A MONTE CARLO MODEL TO SYNTHETICALLY GENERATE THE PEM BOUNDARY CONDITIONS FOR CHAIN BRIDGE

by

H. Carlton Haywood
Kathleen B. Hogan

October 1986
ICPRB Report No. 86-15

Interstate Commission on the Potomac River Basin
6110 Executive Boulevard, Suite 300
Rockville, Maryland 20852

This report has been prepared by the Interstate Commission on the Potomac River Basin. Funds for this report are provided by the United States Government, the U.S. Environmental Protection Agency, and the signatory bodies to the ICPRB: Maryland, Pennsylvania, Virginia, West Virginia, and the District of Columbia. The opinions expressed are those of the authors and should not be construed as representing the opinions or policies of the United States or any of its agencies, the several states, or of the Commissioners of the Interstate Commission on the Potomac River Basin.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

In 1985, following the recommendations of the 1983 Algae Bloom Expert Panel (Ref 1), a multi agency "Long Term Workplan" for Potomac Estuary technical studies was established by the Potomac Strategy State/EPA Management Committee to investigate the causes, and possible control, of eutrophication in the Potomac Estuary. Part of the Long Term Workplan is to quantify the frequency of occurrence, and severity, of algae blooms with each of several pollution control options (Ref 2). Under Task 4 of the Workplan are several activities undertaken cooperatively by the Interstate Commission on the Potomac River Basin (ICPRB) and the Metropolitan Washington Council of Governments (MWCOG), jointly intended to describe the frequency of occurrence of algae blooms in the Potomac Estuary. As part of the ICPRB contribution to Task 4, this report describes a probabilistic computer model of the frequency distribution of Potomac Estuary boundary conditions. This model, after further development in cooperation with MWCOG, will be used in conjunction with the Potomac Eutrophication Model (PEM, Ref 3) in a Monte Carlo procedure to estimate the frequency of occurrence of algae blooms.

THE MONTE CARLO PROCEDURE FOR ESTIMATING ALGAE BLOOM FREQUENCY

Nuisance algae blooms occur as a result of a combination of factors including climate conditions (flow, water temperature, etc.); and the availability of nutrients (Nitrogen, Phosphorus, etc.) from point and nonpoint pollution sources (STPs, urban runoff, upstream sources, sediment). Reductions in the nutrient load from point and nonpoint sources, as a result of adopting various pollution control options, should reduce the frequency and severity of blooms. The joint impact of boundary conditions (flow, temperature, sunlight, etc.) and STP loads on algae growth is complex and not well understood, but may be estimated using the PEM model.

PEM is a time variable, deterministic, model which computes chlorophyll-a concentrations, from which algae concentrations are estimated, given a set of boundary conditions and input loads. In the past, PEM has been run as a historical simulation of the events of a given year. Used in this way, the likelihood of recurrence of the chlorophyll levels predicted by PEM is not known because the joint frequency of occurrence of the combination of boundary conditions and loads is unknown. As such PEM by itself does not provide an estimate of the frequency of occurrence of algae blooms. Still, PEM can serve as a transformation function between the STP loads plus boundary conditions as input to the model, and chlorophyll as output from the model. (How well PEM predictions of chlorophyll correlate with actual algae in the Estuary is a problem to be addressed by other Task 4 projects.)

To make the relationship between inputs and output more clear, an abbreviated version of PEM can be set up to run in a steady state, rather than time variable, mode with fixed STP loads and fixed boundary conditions. Just as this version of PEM can be used to transform any particular combination of boundary conditions and loads into chlorophyll, it can be used to transform the joint frequency distribution of boundary conditions and loads into a frequency distribution of chlorophyll. This distribution of chlorophyll can be approximated using a Monte Carlo procedure in which sets of boundary conditions are randomly selected from their joint distribution. For each set of boundary conditions a steady state PEM run is made to determine the associated chlorophyll levels. This procedure allows the simulation of more combinations of boundary conditions than is possible from observation of the historical record, and with a sufficiently large number of runs the joint frequency distribution of inputs to PEM is transformed into a frequency distribution of chlorophyll (and by implication, algae).

If one assumes that STP loads are fixed, at levels determined by the control options: then the distribution of chlorophyll, for that control option, is a function of the joint distribution of just the boundary conditions and pollution control options may be evaluated by repeating the Monte Carlo procedure for each option and comparing the resulting chlorophyll distributions.

DEVELOPMENT OF THE MONTE CARLO MODEL

The Monte Carlo model is a FORTRAN language computer program (Appendix I) which repeatedly generates sets of PEM boundary conditions for Chain Bridge (see Figure 1) by randomly selecting values from distributions of individual parameters (listed in Table 1). The model uses distributions for each of the output parameters except carbon and total daily solar radiation, which are computed based on distributions of other parameters. Calculation of values for four parameters (temperature, flow, solar radiation, and carbon) is dependent on season, so a day of year must be specified for each set of boundary conditions generated. This version of the model assumes that individual parameter's values are independent (see Discussion, p. 8).

Distributions were developed by combining the available data into a single data set for each parameter. The distributions were determined by ranking these data in ascending order and then calculating an empirical cumulative probability for each value. That is, the probability that a parameter will have a value less than or equal to some specified value is,

$$P\{X<=x\} = r / (n + 1) \tag{1}$$

where

P = a cumulative probability function for parameter X,
x = a specified value for X,
r = rank of parameter value x,
n = number of observations in the observed data set,

The resulting distributions are plotted in Figures 3-24, Figures 26-37, and Figure 39. These plots use a Normal probability scale which has the property that Normally distributed data will plot as a straight line. A Normally distributed approximation of each data set, calculated as

$$z = (x - xmean)/sd \tag{2}$$

where

z = standard Normal deviate,
$xmean$ = a specified value for the parameter X,
x = sample mean X,
sd = sample standard deviation for X,

was also plotted (dashed line) on each Figure. By comparing the empirical distributions with the Normal approximations, it can be seen that: a) the empirical distributions tend not to be "smooth"; and b) that even when the Normal approximation is "close" to the empirical distribution (e.g. Figure 9) the greatest errors are in the tails of the distributions which are precisely the regions of most concern. For these reasons it was decided to base the Monte Carlo model on the empirical distributions, rather than on some parametric distribution, Normal or otherwise.

This approach has the advantage that predictions of boundary condition values are based solely on the observed data sets, and are not dependent on any assumptions about distributional characteristics. Using empirical distributions, however, means that the probability of occurrence of values outside the range of the observed data cannot be exactly calculated. Rather, only the probability of occurrence of <u>all</u> values greater than the maximum observed, or less than the minimum observed, can be calculated as

$$P\{X>=xmax\} = 1 - n/(n+1) \tag{3}$$

$$P\{X<=xmin\} = 1/(n+1) = P\{X>=xmax\}$$

where

$xmax$ = largest observed value for parameter X,
$xmin$ = smallest observed value for X,
n = number of observations.

It can be seen from Equation (3) that the ability to predict the frequency of occurrence of extreme events using empirical distributions is limited by the sample size. Unfortunately, the sample size for some of the boundary condition parameters is quite small.

In operation, the program has stored in memory the empirical distribution for each parameter listed in Table 1 (except carbon and solar radiation). Each distribution is an array of pairs of numbers, x_i and $P\{X<=x\}$, where the x_i are the observed data and $P\{X<=x\}$ are the associated probabilities as defined in Equation 1. To generate a set of boundary condition values, the program reads, either from a file or from the computer terminal, a list of numbers representing Days of Year. For each Day of Year: 1) a Uniformly distributed random number is generated for each parameter except carbon and solar radiation (I); 2) for each of these parameters a value is selected with the lowest probability greater than or equal to the random number; 3) an I value is derived by multiplying the random %CSR value by the Clear Sky Radiation for that Day of Year (Table 8); 4) carbon is derived using Equations 4-6 from the Chla and I values; and 5) the set of generated boundary condition values is printed. The list of Days of Year may include only a single day repeated many times, or may vary within a narrow range, to generate boundary conditions (and ultimately Chlorophyll from PEM) at specific seasons of the year; as for example, during August.

DESCRIPTION OF DATA FOR PARAMETER DISTRIBUTIONS

Input parameter distributions were developed by combining data obtained from several sources including some long term monitoring programs and short term studies. These sources are described in Table 2. Data are available from the authors as well as from the original sources.

The amount of data available for each parameter varies substantially, from 20638 daily Little Falls flows to 124 Chla observations. The number of monthly observations for each parameter is provided in Table 3. Although all months of the year are represented, for most parameters there are approximately twice as many observations in the April through September period as during October through March.

Seasonal trends in each parameter were considered. However, only the flow and temperature data sets have enough observations to support monthly distributions. Summer versus Winter comparisons were made for the other parameters, except for Chla. Means and standard deviations for Summer versus Winter data, listed in Table 4, did not vary sufficiently to justify separate distributions for each season for parameters other than flow and temperature. There were too few Winter observations to justify separate seasonal distributions for Chla.

Distributions for Q (abbreviations defined in Table 1), T, DO, BOD, NH3, NOX, and Chla, were developed from data consisting of direct measurements of the parameters. Distributions for the phosphorus species (DOP, POP, DIP, and PIP), TON, and %CSR, were developed by transforming other measurements. Development of each parameter distribution is described below.

Streamflow (Q): Streamflow data are daily mean flow (cfs) from the USGS gage at Little Falls, MD, unadjusted for water supply withdrawals (Ref 4). The period of record is from 1 January 1930 to 22 October 1986, including 20638 daily mean flows. Because stream flow shows a significant seasonal trend, separate distributions are computed for each month of the year. Monthly variation in the log of stream flow is shown Figure 2 (means, standard deviation, minimum, and maximum) and summarized in Table 6. The distributions of monthly flows are plotted in Figures 3 - 14.

Water Temperature (T): PEM has one temperature function for the upper estuary, above River Mile (RM) 38.5. For PEM calibration, verification, and post-audit, (Ref 3,5) runs this function was determined by averaging all observations in the upper estuary. For this model, however, the temperature data set used is a composite of data collected by various agencies at stations from Chain Bridge to Wilson Bridge, RM 12.1. This data set was used because it was already available (Ref 6), and the majority of temperature data available for the upper 38 miles are actually from stations above RM 12. Comparison of this data set with data used in the 1983 PEM Post Audit Run show the average temperature between RM 0.0 and 12.1 is not significantly different from the average between RM 0.0 and RM 38.5. Although the data were collected over a twenty year period, this composite record is heavily weighted to the climate of the 1980 - 1986 period, with daily average temperatures for: 2218 days from June 1980 through June 1986, 584 days June 1978 to May 1980, and 302 days from 1966 to May 1978.

Daily Solar Radiation and Percent Clear Sky Solar Radiation (%CSR): The solar radiation data set is based on NOAA's daily edited global radiation for Sterling, Virginia (Dulles Airport) for the period January 1977 through June 1980 (Ref 7). This same data set was used to provide solar radiation values for calibration and verification runs of PEM. The strong seasonal trend in daily solar radiation was removed by converting the received solar radiation in the data set into a percentage of the potential "clear sky" radiation (%CSR) for that day of year at Sterling. A daily solar radiation is computed by the Monte Carlo Model from the distribution of %CSR by randomly selecting a %CSR and multiplying that number by the "clear sky" radiation for that day of year. For this reason a day of year must be specified for each set of boundary conditions generated by the model.

As indicated in Figure 38 and Table 4, there does not appear to be a seasonal trend in %CSR so a single distribution for the entire year was used. The absence of very low %CSR values in August is suggestive, but of limited significance because of the lack of August data. Summary statistics for the distribution are in Table 5.

Carbon and Chlorophyll-a (Chla): Carbon data, from which a carbon distribution can be estimated, were not available. However, carbon can estimated using a method described in the PEM Calibration and Verification Report (Ref 3), pp. 235-237, in which carbon is a function of solar radiation, temperature, minutes of daylight, and the Chla concentration. The relationship is

$$R = (68 * I) / (min * 1.068{*}{*}(T - 20)) \tag{4}$$

$$R >= 20.0 \tag{5}$$

$$Carbon = R * Chla / 1000. \tag{6}$$

where

R = ratio of carbon to chlorophyll-a, which is always greater than 20.0

I = total global incident radiation (langley/day)

min = minutes of daylight for that day of year

T = water temperature (oC)

Carbon = algal carbon (ug/l)

Chla = chlorophyll-a concentration corrected for phaeophytin (ug/l)

Equations (4-6) are used in the model to compute the carbon parameter after I, T, and Chla, are randomly selected. Minutes of daylight is determined by the day of year which is specified for each carbon value to be generated. The Chlorophyll distribution was developed from data obtained from the EPA and the MWCOG (Table 2). Data from MWCOG are corrected for phaeophytin. EPA data are not corrected, but are used because there is so little data available. The Chla distribution is plotted in Figure 20. No attempt was made to test separate seasonal distributions of Chla because of the small size of the data set.

Nitrogen Fractions: Nitrogen is most commonly analyzed and reported in the forms of total Kjeldahl nitrogen (TKN), ammonia (NH3), and nitrite plus nitrate (NOX). TON was calculated by subtracting NH3 from TKN. Nitrogen distributions are plotted in Figures 17-19.

Phosphorus Fractions: PEM requires that phosphorus be input as dissolved organic (DOP), dissolved inorganic (DIP), particulate organic (POP), and particulate inorganic (PIP), fractions. However, phosphorus is usually not analyzed into these fractions, and different agencies record their phosphorus data in different ways. To develop distributions for the required PEM parameters, the available phosphorus data were divided into DIP, DOP, PIP, and POP, as described below.

Data reported as mg PO4/l were converted to mg P/l. DIP was set equal to orthophosphorus (OP). For sets of daily phosphorus observations which included total phosphorus (TP), the fractions were determined differently depending on the additional phosphorus forms reported and when the measurements were made. When total soluble phosphorus (TSP) and OP values were present, then:

DIP = OP
DOP = TSP - DIP
PIP = f * (TP - TSP)
POP = 1 - PIP

where f partitions the particulate phosphorus into inorganic and organic components. f was obtained from the PEM Calibration Report (p. 90) and equals .13 for observations in the 1960's and .1 for 1970 until present.

When only OP was present in addition to TP, then:

1960's: DIP = OP
DOP = .222 * (TP - DIP)
PIP = .113 * (TP - DIP)
POP = .665 * (TP - DIP)

1970's, 1980's: DIP = OP
DOP = .091 * (TP - DIP)
PIP = .091 * (TP - DIP)
POP = .818 * (TP - DIP)

where these fractions are also obtained from the PEM Calibration Report.

When TP alone was reported, then:

1960's: DIP = .391 * TP
DOP = .135 * TP
PIP = .069 * TP
POP = .405 * TP

1970's, 1980's: DIP = .175 * TP
DOP = .075 * TP
PIP = .075 * TP
POP = .675 * TP

For the Chain Bridge data supplied by MWCOG, used in PEM calibration and verification, TP and inorganic phosphorus (IP) were recorded. In this case DIP = IP and the fractions shown above when TP and OP are present were used. Lastly, for the regional monitoring data obtained from MWCOG, TP, TSP, and OP were reported. The first set of fractions was used to transform and complete this data.

There were no significant differences between summer and winter values for phosphorus fractions (Table 4) therefore annual distributions are used in the model. The distributions are plotted in Figures 21-24, with summary statistics in Table 5.

Dissolved Oxygen (DO): DO data have been collected by various agencies at Chain Bridge. The data we obtained were collected at different times by the USGS, EPA, MWCOG, and the DCDSE (see Table 2). The data are from grab samples collected at approximately weekly to biweekly intervals, with all years from 1965 through 1985 represented. As shown in Table 4, Winter DO values are lower than Summer values, which is expected since the DO saturation level is inversely related to temperature. However, DO at Chain Bridge is rarely low, with only 5 values out of 427 in the data set below 6.0 mg/l. Therefore this version of the Monte Carlo program has only one DO distribution for the entire year.

DISCUSSION

The program described in this report is a draft version of the program that will ultimately be used with the PEM model to develop Chlorophyll distributions for various pollution control options. This draft is intended to demonstrate and explain the Monte Carlo methodology for generating inputs to PEM. A final version will be written following discussions with MWCOG to refine the definitions of some of the input parameters in ways most appropriate for the PEM model, and with the addition of additional data as they become available.

The approach used in the program to generate parameter values, random sampling from the set of observed data, depends on having the largest possible set of observed data. An underlying premise is that the observed data are a representative sample of the "true" distribution of these parameters. For data sets with few observations, or observations collected in only a few years, there is greater uncertainty that the data are representative. In addition, the ability to estimate the frequency of occurence of extreme events is limited by the size of the observed data set, as shown by Equation 3. However, the only alternative to this approach is to assume that parameter values can be estimated from some function (Normal distribution, etc.). As can be seen by inspection of the data plotted in the Figures, the distributions tend not to be smooth, they vary among the parameters, and any functional approximation of the empirical distributions is most likely to be violated at the extreme values which are the regions of most interest. Therefore, the risks of relying on small data sets seem preferable to making assumptions about the functional nature of the distributions. The best way to assure the representativeness of the sample data sets is to increase their size. During preparation of the next version of this program, every effort will be made to add additional data.

This version of the program assumes that individual parameters are independent and that there is no serial correlation. For at least some of parameters this is known to be not true. For example, both flow and temperature have a significant one day serial correlation. Serial correlation complicates the estimate of the frequency of occurence of given values by making it dependent on past history as well as on the total distribution of the parameter. Several techiniques will be evaluated to limit the potential influence of serial correlation on flow and temperature values generated by the next version of the program. The most likely approach will be to redefine these parameters to be estimates of flow or temperature for a period of more than a day; for example weekly or monthly mean temperature. Cross correlation of parameter values will be accounted for in the next version of the program by: a) computing the covariance of parameters in the observed data sets, and b) where the covariance is significant between parameters, conditioning the random selection of one parameter value by the covariance and the value selected for the other parameter.

References

1. Algal Bloom Expert Panel (1985), The 1983 Algal Bloom in the Potomac Estuary, Potomac Strategy State/EPA Management Committee.

2. Potomac Strategy State/EPA Committee (1985), Potomac Strategy Work Program.

3. R.V. Thomann, J.J. Fitzpatrick (1982), Calibration and Verification of a Mathematical Model of the Eutophication of the Potomac Estuary, Hydroqual, Inc.

4. U.S. Geological Survey (1930-1986), Daily Streamflow for Little Falls, MD, U.S.G.S. gage 01646500, obtained in machine readable form from the U.S. EPA's STORET database.

5. Metropolitan Washington Council of Governments (1984), The Potomac Eutrophication Model (PEM): Results of a Post-Audit Run for 1983 Conditions, Metropolitan Washington Council of Governments.

6. H.C. Haywood (1986), An Analysis of the Distribution of Daily Flow and Water Temperature for the Upper Potomac Estuary, ICPRB Technical Report 86-14, Interstate Commission on the Potomac River Basin.

7. National Oceanic and Atmospheric Administration (1977-1980), NOAA Monthly Summary Daily Radiation Reports for Sterling, Va., (monthly).

8. S. Freudberg (1983), Data used for Calibration and Verification of PEM, Metropolitan Washington Council of Governments, Personal Communication.

9. U.S. EPA STORET Data Base, Chain Bridge Water Quality Data. Data originally collected by various agencies are stored on this data base and were obtained in machine readable form.

10. Wendy Chittendon (1985,1986), Metropolitan Washington Council of Governments Chain Bridge Water Quality Monitor Data, Personal Communication.

Table 1
PEM parameters generated by Monte Carlo Model

	Parameter	Abbrev.	Units
1	mean daily streamflow	Q	cfs
2	mean daily water temperature	T	C
3	dissolved oxygen	DO	mg/l
4	biochemical oxygen demand	BOD	mg/l
5	ammonia	NH3	mg N/l
6	total organic nitrogen	TON	mg N/l
7	nitrite + nitrate, NO2+NO3	NOX	mg N/l
8	algal carbon		ug C/l
9	dissolved organic phosphorus	DOP	mg P/l
10	particulate organic phos.	POP	mg P/l
11	dissolved inorganic phos.	DIP	mg P/l
12	particulate inorganic phos.	PIP	mg P/l
13	daily solar radiation	I	langley/day

Additional Input Distributions

	Parameter	Abbrev.	Units
14	Chlorophyll-a	CHLA	ug/l
15	Percent Clear Sky radiation	%CSR	%

Table 2
Sources of Data for PEM Chain Bridge Boundary Parameters.

Ref. (1)	Agency (2)	Number Obs.	Period of Record	Parameters in Record
9	DCDSE	226	7/69 - 1/76	DO,BOD,NH3,NO2,NO3,TKN,TP
9	EPA	18	1/74 -11/77	DO,BOD,NH3,TKN,NOX,TP,CHLA
8	EPA	36	8/65 - 3/67	DO,BOD,TP,IP,TKN,NH3,NO3
6	various	3104	2/63 - 6/86	T
9	MDDNR	60	9/69 - 6/81	DO,NH3,NO2,NO3,TKN
10	MWCOG	192	1/83 -12/85	DO,BOD,TP,TSP,OP,TKN,NH3,NOX,CHLA
9	USGS	558	1/73 - 5/86	DO,BOD,NH3,TKN,NOX,TP,OP
8	USGS	35	7/77 -10/79	DO,BOD,TP,IP,TKN,NH3,NO3,CHLA
4	USGS	20638	10/30 -10/86	Q
7	NOAA	588	1/76 - 6/80	I

NOTES

(1) Numbers refer to References cited on page 11.

(2) Agencies refer to the agency that collected the data.
Abbreviations are:

DCDSE	--	D.C. Department of Sanitary Engineering
MDDNR	--	State of Maryland, Department of Natural Resources
EPA	--	United States Environmental Protection Agency
MWCOG	--	Metropolitan Washington Council of Governments
USGS	--	United States Geological Survey
various	--	See Ref 6 for information on sources for temperature data

Table 3
Number of Parameter Observations by Month

Month	DO	BOD	NH3	TON	NOX	CHLA	Phos(1)	Q	T	I(2)
Jan	17	16	20	17	30	6	25	1730	260	67
Feb	14	13	29	24	61	2	28	1577	236	76
Mar	25	21	31	21	70	7	38	1760	236	77
Apr	27	21	28	26	66	8	34	1705	232	88
May	35	23	38	31	53	11	32	1762	258	59
Jun	42	27	42	34	67	16	34	1707	294	55
Jul	55	37	51	38	63	23	42	1761	268	38
Aug	63	50	68	51	80	23	54	1763	265	12
Sep	57	39	59	47	88	16	46	1705	268	30
Oct	37	23	38	32	54	4	40	1752	266	28
Nov	33	26	45	32	65	5	37	1680	255	27
Dec	22	15	25	33	35	3	22	1736	266	31
Total	427	311	474	376	732	124	432	20638	3104	588

(1) Identical for each phosphorus fraction.
(2) Incident radiation, I, data were used to develop the distribution of %CSR.

Table 4
Comparison of Parameter Values in Summer vs Winter

Parameter	Summer			Winter		
	n	mean	sd	n	mean	sd
DO (mg/l)	279	8.4	1.57	148	11.8	1.74
BOD (mg/l)	197	2.95	1.81	114	2.69	2.22
NH3 (mg N/l)	286	0.259	0.366	188	0.266	0.365
TON (mg N/l)	227	0.817	0.574	149	0.793	0.609
NOX (mg N/l)	417	0.764	0.471	315	1.24	0.524
DIP (mg P/l)	242	0.025	0.035	190	0.032	0.031
PIP (mg P/l)	242	0.006	0.006	190	0.009	0.022
DOP (mg P/l)	242	0.010	0.012	190	0.010	0.012
POP (mg P/l)	242	0.053	0.051	190	0.081	0.204
%CSR (%)	282	64.3	25.6	306	64.6	29.9

where summer = April through September
winter = October through March
n = number of observations
mean = mean of observations
sd = standard deviation

Table 5

Summary Statistics for the Chain Bridge Parameters (except Flow and Temperature)

Parameter	n	max	min	mean	sd
DO (mg/l)	427	15.5	0.1	9.6	2.3
BOD (mg/l)	311	15.0	0.0	2.9	2.0
NH3 (mg N/l)	474	2.49	0.0	0.26	0.37
TON (mg N/l)	376	3.9	0.0	0.81	0.59
NOX (mg N/l)	732	5.03	0.0	0.97	0.55
CHLA (ug/l)	124	142.0	0.5	23.1	25.8
DOP (mg P/l)	432	0.1	0.0	0.01	0.01
POP (mg P/l)	432	2.1	0.0	0.06	0.14
DIP (mg P/l)	432	0.44	0.0	0.028	0.033
PIP (mg P/l)	432	0.235	0.0	0.007	0.016
%CSR (%)	588	104.0	4.0	64.4	27.9

where

n = number of observations
min = minimum of observations
max = maximum of observations
mean= mean of observations
sd = standard deviation

Table 6
Summary statistics for Little Falls Flow (Unadjusted) by Month, for Period of Record 1930-1986

Flow, cfs

Month	n	max	min	mean	med
January	1730	118000	434	12887	8820
February	1577	209000	763	17211	12300
March	1760	426000	2600	22585	16100
April	1705	310000	3350	19791	14000
May	1762	163000	2300	13876	10100
June	1707	334000	975	8883	6050
July	1761	82300	309	4845	3680
August	1763	207000	255	4828	2870
September	1705	186000	121	4244	2290
October	1752	407000	488	6216	2420
November	1680	293000	467	7165	3640
December	1736	126000	632	10536	6890

Log(10) Flow

Month	n	max	min	mean	sd
January	1730	5.072	2.637	3.951	.371
February	1577	5.320	2.883	4.085	.358
March	1760	5.629	3.415	4.235	.308
April	1705	5.491	3.525	4.182	.293
May	1762	5.212	3.362	4.030	.293
June	1707	5.524	2.989	3.808	.301
July	1761	4.915	2.490	3.567	.304
August	1763	5.316	2.407	3.486	.371
September	1705	5.270	2.083	3.396	.376
October	1752	5.610	2.688	3.473	.423
November	1680	5.467	2.669	3.636	.399
December	1736	5.100	2.801	3.832	.401

where

n = number of observations
min = minimum flow
max = maximum flow
mean= mean flow
med = median flow
sd = standard deviation

Table 7
Summary statistics for Water Temperature above Wilson Bridge by Month

		Temperature, C			
Month	n	max	min	mean	sd
January	260	0.0	9.2	3.03	1.90
February	236	0.0	8.8	3.64	2.26
March	236	1.0	14.4	7.36	2.24
April	232	6.1	22.0	13.27	3.25
May	258	12.1	24.9	19.74	2.63
June	294	16.2	30.3	24.54	2.39
July	268	20.7	33.0	27.51	1.64
August	265	23.4	31.0	27.18	1.65
September	268	16.9	29.4	24.07	2.67
October	266	10.9	22.9	17.32	2.93
November	255	4.6	20.4	11.18	2.53
December	266	0.6	13.4	5.61	2.60

where

n = number of observations
min = minimum temperature
max = maximum temperature
mean= mean temperature
sd = standard deviation

Table 8
Clear Sky Total Daily Solar Radiation[1] for Washington, DC,

	Langleys/Day											
Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	232	311	437	586	697	754	745	687	587	462	331	244
2	233	314	442	590	699	754	744	684	584	457	327	242
3	235	318	446	594	702	755	743	681	580	453	323	240
4	236	322	452	599	705	756	741	679	575	448	320	239
5	238	326	457	603	707	756	740	676	571	444	316	237
6	239	331	462	608	709	756	738	673	568	440	313	236
7	241	334	467	612	712	756	736	670	563	436	309	235
8	243	339	472	616	714	756	734	667	559	431	306	234
9	245	343	477	620	717	756	732	664	555	427	302	233
10	247	347	482	624	719	756	730	661	551	423	299	232
11	249	352	487	628	721	756	728	658	547	418	296	231
12	251	356	493	632	723	756	726	655	542	414	293	230
13	253	360	498	636	726	756	724	651	538	409	289	229
14	255	365	503	640	728	756	723	648	534	405	286	229
15	258	370	508	644	730	756	721	645	530	401	283	228
16	260	374	513	648	732	755	719	642	525	396	280	228
17	263	379	517	651	733	755	717	639	521	392	277	227
18	265	383	522	655	735	755	715	635	517	388	274	227
19	268	388	527	658	737	754	714	633	513	383	272	227
20	271	393	532	662	739	754	712	629	509	379	269	226
21	273	398	536	665	740	754	710	626	505	375	266	226
22	276	402	540	669	742	753	709	623	501	371	264	226
23	279	407	545	672	743	753	707	619	496	366	261	226
24	282	412	549	676	745	752	705	616	492	362	259	226
25	285	417	554	679	746	751	703	613	488	358	256	227
26	289	422	559	682	747	750	701	609	483	354	254	227
27	292	427	563	685	748	750	699	606	479	350	252	228
28	296	432	568	688	749	749	697	602	475	346	250	228
29	299	0	572	691	751	748	695	599	471	342	248	229
30	303	0	577	694	752	747	692	595	466	338	246	230
31	306	0	581	0	753	0	689	591	0	334	0	231

---

1) Computed from a program supplied by Stuart Freudberg (Personal Communication), Metropolitan Washington Council of Governments, 20 December 1984.

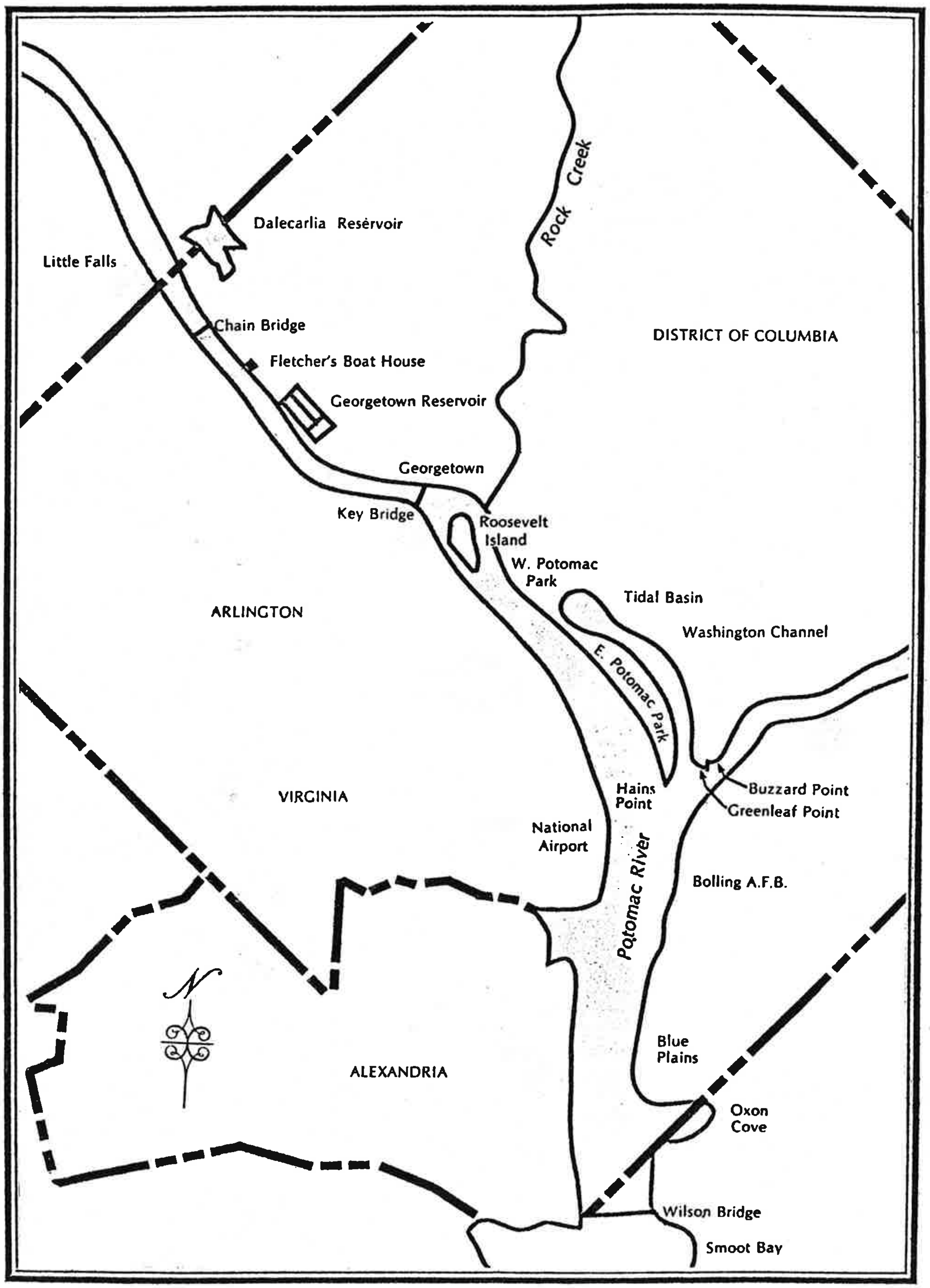


FIGURE 1
Location of Potomac Estuary Boundary at Chain Bridge

FIGURE 2

Log of Little Falls Flow (unadjusted) by Month
Mean, Standard Deviation, Range, and Number Daily Values

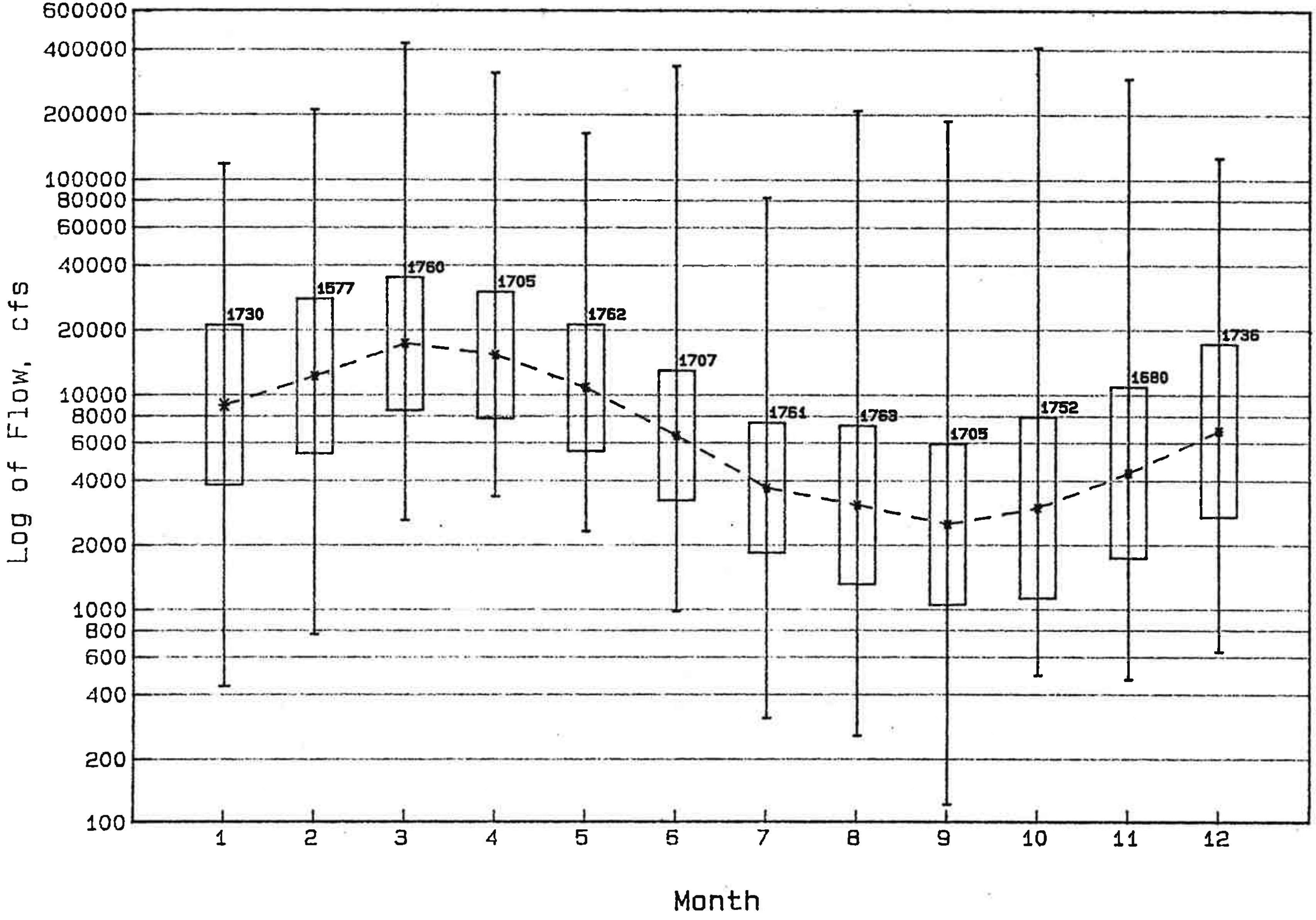


FIGURE 3

Distribution of Little Falls Flow in January
Number of Observations = 1730

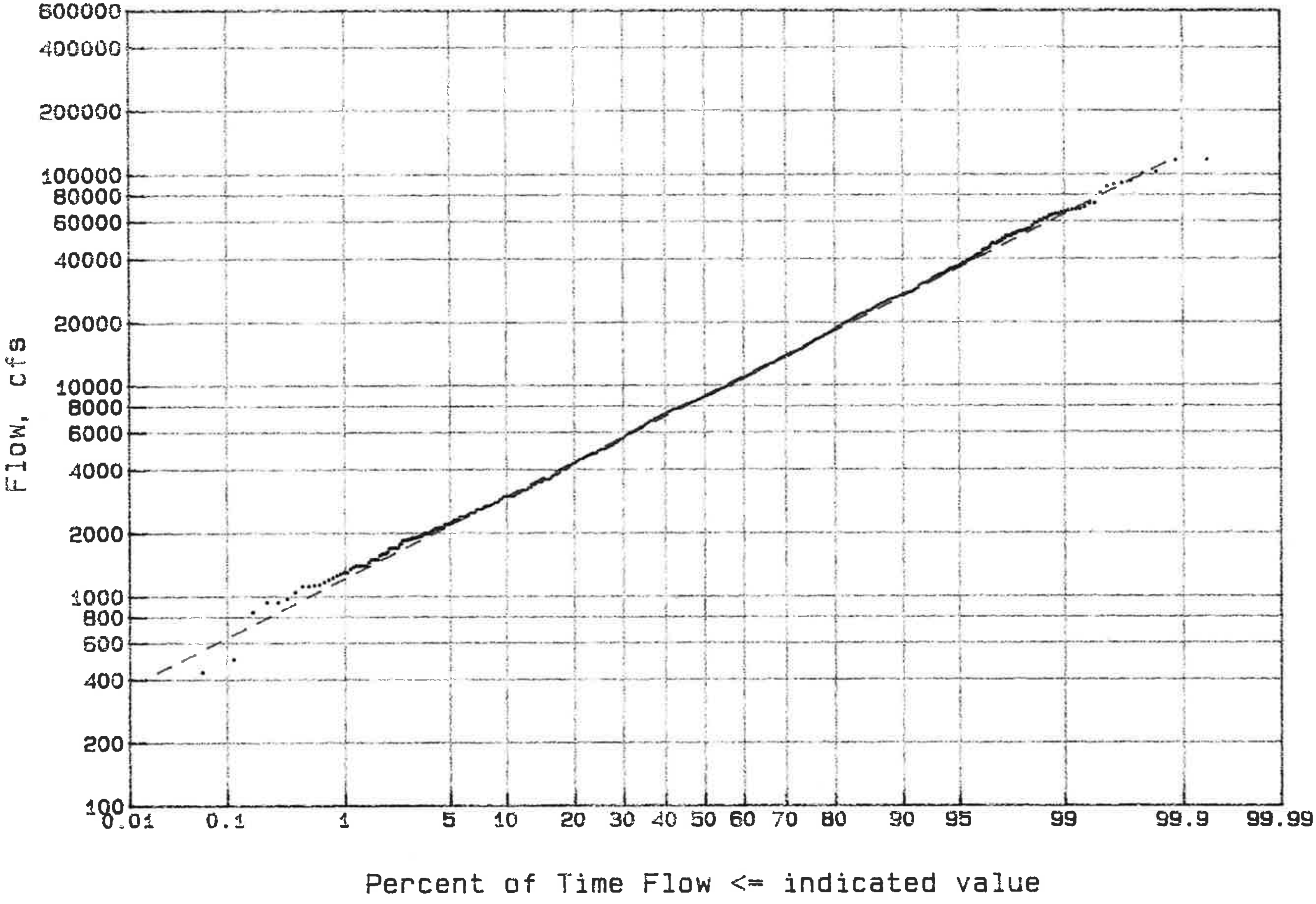


FIGURE 4

Distribution of Little Falls Flow in February
Number of Observations = 1577

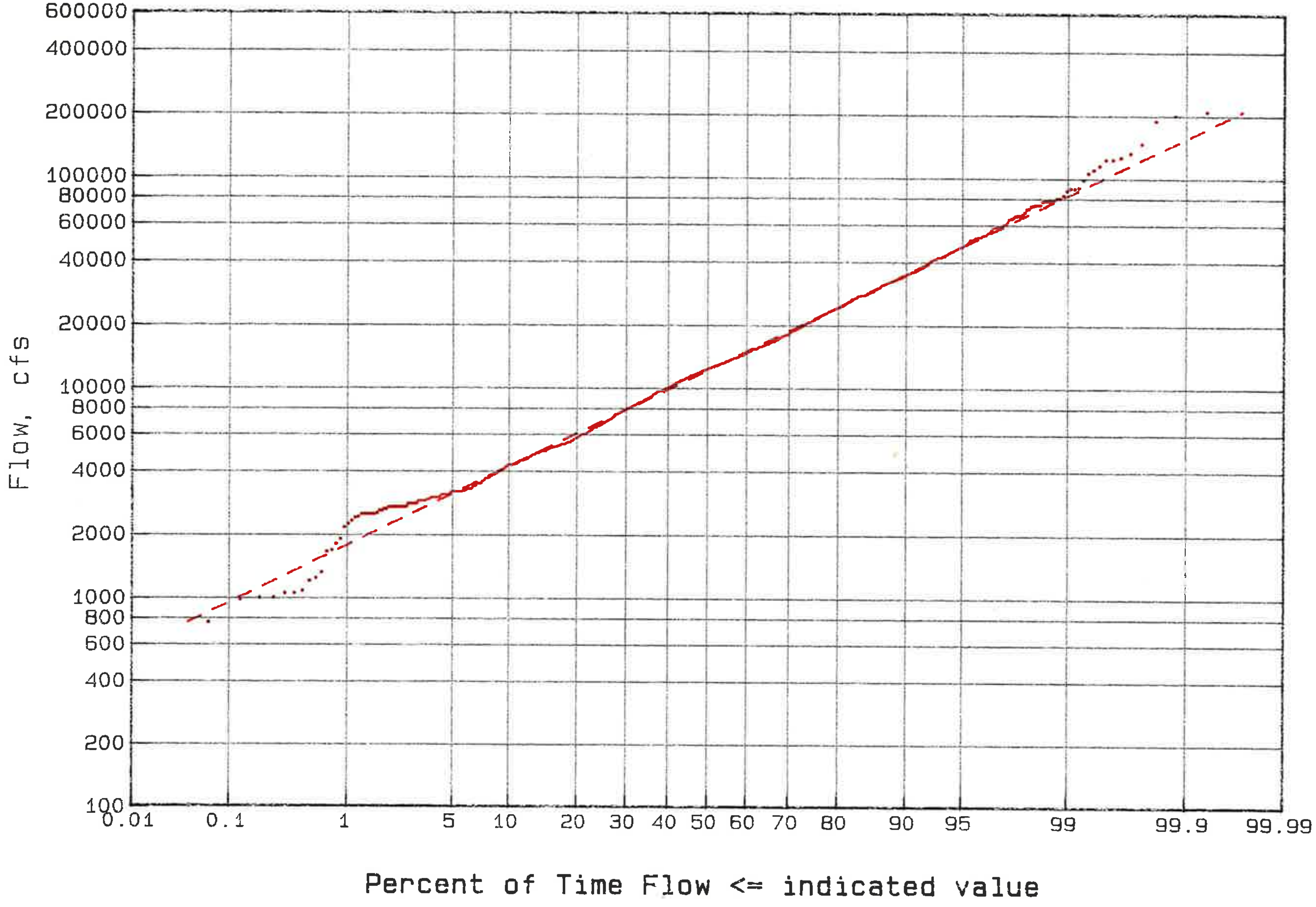


FIGURE 5

Distribution of Little Falls Flow in March
Number of Observations = 1760

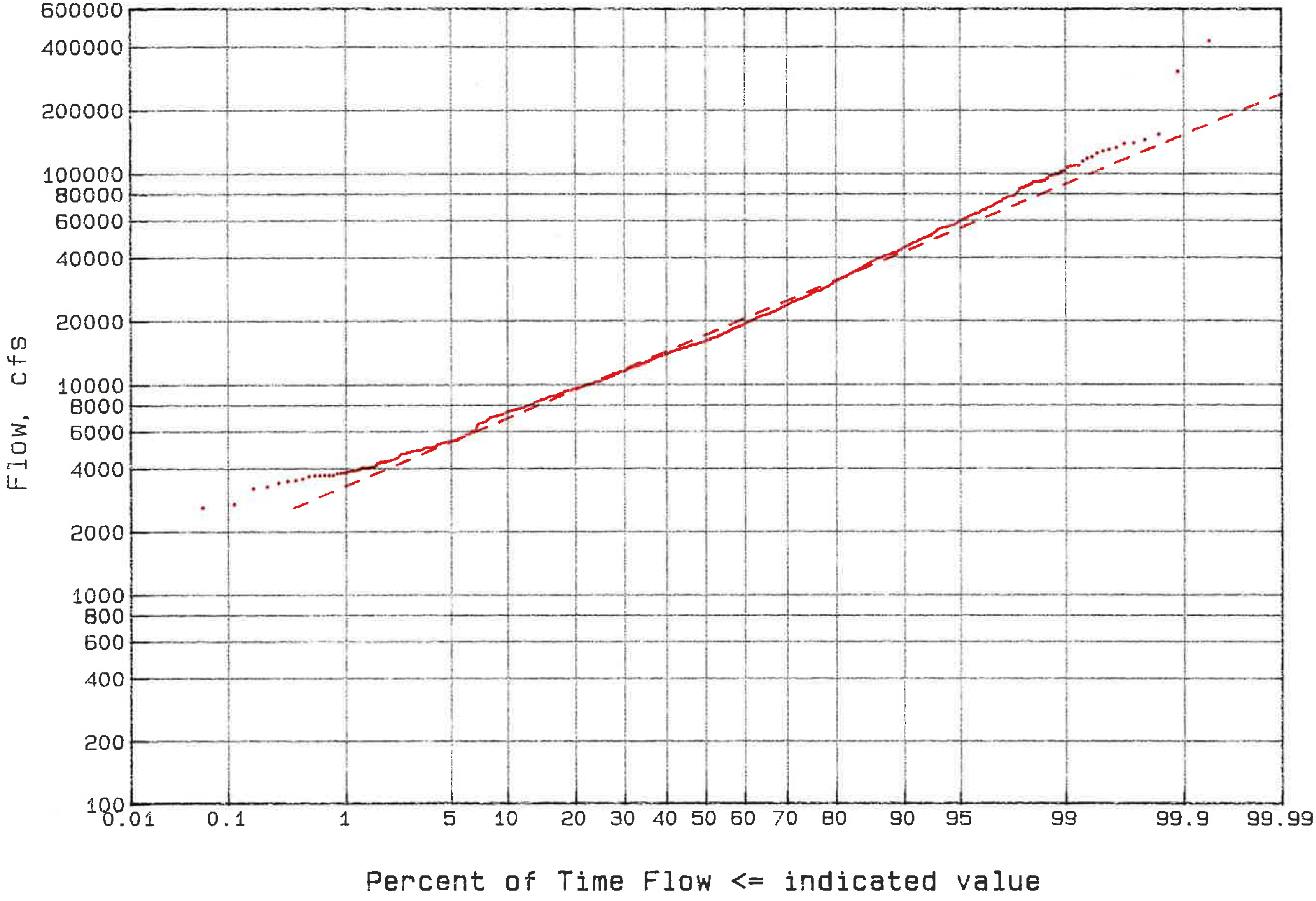


FIGURE 6

Distribution of Little Falls Flow in April
Number of Observations = 1705

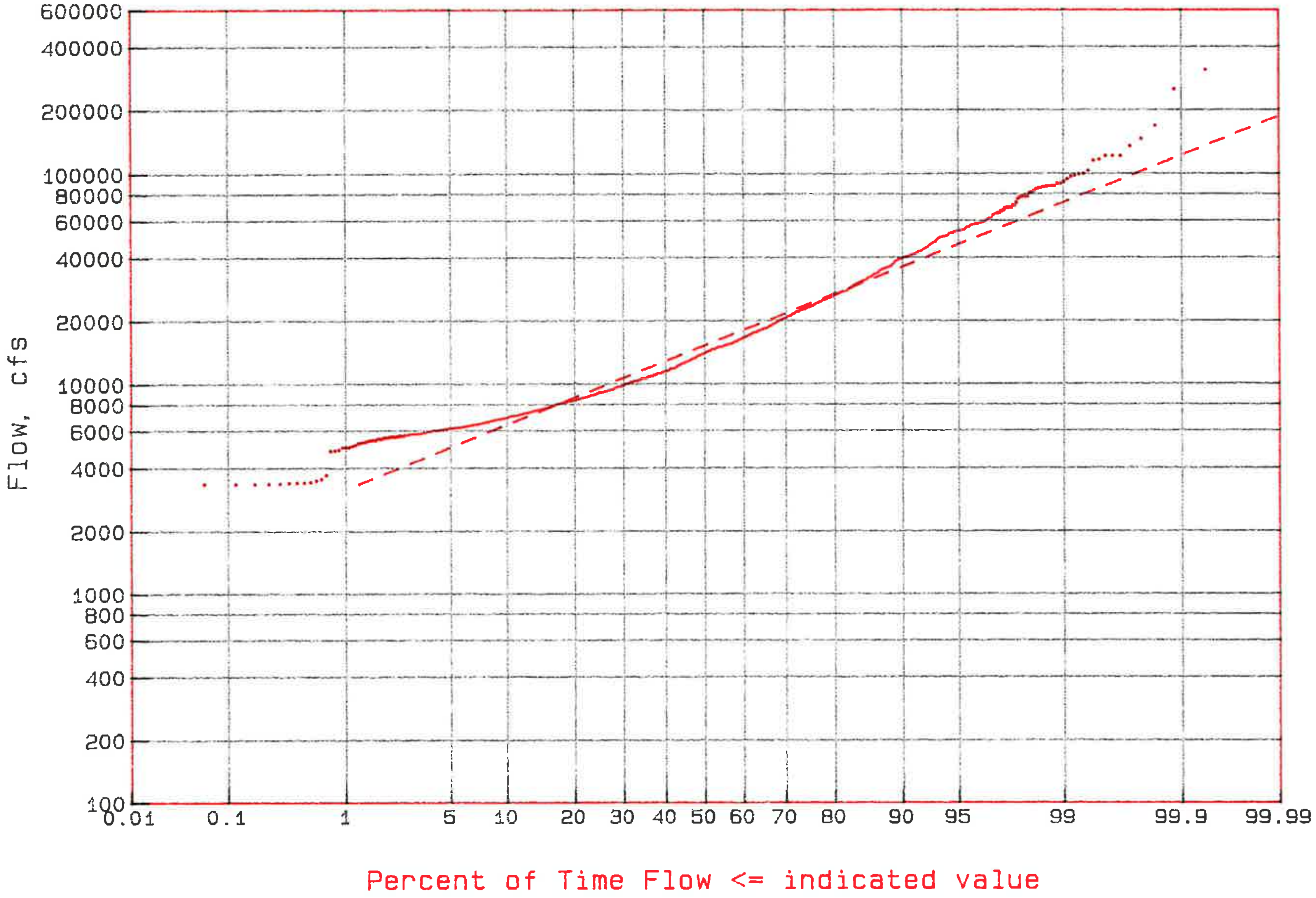


FIGURE 7

Distribution of Little Falls Flow in May
Number of Observations = 1762

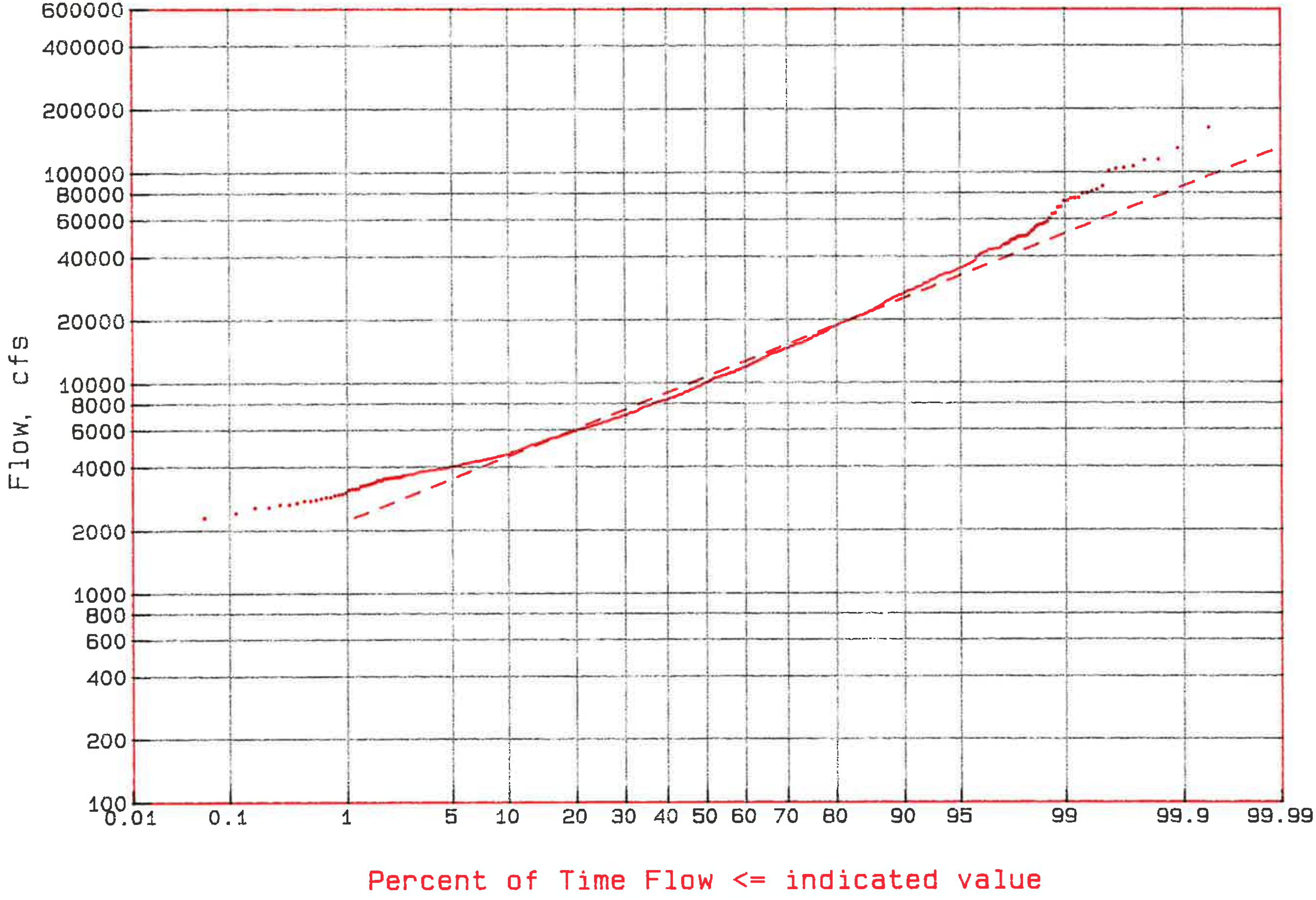


FIGURE 8

Distribution of Little Falls Flow in June
Number of Observations = 1707

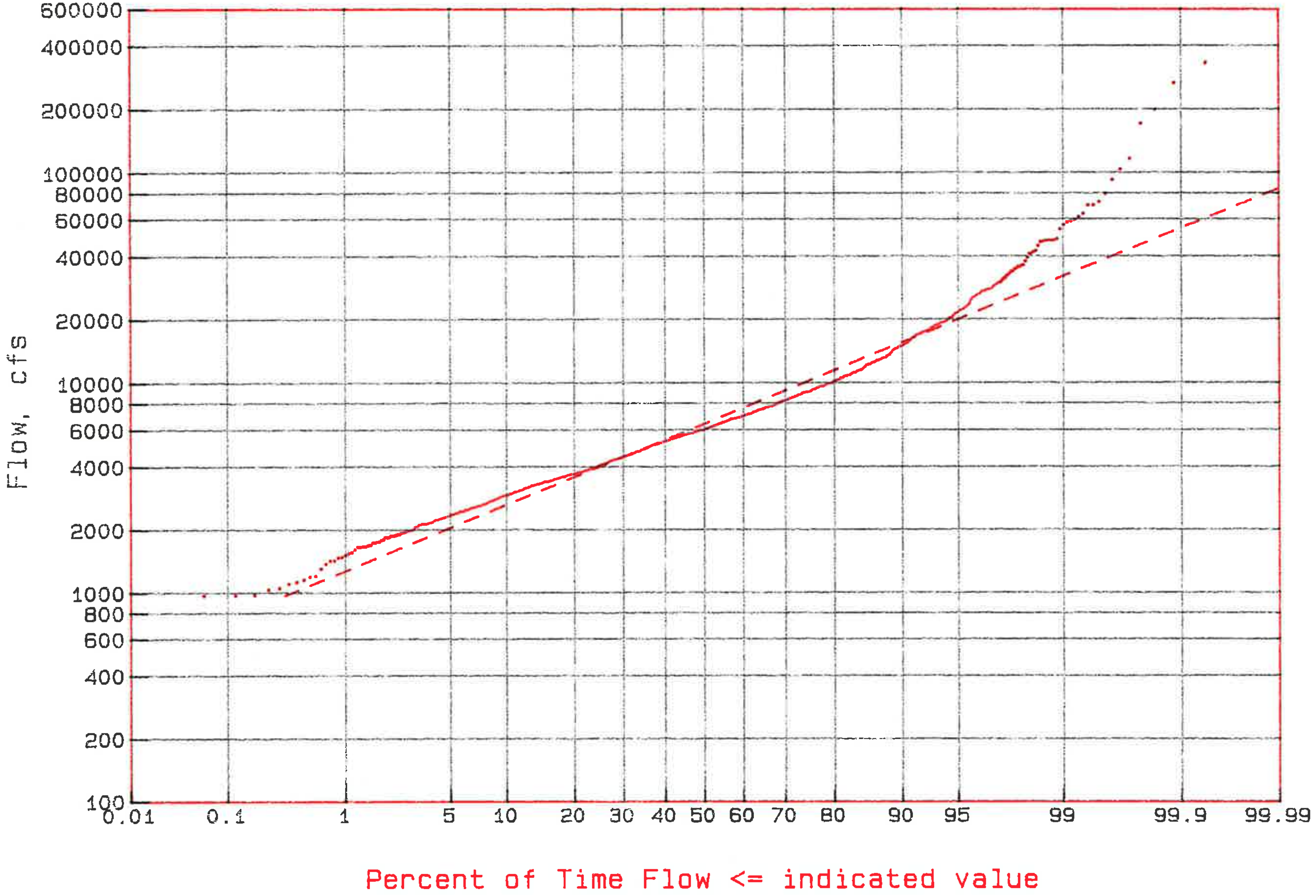


FIGURE 9

Distribution of Little Falls Flow in July
Number of Observations = 1751

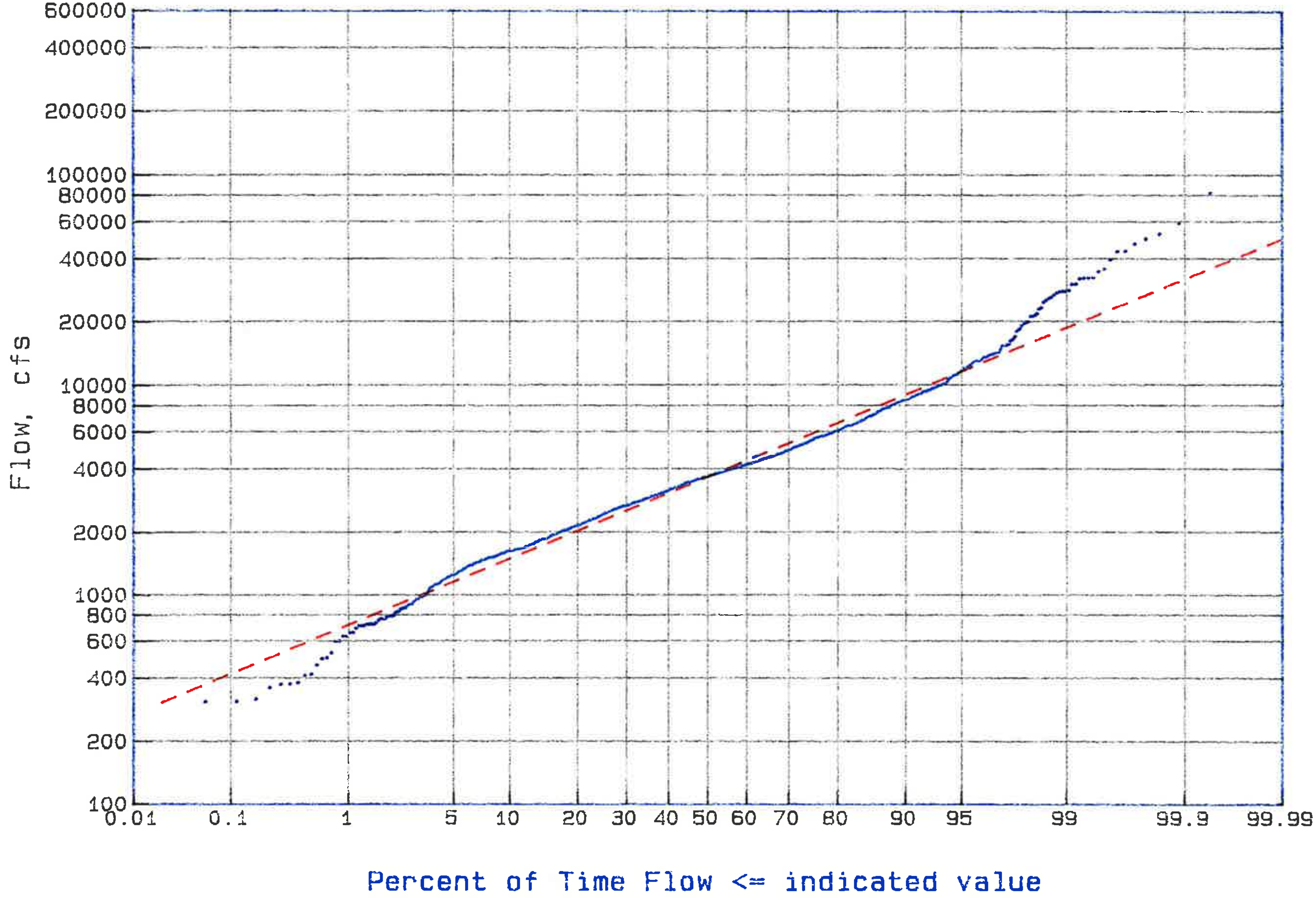


FIGURE 10

Distribution of Little Falls Flow in August
Number of Observations = 1763

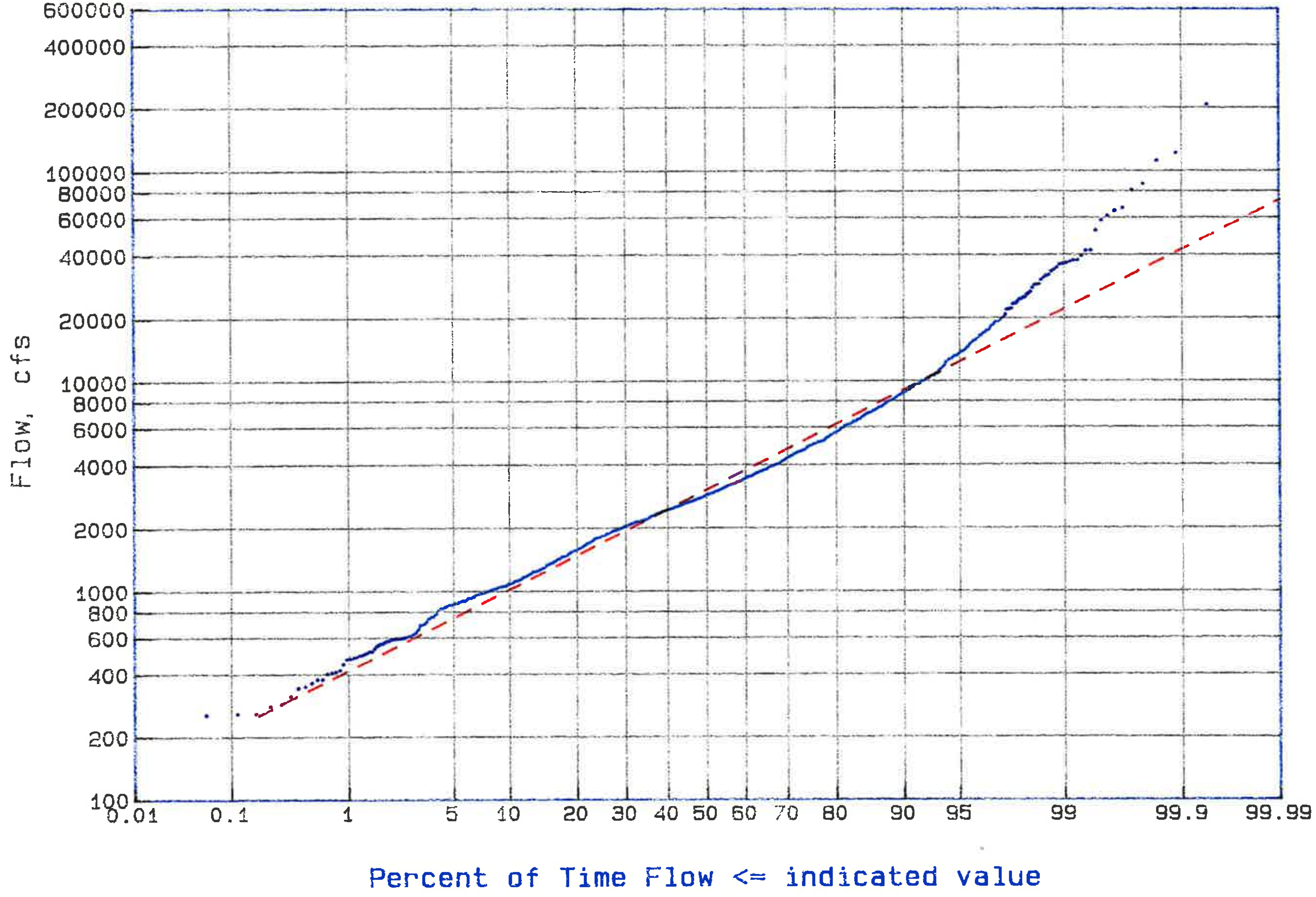


FIGURE 11

Distribution of Little Falls Flow in September
Number of Observations = 1705

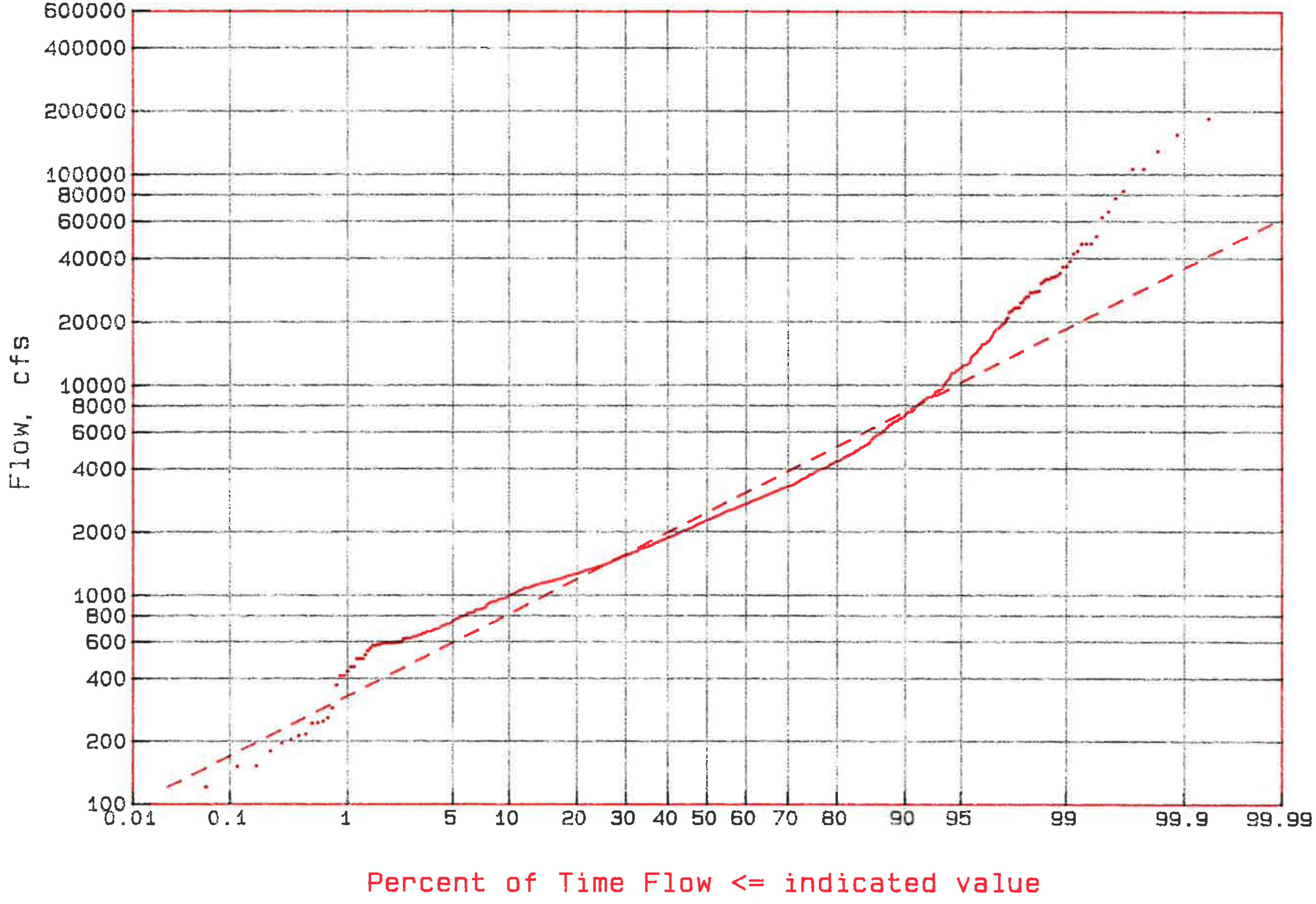


FIGURE 12

Distribution of Little Falls Flow in October
Number of Observations = 1752

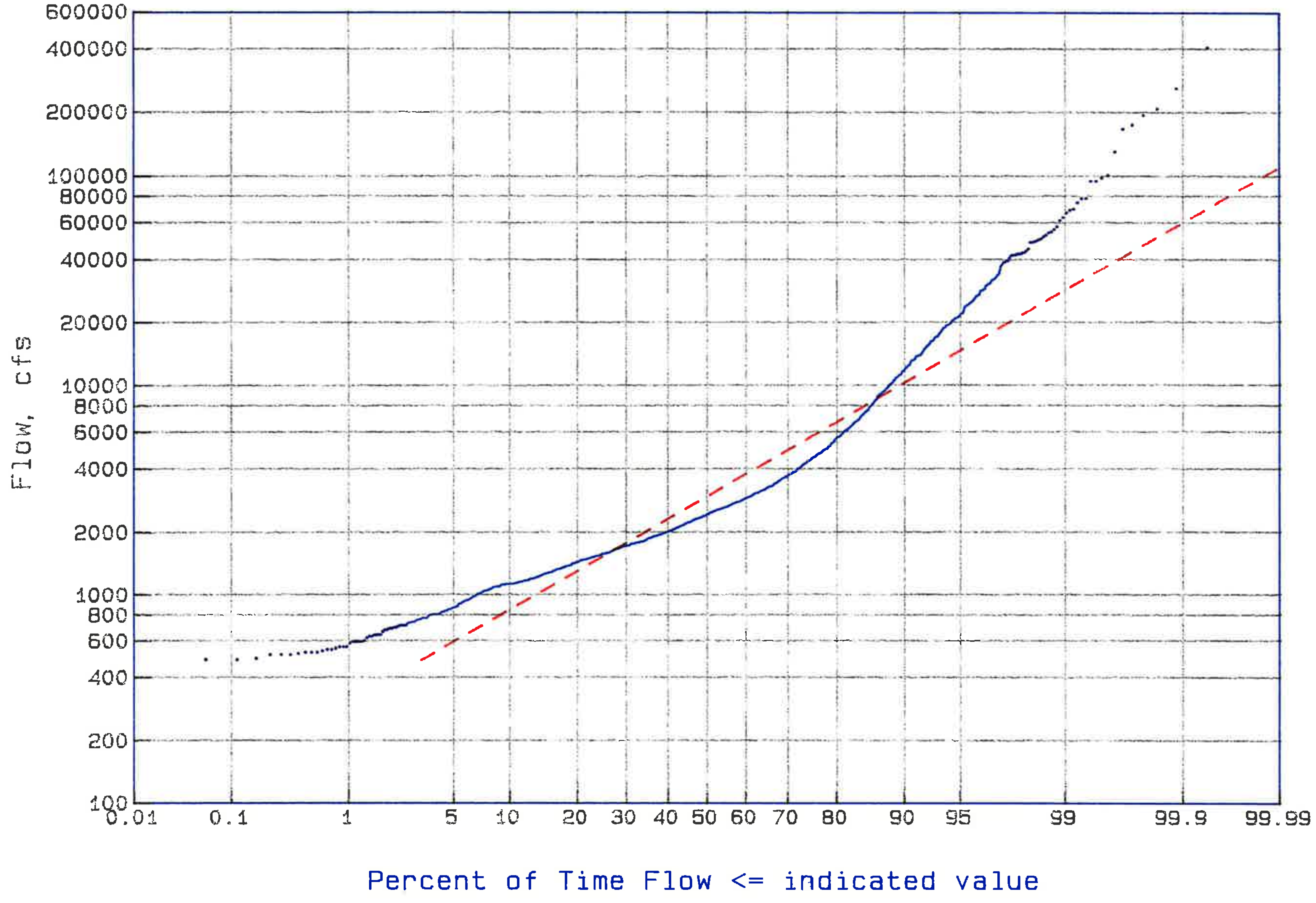


FIGURE 13

Distribution of Little Falls Flow in November
Number of Observations = 1680

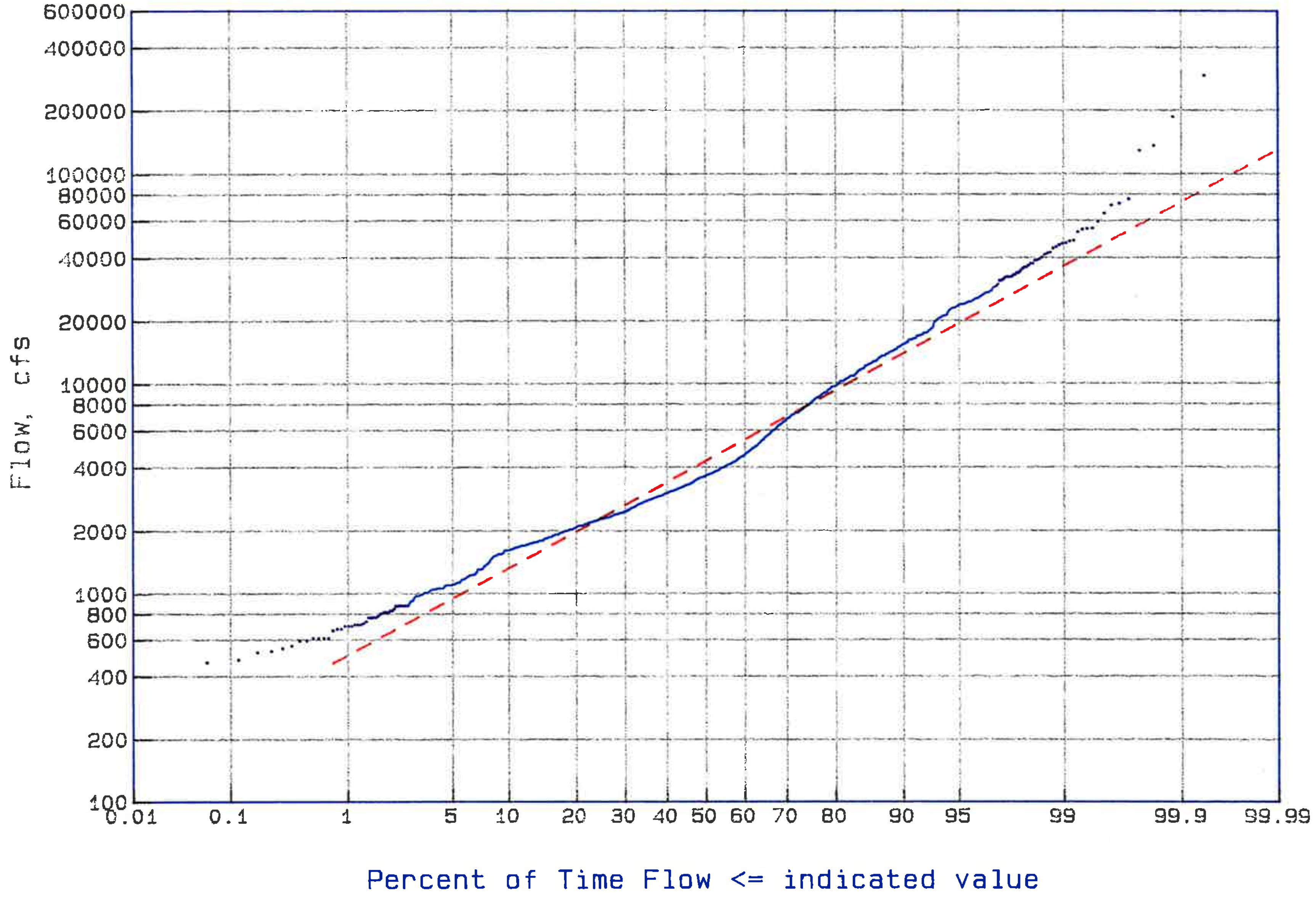


FIGURE 14

Distribution of Little Falls Flow in December
Number of Observations = 1736

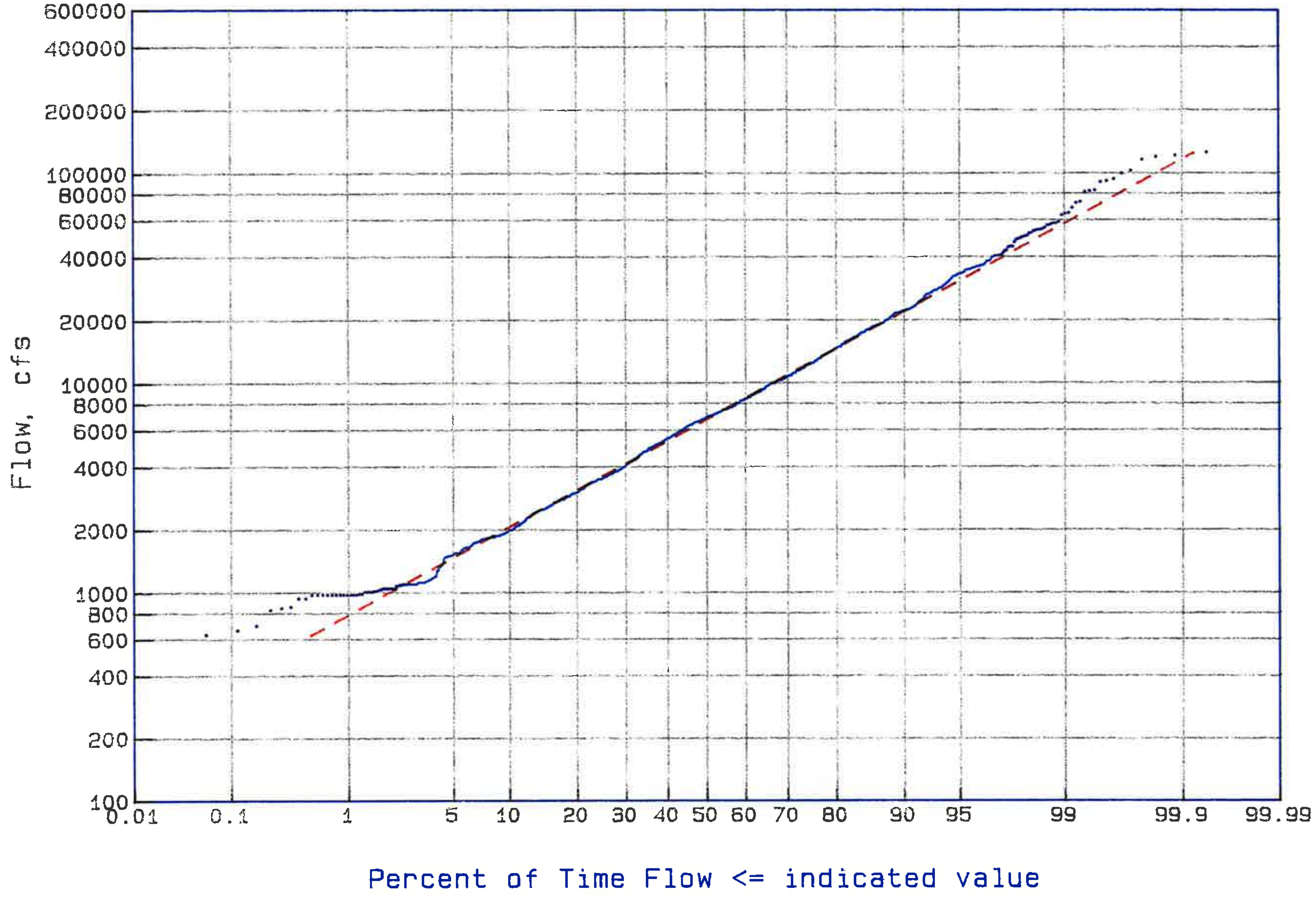


FIGURE 15

Distribution of Dissolved Oxygen at Chain Bridge
Number of Observations = 427

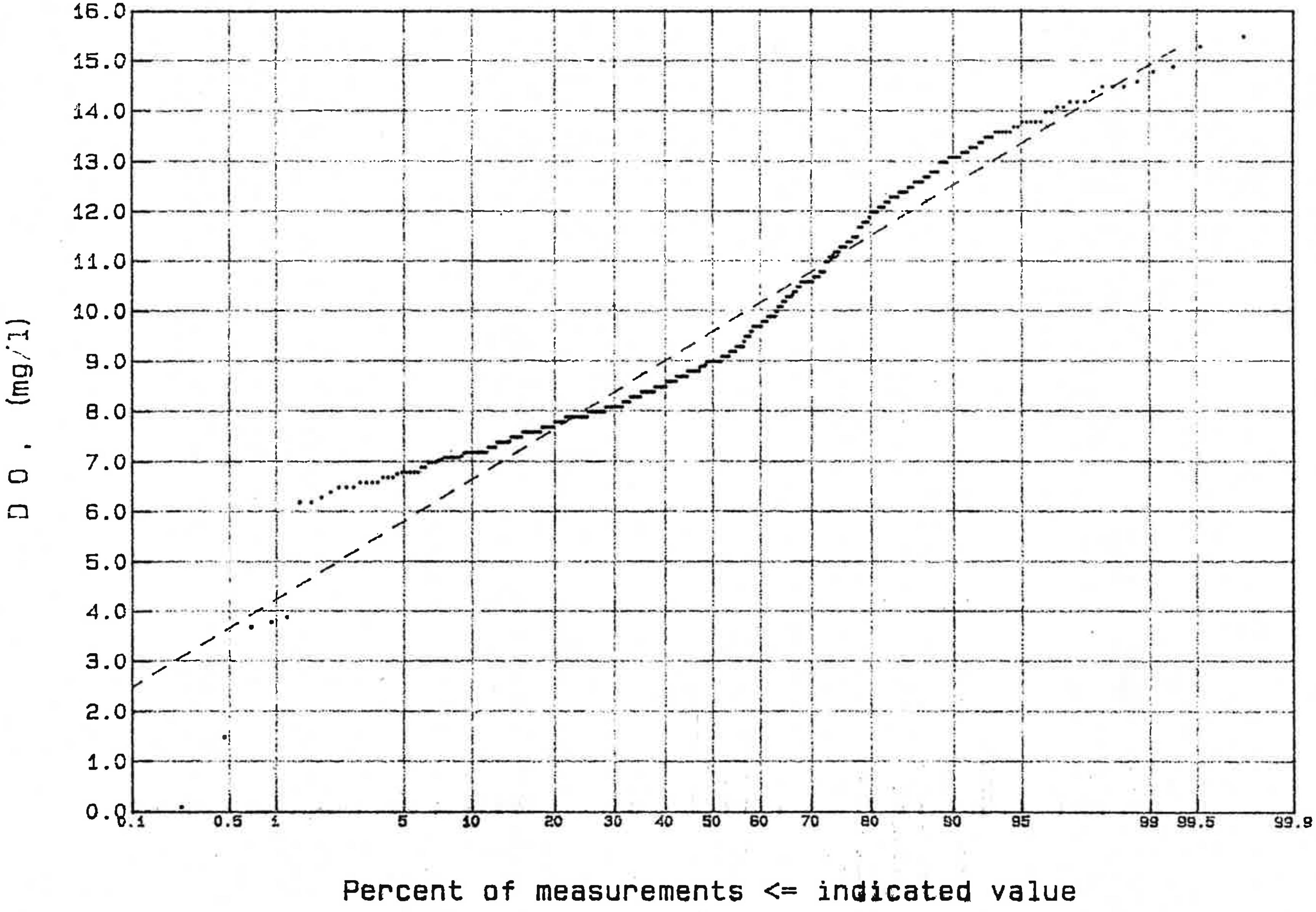


FIGURE 16
Distribution of Biological Oxygen Demand (BOD)
Number of Observations = 311

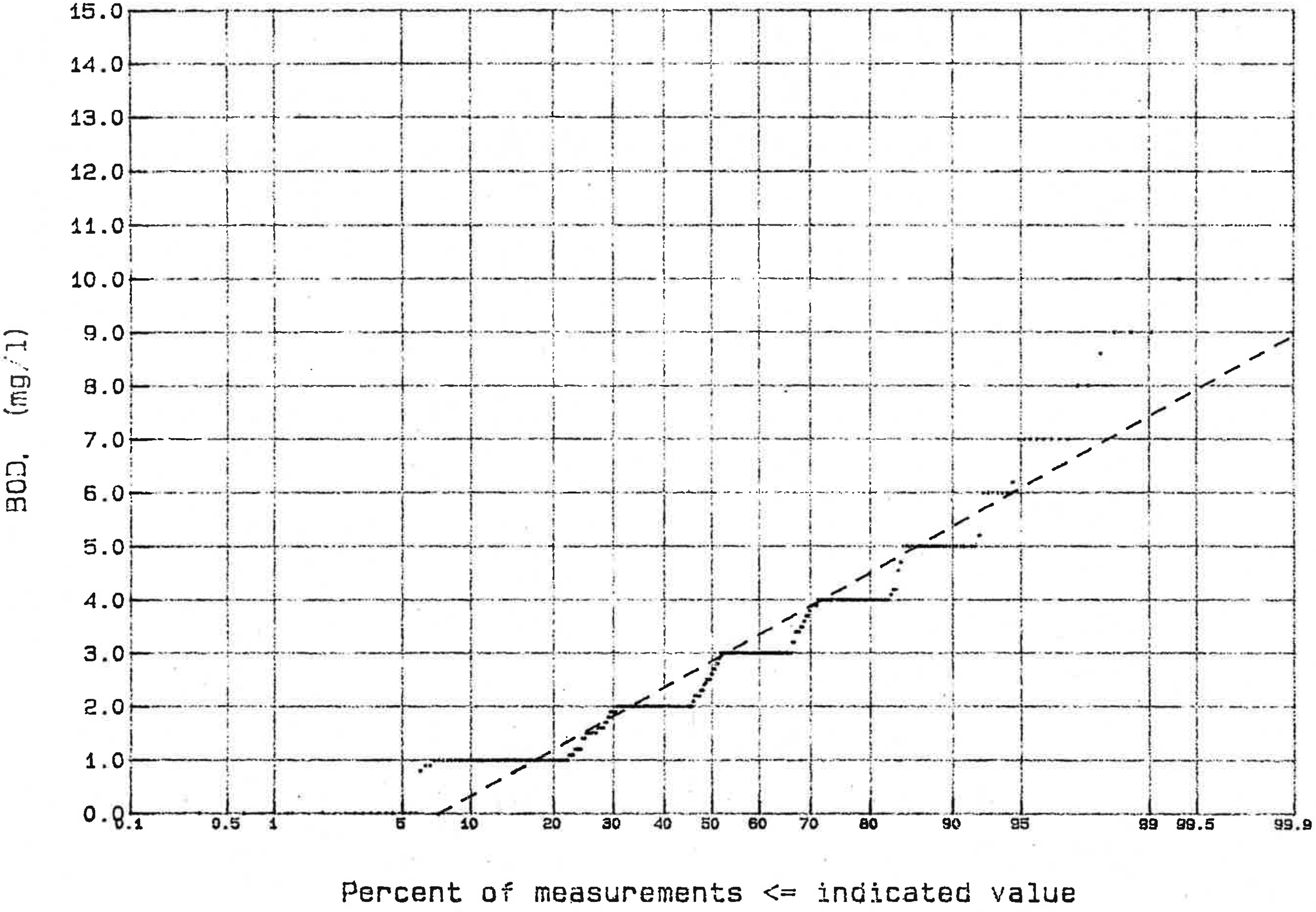


FIGURE 17

Distribution of Ammonia (NH3) at Chain Bridge
Number of Observations = 474

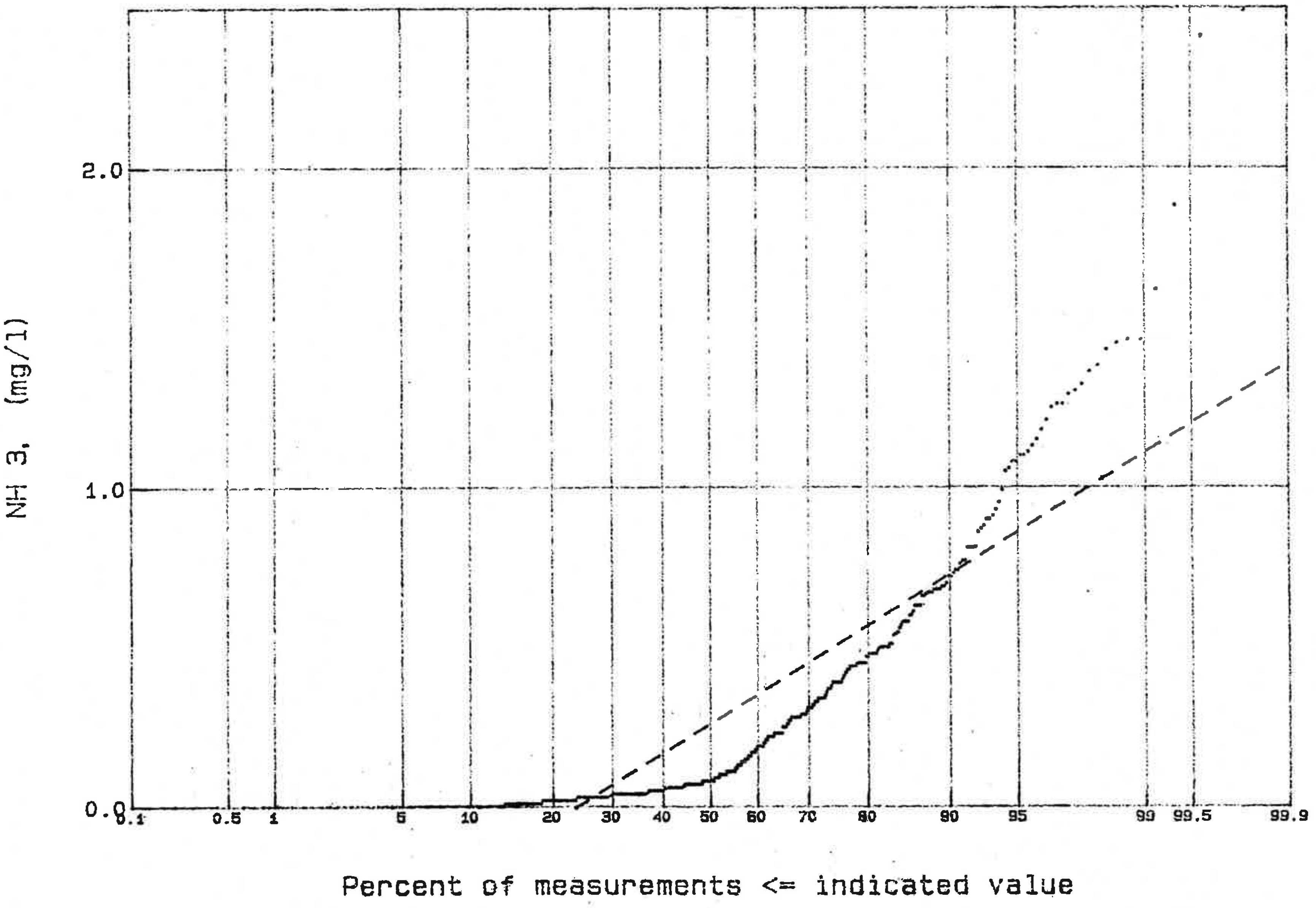


FIGURE 18

Distribution of Total Organic Nitrogen (TON) at Chain Bridge
Number of Observations = 376

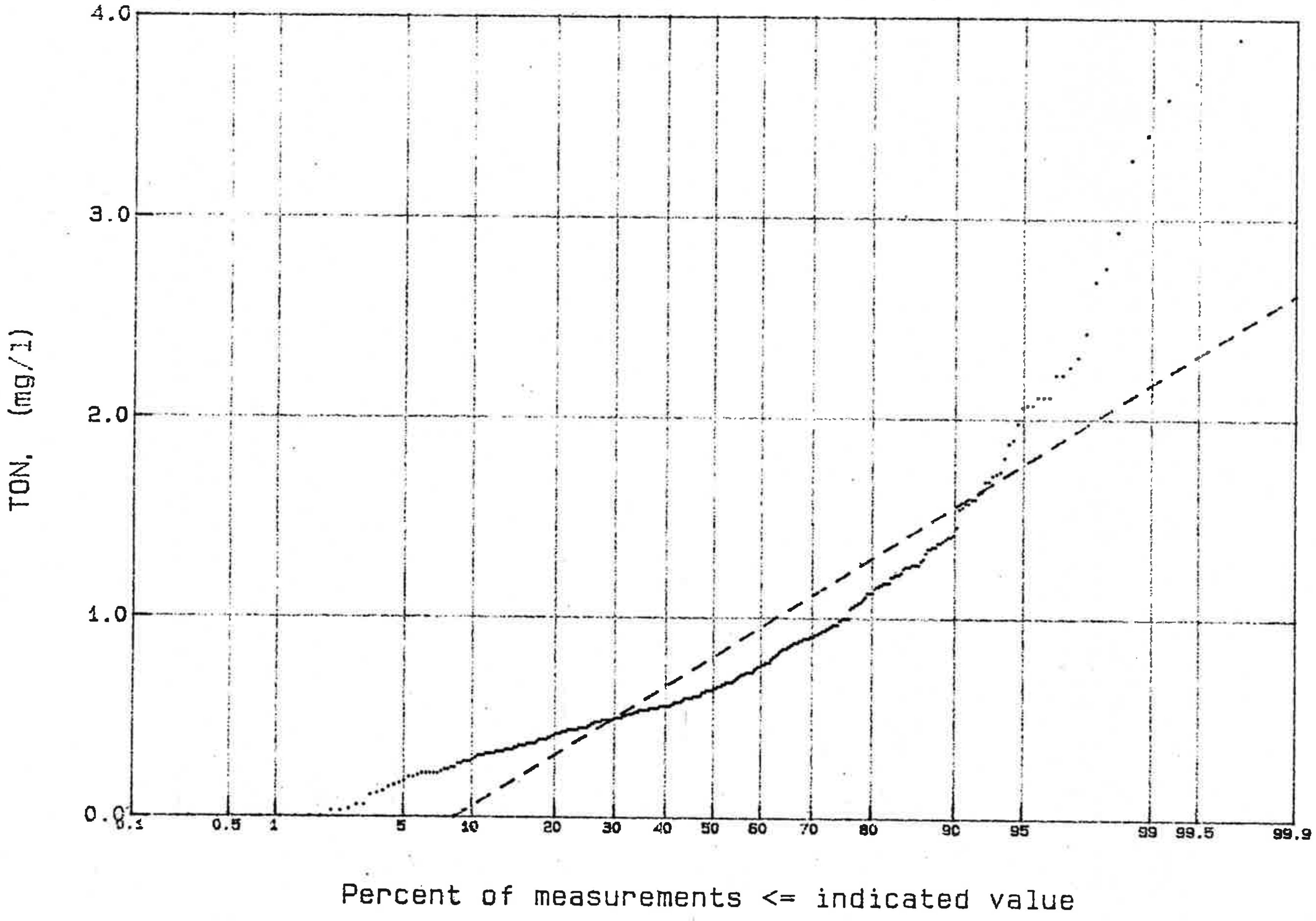


FIGURE 19

Distribution of Nitrate and Nitrite (NO3/NO2) at Chain Bridge
Number of Observations = 732

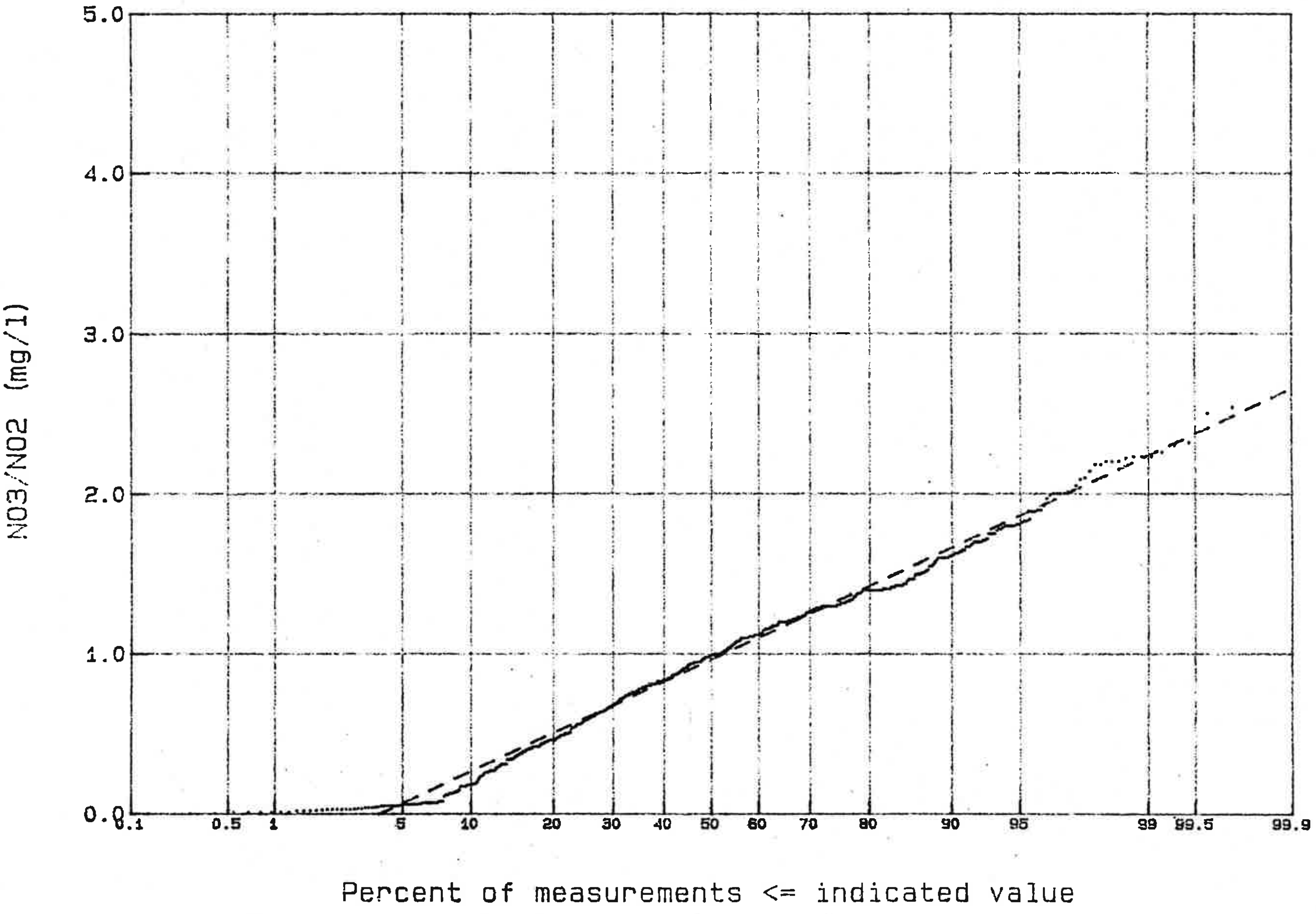


FIGURE 20

Distribution of Chlorophyll-a at Chain Bridge
Number of Observations = 124

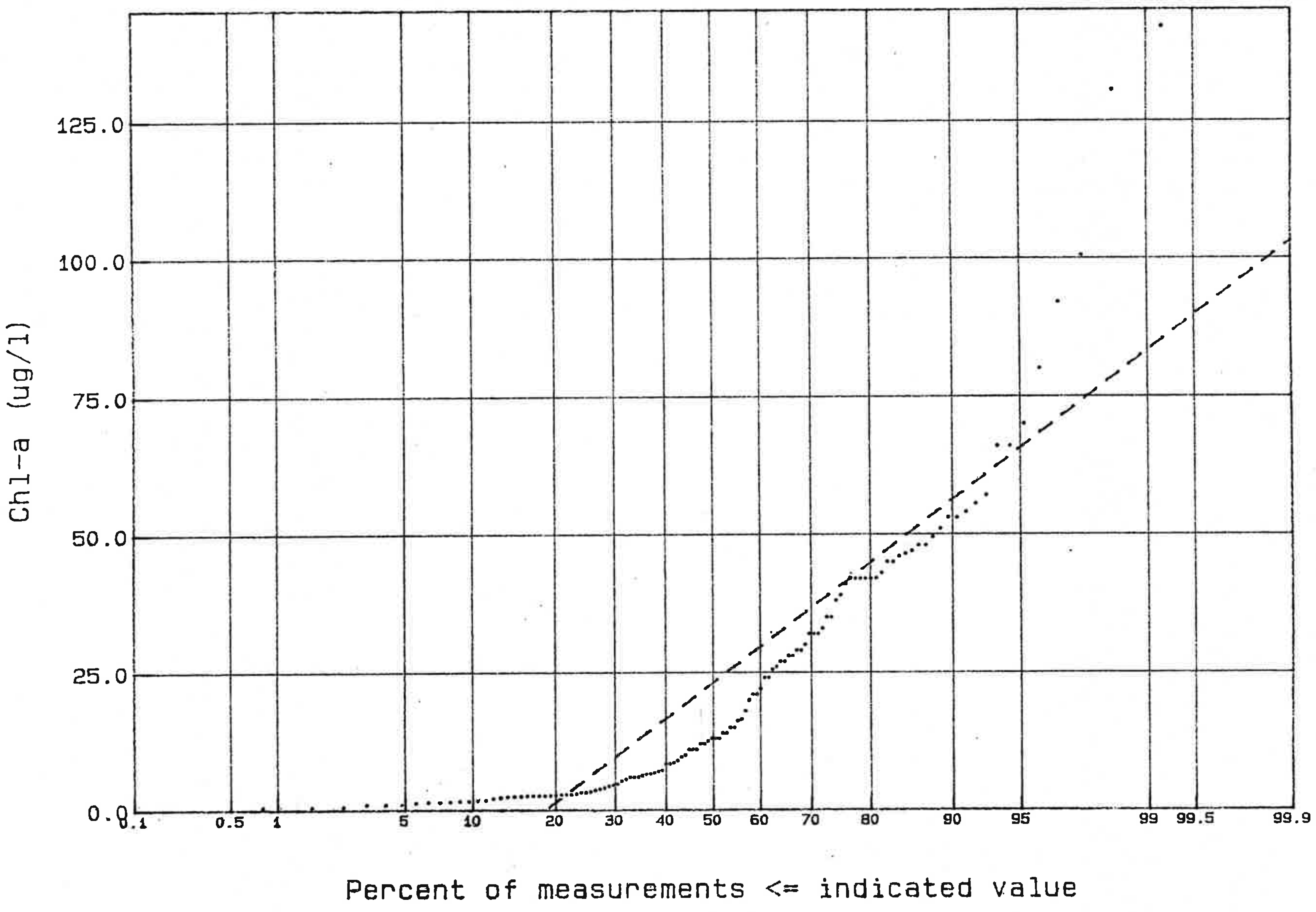


FIGURE 21

Distribution of Dissolved Organic Phosphorus (DOP)
at Chain Bridge -- Number of Observations = 432

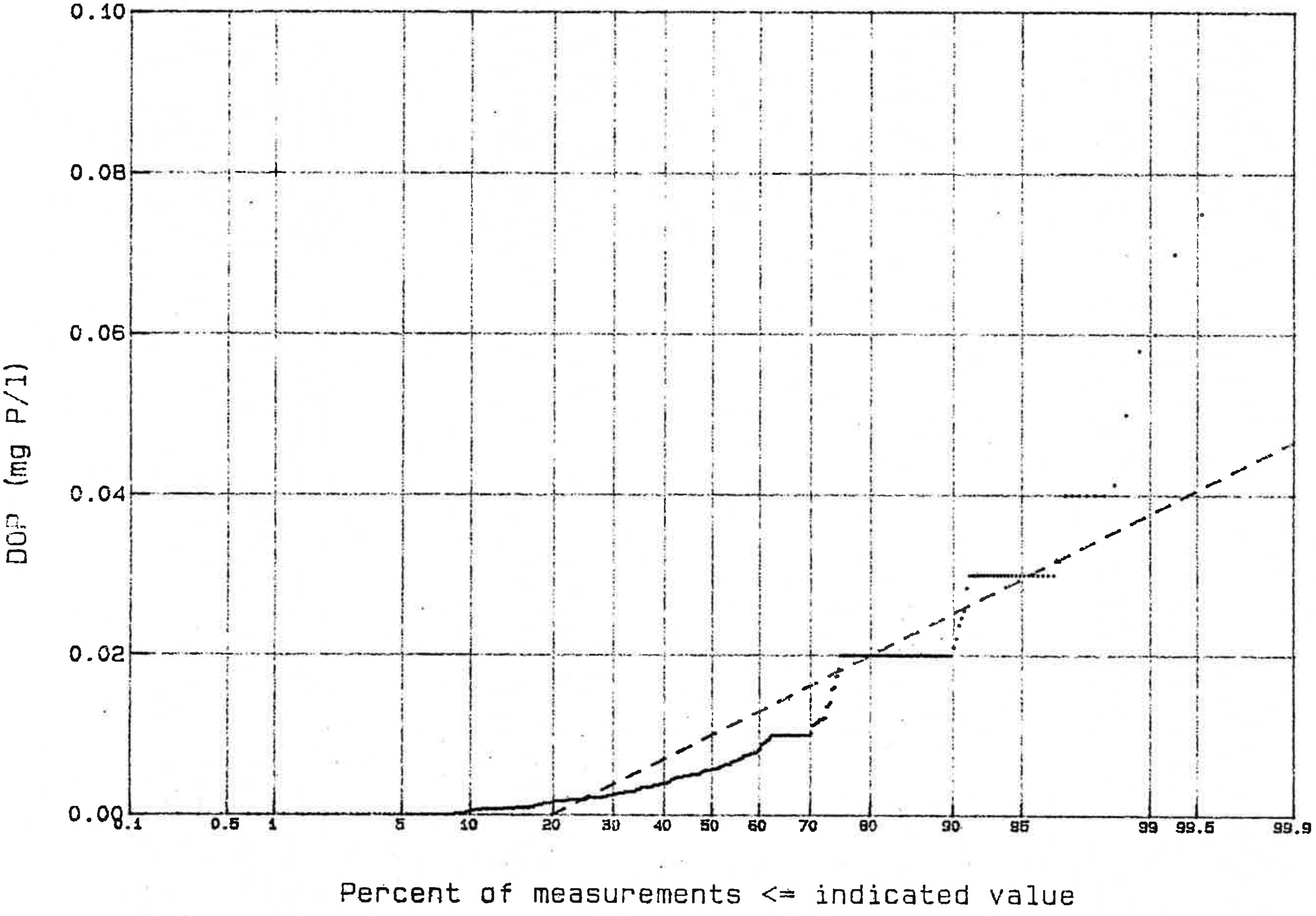


FIGURE 22

Distribution of Particulate Organic Phosphorus (POP)
at Chain Bridge -- Number of Observations = 432

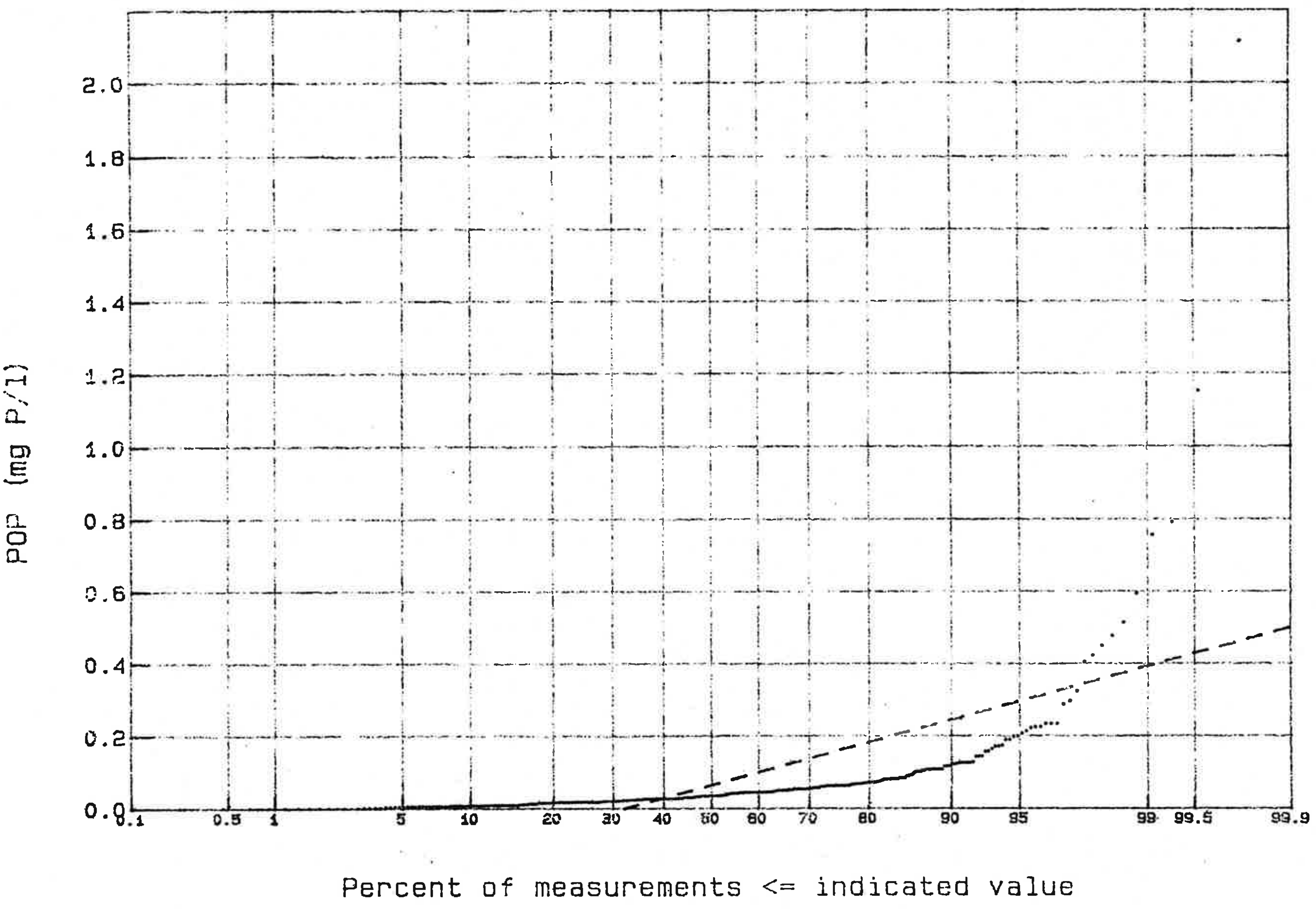


FIGURE 23

Distribution of Dissolved Inorganic Phosphorus (DIP)
at Chain Bridge -- Number of Observations = 432

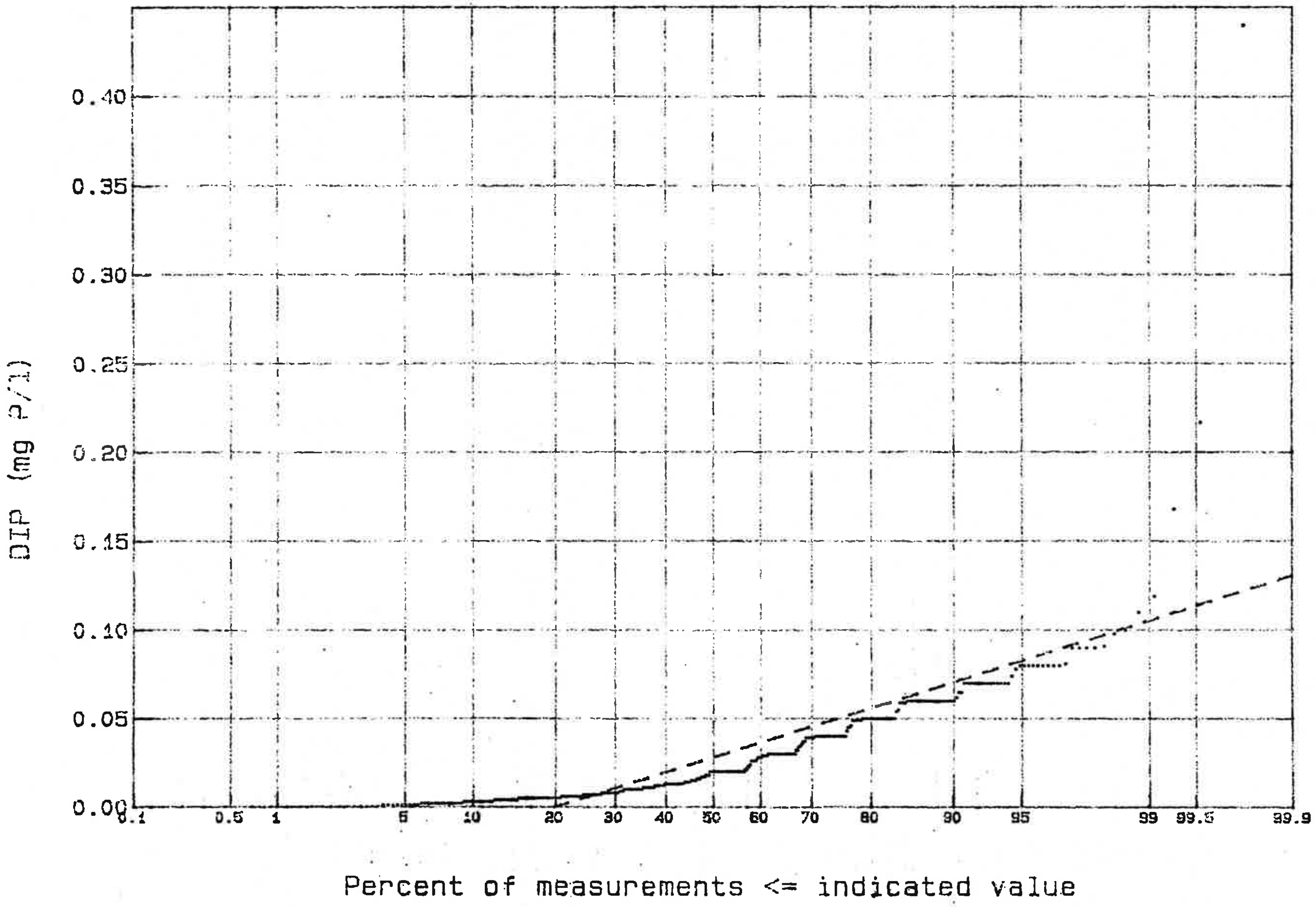


FIGURE 24

Distribution of Particulate Inorganic Phosphorus (PIP)
at Chain Bridge -- Number of Observations = 432

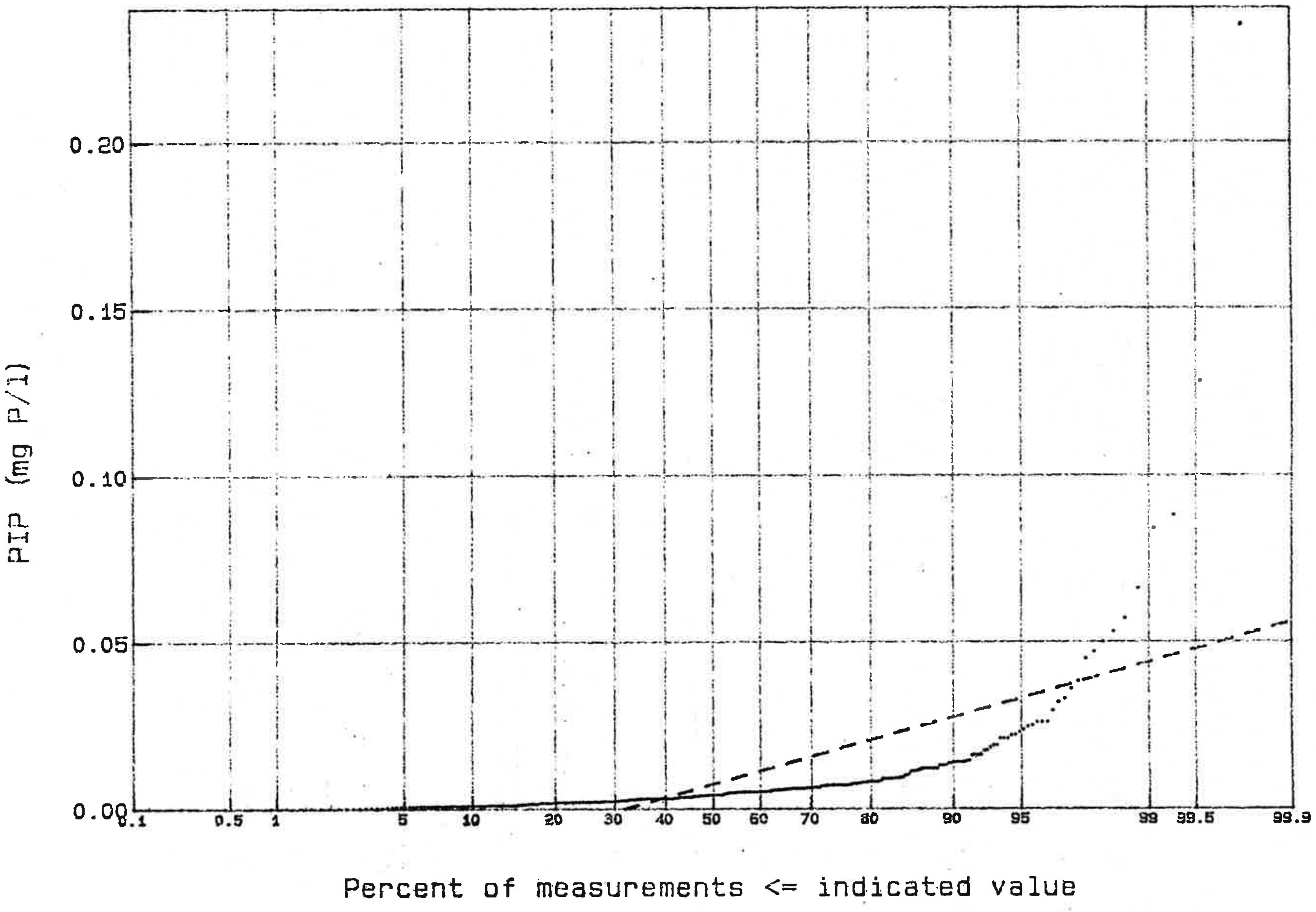


FIGURE 25

Water Temperature by Month
Mean, Standard Deviation, Range, and Number Daily Values

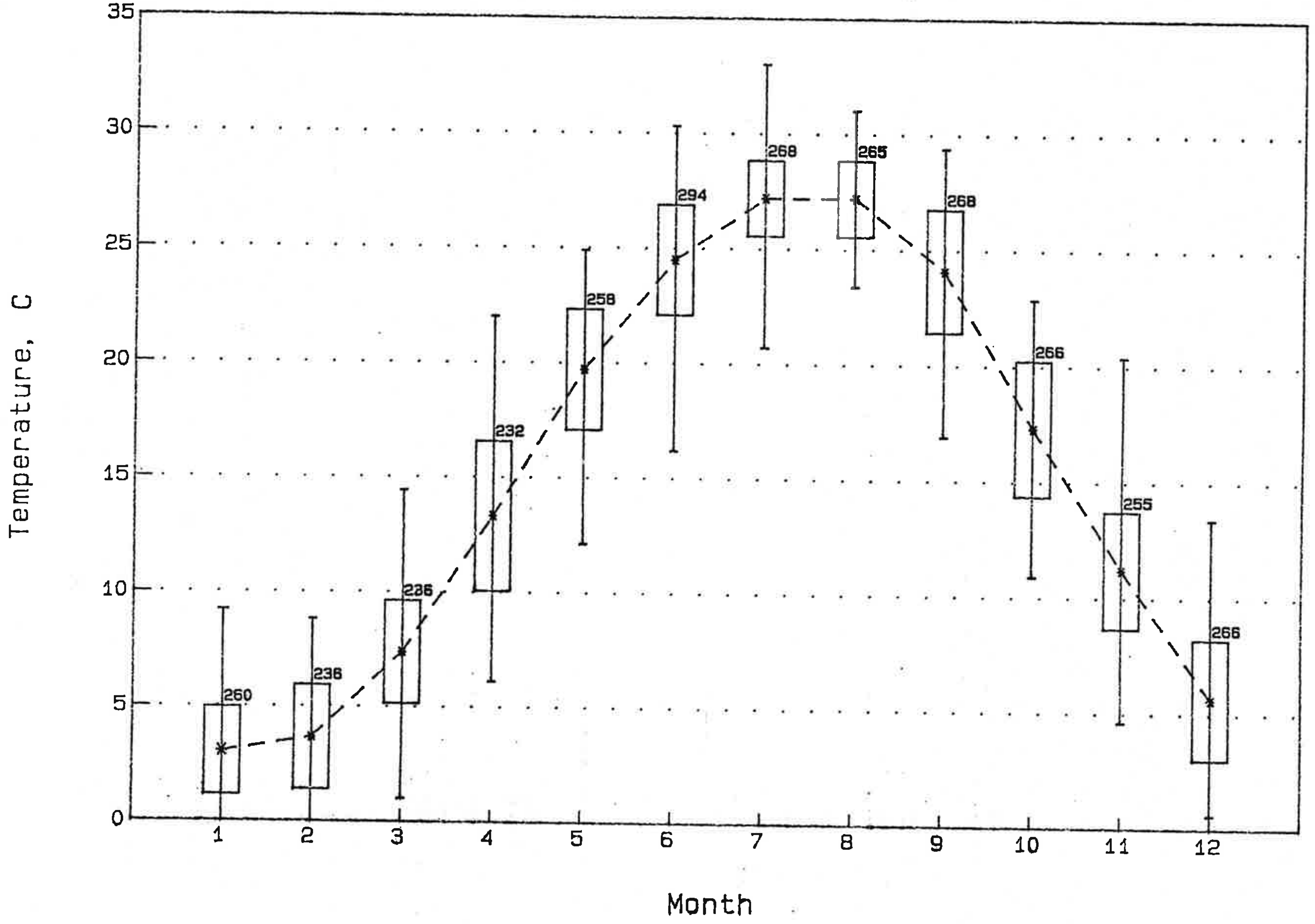


FIGURE 26

Temperature Duration Curve for January

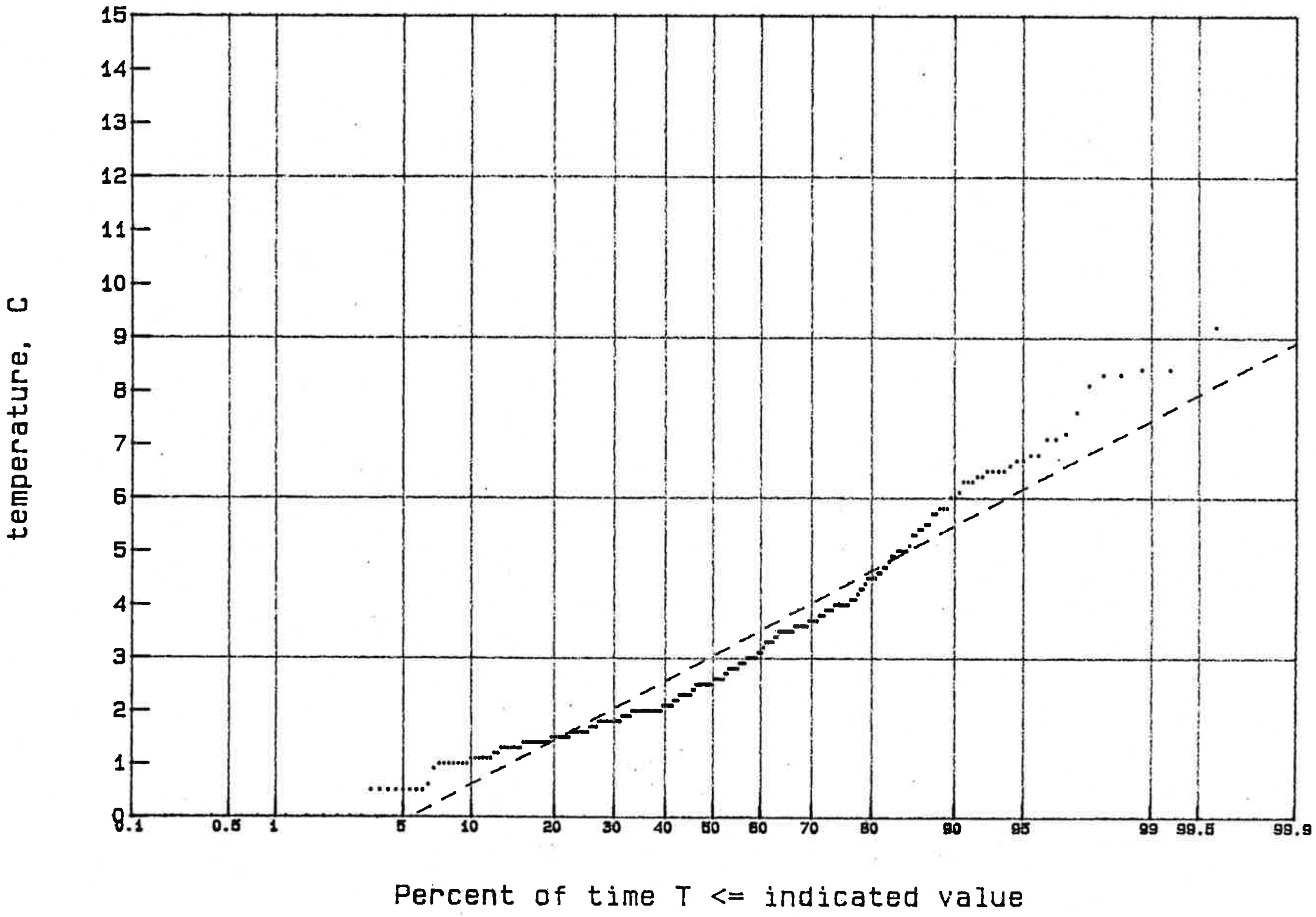


FIGURE 27

Temperature Duration Curve for February

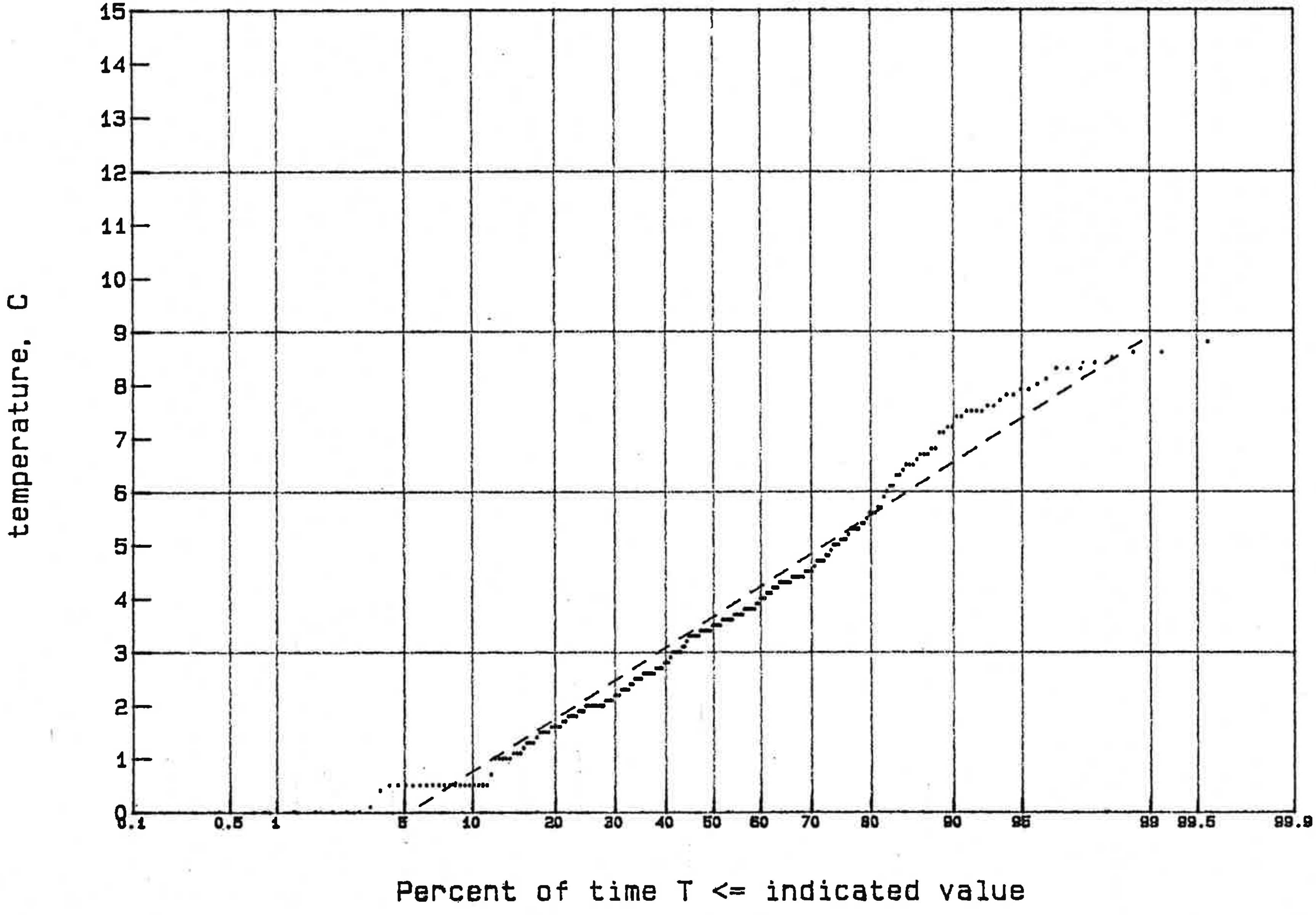


FIGURE 28

Temperature Duration Curve for March

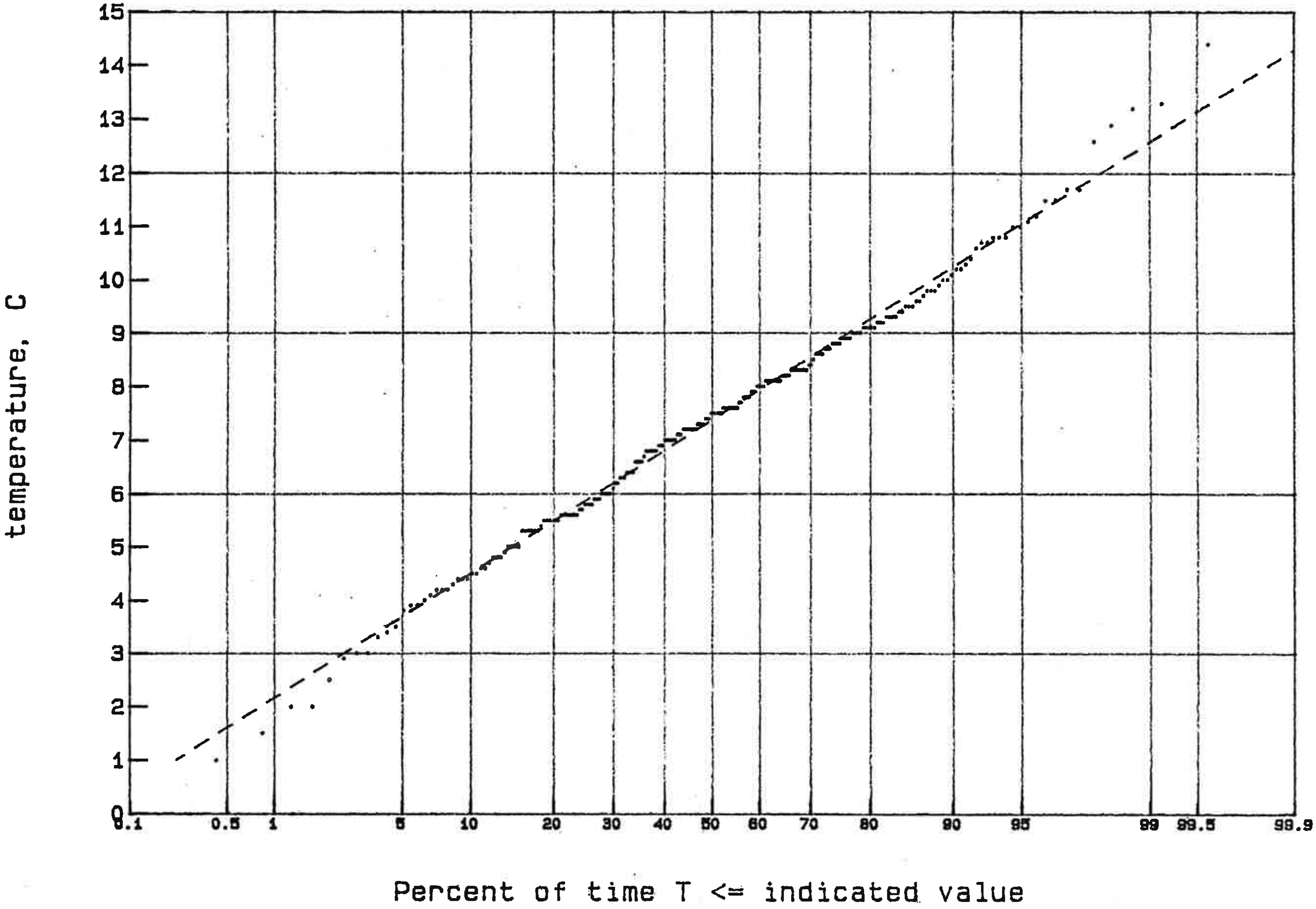


FIGURE 29

Temperature Duration Curve for April

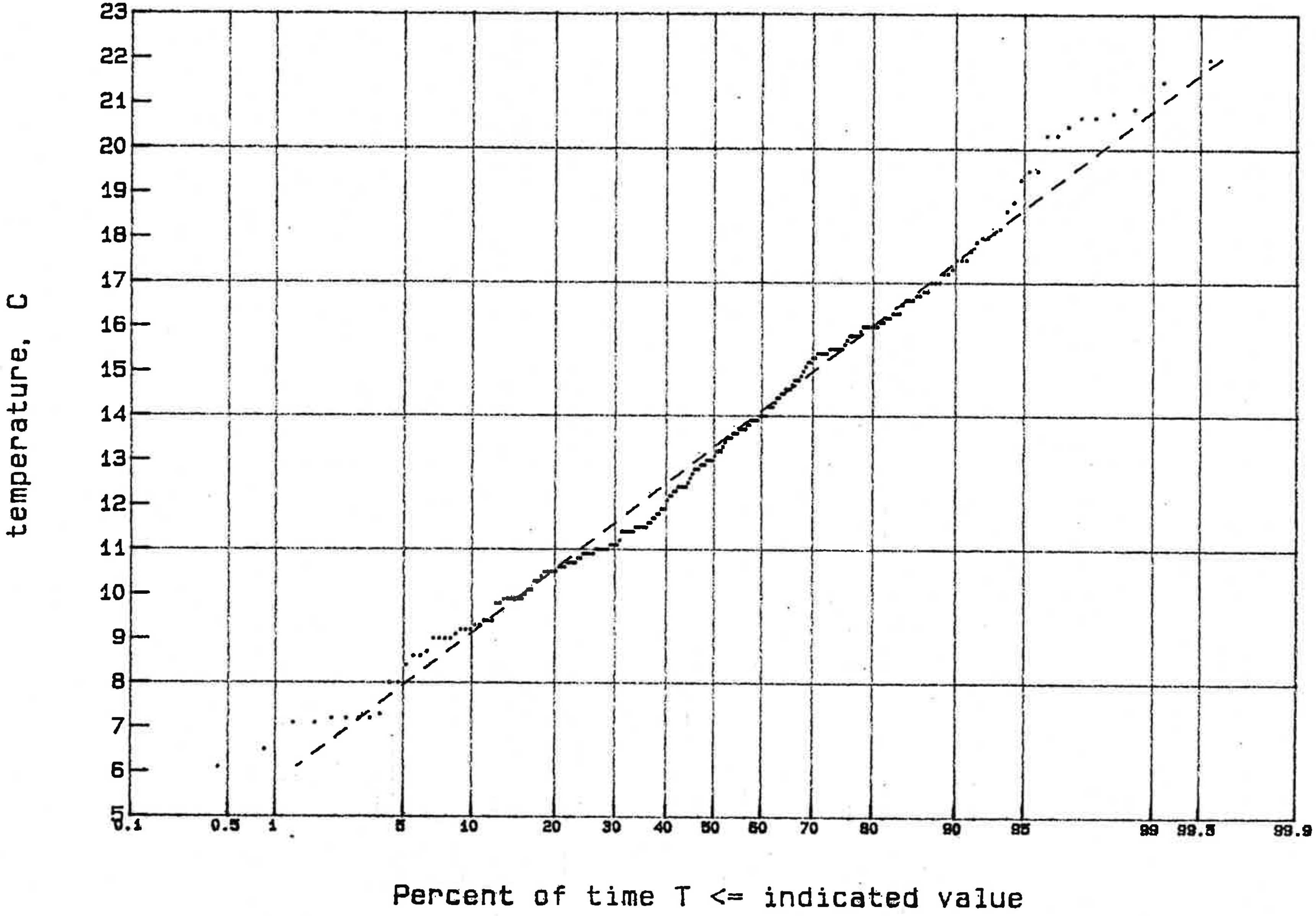


FIGURE 30

Temperature Duration Curve for May

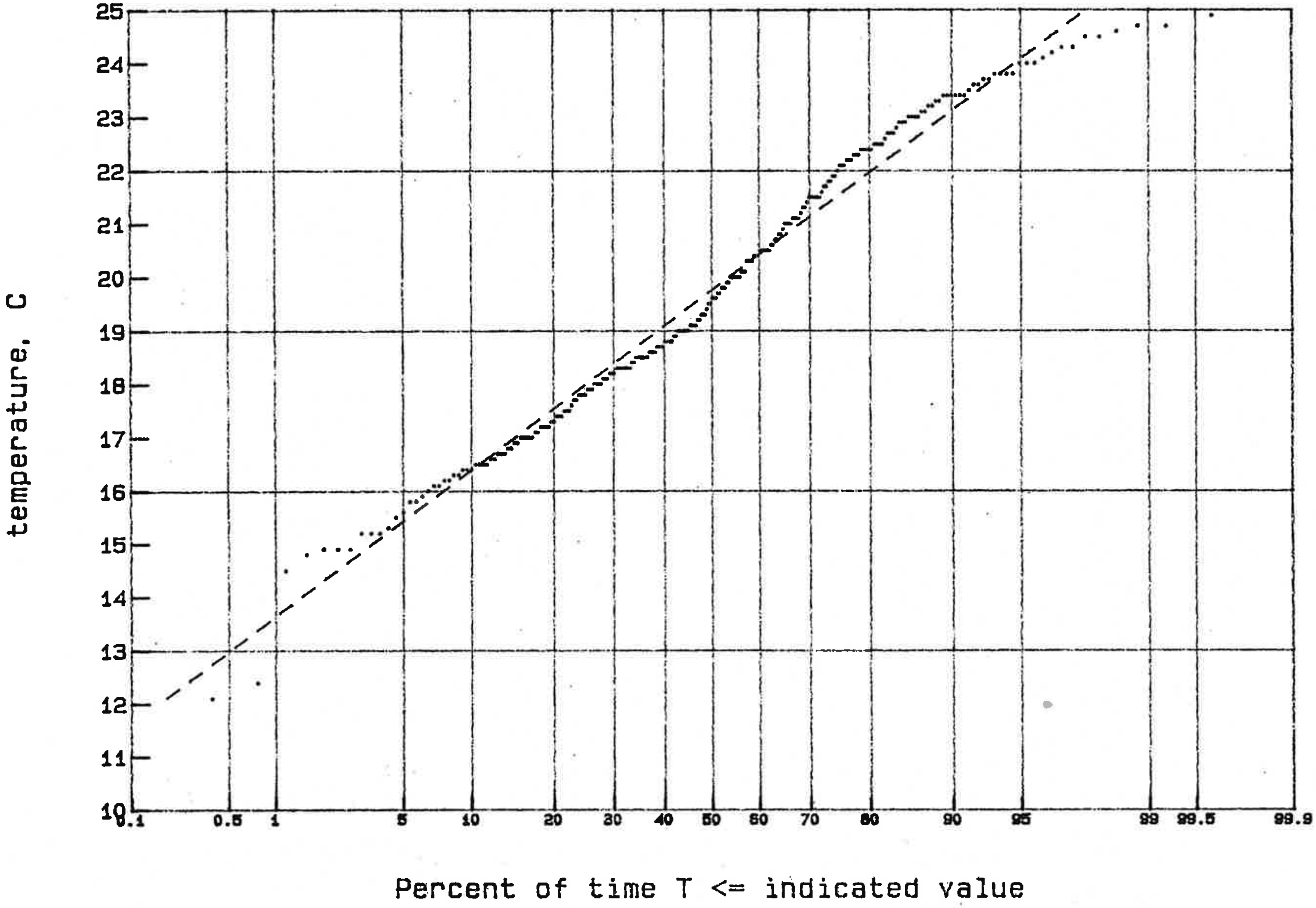


FIGURE 31

Temperature Duration Curve for June

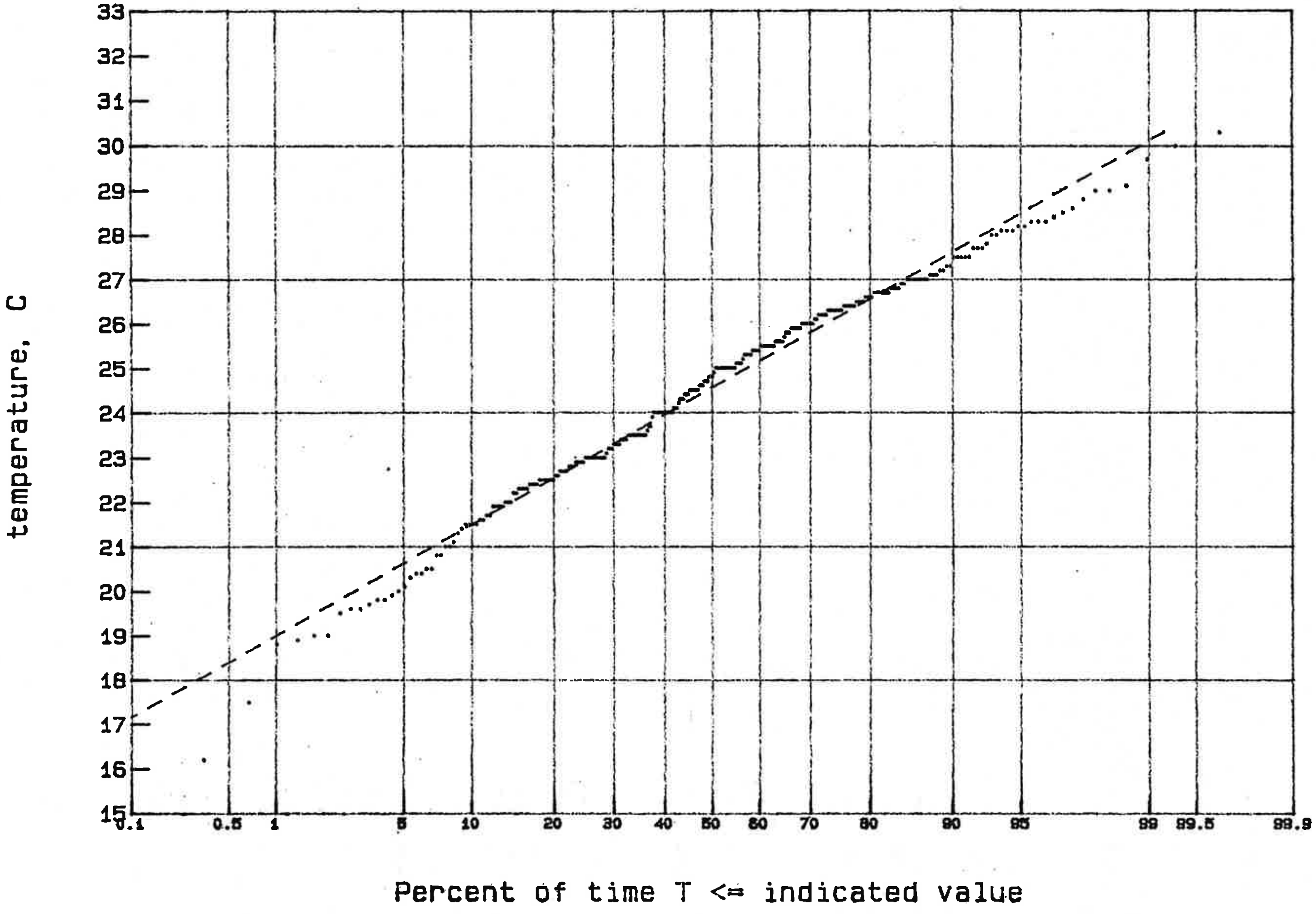


FIGURE 32

Temperature Duration Curve for July

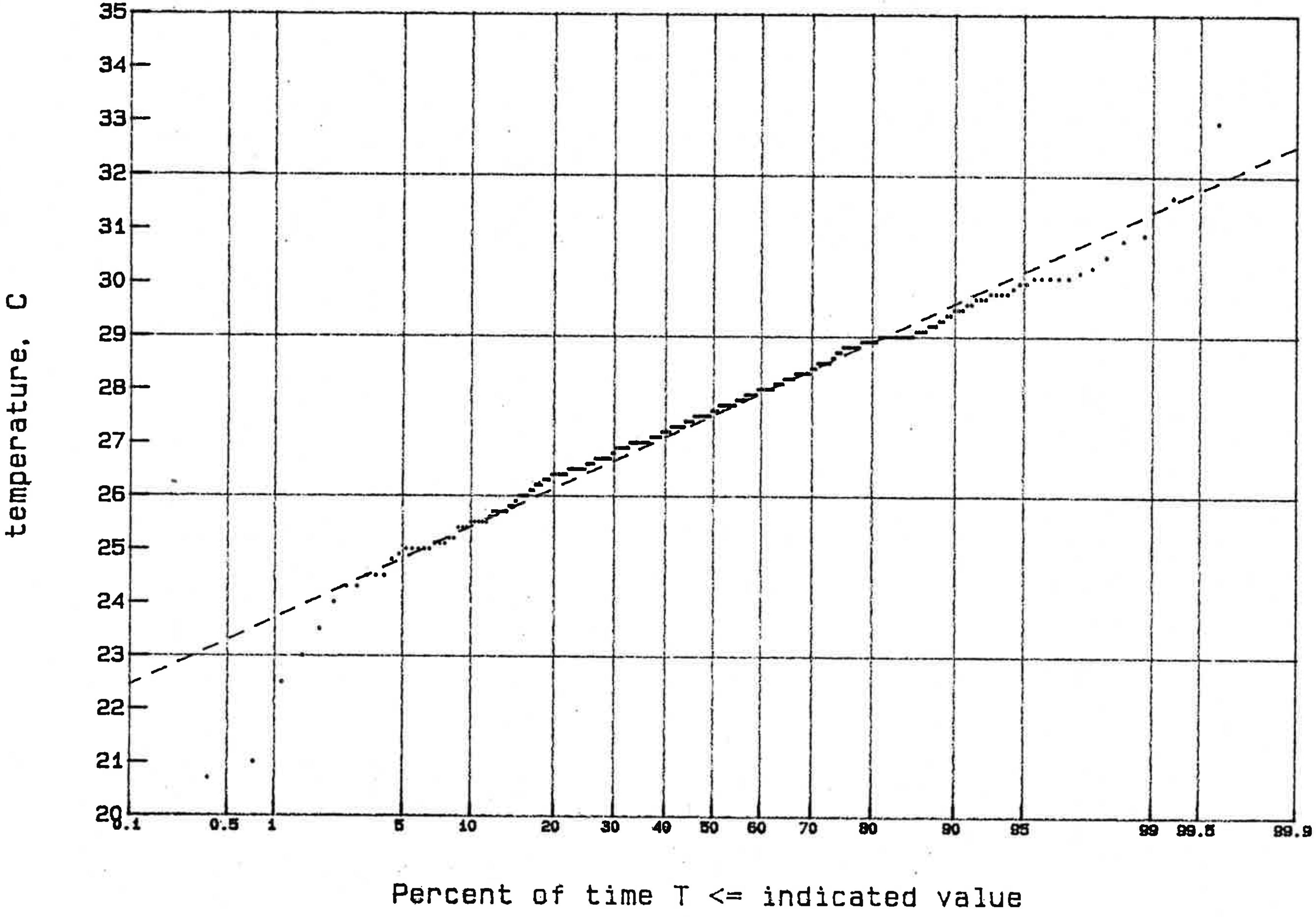


FIGURE 33

Temperature Duration Curve for August

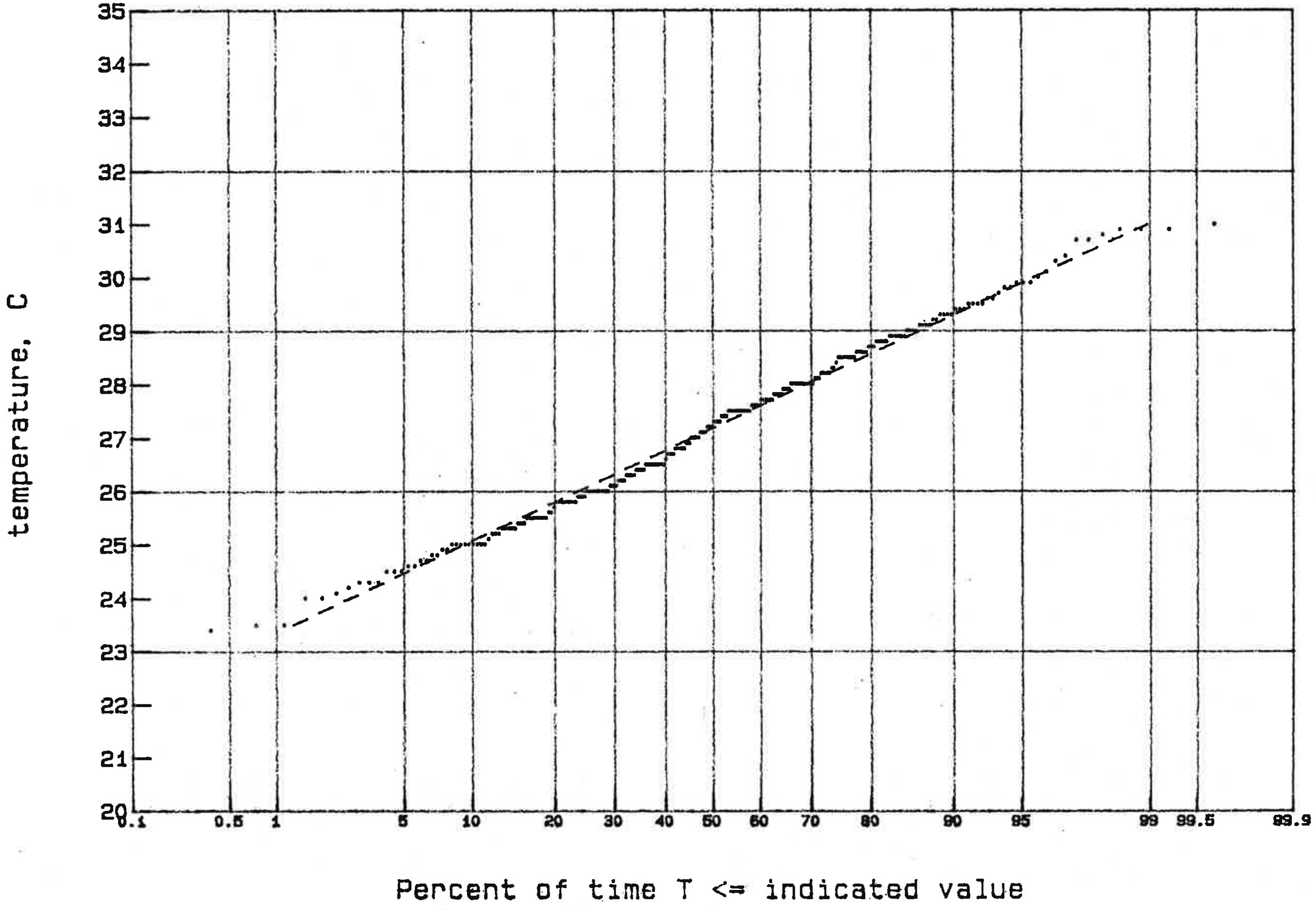


FIGURE 34

Temperature Duration Curve for September

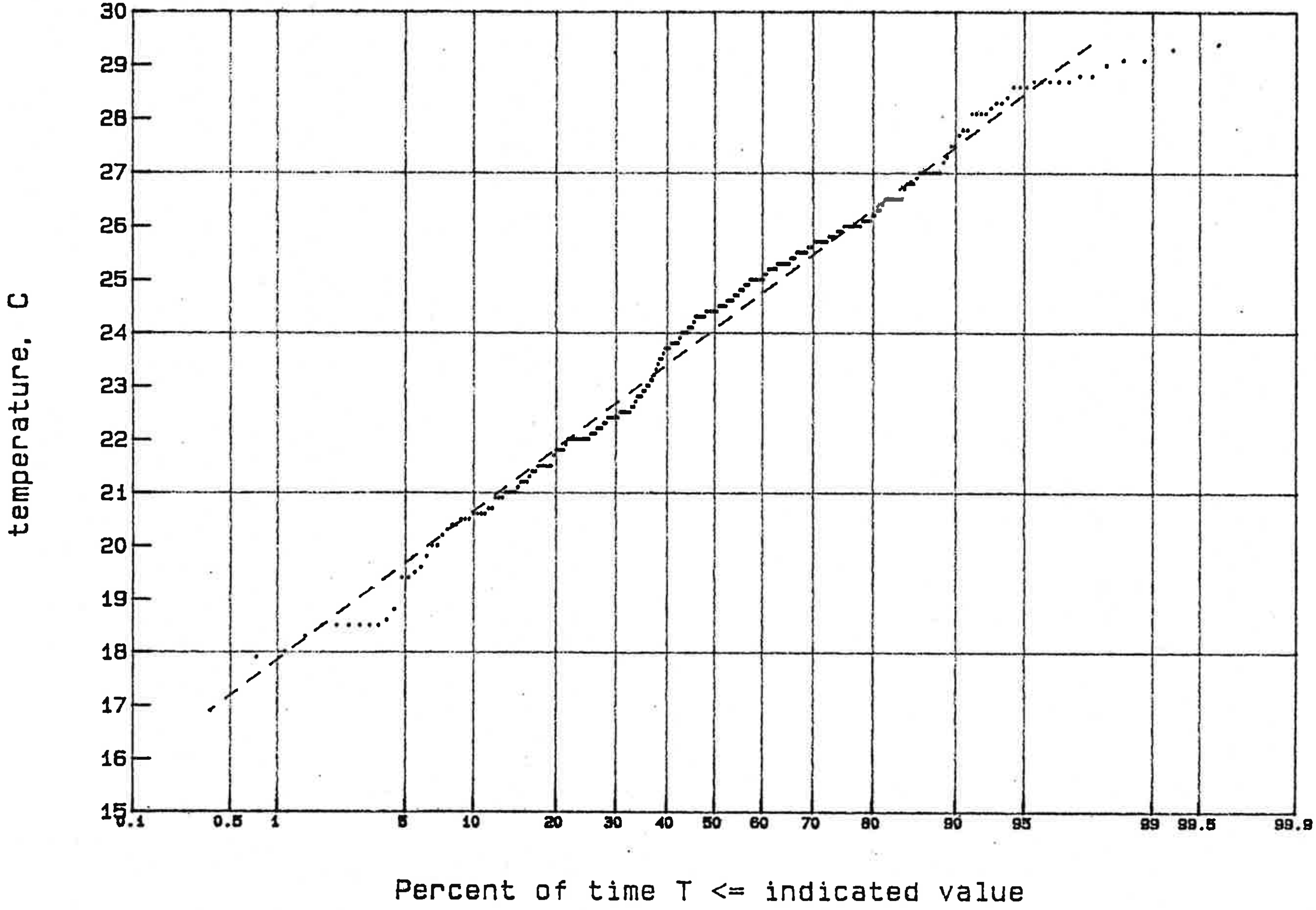


FIGURE 35

Temperature Duration Curve for October

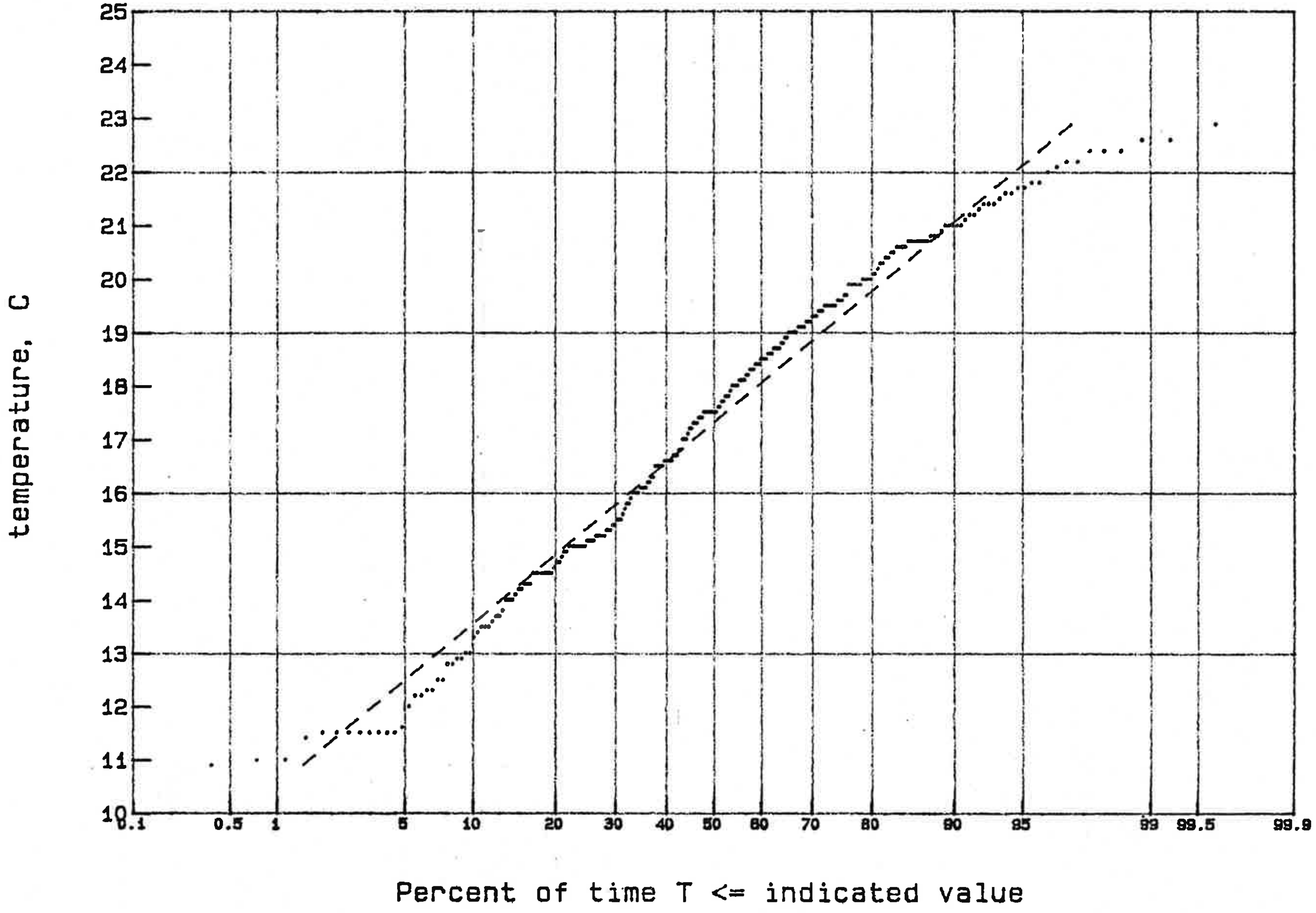


FIGURE 36

Temperature Duration Curve for November

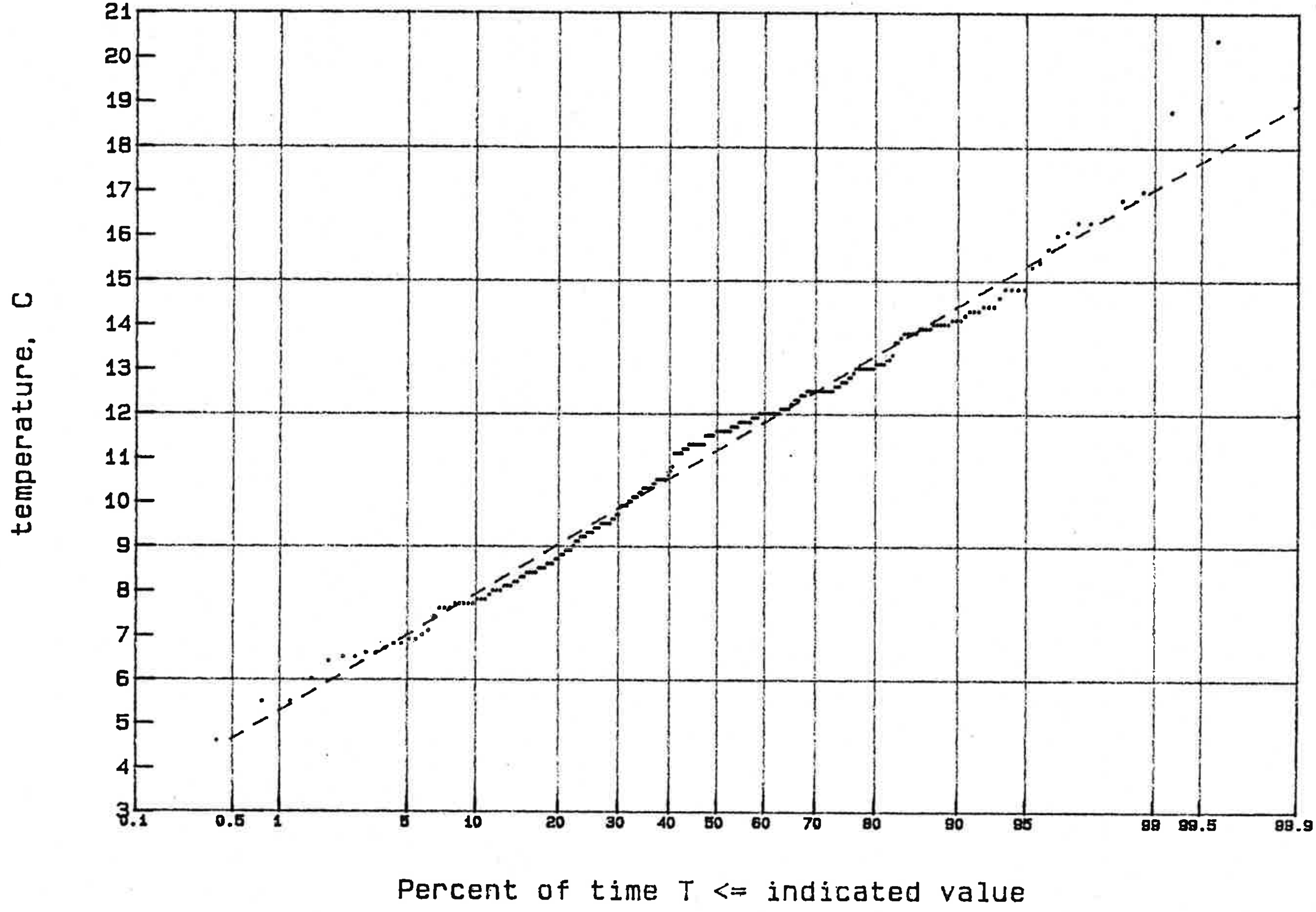


FIGURE 37

Temperature Duration Curve for December

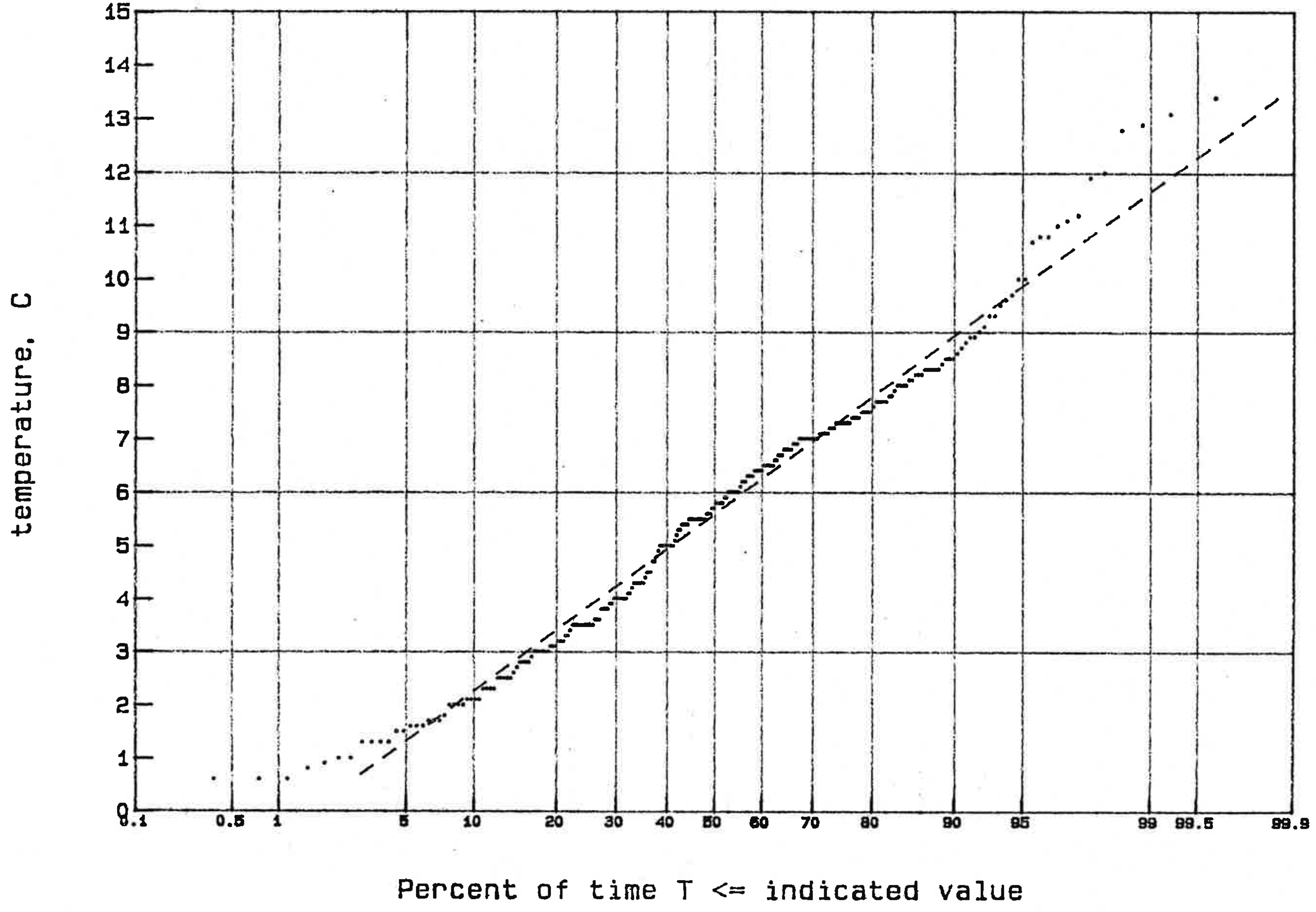


FIGURE 38

Solar Radiation versus Day of Year
588 Days from 1Jan77 to 28Jun80

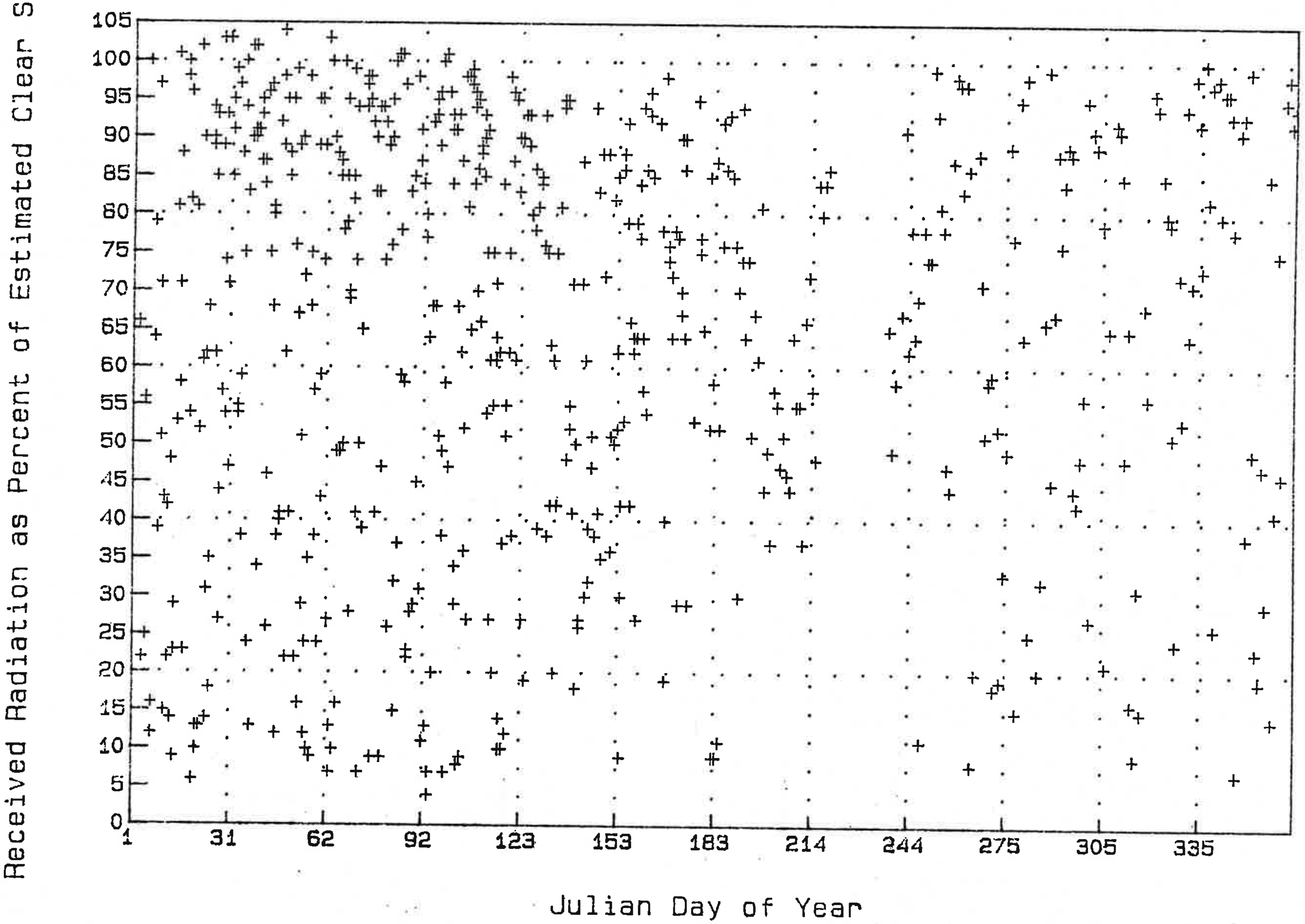
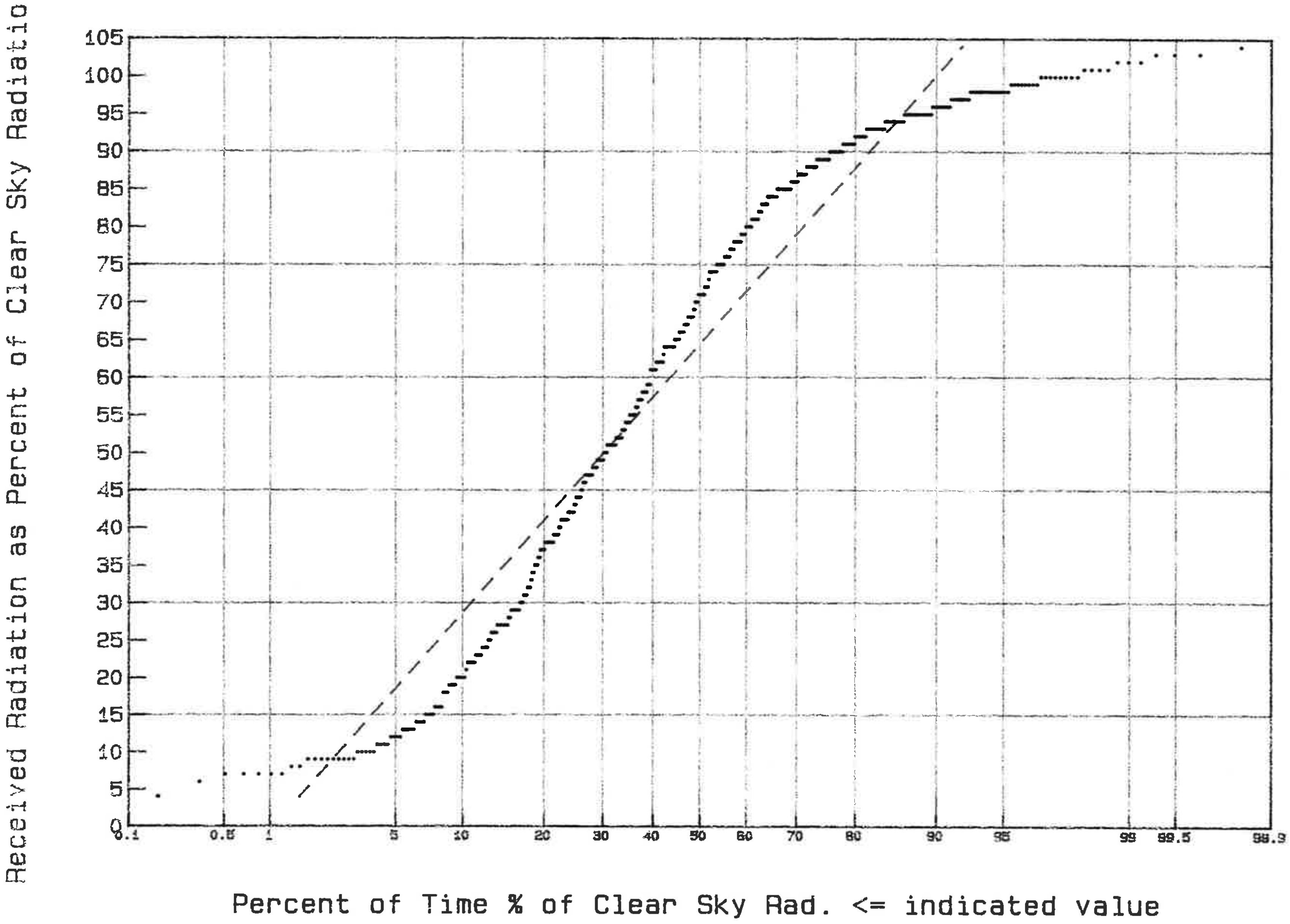


FIGURE 39

Distribution of Solar Radiation
As a Percent of Potential Clear Sky Radiation
Number of Observations = 588



FORTRAN 77 SOURCE CODE FOR PROGRAM DAYWQ WHICH GENERATES VALUES FOR PEM BOUNDARY CONDITION PARAMETERS.

```
c daywq generates independent, random, values for 14 water quality
c        parameters at Chain Bridge.
c     To compile on Fortune computer:  f77 -ffp daywq.f tdate.o -o daywq
c       Written by k. hogan & c. haywood, 28oct86
c       Modified 21nov86:  read parameter distn's from internal arrays
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c  Variable definitions:
c   -anum:  dummy real variable
c   -csky:  name of file with clear sky solar radiation
c      and minutes of daylight (for carbon calculation) by DOYR
c   -dfile:  names of files w/ distns for param 1-11.
c   -doyr:  the day of year, read from std input
c   -fdistn:  the 12 monthly flow distributions, each observation
c               includes a flow value and a probability
c   -flim:  max number data points in each flow dist'n
c   -flonum:  actual number data points read for each flow dist'n
c   -flows:  names of files for 12 monthly flow dist'ns
c   -fmax:  maximum probabilities for each flow dist'n
c   -fmin:  minimum probabilities for each flow dist'n
c   -i, j:  loop counters and array indices
c   -ifnum,itnum,ipnum:  arrays holding the number of data values
c       stored for each array of fdistn(), tdistn(), and pdistn()
c   -imon:  month of year corresponding to doyr
c   -ix,iy,iz:  seeds for the random number generator
c   -month:  array of integers used to calculate imon
c   -param:  array for storing parameter values, in this order
c         1. do         6. chl-a      11. solar radiation
c         2. bod        7. dop        12. flow
c         3. nh3        8. pop        13. water temp, C
c         4. ton        9. dip        14. Carbon
c         5. nox       10. pip
c   -pdistn, plim, pnum:  for remaining parameters, similar to flow
c   -plim:  max number data points in each dist'n for param. 1-11
c   -pmax:  maximum probabilities in each dist'n for param. 1-11
c   -pmin:  minimum probabilities in each dist'n for param. 1-11
c   -sclr:  array contains clear sky radiation, and minutes of daylight
c             for each day of year.
c   -tdistn, tlim, tmpnum,:  for temperature, similar to flow variables
c   -temps:  names of files for 12 monthly temperature distns
c   -tlim:  max number data points in each temperature dist'n
c   -tmax:  maximum probabilities for each temperature dist'n
c   -tmin:  minimum probabilities for each temperature dist'n
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
        real sclr(365,2),fdistn(1800,12),tdistn(300,12),pdistn(750,11)
        real rprob, param(14),anum
        real pmin(11),pmax(11),fmin(12),fmax(12),tmin(12),tmax(12)
        integer ix,iy,iz,doyr, month(12),ifnum(12),itnum(12)
        integer flim, tlim, plim, ipnum(11), imon
        character*13 pfile(11),flows(12), temps(12), csky
```

```
c
      common /rand/ ix, iy, iz
c
      data pfile /'rdo1','rbod1','rnh31','rton1','rnox1',
     *     'rchla', 'rdop1','rpop1','rdip1','rpip1','rsolar'/
c
      data flows/'r1fu01','r1fu02','r1fu03','r1fu04','r1fu05',
     * 'r1fu06','r1fu07','r1fu08','r1fu09','r1fu10','r1fu11','r1fu12'/
c
      data temps/'tmp01','tmp02','tmp03','tmp04','tmp05',
     * 'tmp06','tmp07','tmp08','tmp09','tmp10','tmp11','tmp12'/
      data csky /'clrsky'/
      data month /31,59,90,120,151,181,212,243,273,304,334,365/
      data flim/1800/, tlim/300/, plim/750/
c
c--PART 1:  SEED RANDOM NUMBER GENERATOR
c---function getnbrs uses system time to provide seeds for random()
      call getnbrs(ix, iy, iz)
c
c---throw away the first 50 random numbers
      do 5 i = 1,50
        rprob = random(1)
5     continue
c
C--PART 2: READ DATA FILES
C--Read clear sky solar radiation in Ly/day, and minutes of daylight
      open(2,file=csky,status='old')
      rewind 2
      do 10 i=1,365
        read(2,*) j,sclr(j,1),sclr(j,2)
10    continue
      close(2)
C
C--READ FLOW DISTRIBUTIONS INTO FDISTN(FLIM,12)
      do 15 i=1,12
        open(2,file=flows(i),status='old')
        rewind 2
        do 12 j=1,flim
          read(2,*,end=13)fdistn(j,i)
12      continue
13      close(2)
        ifnum(i) = j
        fmin(i) = 1.0/(1.0 + ifnum(i))
        fmax(i) = 1.0 - fmin(i)
15    continue
c
C--READ TEMPERATURE DISTRIBUTIONS INTO TDISTN(TLIM,12)
      do 20 i=1,12
        open(2,file=temps(i),status='old')
        rewind 2
        do 17 j=1,tlim
          read(2,*,end=18)tdistn(j,i)
17      continue
18      close(2)
```

```
          itnum(i) = j
          tmin(i) = 1.0/(1.0 + itnum(i))
          tmax(i) = 1.0 - tmin(i)
20      continue
c
C---READ ALL OTHER PARAMETER DISTRIBUTIONS INTO PDISTN(PLIM,11)
        do 25 i=1,11
          open(2,file=pfile(i),status='old')
          rewind 2
          do 22 j=1,plim
            read(2,*,end=23)pdistn(j,i)
22        continue
23        close(2)
          ipnum(i) = j
          pmin(i) = 1.0/(1.0 + ipnum(i))
          pmax(i) = 1.0 - pmin(i)
25      continue
c
C----PART 3:  MAIN LOOP STARTS HERE.
c--- A day of year is read from std input for
c--- each set of parameters to be generated.
50      read(*,*,end=999)doyr
c--- determine the month for flow and temperature, save in imon
        do 55 imon=1,12
          if(doyr .le. month(imon))goto 60
55      continue
c
c---for each of parameters 1-13 the procedure is:
c--- since Prob = Rank/(m+1)
c--- and   Rank = Prob * (m+1)
c--- where prob = probability of exceedence
c---       Rank = rank order statistic for the value of a parameter
c---       m    = number of values in the data set
c--- then  to select a random value for each parameter,
c---       generate a random exceedence value, rprob
c---       calculate rank = nint(rprob * (m+1))
c---       and value(rank) = the randomly selected value
c
60      do 80 i = 1,11
          j = rank(pmin(i),pmax(i),ipnum(i))
          param(i) = pdistn(j,i)
80      continue
c
c---solar radiation is determined as sclr * %sunlight
        param(11) = param(11) * sclr(doyr,1) / 100.0
c
c---select flow from the distn for month imon
        j = rank(fmin(imon),fmax(imon),ifnum(imon))
        param(12) = fdistn(j,imon)
c
c---select temperature from the distn for month imon
        j = rank(tmin(imon),tmax(imon),itnum(imon))
        param(13) = tdistn(j,imon)
c
c---carbon is derived from temperature and solar rad. as follows
```

```
c--- this equation derived from Eq. 17 & 21 in PEM Calibration, p235-237
      anum = sclr(doyr,2) * 1.068**(param(13) -20)
      ratio = (68. * param(11)) / anum
      if(ratio .lt. 20.) ratio = 20.
      param(14) = ratio * param(6)
c
c---print the results
      print 900,doyr,param
c
c---go back to begining of main loop
      goto 50
999   stop
900   format(1x,i4,2f6.1,3f6.2,f6.0,4f6.3,f6.0,f8.0,f6.1,f7.0)
      end
c
c
c  this psuedo random number generator is a modified version of
c    Algorithm AS 183, from Journal of Applied Statistics, 1982
      function random(1)
      integer ix, iy, iz
      real arand
      common /rand/ ix, iy, iz
c
      ix = 171 * mod(ix,177) -  2*(ix/177)
      iy = 172 * mod(iy,176) - 35*(iy/176)
      iz = 170 * mod(iz,178) - 63*(iz/178)
      if (ix .lt. 0) ix = ix + 30269
      if (iy .lt. 0) iy = iy + 30307
      if (iz .lt. 0) iz = iz + 30323
      arand = real(ix)/30269. + real(iy)/30307. + real(iz)/30323.
      random = arand - aint(arand)
      return
      end
c
c
c  rank() calculates the rank order statistic given a dist'n and a
c    probability.
      integer function rank(min,max,num)
      real prob,min,max
      integer num
c
c---get a uniform (0,1) psuedo random number
        prob = random(1)
c  if random number <= minimum probability then rank is 1
      if(prob .le. min)then
        rank = 1
c  else if random number >= maximum prob. then rank is num
      else if(prob .ge. max)then
        rank = num
c---else calculate the rank
      else
        rank = nint((prob * (1 + num)))
      endif
      return
      end
```