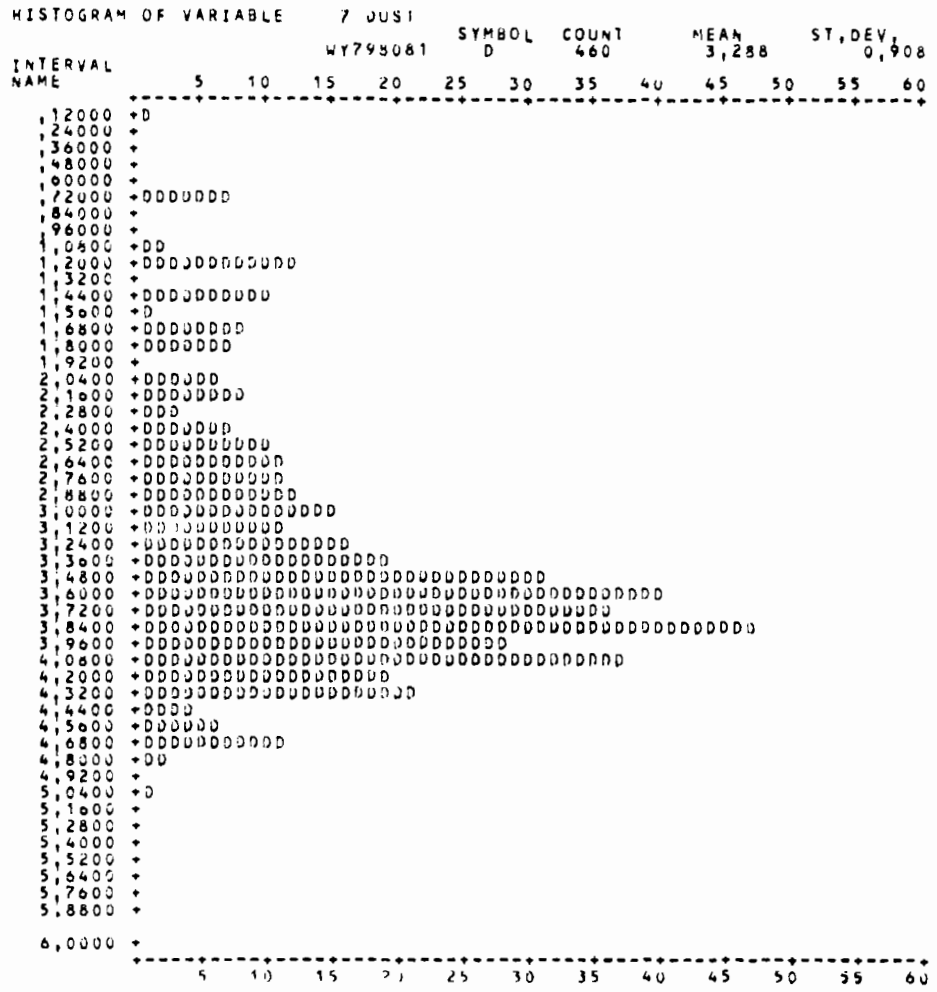


Figure A.3.13.:Histogram of logarithmic concentrations of suspended particulate matter. Borovoe station, warm seasons, 1979–81.

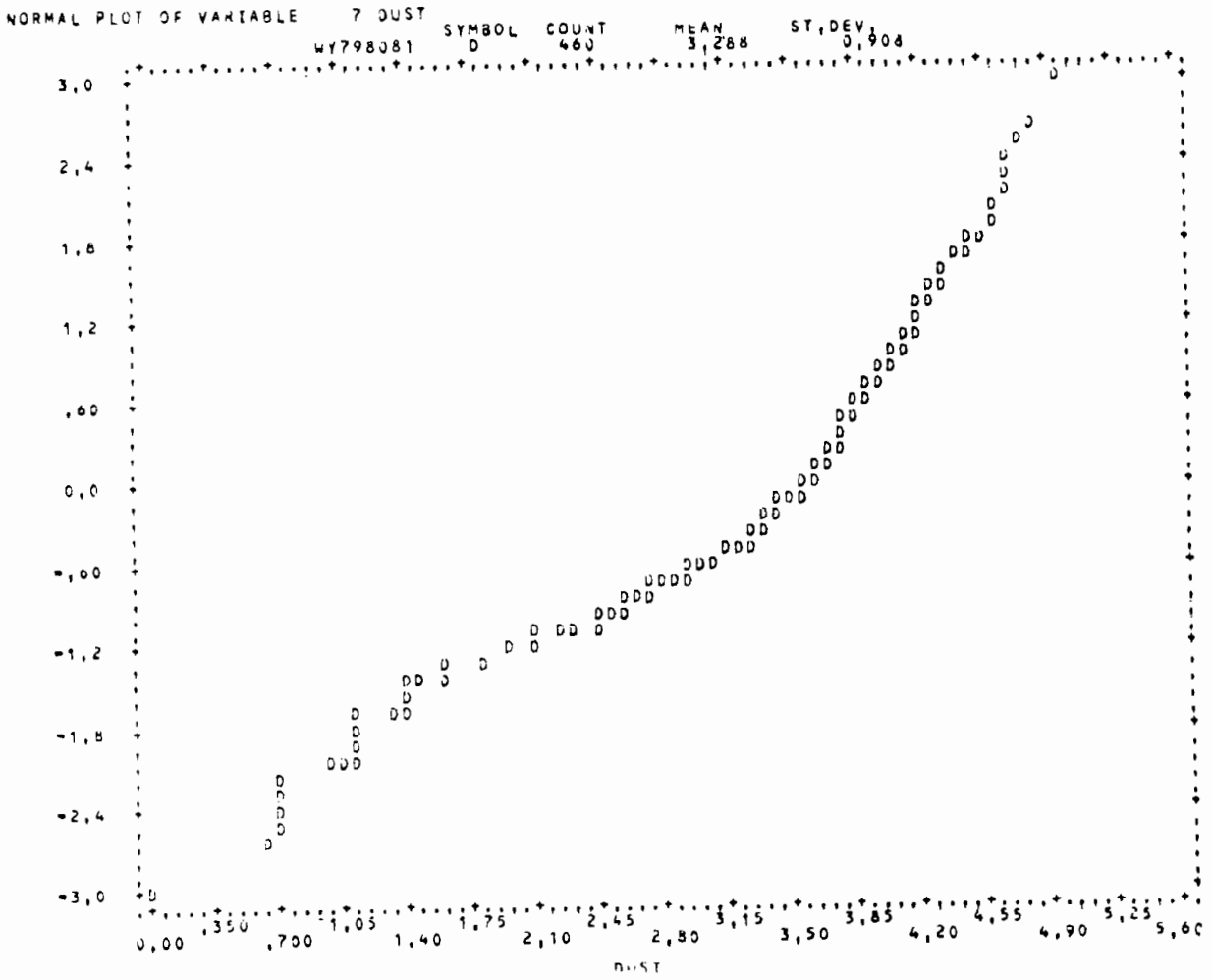


Figure A.3.14.:Normal plot of cumulative concentrations of suspended particulate matter. Borovoe station, warm seasons, 1979-81.

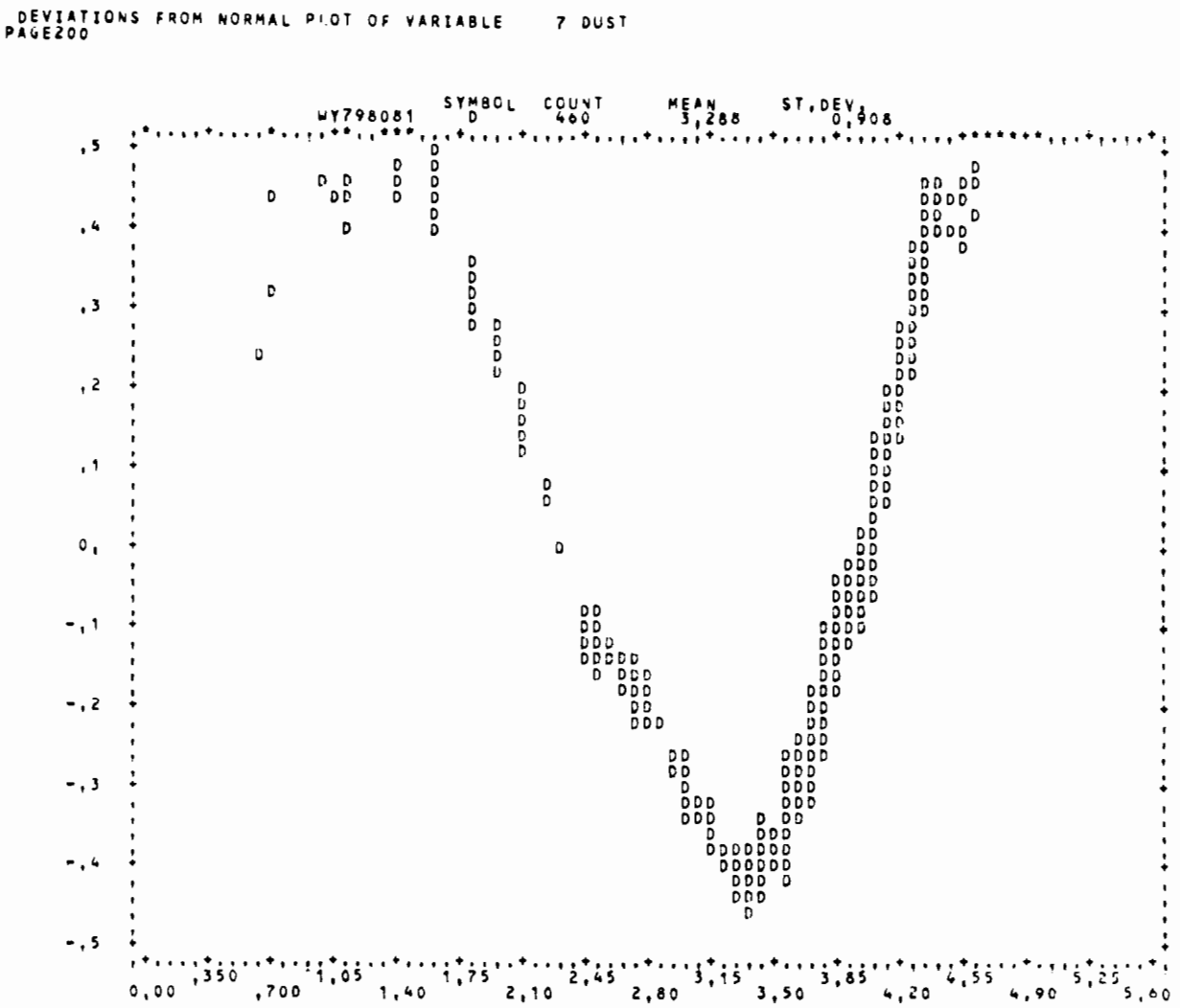


Figure A.3.15.:Deviations from normal plot of logarithmic concentrations of suspended particulate matter. Borovoe station, 1979-81.

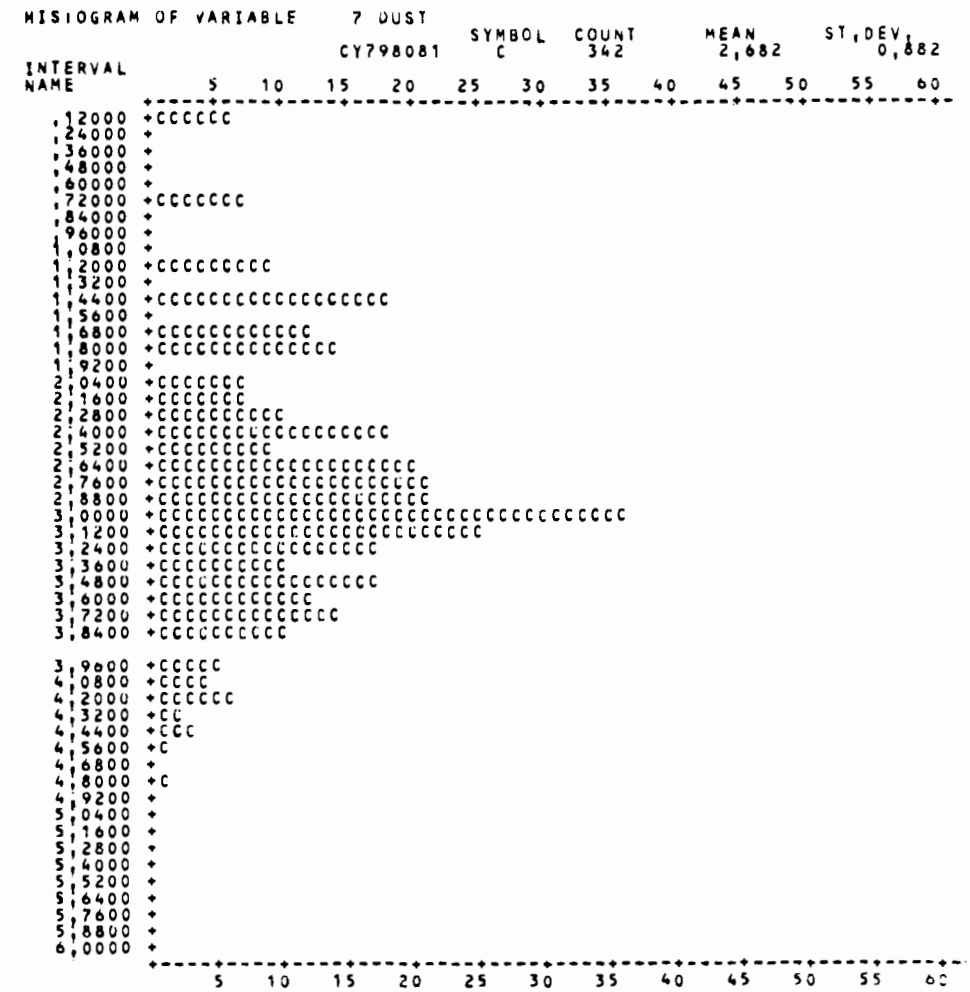


Figure A.3.16.:Histogram of logarithmic concentrations of suspended particulate matter. Borovoe station, cold seasons, 1979-81.

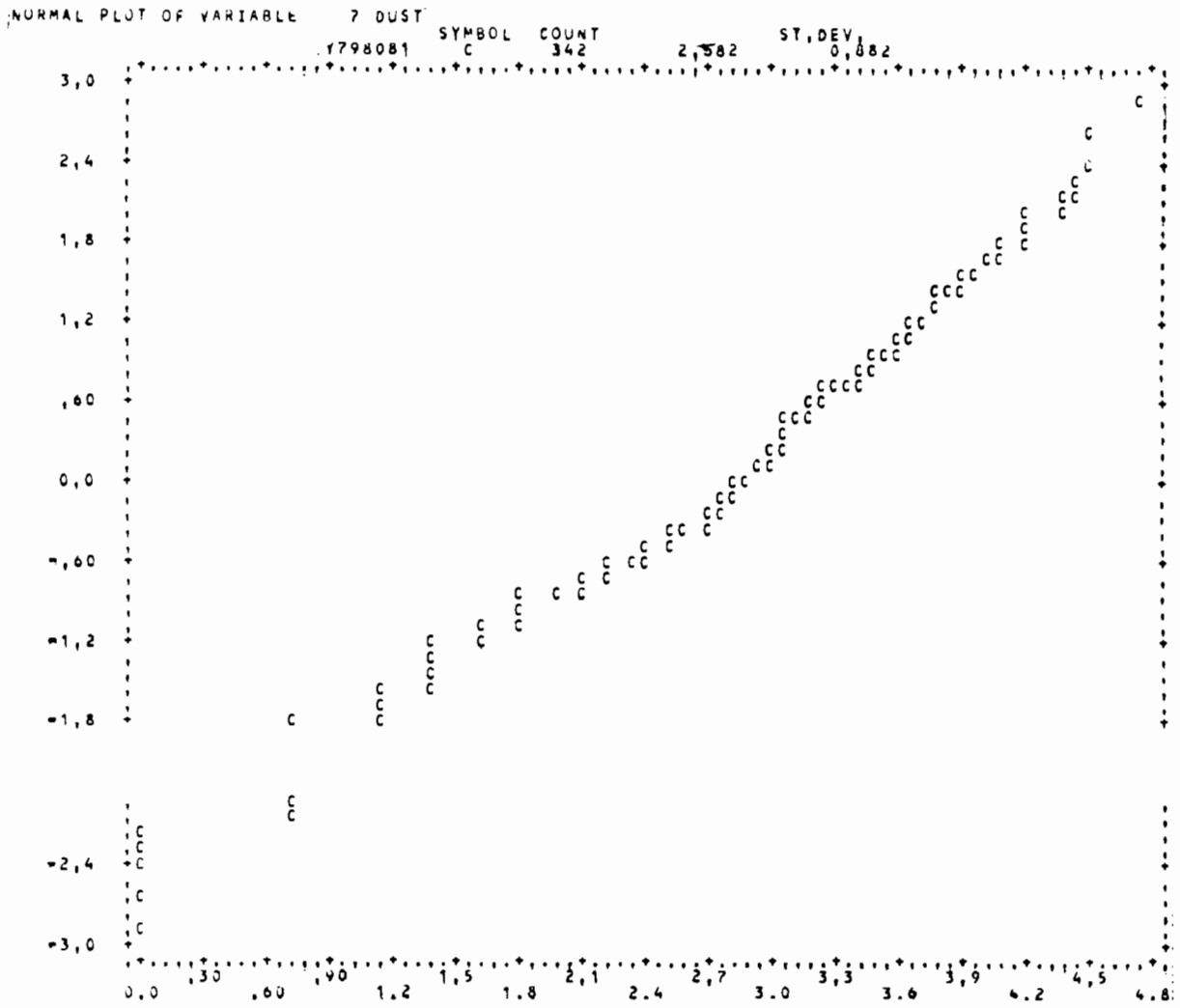


Figure A.3.17.:Normal plot of logarithmic concentrations of suspended particulate matter. Borovoe station, cold seasons, 1979-81.

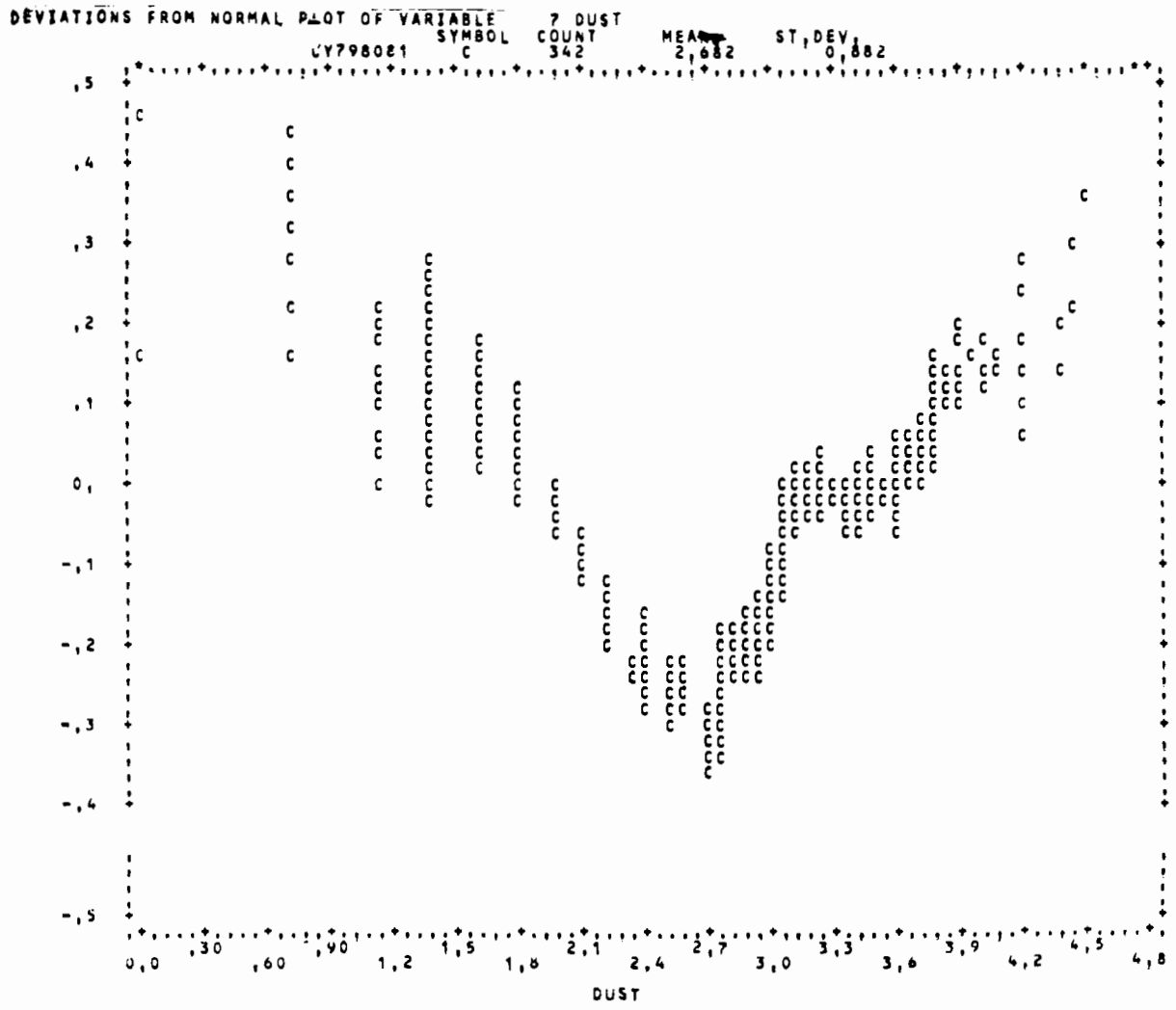


Figure A.3.18.:Deviations from normal plot of logarithmic concentrations of suspended particulate matter. Borovoe station, cold seasons, 1979-81.

APPENDIX TO CHAPTER 4.

Modeling to distributions. Methods of estimations of parameters of mixed distributions.

```
PROGRAM CONTROL INFORMATION
        /PROBLEM COMMENT  IS 'МОДЕЛИРОВАНИЕ СМЕСЕЙ'.
              TITLE IS 'КОМПОЗИЦИЯ НОРМАЛЬНЫХ ЗАКОНОВ'.
        /INPUT   COMMENT IS ' ВВОД ДАННЫХ '.
              VVARIABLES ARE 9.
              UUNIT IS 9.
              FFORMAT IS'(3F2.0,F5.1,2F5.2,2F5.1,F5.3)'.
               UNIT IS 8.
               CODE IS    SFRNG.
               CONTENT IS DATA.
               LABLE IS '1301 РЕАЛИЗ ИЗ N(0,1)'.
        /VARIABLE COMMENT IS '3 - ДАТА,6 - ПРИНЕСИ, 1 - КОД, 1 - RNDG,2 - МОДЕЛ
               NAMES ARE YEAR,MON,DAY,PB,KD,SO2,DUST,MG,BP,COD,RNG,RNCL,RNMOD.
               ADD ARE 2.AADD IS 1.
              UUSED ARE 1,2,3,4,5,6,7,8,9,10,11,12,13.
        MMAXS ARE 83,12,31,999,999,999,999,999,999,1111,5,999,999.
        MMINS ARE 76,1,1,-6.9,-6.9,-6.9,-6.9,-6.9,-6.9, 0.
               IINTERVALS ARE 10,10.
               AFTERT.
              GGROUPING IS COD./
        CCATEGORY COMMENT IS 'ГРУППИРОВКА'.
           CCUTPOINTS (2) ARE 4.,9.
               CODES (10) ARE 800,810,820,801,802,811,812,821,820.
            NAMES (10) ARE Y80,Y81,Y82,LL80,ZZ80,LL81,ZZ81,LL82,ZZ82.
        /TTRANSFORM
                    RNG = 0.71994.
                    L1  = KASE EQ 1.
                    R0  = RNG IF L1.
                    R0  = R0 * 100000. R0  = R0  + 1.  R0  = INT ( R0  ).
                    RNG = RNDG  ( R0  ) .
                    R0  = RNG.
                    COD = YEAR.
         /TRANSFORM
                   L = KASE EQ 1.
           M                =              0,0.
           S                =              0,0.
              OMIT = 1.
              DELETE = 1.
              R  = 1308 / 55  , R = INT (R) .
              L1 = 1 IF L. R11 = R IF L1. R11 = R11 - 1. L1 = R11 EQ 0.
              R  = 1308 / 71  , R = INT (R).
              L2 = 1 IF L. R22 = R IF L2.  R22 = R22 - 1. L2 = R22 EQ 0.
              R  = 1308 / 0.5, R = INT (R).
              L  = 1 IF L. R   = R  IF L . R   = R   - 1. L  = R   EQ 0.
              M  =  50 IF L . S =  1 IF L .
              M  =  27 IF L2. S =  5 IF L2.
              M  =  23 IF L1. S =  1 IF L1.
        RNCL                   =              0,0.
              S  = S.
           RNCL = RNG * S.
           RNCL = RNCL + M.
        USE    = L1 OR L2.
        /TTRANSFORM
              M = 10.
              S = 3 . S = S + 3.
              P11 = 0 IF L.P12 = 0 IF L.P21 = 0 IF L.P22 = 0 IF L.
              B = M-S.  R = RNCL LT B.
```

Figure A.4.1: Program of modeling composite of two distinct distributions on language of management of programs PPPBMDP.

```
      L = L1 AND R.
      RNMOD = 1 IF L.
      L = L2 AND R.
      RNMOD = 0 IF L.
      L = R EQ 0.
      B = M.     R = RNCL LT S. R = R AND L.
      L = L1 AND R.
      B = 0.8 * L. P11 = P11 + B. L = P11 GT 1. P11 = P11 - L.
      RNMOD = 1 IF L.
      L = L2 AND R.
      B = 0.2 * L. P22 = P22 + B. L = P22 GT 1. P22 = P22 - L.
      RNMOD = 1 IF L.
      L = R EQ 0.
      B = M + S. R = RNCL LT B. R = R AND L.
      L = L1 AND R.
      B = 0.2 * L. P12 = P12 + B. L = P12 GT 1. P12 = P12 - L.
      RNMOD = 1 IF L.
      L = L2 AND R.
      B = 0.8 * L. P21 = P21 + B. L = P21 GT 1. P21 = P21 - L.
      RNMOD = 1 IF L.
      L = R EQ 0.
      B = M + S. R = RNCL GT B. R = R AND L.
      L = L1 AND R.
      RNMOD = 0 IF L.
      L = L2 AND R.
      RNMOD = 1 IF L.
   L = RNMOD.
      RNMOD = 1000 IF R.
      RNMOD = RNCL IF L.
   RNG = RNG.
   RNCL = RNCL.
   RNMOD = RNMOD./
SSAVE COMMENT IS 'СОХРАНЕНИЕ'.
         CODE IS   SFRNG. UNIT IS 8.
         LABEL IS '1301 РЕАЛИЗ ИЗ N(0,1)'.
         CONTENT IS DATA.
         NEW./
 PLOT COMMENT IS 'ГИСТОГРАММЫ      '.
       VARIABLES ARE    12    .
       TYPES ARE HIST,NORM,DNORM.
      GGROUP IS 1 GGROUP IS 2.GGROUP IS 3.GGROUP IS 4. GGROUP IS
      GGROUP IS 6.GGROUP IS 7.GGROUP IS 8.GGROUP IS 9.
      GGGROUP IS  4,6     .
       SSTATIST./
PPRINT COMMENT IS 'ПЕЧАТЬ ДЛЯ РАЗН. ПР.'. MAXIMUMS.MINIMUMS./
       DDATA.MISSING,CORR.MEAN,REGULAR.
       WINSOR.VERTICAL.NO STEP.CLASS IS 1,2,3,4,5.
HHIST COMMENT IS 'ГИСТОГР. ГР. '.
         GROUPING IS DATA.
         VARIABLES ARE 2,3,4,5.
           COMMENT IS 'КЛ. АНАЛИЗ   '.
         SUMOFP=3.STAND.
 /END
```

Figure A.4.2: Continuation of Figure A.4.1.

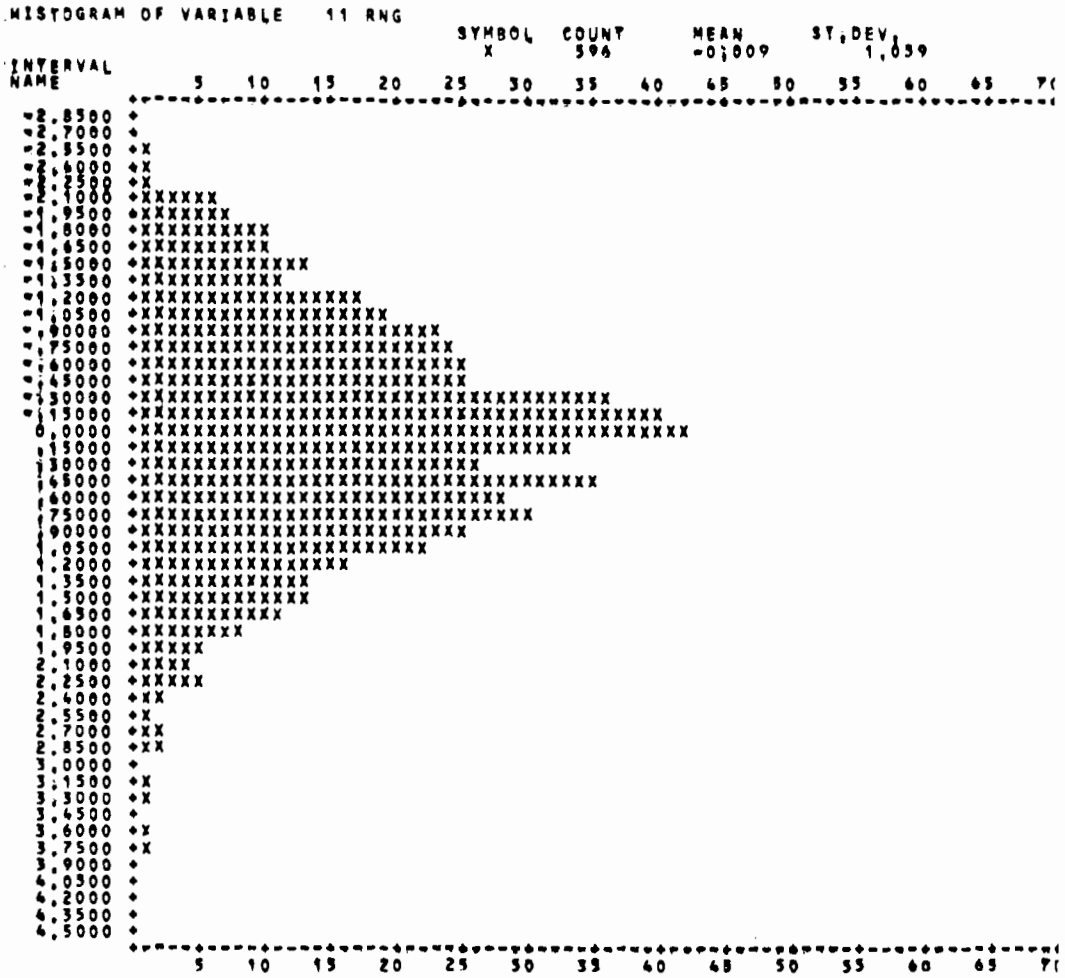


Figure A.4.3: Histogram of normally distributed random variation, RNG (model), from which every distribution is modeled further.

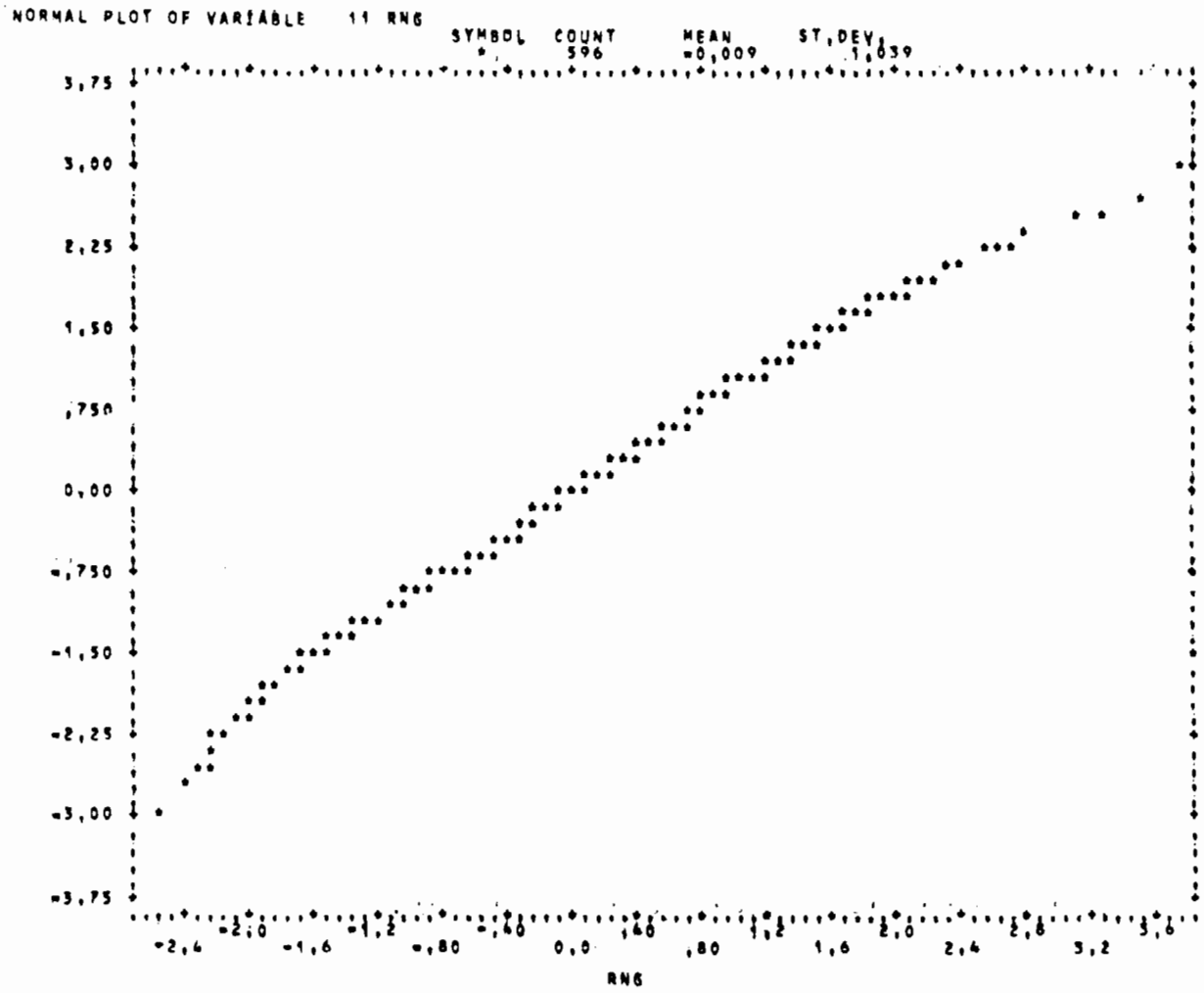


Figure A.4.4: Normal plot of random variable RNG.

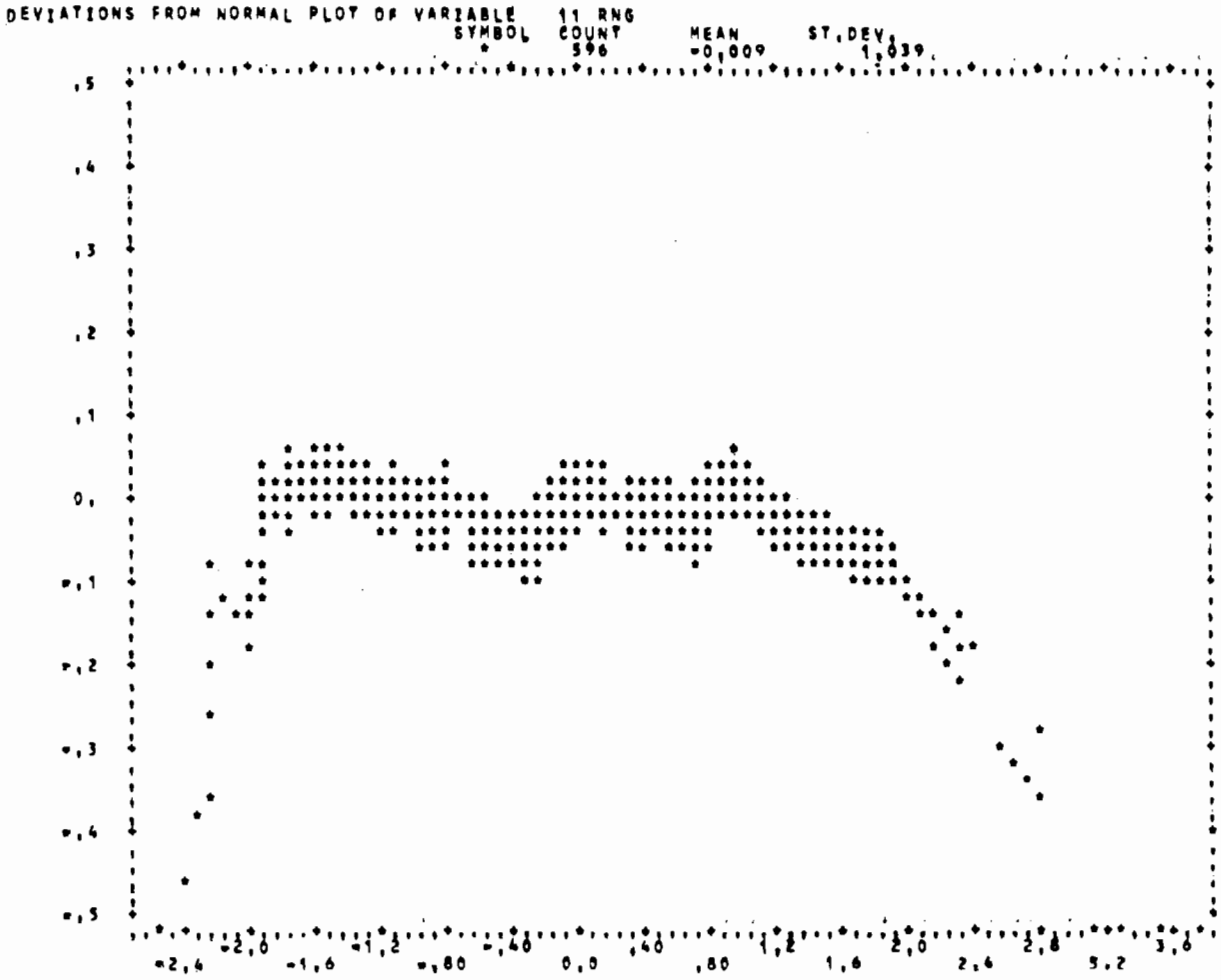


Figure A.4.5: Deviations from normal plot of random variable RNG.

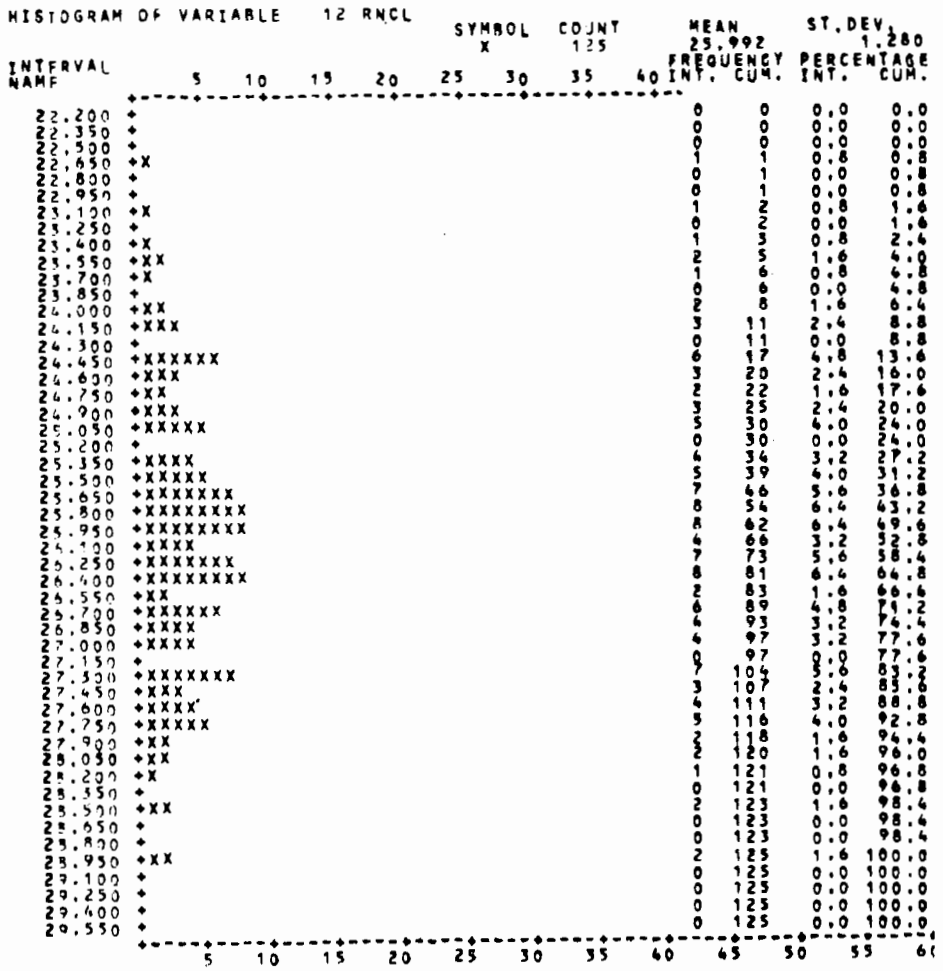


Figure A.4.6: Histogram of mixed distributions 70N/27,1/ and 55N/25,/1 (modeling).

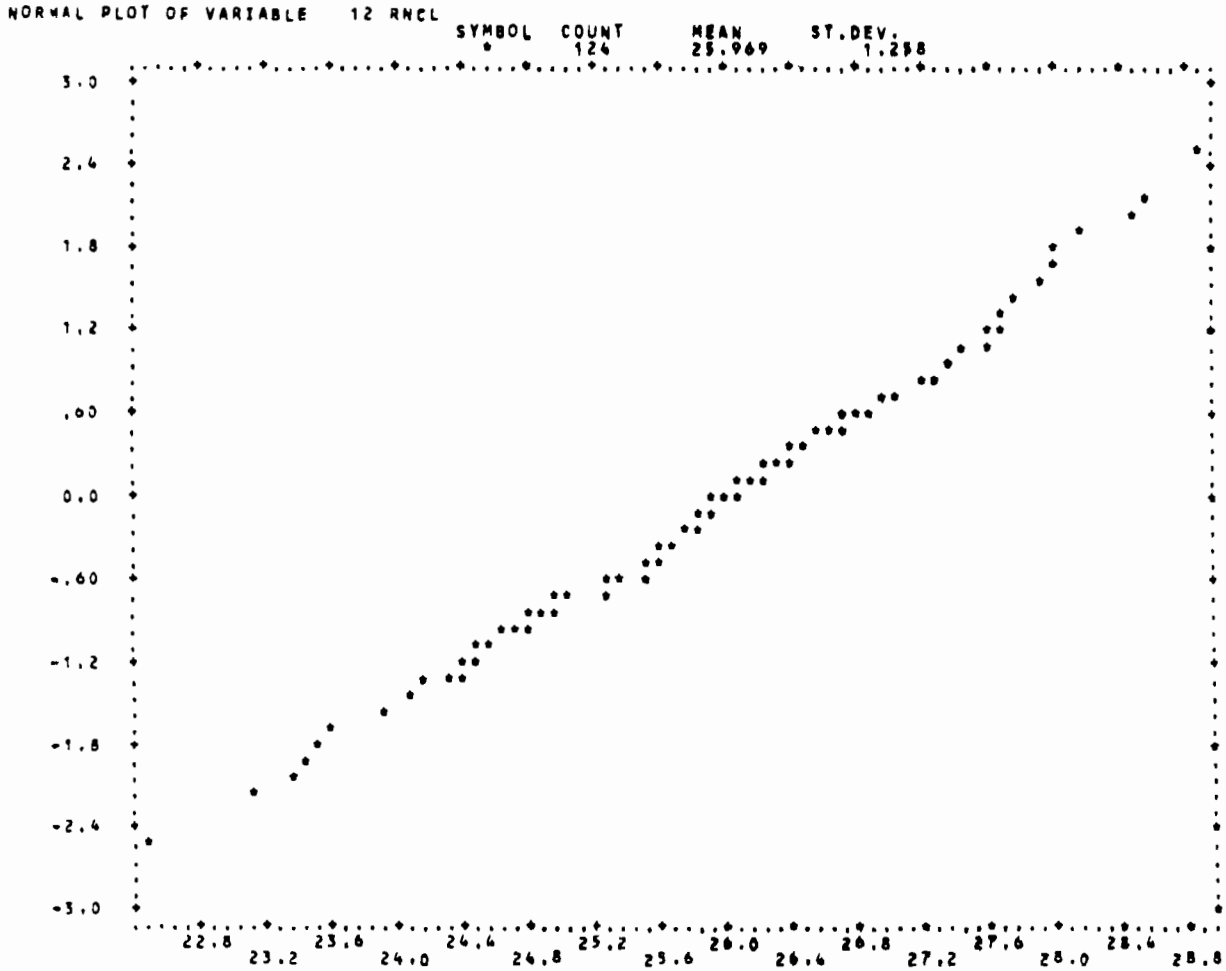


Figure A.4.7: Normal plot of mixed distributions 70N/27,1/ and 55N/25,1/

Figure A.4.8: Deviations from normal plot of distributions 70N/27,1/ and 55N/25,1/ (modeling).

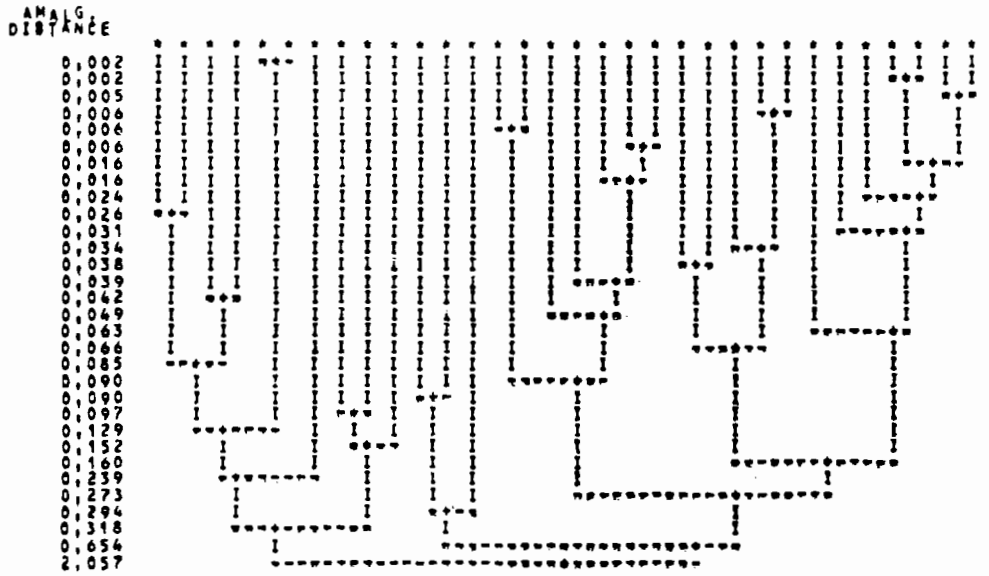


Figure A.4.9: Hierarchical cluster analysis on data of simulation.

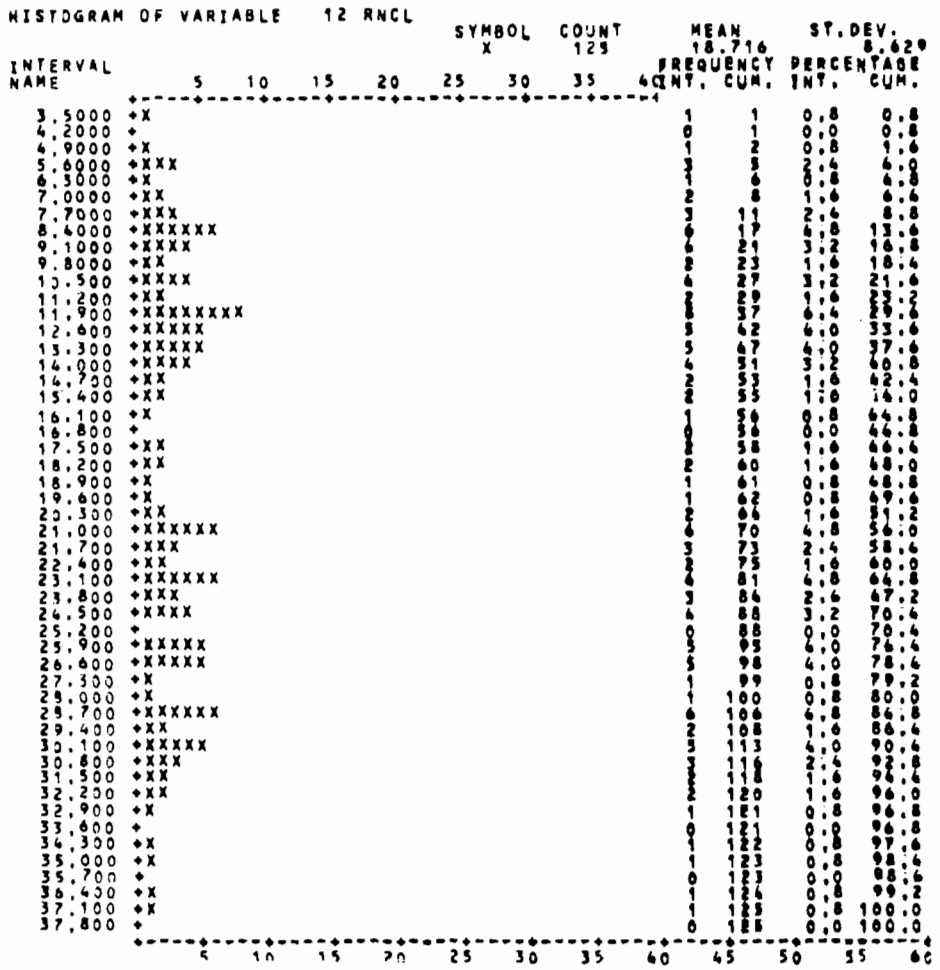


Figure A.4.10:Histogram of mixed distributions 55N/10,3/ and 70N/27,5/ (modeling).

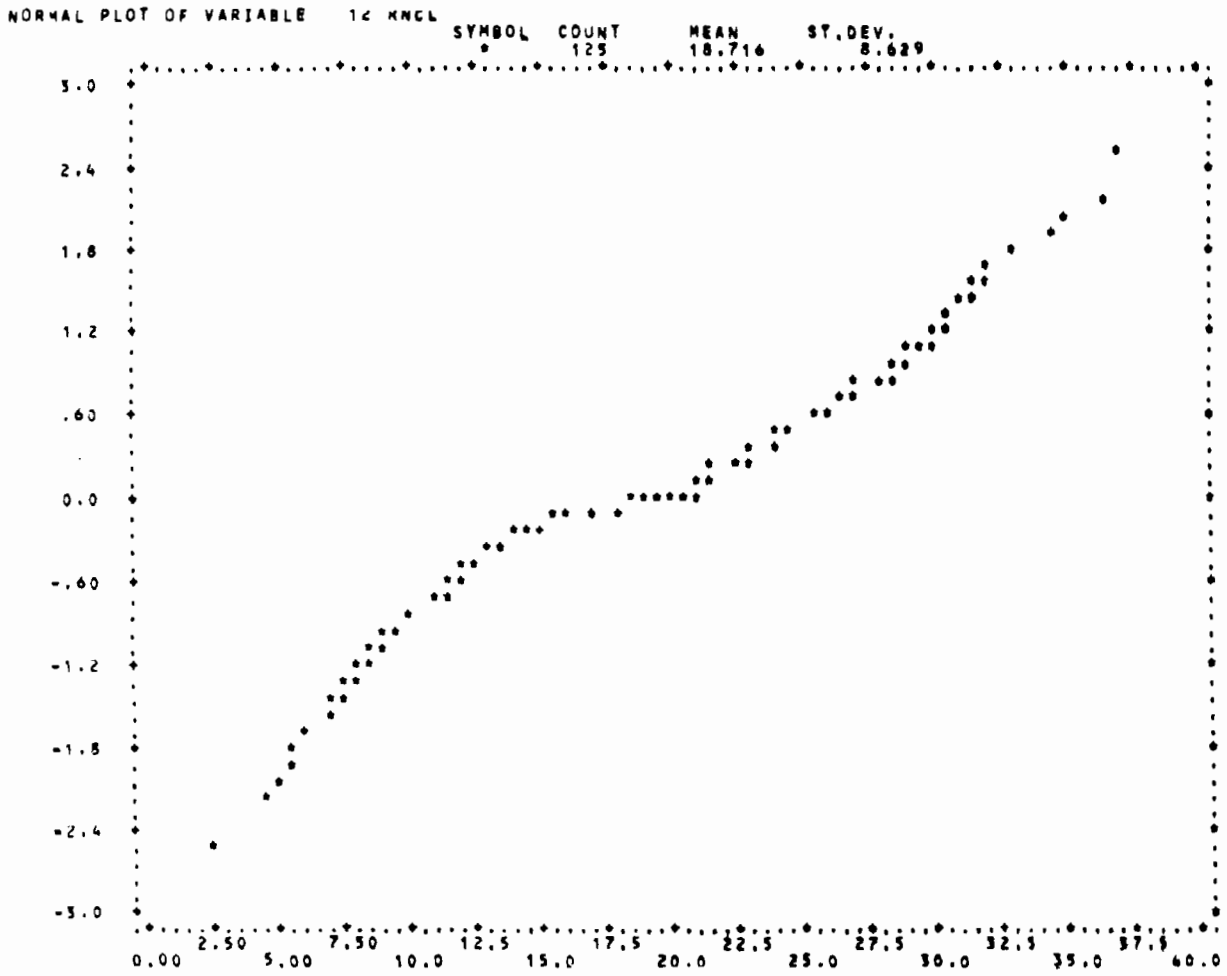


Figure A.4.11:Normal plot of mixed distributions 55N/10,3/ and 70N/27,5/ (modeling).

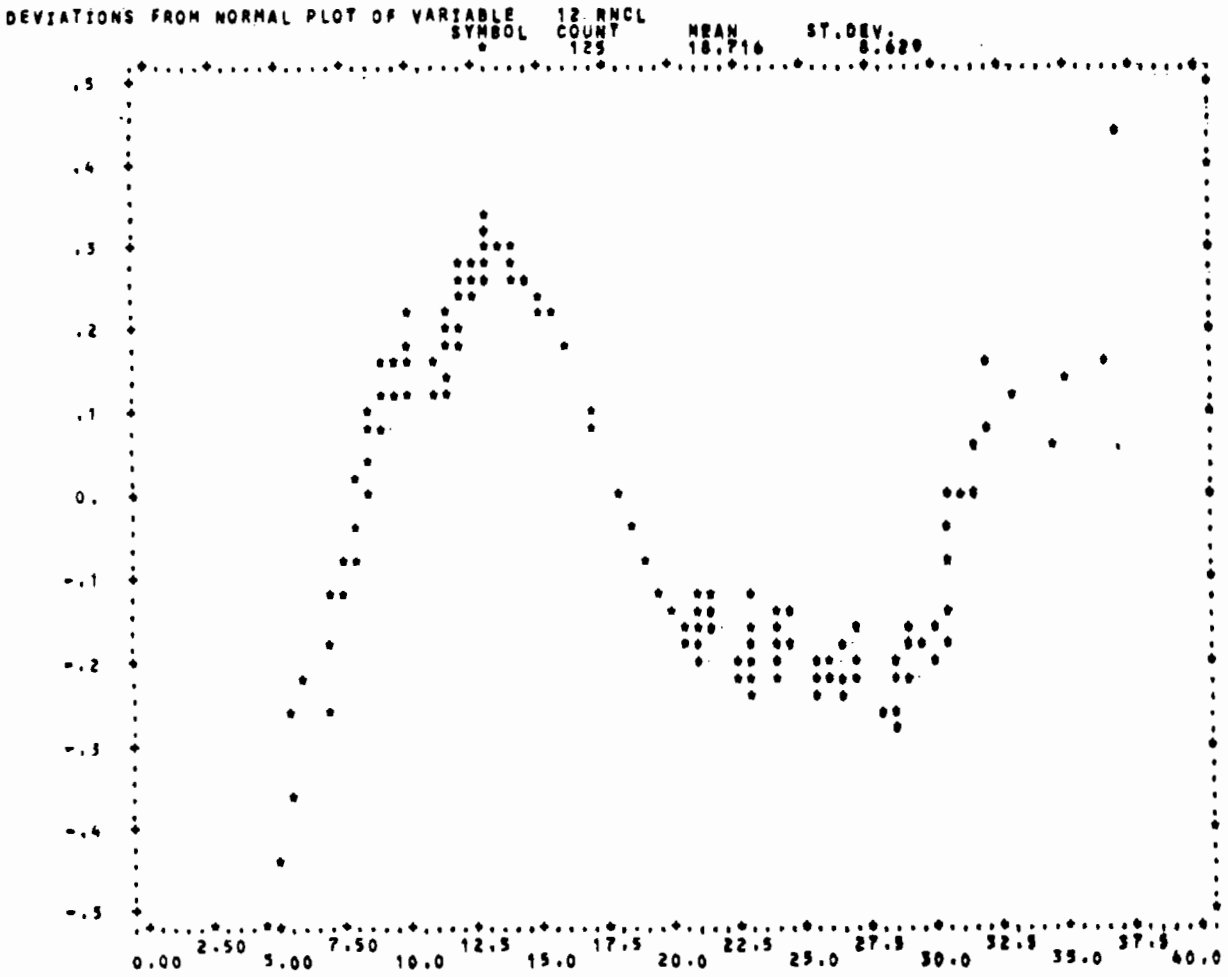


Figure A.4.12:Deviations from normal plot of mixed distributions 55N/10,3/ and 70N/27,5/ (modeling).

APPENDIX TO CHAPTER 4 (continuation)
Smoothing of series of observations.

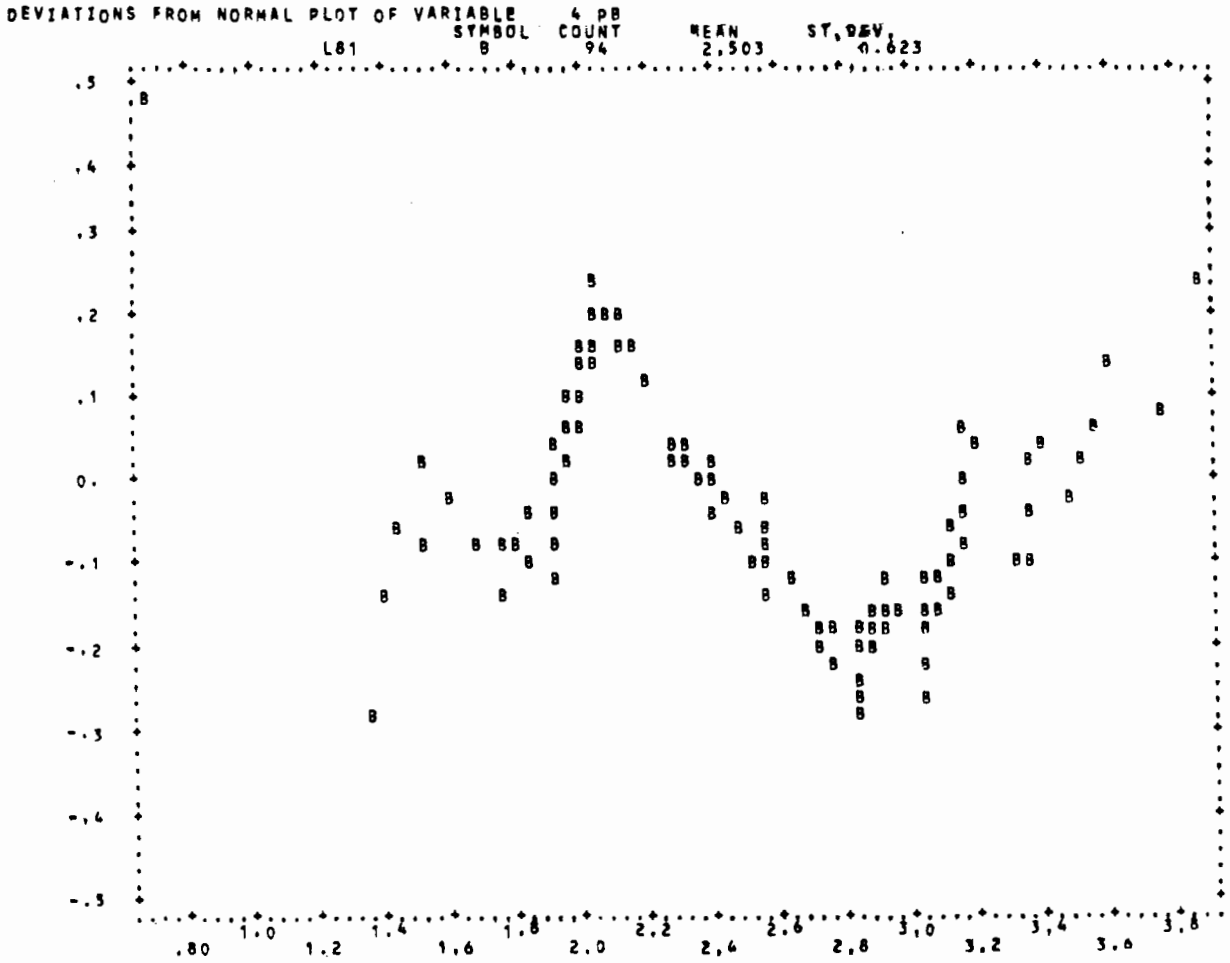


Figure A.4.13:Deviations of normal plot of logarithmics of concentrations of lead. Berezina, warm season, 1981.

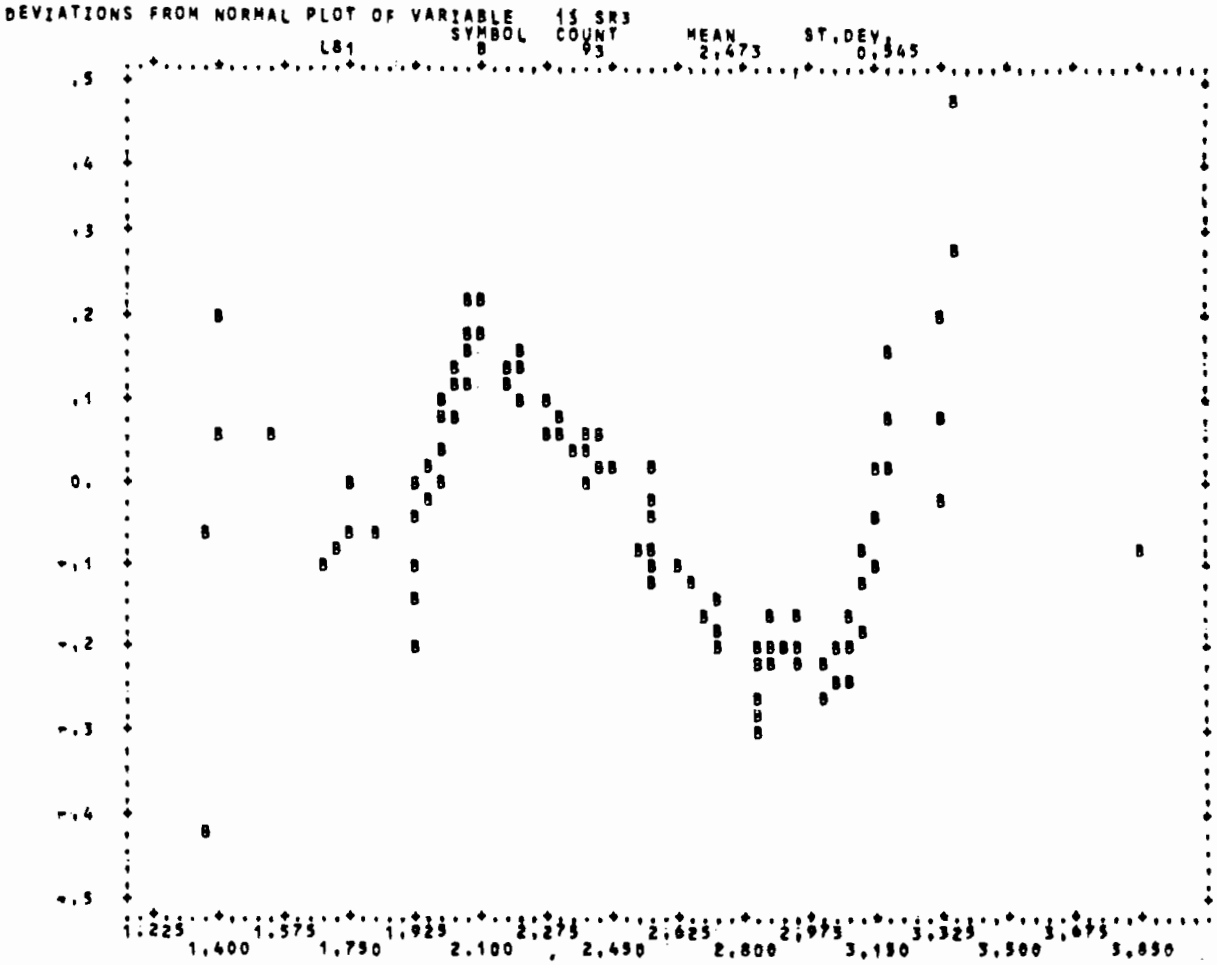


Figure A.4.14:Deviations of normal plot of smoothing series of logarithmic concentrations of lead. Berezina, warm season, 1981.

APPENDIX TO CHAPTER 5.

Construction of component of distributions of concentrations on seasonal series of observations.

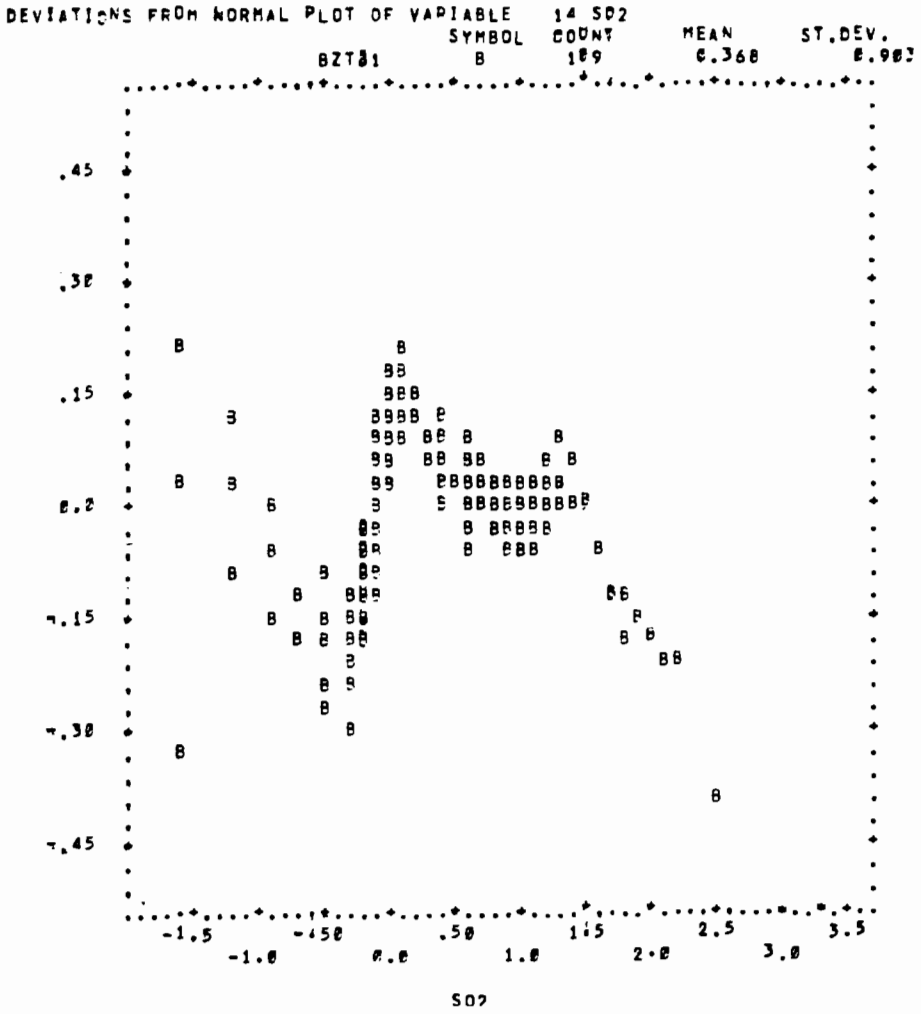


Figure A.5.1: Deviations from normal plot of logarithmic concentrations of sulfur dioxide, 1977-1979. Berezina, warm season, 1981.

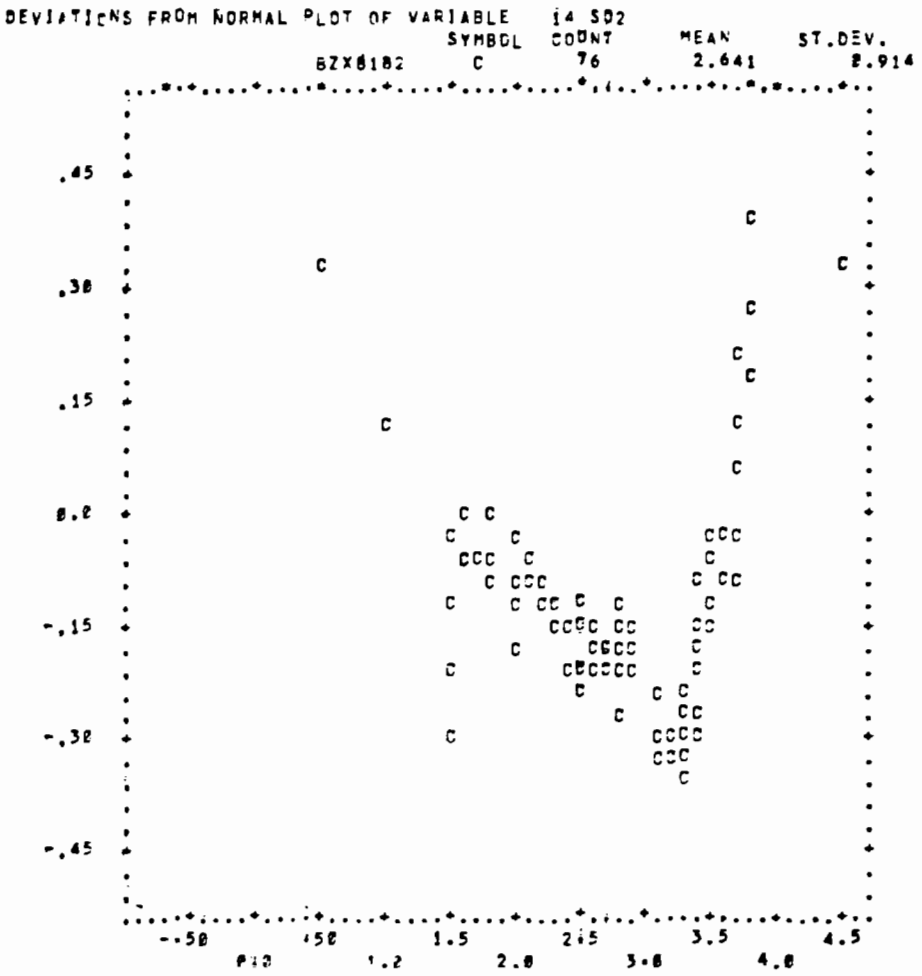


Figure A.5.2: Deviations from normal plot of logarithmic concentrations of sulfur dioxide, Berezina, cold season, 1981-1882.

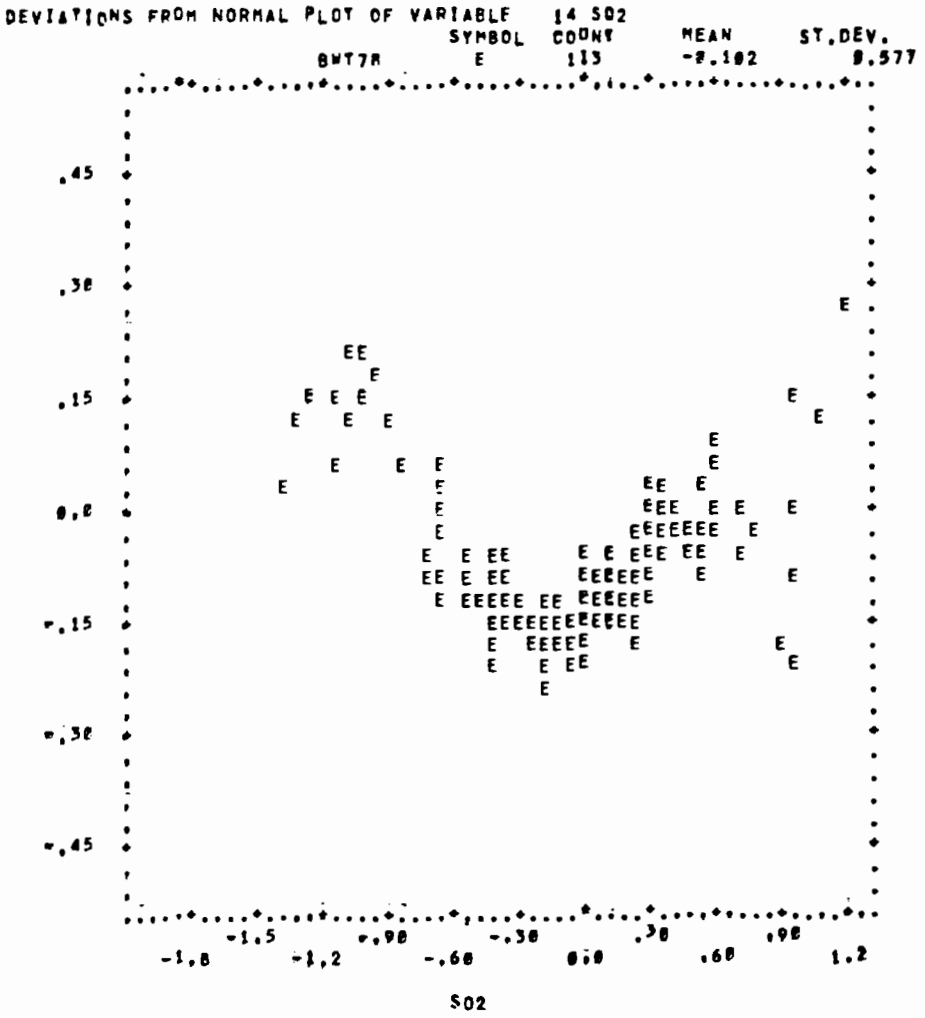


Figure A.5.3: Deviations from normal plot of logarithmic concentrations of sulfur dioxide, Borovoe, warm season, 1978.

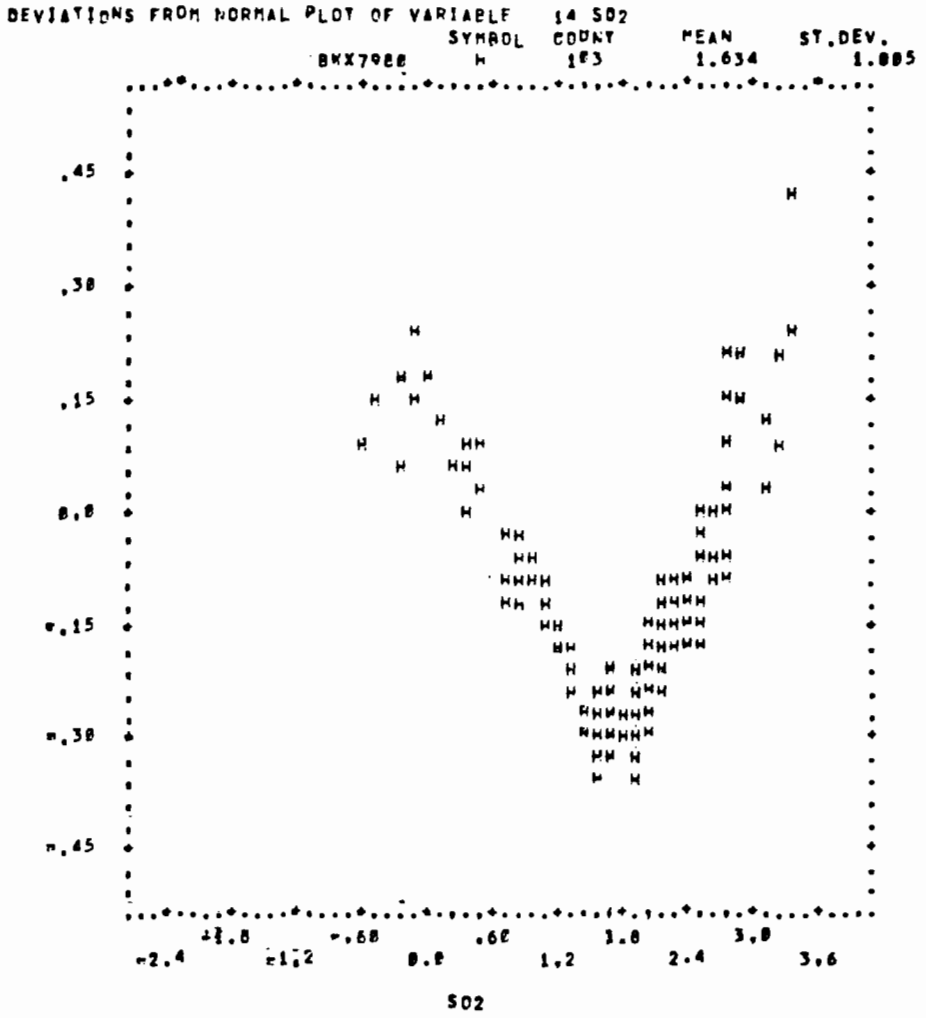


Figure A.5.4: Deviations from normal plot of logarithmic concentrations of sulfur dioxide, Borovoe, cold season, 1979-1980.

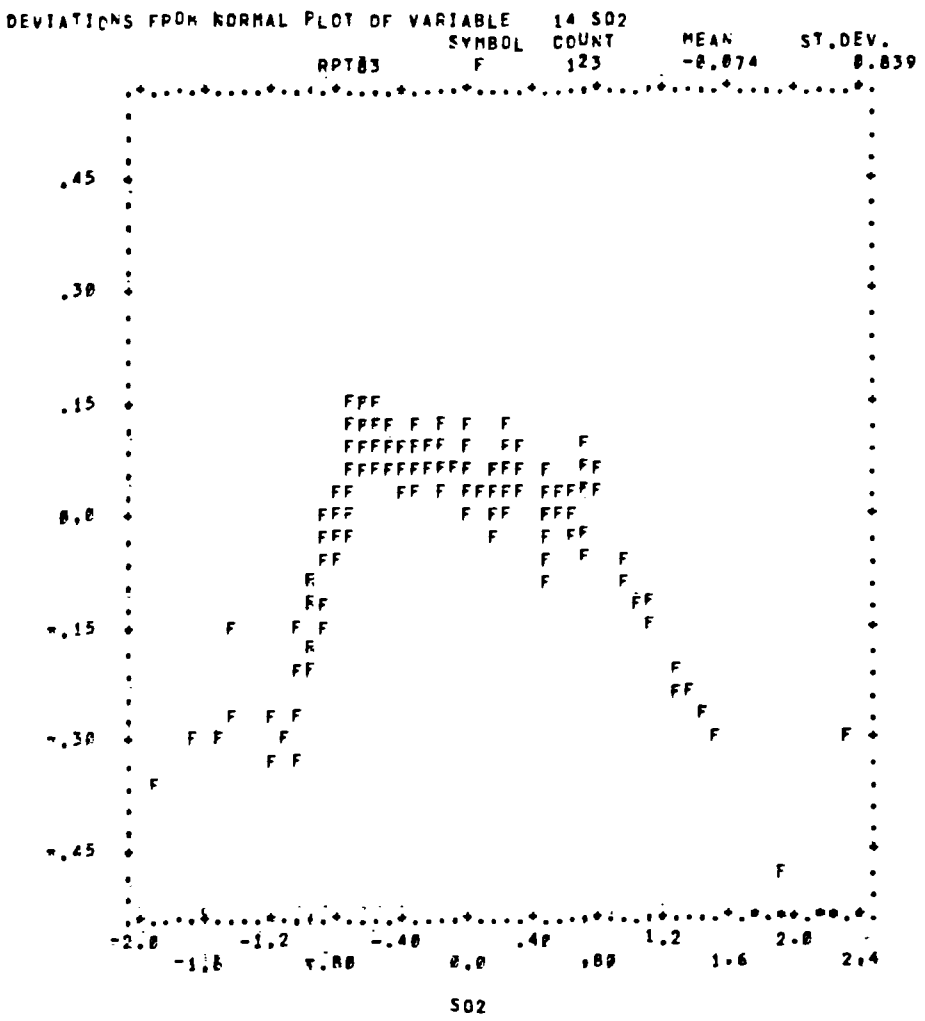


Figure A.5.5: Deviations of normal plot of logarithmic concentrations of sulfur dioxide, Repetek, warm season, 1983.

APPENDIX TO CHAPTER 5 continued.

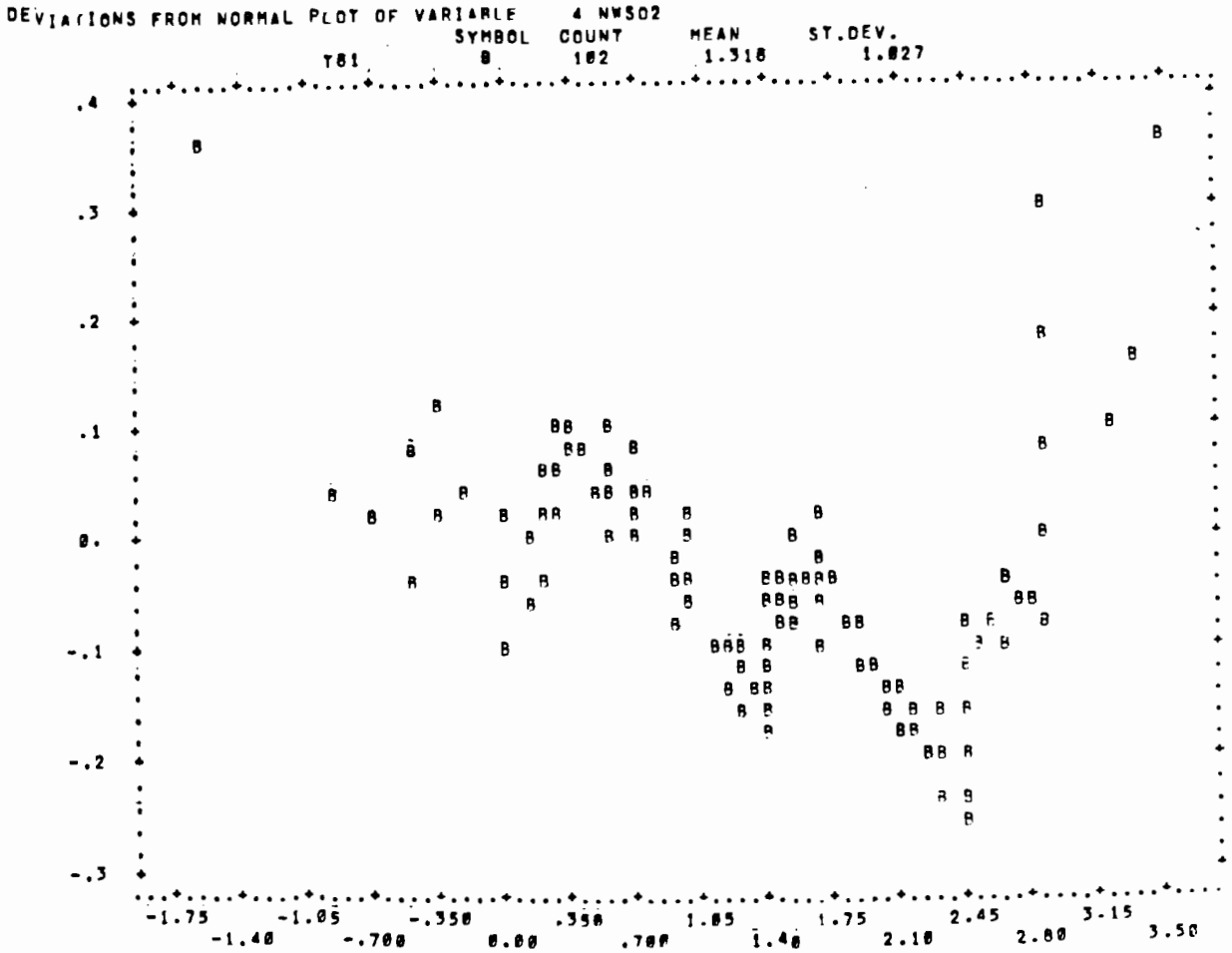


Figure A.5.6: Deviations of normal plot of logarithmic concentrations of sulfur dioxide, Norway, warm season, 1981.

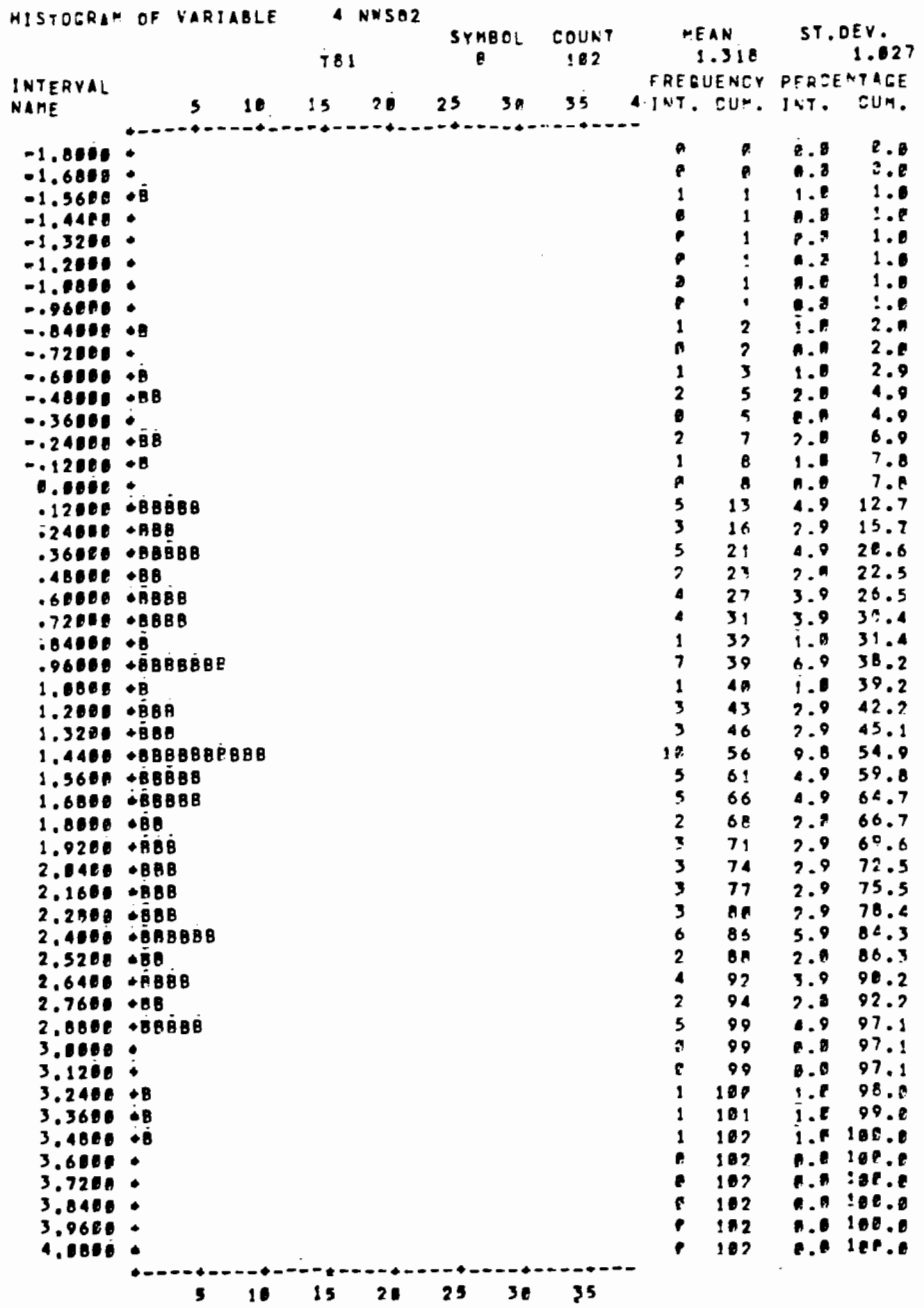


Figure A.5.7: Histogram of logarithmic concentrations of sulfur dioxide, Norway, warm season, 1981.

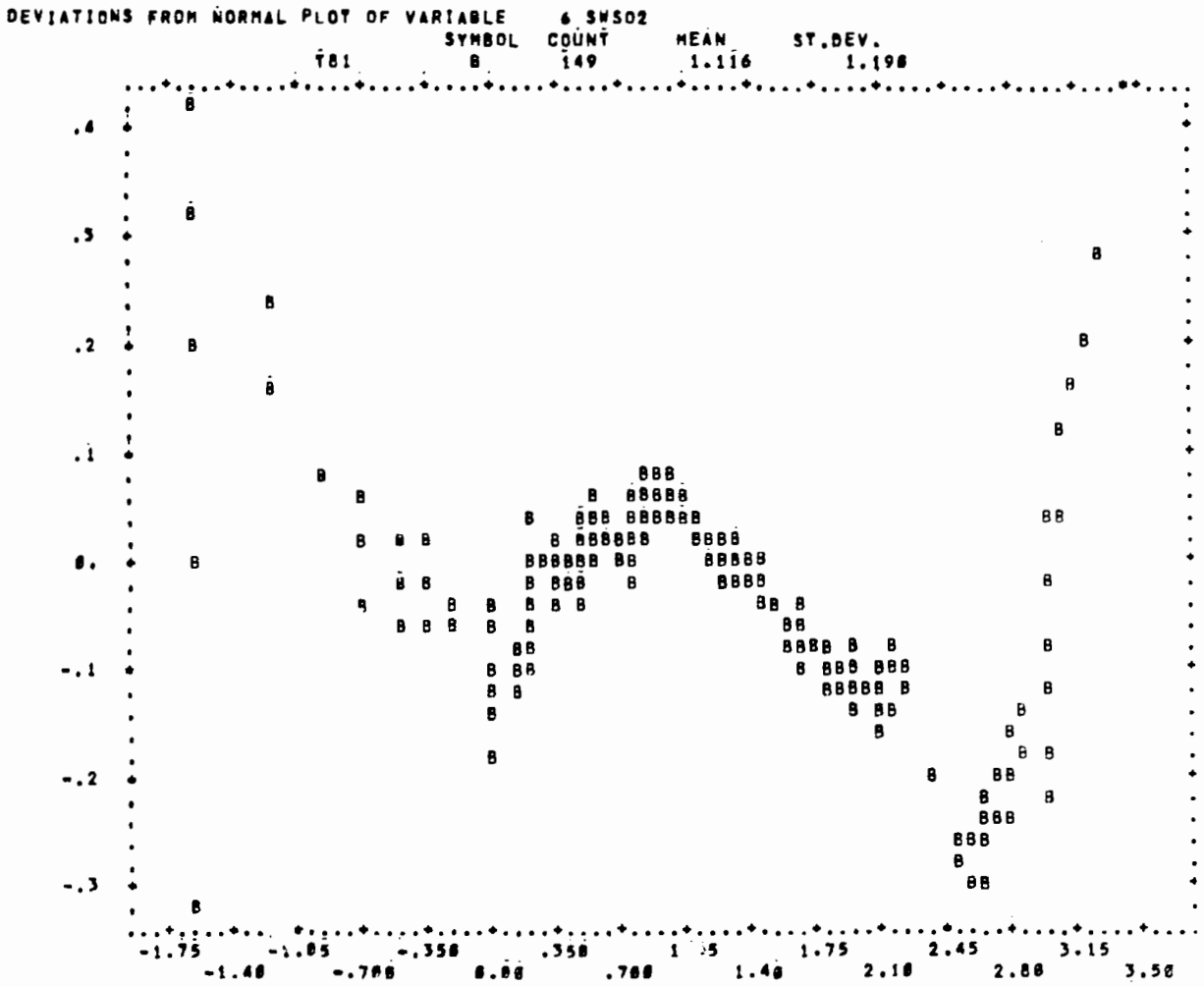


Figure A.5.8: Deviations of normal plot of logarithmic concentrations of sulfur dioxide, Sweden, warm season, 1981.

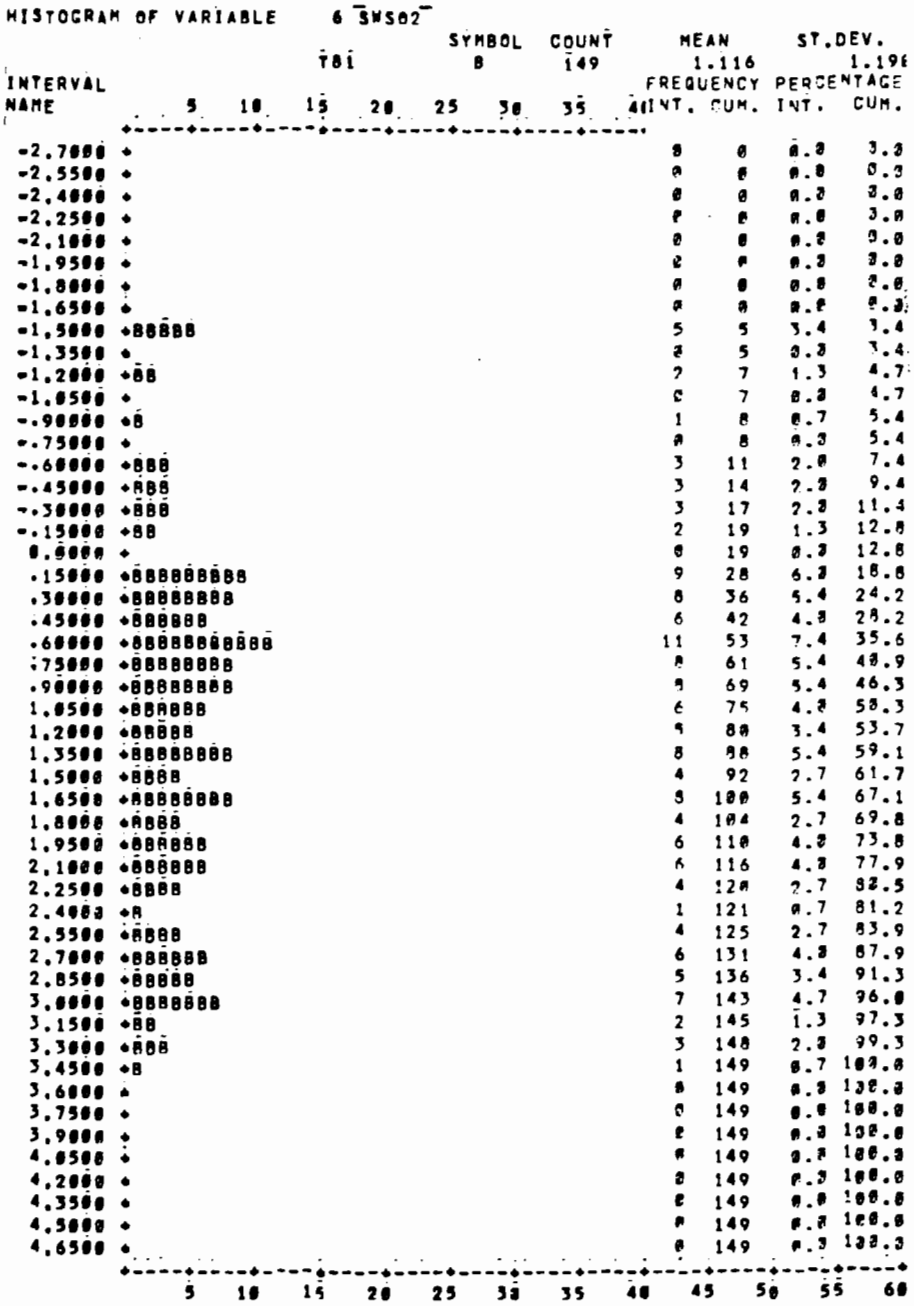


Figure A.5.9: Histogram of logarithmic concentrations of sulfur dioxide, Sweden, warm season, 1981.

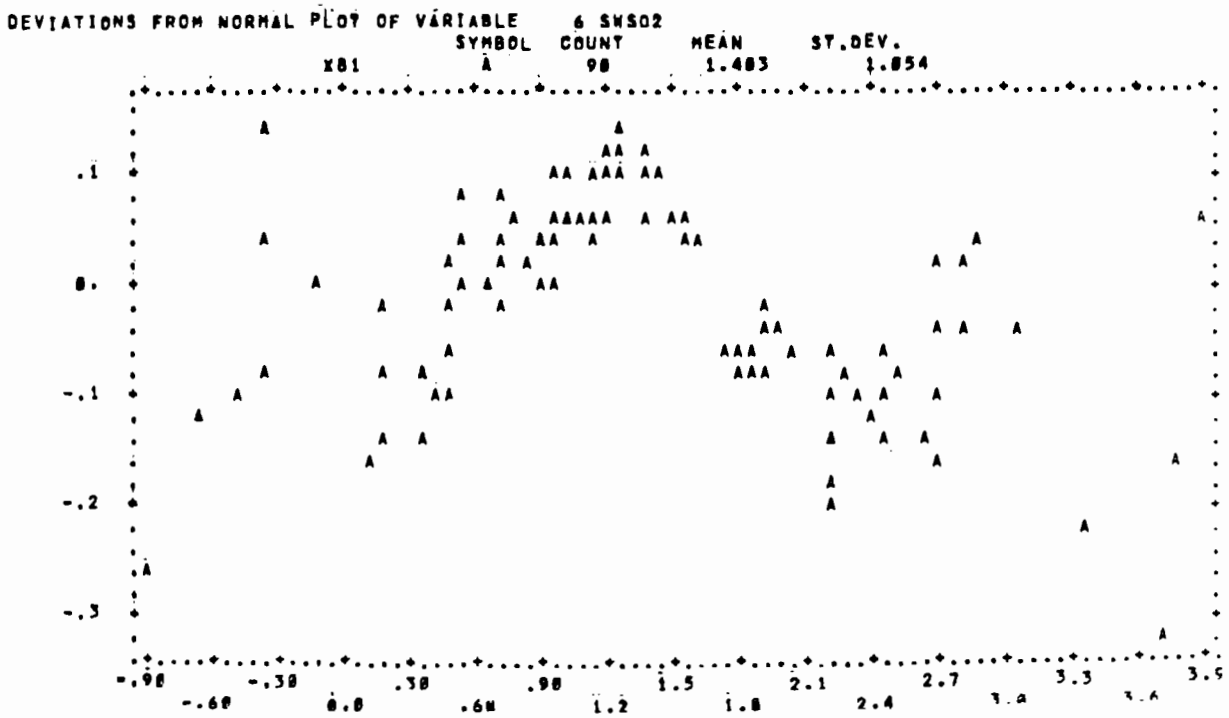


Figure A.5.10:Deviations of normal plot of logarithmic concentrations of sulfur dioxide, Sweden, cold season, 1981.

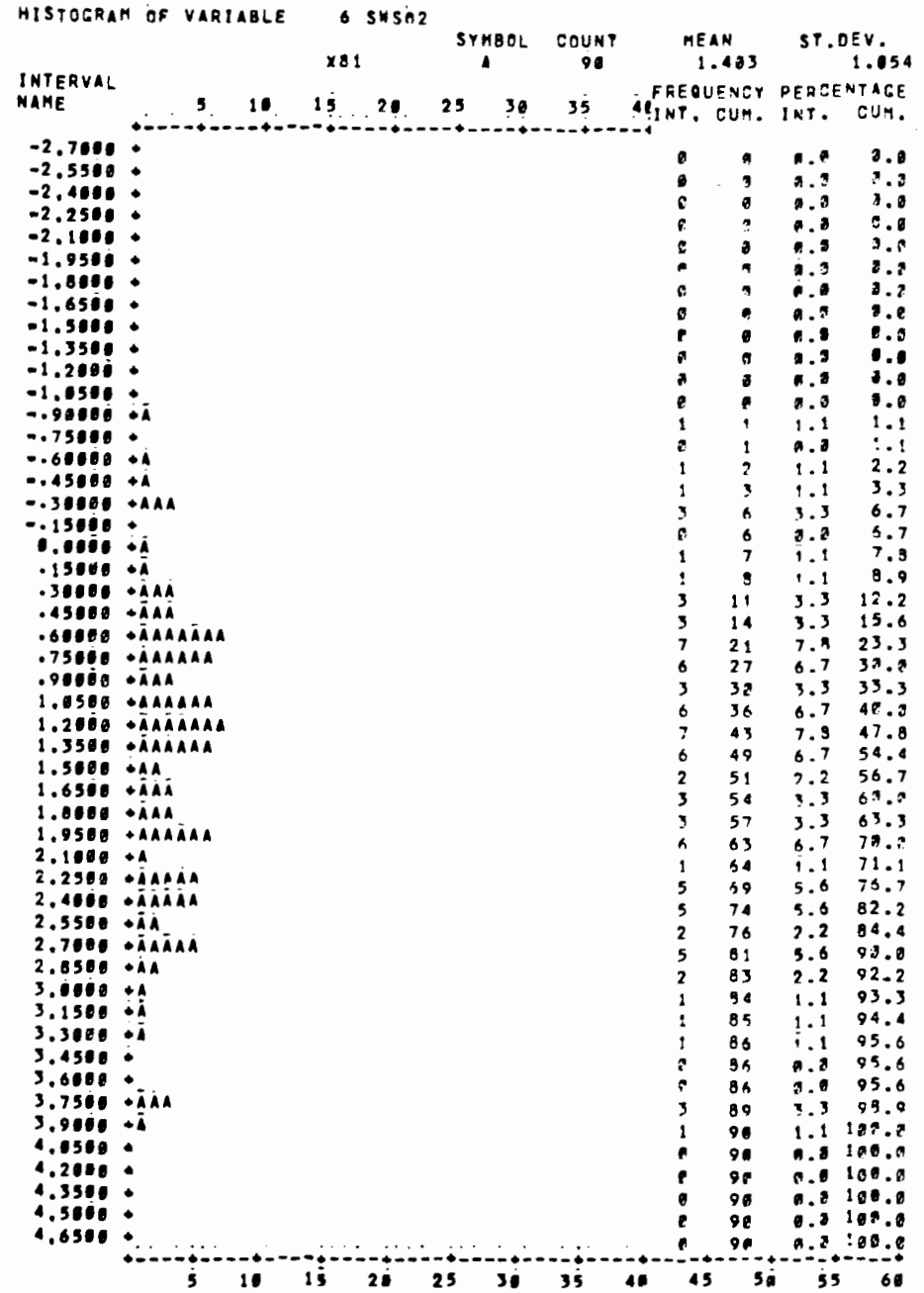


Figure A.5.11:Histogram of logarithmic concentrations of sulfur dioxide, Sweden, cold season, 1980-1981.

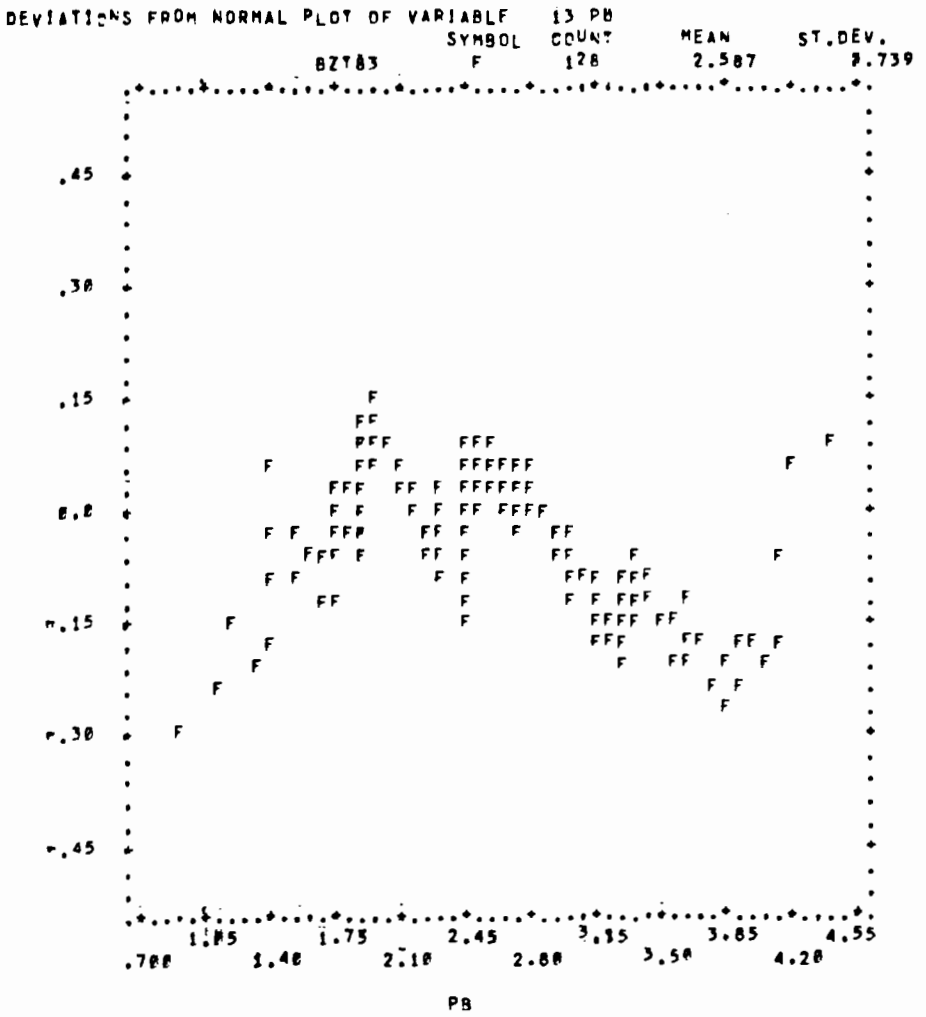


Figure A.5.12:Deviations of normal plot of logarithmic concentrations of lead, Berezina, warm season, 1983.

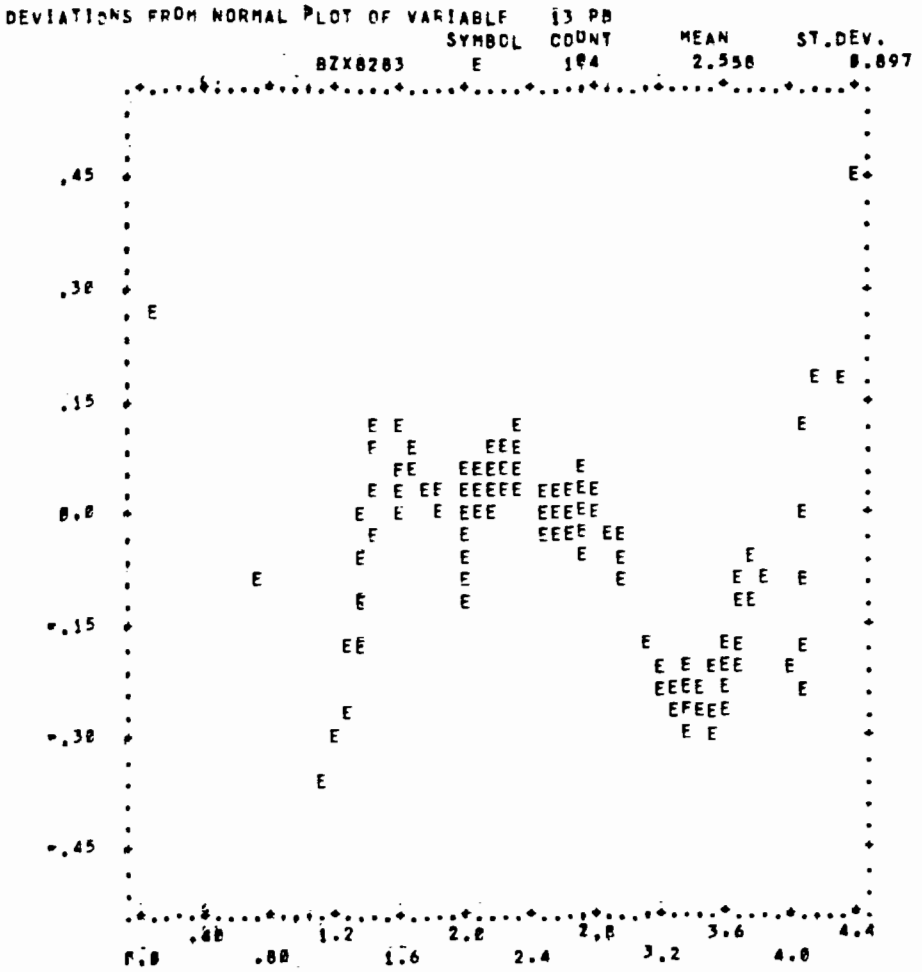


Figure A.5.13:Deviations of normal plot of logarithmic concentrations of lead, Borovoe, warm season, 1979.

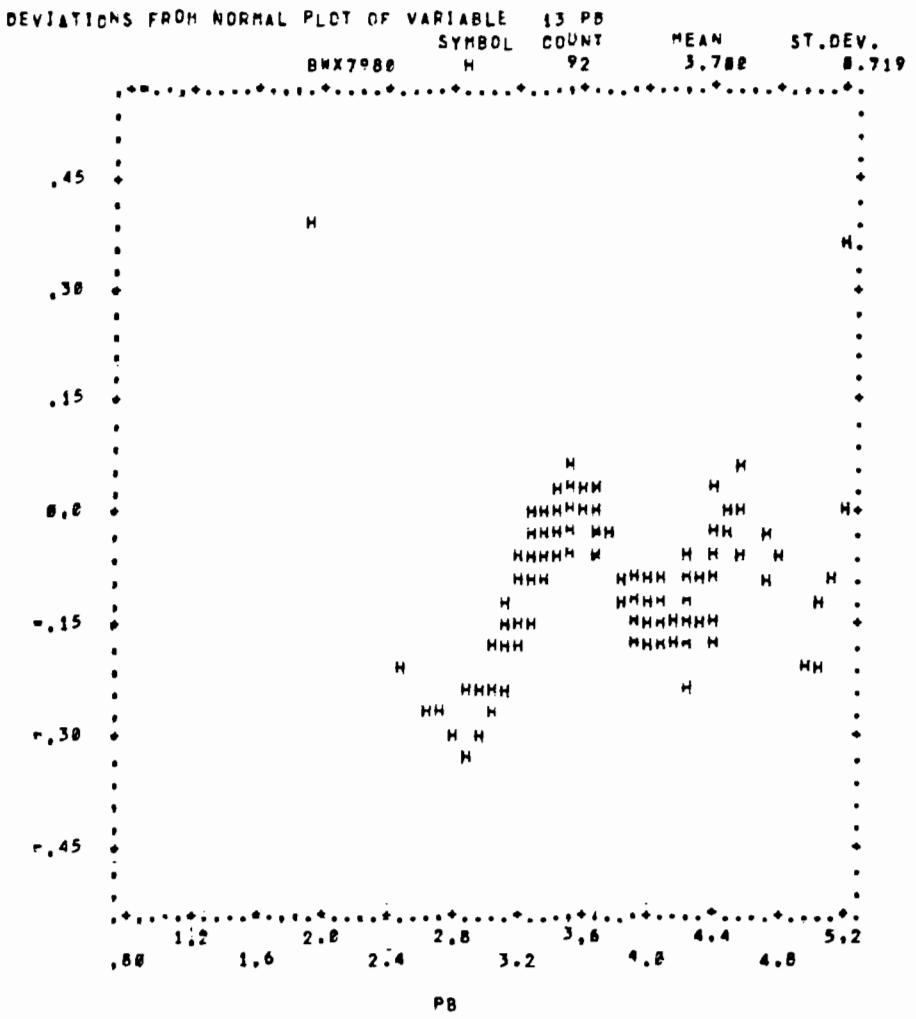


Figure A.5.14:Deviations of normal plot of logarithmic concentrations of lead, Repetek, warm season, 1981.

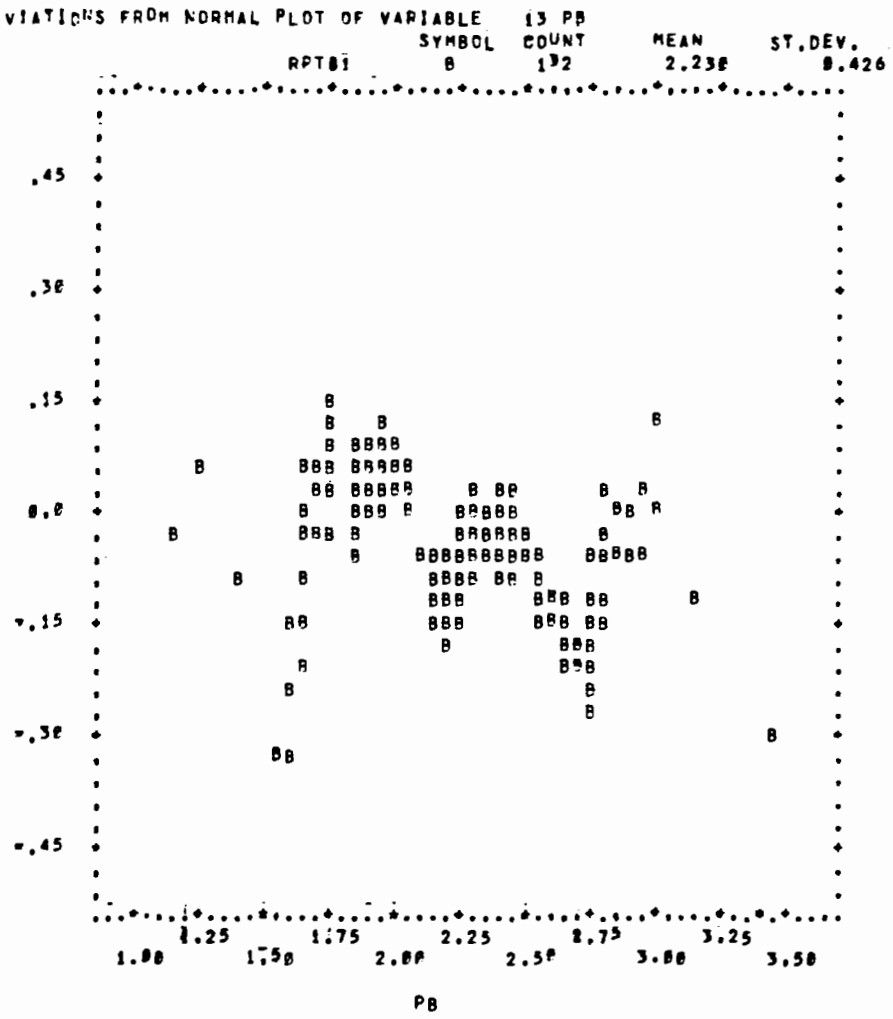


Figure A.5.15:Deviations of normal plot of logarithmic concentrations of lead, Repetek, cold season, 1981-1982.

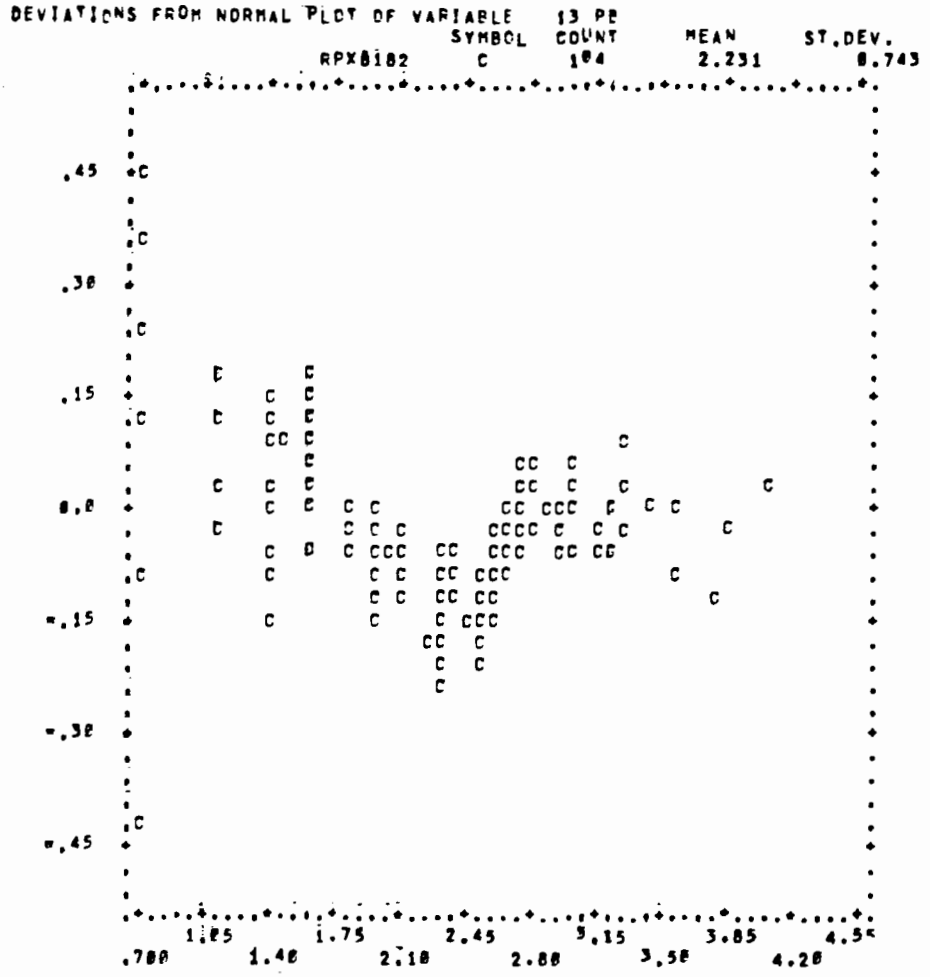


Figure A.5.16:Deviations of normal plot of logarithmic concentrations of lead, suspended particulate matter, Berezina, warm season, 1982.

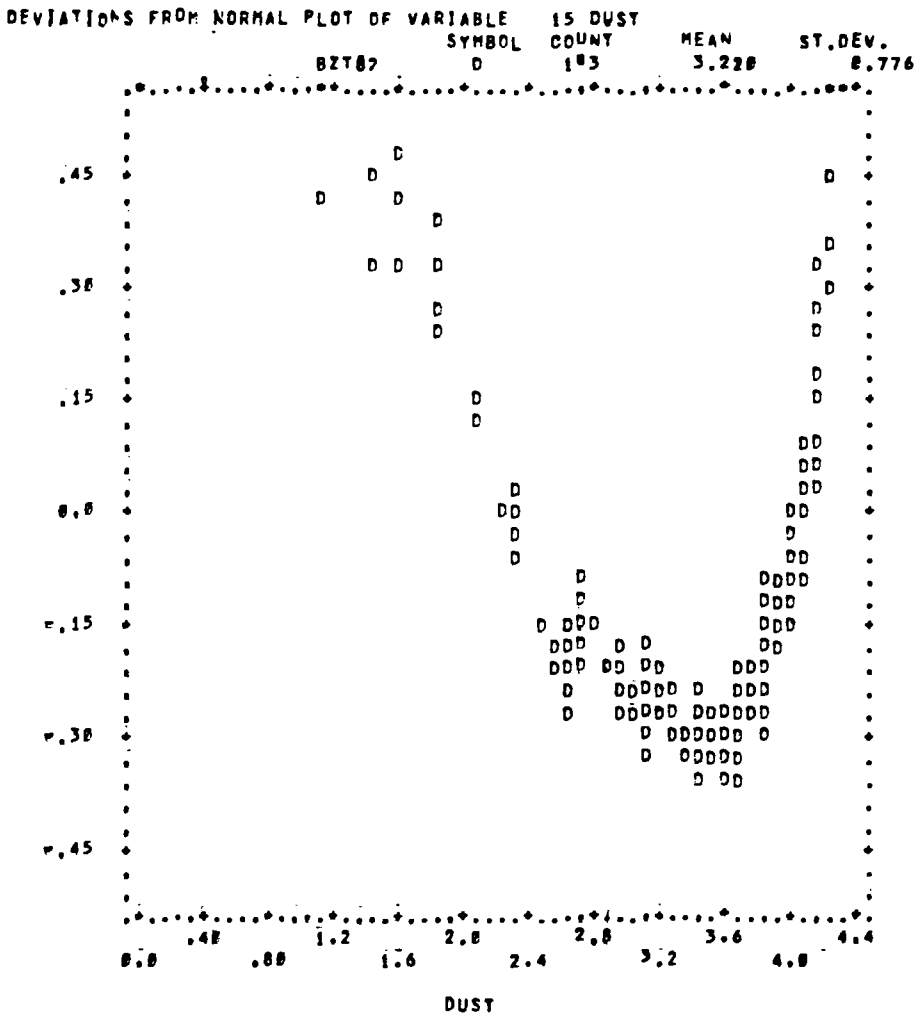


Figure A.5.17:Deviations of normal plot of logarithmic concentrations of suspended particulate matter, Berezina, cold season, 1982-1983.

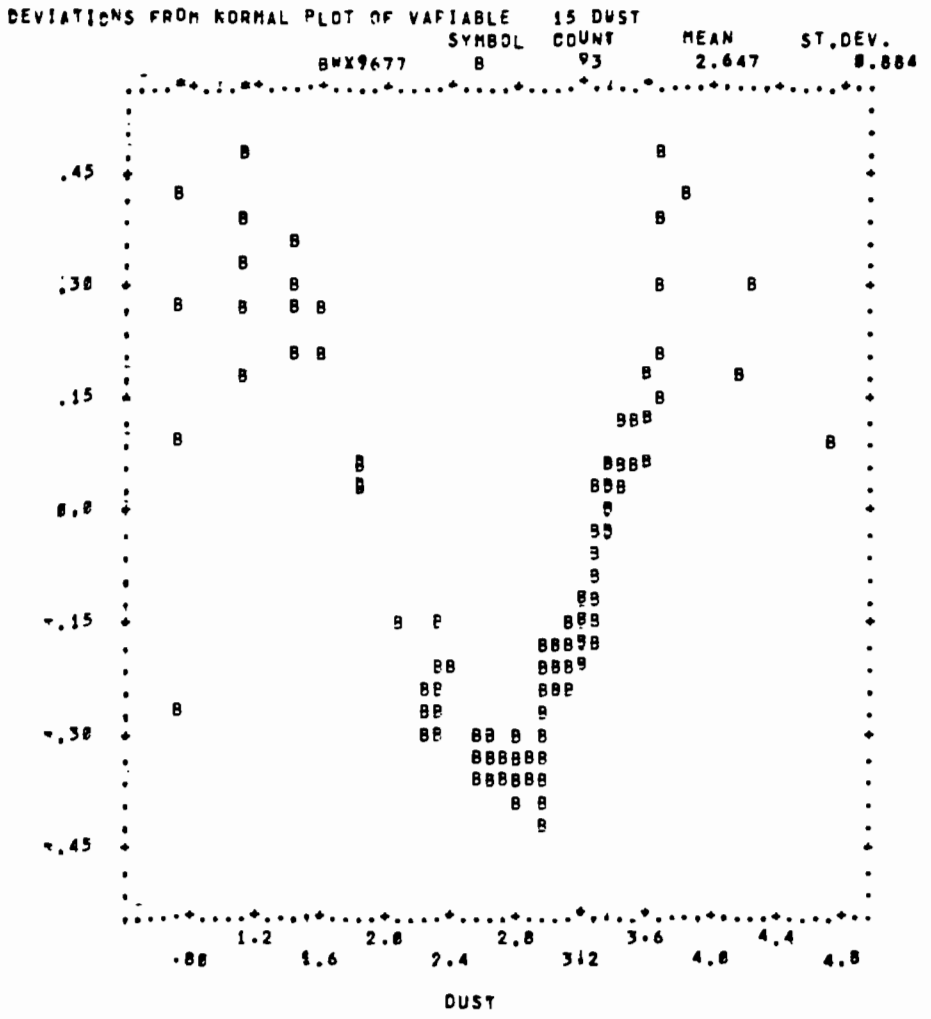


Figure A.5.18:Deviations of normal plot of logarithmic concentrations of suspended particulate matter, Borovoe, warm season, 1979.

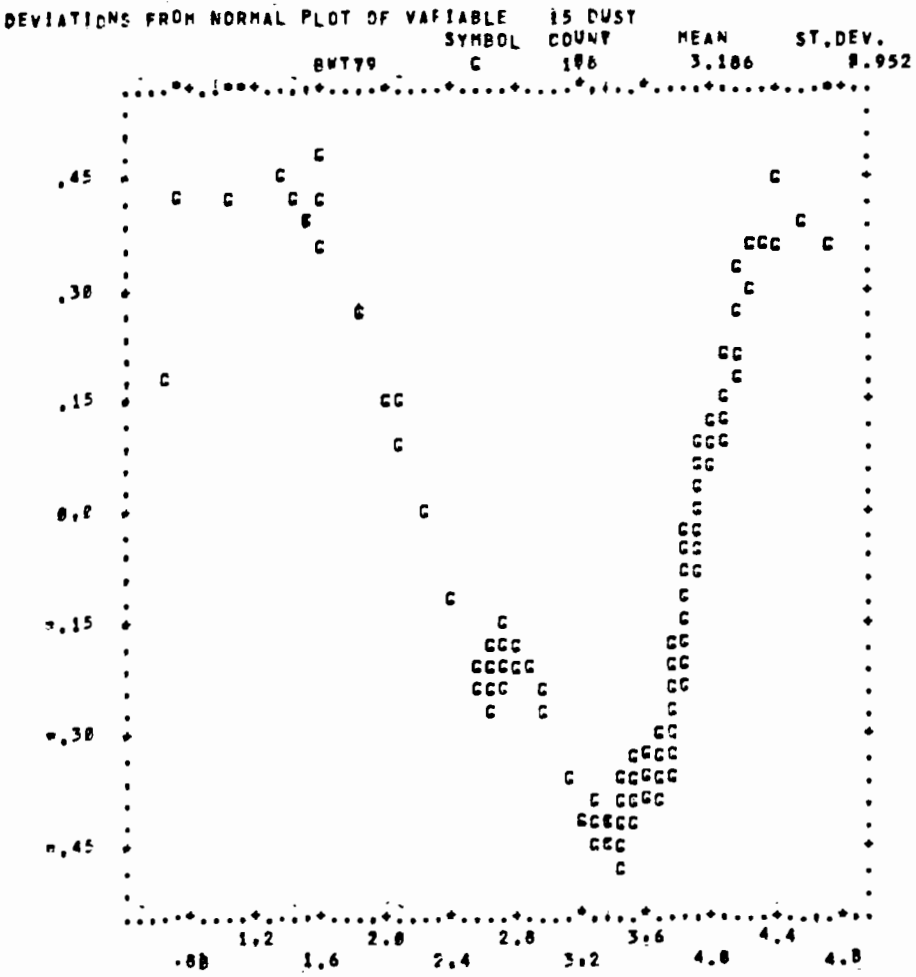


Figure A.5.19:Deviations of normal plot of logarithmic concentrations of suspended particulate matter, Repetek, warm season, 1982.

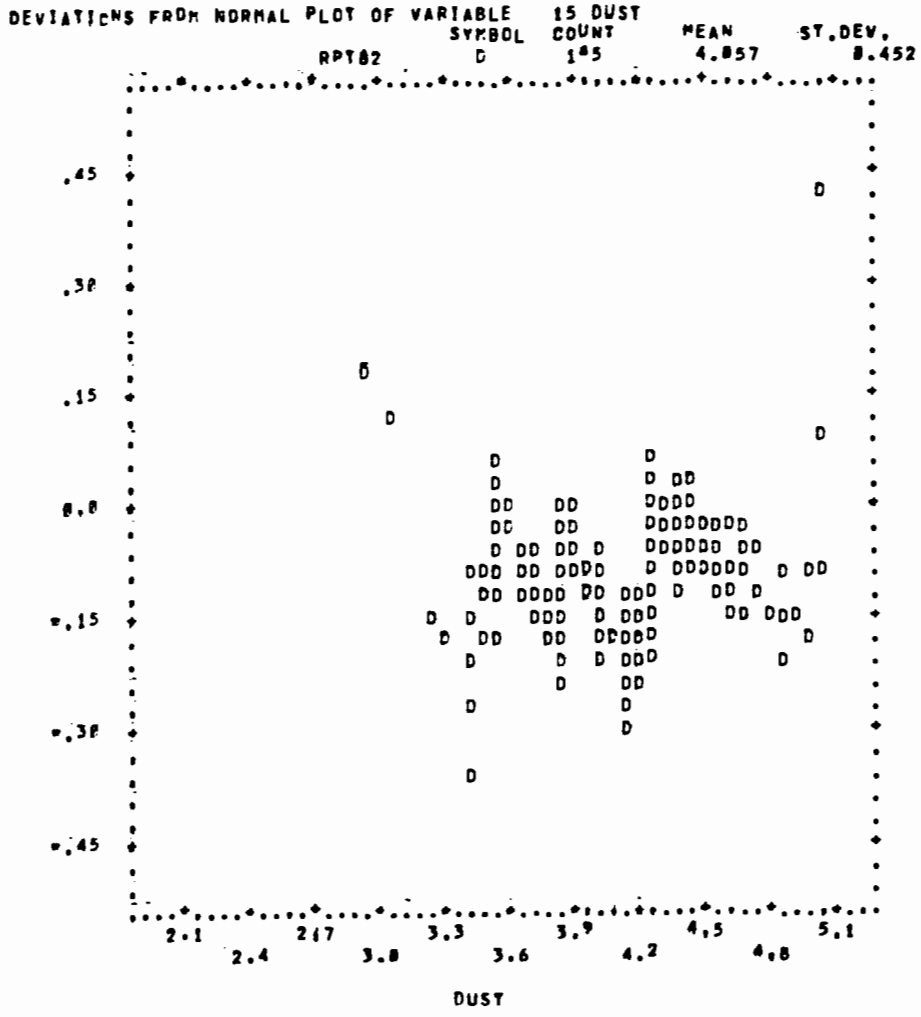


Figure A.5.20:Deviations of normal plot of logarithmic concentrations of suspended particulate matter, Repetek, warm season, 1980-1981.