BNL-26188

CONF-790567-

MASTER

INTERRELATIONSHIPS BETWEEN MAN, ENERGY, AND WATER QUALITY: A NEW METHODOLOGY FOR INTEGRATIVE ANALYSES*

Edward Kaplan and Henry C. Thode, Jr.

May 9, 1979

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

Presented at the Fourth EPA/STORET User Conference,
Bethany, West Virginia, May 7-10, 1979

BIOMEDICAL AND ENVIRONMENTAL ASSESSMENT DIVISION
NATIONAL CENTER FOR ANALYSES OF ENERGY SYSTEMS
BROOKHAVEN NATIONAL LABORATORY
ASSOCIATED UNIVERSITIES, INC.
UPTON, NEW YORK 11973

*Research carried out under auspices of Division of Regional Assessments, Office of Technology Impacts, United States Department of Energy, under contract no. EY-76-C-02-0016.

CONTENTS

Abstract..... ii

Acknowledgments..... iii

1.0 Introduction..... 1

2.0 Introduction to Part I 2

2.1 Data 3

2.2 Retrieval and Aggregation 3

2.3 Explanation of Program 4

2.4 Qualification of Data 6

2.5 Summary (Part I) 8

3.0 Introduction to Part II 9

3.1 Chemical Relationships 9

3.2 Path Analyses 11

4.0 Summary and Conclusions 15

5.0 Improvements and Future Applications 16

Appendix A - Tables 18

Appendix B - Figures 28

ABSTRACT

A. Qualification of STORET Data For Use in Regional Analyses, H. C. Thode, Jr.

The STORET/MSP option was used to obtain county aggregated information on ambient water quality for sixty parameters during the period 1950-1978. Masks, extended EXTRACT specifications and bounds on allowable values limited inclusion of erroneous data. Remark codes were required to aggregate STORET parameters to obtain increased numbers of observations. Numerous statistical analyses led to the conclusions that medians were more useful than means, that trimming on number of observations was required to eliminate counties with extreme values, and that many parameters required logarithmic transformation to be useful in regional analyses.

B. An Innovative Technique For Understanding Relationships Between Energy and Water Quality, E. Kaplan

County aggregated data for nineteen water quality parameters were examined in terms of their ability to describe qualitative chemical characteristics of water. Anion-cation balances as well as expected relationships between conductivity and other parameters were correctly accounted for. Factor analysis indicated the existence of three principal components describing patterns between metal ions, non-metal ions, and alkalinity-bicarbonate, respectively. These factors were used in place of the original complete set of water quality parameters in a structural equation approach describing relationships between various of mans activities. It was found that counties with high industrial electric consumption, farming and mineral shipments tended to have increased levels of most water quality parameters. It was also found that simpler path diagrams may be indicated to reduce the effects of redundancy in adequately describing energy-water relationships.

ACKNOWLEDGMENTS

The authors wish to thank the staff of the STORET/User Assistance Section for their encouragement and support during the past years. This is especially so for those trying times when we were allowed to use STORET programs and options which were not fully documented nor were available to the general STORET user community.

Appreciation is also given to Dr. Stephen Gage, EPA Assistant Administrator, Office of Research and Development, for allowing the authors to gain access to the STORET system.

1.0 Introduction

Our presentation will describe a framework for understanding relationships between many of people's activities and the quality of their water. Studies of this type are performed under the rubric of 'integrative analyses', and are receiving increased attention with the Department of Energy. Two such ongoing studies by both DOE and the EPA typify the need for increased understanding of how water quality and energy relate to one another.

DOE's Regional Issue Identification Assessment (RIIA) seeks to understand how future energy siting will be affected by water quality. Scenarios include projected energy supply and, in some cases, projected energy demand. One requirement to understand how water - both quality and quantity - may both constrain and, in turn, be affected by the scenarios.

In another study, both DOE and EPA seek to identify those regions most susceptible to the effects of acid precipitation, a phenomenon with roots in both natural and manmade combustion.

Both studies require a knowledge of water quality on a useful geologic unit. Both require the use of this water quality data together with energy and socioeconomic information, woven together in a fabric, heretofore unavailable, which adequately describes their complex interrelationships.

A description of how we assembled water quality data on a county basis will be discussed at this time by Henry Thode, a mathematics associate in our Biomedical and Environmental Assessment Division. Following Henry's presentation, I will describe how we determined the adequacy of this database, as well as presenting a new methodology for integrative water-energy-socioeconomic analyses.

2.0 Introduction to Part I

Our STORET retrievals were made out of a desire to create a data base consisting of water, energy, and socio-economic variables with which we could perform a regional analysis of the relation between chronic water quality and energy systems on a county-level basis. The reason for aggregating on a county-level basis is that health data at Brookhaven is on a county level as are these other variables, and we had to have water data which would be amenable to comparisons with them. In order to meet this objective, some type of descriptors of water quality was needed for each of the counties in the region we wish to cover (New England, Middle Atlantic States, and Ohio). Most water quality analyses use the standard statistics of mean and standard deviation. We wanted to look at the chronic situation. Our first look at the date showed that much of the data was taken over short periods of time, indicating that the

measurements were taken in response to short-term crisis situations (such as chemical spills). More will be said of this later. It was decided that we should consider order statistics (median and other percentiles) as possible descriptors of average water quality because they would be more resistant to outliers (crisis situations) than would means and standard deviations.

2.1 Data

The 65 water parameters (Table 1,2) for each county that were extracted from the STORET data base were subject to the following constraints (Fig. 1):

1. The measurements should come from ambient surface water;
2. should have been made between 1955-1978;
3. should have been made at state or federal measuring stations;
4. should not have been remark coded, as these data were found to have too many caveats.

2.2 Retrieval and Aggregation

Retrievals from STORET were made using program MSP, some of our own masks, and some changes in the standard JCL.

Masks were used to help keep computer costs down and to promote efficient searches. The changes in the JCL were additions and changes in time and track allocation parameters and requests due to the extremely large size of some of the retrievals.

MSP was used because it was the only STORET program that would extract the order statistics that we wanted (15th, 50th, and 85th percentiles). Using MSP required that we ask for a large amount of superfluous information; for example, MSP makes plots of all the statistics extracted for each parameter, which were useless for our purposes.

2.3 Explanation of Program (Fig. 2)

This particular retrieval program was used to extract toxic metals and non-metals (extract codes) for type A, ambient data in lakes and streams, by county, for the states shown.

The R= Z parameter indicates that we do not want remark coded data extracted from the data base for the parameters specified. This would eliminate data with remark codes such as C (calculated), J (estimated value), K (actual value known to be less than value given) and L (actual value is known to be greater than the value given).

The P and PM parameters were used in order to increase the number of observations of the water parameters in each county.

That is, when a P=N1 is followed by PM=N2, Pm=N3,...,Pm=Nn, all of the parameters will be retrieved as being equivalent variables and will be listed under the code given by the last PM= parameter in the string.

The RR code indicates that the parameter preceeding the RR should be referred back to the R parameter, i.e., the parameter is subject to the remark code restrictions (R= Z). (Note: P=745, P=746, P=71875 were supposed to be combined under code number 71875, but due to typing errors were extracted as separate parameters. The correct string is P=745, PM=746, PM=71875. These are various sulfides, and were not used in later analyses.)

The MSPARMS=MODEL indicates the statistics that we were after: 15th, 50th, 85th percentiles (50th percentile is the median, while 85th-15th percentile is a measure of data dispersion analagous to standard deviation), mean, mean + 1 standard deviation, maximum, and minimum.

The plot options gave us the choice minimum and maximum allowable values for extractions. This helped eliminate very high, possible erroneous, values that have been entered into the STORET data base. These limits were decided upon based on extensive literature searches to determine a limit, for each variable, above which a value could reasonably be considered spurious.

2.4 Qualification of Data

After STORET retrievals were obtained on tape, the data was transferred to our CDC computer and transformed into a county-level data base so that the capability for using the means or medians of the water parameters as the descriptors of water quality could be studied. After putting this data together, it was noted that even with 65 water parameters extracted, 16 counties have no data at all (Table 3). Later it will be seen that by looking at 19 parameters simultaneously, only 138 counties have complete sets of data. This suggests the need for a more complete and comprehensive water quality measurement system.

While means have been used in most water studies, our conjecture was that the medians would be less affected by outliers which might indicate a measurement error, incorrect entry into the data base, or critical short-term situation.

The criteria for the choice of statistic was comparison of frequency distributions between the mean and median (or some transformation of them). It was desired to have the distributions as well behaved and as similar as possible. Results of analyses would be more reliable the closer to the Normal distribution we could get, because statistical analyses (such as regressions, F-tests) are designed to be more efficient when the distributions are Normal or near Normal.

A graph of the arsenic frequency distributions (Fig. 3) of mean, median, started log of the median (to take care of zero values), and started square root of the median show that the most nearly Normal frequency distribution is the log transformed data. The two curves on each graph represent the actual distribution of the arsenic values and the expected distribution if the distribution was actually Normal with mean and standard deviation equal to those of the actual data (0 are expected, X are actual). Since the expected and actual curves of the log transformed medians are the most nearly coincidental, this is the form of the arsenic data that we used in later analyses.

The graphs of the frequency distributions (Fig. 4,5) of lead and selenium show that the same choice could be made in these cases. Later work using 19 of the water quality variables indicated that some needed transforming while some were better using raw median values (Table 4).

One more technique was used in creating our present data base. Graphs of the median values vs. the number of observations of a water parameter in the counties showed that many parameters had high median values associated with low numbers of observations (see Fig. 6,7). This is due possibly to acute situations which had to be monitored for short periods of time and then discontinued after the pollutants involved had dissipated. Since this was not representative of long-term water quality, we trimmed the data on the basis of number of observations.

This trimming had to be done to limit these occurrences without arbitrarily removing counties which had high values.

Trimming on the basis of five and ten observations was attempted (Tables 5,6). Trimming in most cases lowered the maximum value and the mean of the distributions. In most cases, it also improved the skewness and kurtosis of the frequency distributions (measures of normality). We took this to be an indication that some counties did have situations that required short-term observation periods.[1] The difference between trimming on five observations and ten observations as indicated by the means, standard deviations, skewness, and kurtosis are negligible. However, trimming on ten observations reduces the number of counties (on the eight parameters shown) by between 32 and 44. We chose to trim on five observations, therefore, so that we would have a larger data base to work with.

2.5 Summary

STORET retrievals using program MSP were run to help create a county-level data base consisting of water, energy, and socioeconomic variables. Comparisons of means and medians of the water variable were made to determine which would be more appropriate as estimates of long-term water quality. These comparisons indicated that medians or the log transformation of

[1]This was in fact the case for several toxic metals in Maine and New Hampshire counties where potato farming was the predominant economic activity. It was ascertained that the agencies making the measurements did so in response to reported run off of toxic materials used in pesticides.

the medians were appropriate. Trimming the data on five observations was also used to help prevent short-term high pollution situations from being included in the data base.

3.0 Introduction to Part 2

Let us remember what it is we seek: a framework which can integrate energy, socioeconomic and water quality into a package which describes how each relates to one another. To do this one must first have confidence in one's data.

Henry began to discuss the rationale for our county aggregation process and described some statistical techniques which were required. Now let me turn to a comparison of what the water quality literature requires of our data, and to illustrations of how well our manipulations reflect these reality constraints.

3.1 Chemical Relationships

We found 138 counties in the New England and middle Atlantic region with a complete set of data for 19 water quality parameters (Fig. 8). These parameters were found after a number of tests to maximize both the number of variables we could use as well as the number of counties with a complete set of data for these variables.

First, we require charge balances. This is accomplished by re-expressing ionic concentrations (in mg/l) into a form recognizing the charge and atomic weight of each specie (in milliequivalents/l). Fourteen of these parameters represent ionic species: 8 cations and 6 anions (Table 7). Clearly the data pass this test (Fig. 9).

Next we require the data to adequately describe previously known equilibrium relationships. For example, the sum of concentrations of the major ionic species is often used as an approximation to total dissolved solids concentration. This was indeed found to be the case (Fig. 10).

Total alkalinity, the capacity of a solution to neutralize acid, is primarily due to carbonate, bicarbonate and orthophosphate ions. Regressions of total alkalinity on the set of dissolved species which we used explains almost all the variance in the alkalinity data. We found (as expected) that HCO_3 was clearly the most important component (Fig. 11 and Table 8).

Relationships between ionic concentrations and specific conductance have been described at length in the literature. In general, we expect linearity over extended concentrations. This is indeed what we find (Figures 12-15).

Conductance is often compared with measured or calculated total dissolved solids (TDS), where the relationship is approximately linear with a reported slope in the range 1.3-1.8. We found a slope of approximately 1.4 (Fig. 16).

These are a few examples of the tests applied to the county aggregated water quality data. We are presently comparing our information to what might be expected from underlying soils and rocks, as well as land use. The latter information has recently been added to our data files, the importance of which will be described later if time permits.

3.2 Path Analyses

The existence of our Division within BNL is predicated upon the necessity to understand how energy relates to man and his environment. Ideally such an understanding should emerge from an approach treating all components in the context of a single system. This would be in contrast to the more prevalent situation in which a single source of some type of energy is analysed for the ways in which IT affects one area of the biosphere, such as air.

For some time now geneticists have used a structural equation approach to describe causal relationships between

various factors relating to the 'genetic package'. The technique is also known as path analysis, and has been used as the vehicle which allows us the use of our water data together with other county level information such as energy production and use, as well as several demographic and economic indicators.

Path analysis begins with an analyst postulating a plausible set of linear, additive relationships between a set of variables. These relationships are expressed in terms of a path diagram. Each variable occupies a box located in a hierarchical setting. One-way arrows are used to show causality: they lead <u>from each</u> determining variable <u>to those variables</u> depending upon it. Associated with each arrow is a coefficient (and a respective significance level).

To overcome the problem of dimensionality (i.e., some variables in the analysis have dimensions that are different from other variables in the analysis) all variables are standardized. That is,

$$x_{std} = \frac{x - x\ mean}{\sigma_x}$$

Thus the distribution of each standardized--and therefore dimensionless--variable has mean zero and standard deviation = 1.

The coefficients are determined via a multiple regression of the standardized dependent variable on the standardized determining variables. The coefficients are thus the more familiar BETA COEFFICIENTS of multiple linear regression.

It is usually not possible to build a completely self contained system. Complete determination will not hold because of several reasons: data for known causes may be lacking and unknown causes are always present. In such an event residual variables are introduced which are assumed to be uncorrelated with other determining variables and which account for the remaining variance in the model.

Consider a framework (i.e., MODEL) in which it is postulated that water quality is determined by both the production of raw fuels and electricity, by industrial, commercial and residential activities, and by farming. One may construct a diagram of the type shown in Fig. 17.

The diagram is obviously formulated in terms of available data. Consider the highest level in the diagram: water quality. The question arises as to how to use our county level data.

Instead of postulating a path diagram for EACH parameter of interest, we used our 138 county subset of 19 variables and performed a factor analysis. This is another statistical technique which is used to elucidate patterns among groups of variables based upon the idea that groups of variables may indeed be formed on the basis of shared variance.

Three such factors were found to best describe our data (Figures 18-20):

1. Conductivity related (Cond, TDS, Na, K, Cl, F, SO_4).
2. Hardness-Alkalinity (T.Alk., HCO_3, Ca, Mg, TSS).
3. Metals related (Fe, Mn, SO_4, pH).

ONLY THREE PATH DIAGRAMS WERE NOW REQUIRED, ONE FOR EACH FACTOR.

Returning to the path diagram (Fig. 21), we desired to use additional information available to us describing both economic (demand) activities and the energy they consume. The data derives from the Bureau of Census, other federal and state agencies, as well as the BNL energy atlas of the U.S.

The complete set of path diagrams (Figures 22-24) may now be constructed and they lead to a number of interesting conclusions. One must be cautioned to bear in mind that they relate only to the postulated path diagram. Using the parsimony

principle of Occam's Razor, the analyst further refines the path diagram to optimize the amount of explained variance described by the model.

4.0 Summary and Conclusions

1. A county level water quality file was created using STORET data for approximately 65 variables spanning the period 1955 to the present.

2. Screening procedures and logarithmic transformations were necessary to obtain better behaved frequency distributions for many parameters.

3. Use of medians is preferred to the use of other statistics describing water quality, especially where integrative analyses are the intent of the analyst.

4. Known relationships describing qualitative chemical characteristics of natural waters were obtainable using the temporally and spatially aggregated data.

5. The structural equation methodology of path analysis is applicable to studies of how energy supply and demand activities relate to water quality.

6. Preliminary path diagrams resulted in explaining 25 to 40% of the variance in several patterns (i.e., factors) describing water quality.

7. Direct production of electricity was not found to significantly affect our three WQ factors.

8. Counties with increased mineral industries tend to have waters with higher Fe, Mn and SO_4 concentrations, as well as lower pH.

9. Counties with increased levels of conductivity and related dissolved ions are also those with high levels of industrial electric consumption.

10. More affluent counties tend toward (expected) higher levels of retail sales but lower levels of mineral and industrial activities -- perhaps a manifestation of the 'bedroom community' concept.

5.0 Improvements and Future Applications

1. Hydrologic connections between counties will hopefully provide a mechanism for describing how one county may affect the water quality of its neighbors.

2. County soil, rock and land use data is presently being analyzed to determine whether a more representative (i.e., natural) background water quality may be ascertained. The residuals which remain when these background concentrations are removed from our present county water quality data would then be more illustrative for use in path diagrams such as those used in this study.

3. Too many counties exist with insufficient WQ data. This situation must be corrected to allow more complete integrative analyses.

4. This data and methodology is being extended to other problems, such as identifying areas most sensitive to impacts from acid precipitation.

Appendix A - TABLES

TABLE 1
WATER QUALITY VARIABLES

FUNCTIONAL GROUP	VARIABLE NAMES	
OXYGEN RELATED	DISSOLVED OXYGEN, B.O.D.	
BACTERIOLOGICAL	COLIFORMS: FECAL AND TOTAL	
NUTRIENTS	NITROGEN (AS N, NO_2, NO_3)	
	PHOSPHORUS (AS P, PO_4)	
HARDNESS RELATED	ALKALINITY	MANGANESE
	BICARBONATE	SODIUM
	CALCIUM	SULFATE
	CHLORIDE	SOLIDS (TDS, TSS)
	IRON	CONDUCTIVITY
	MAGNESIUM	TURBIDITY
TRACE AND TOXIC ELEMENTS	ARSENIC	LEAD
	BARIUM	MERCURY
	BERYLLIUM	NICKEL
	CADMIUM	SELENIUM
	CHROMIUM	SILVER
	COPPER	ZINC

(CONTINUED)

TABLE 1
WATER QUALITY VARIABLES (CONTD)

FUNCTIONAL GROUP	VARIABLE NAMES	
TOXIC NONMETALS	AMMONIA	CYANIDE
	CHLORINE	SULFIDE
ORGANICS	PBB'S	PHTHALATE ESTERS
	PCB'S	OIL AND GREASE
	PHENOLS	
RADIONUCLIDES	ALPHA	BETA
	RA 226	SR 90
PESTICIDES AND HERBICIDES	DDD	2,4-D
	DDE	2,4,5-TP
	DDT	HEPTACHLOR
	ALDRIN	MALATHION
	DIELDRIN	PARATHION
	CHLORDANE	TOXAPHENE
	LINDANE	SILVEX
GENERAL	pH	FLOURIDE
	POTASSIUM	

TABLE 2
BEAD COUNTY LEVEL WATER QUALITY DATA

- SIXTY VARIABLES IN TEN FUNCTIONAL GROUPS (E.G. TRACE METALS; ORGANICS)
- U.S.G.S., E.P.A. AND STATE DATA
- COMPLETE FOR NEW ENGLAND, MIDDLE ATLANTIC STATES PLUS OHIO AND PENNSYLVANIA
- TO BE EXTENDED TO ALL U.S. COUNTIES

APPLICATIONS

- CORRELATIONS WITH ENERGY PRODUCTION (USE BEAD COUNTY ENERGY DATA)
- IMPACTS OF ENERGY USE (INCORPORATE BEAD INDICATORS OF SIX MAJOR ENERGY ENERGY ACTIVITIES, E.G. PRIMARY METALS INDUSTRIES)
- HEALTH AND ENVIRONMENTAL IMPACTS ANALYSES

TABLE 3
COUNTIES IN NEW ENGLAND AND
MIDDLE ATLANTIC STATES

STATE	# OF COUNTIES IN DATABASE	# OF MISSING COUNTIES
CONNECTICUT	8	0
DELAWARE	3	0
MAINE	16	5
MARYLAND*	14	2
MASSACHUSETTS	9	1
NEW HAMPSHIRE	10	1
NEW JERSEY	21	1
NEW YORK	62	4
OHIO	88	1
PENNSYLVANIA	67	0
RHODE ISLAND	5	0
VERMONT	14	1
TOTAL	317	16

OVERALL TOTAL..... 333 COUNTIES

*INCLUDES INDEPENDENT CITIES

Table 4 - Transformed and Untransformed Variables

Variable	Units*	Raw Median	Log (Raw median +1)
Turbidity	JTU		✓
Conductivity	micromhos/cm		✓
Mn	µG/L		✓
Total Suspended Solids	MG/L		✓
Total Dissolved Solids	MG/L		✓
F	MG/L		✓
SO_4	MG/L		✓
Ca	MG/L		✓
Mg	MG/L		✓
Na	MG/L		✓
Cl	MG/L		✓
NO_3	MG/L		✓
PO_4	MG/L		✓
HCO_3	MG/L	✓	
Cu	µG/L	✓	
Fe	µG/L	✓	
K	MG/L	✓	
Total Alkalinity	MG/L	✓	
pH	STD Units	✓	

* of untransformed data

Table 5 - Distribution of Medians for County Aggregate Data (μg/ℓ)

Variable Name	Number of Counties	Water Quality Criteria*	Maximum Value	Mean	Standard Deviation	95% Confidence Interval	Skewness	Kurtosis
Arsenic	230	50	250	3.3	17	1.1-5.6	13	182
Barium	171	1000	1000	81	131	61-100	4	27
Cadmium	243	10	150	3.3	13	1.7-4.9	9	94
Copper	292	1000	200	18	28	15-21	16	257
Lead	268	50	700	12	46	7-18	4	15
Mercury	222	2	550	3.0	37	2-8	12	177
Nickel	230	100	430	26	45	20-32	4	30
Silver	173	50	30	3.9	9.6	2.4-5.3	2	3

* Quality Criteria for Water, U.S. Environmental Protection Agency, Document Number EPA-440/9-76-023 (1976).

Table 6 - Trimming the Distribution of Medians Using Criterion of Minimum Number of Observations in Each County
(units are micrograms/liter)

Variable	Threshold on minimum number of observations in each county	Number of Counties Meeting Threshold	Mean	Standard Deviation	Skewness	Kurtosis
Arsenic	5	178	1.9	2.8	2.9	11
	10	141	1.7	2.8	3.4	15
Barium	5	109	79	82	0.7	-1.3
	10	77	81	82	0.6	-1.4
Cadmium	5	172	2.0	3.8	2.8	9.1
	10	133	1.9	3.9	3.0	10
Copper	5	246	17	28	4.0	1[illegible]
	10	203	17	29	4.1	19
Lead	5	210	9.0	20	6.1	43
	10	166	6.9	12	5.7	40
Mercury	5	141	0.5	1.1	2.3	4.5
	10	110	0.5	1.1	2.1	3.6
Nickel	5	140	24	51	4.9	32
	10	102	24	55	5.1	32
Silver	5	104	5.9	12	1.5	0.4
	10	64	4.4	10	2.0	2.1

TABLE 7
CATION/ANION FACTOR ANALYSIS

1. SHARED VARIANCE (ESTIMATED COMMUNALITIES)

CATIONS		ANIONS	
MN (L)*	.47	F (L)	.63
CA (L)	.86	SO_4 (L)	.44
MG (L)	.87	CL (L)	.23
NA (L)	.50	NO_3 (L)	.52
CU	.05	PO_4 (L)	.29
FE	.41	HCO_3	.35
K	.64		
H^+	.14		

2. FACTORS WITH EIGENVALUES > 1.0 (AND % VARIANCE EXPLAINED)

CATIONS: FACTOR 1 = .43 MG + .26 CA + .25 K + .10 NA (40%)

FACTOR 2 = 1.03 MN (22%)

ANIONS: FACTOR = .50 F + .22 NO_3 + .13 HCO_3 + .12 SO_4 (54%)

TABLE 8
STEPWISE REGRESSION OF TOTAL ALKALINITY ON ANION/CATION GROUPS

STEP 1. $T.ALK. = 0.76\ HCO_3$ ($R^2 = .95$)

STEP 2. $T.ALK. = 0.69\ HCO_3 + 0.37\ CA$ ($R^2 = .96$)

STEP 3. $T.ALK. = 0.68\ HCO_3 + 0.51\ CA - 0.20\ CL$ ($R^2 = .97$)

STEP 4. $T.ALK. = 0.64\ HCO_3 + 0.63\ CA - 0.20\ CL - 0.01\ MN$ ($R^2 = .97$)

USING ALL CATION/ANION SPECIES, THERE RESULTS

$$T.ALK. = .60\ HCO_3 + .43\ CA - .22\ CL + 10.0\ PO_4 + \Sigma\ (10\ \text{LESSER VARIABLES})$$

($R^2 = .98$)

FIGURES

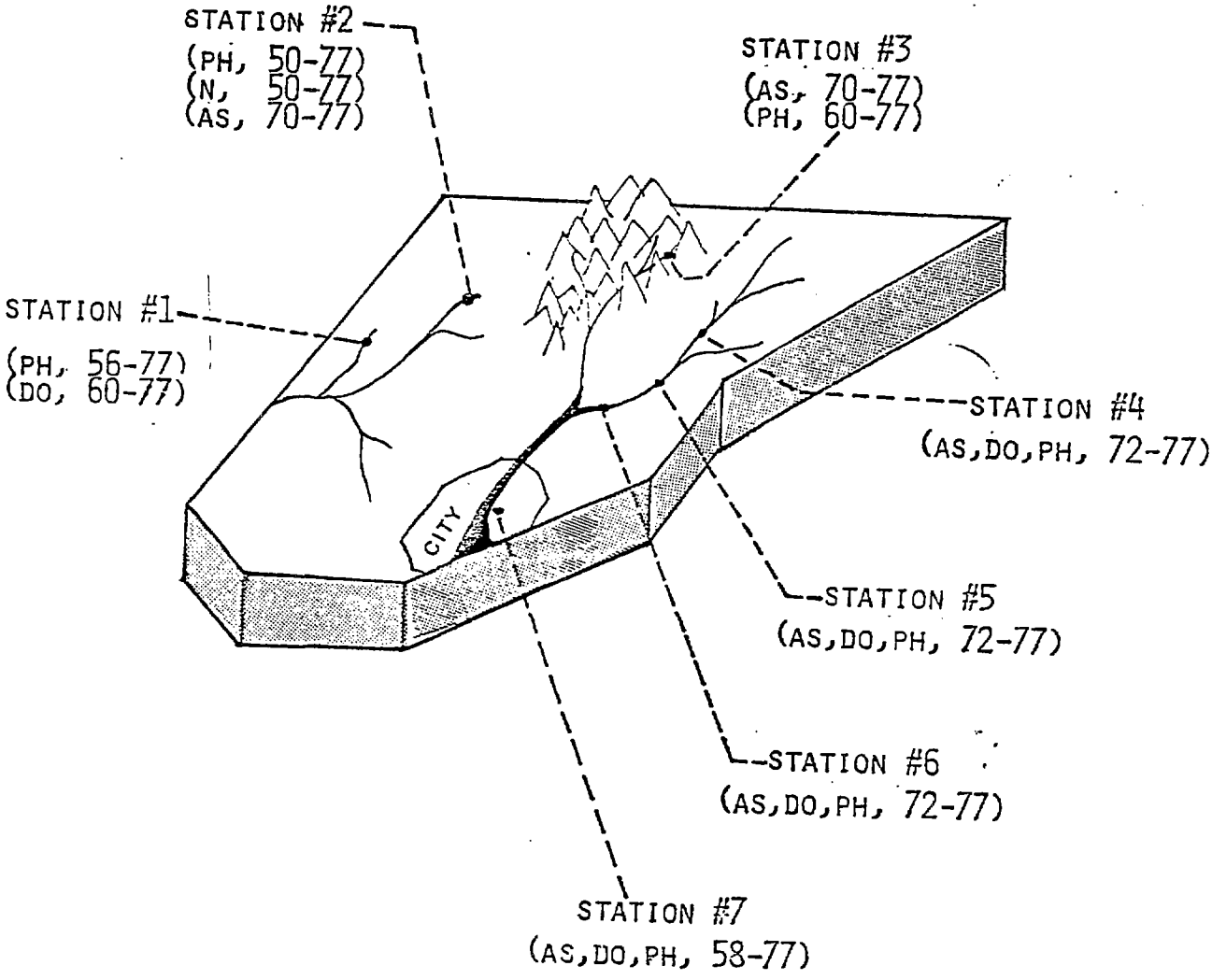

STATION #2
(PH, 50-77)
(N, 50-77)
(AS, 70-77)
STATION #3
(AS, 70-77)
(PH, 60-77)
STATION #1
(PH, 56-77)
(DO, 60-77)
STATION #4
(AS,DO,PH, 72-77)
CITY
STATION #5
(AS,DO,PH, 72-77)
STATION #6
(AS,DO,PH, 72-77)
STATION #7
(AS,DO,PH, 58-77)

FIGURE 1

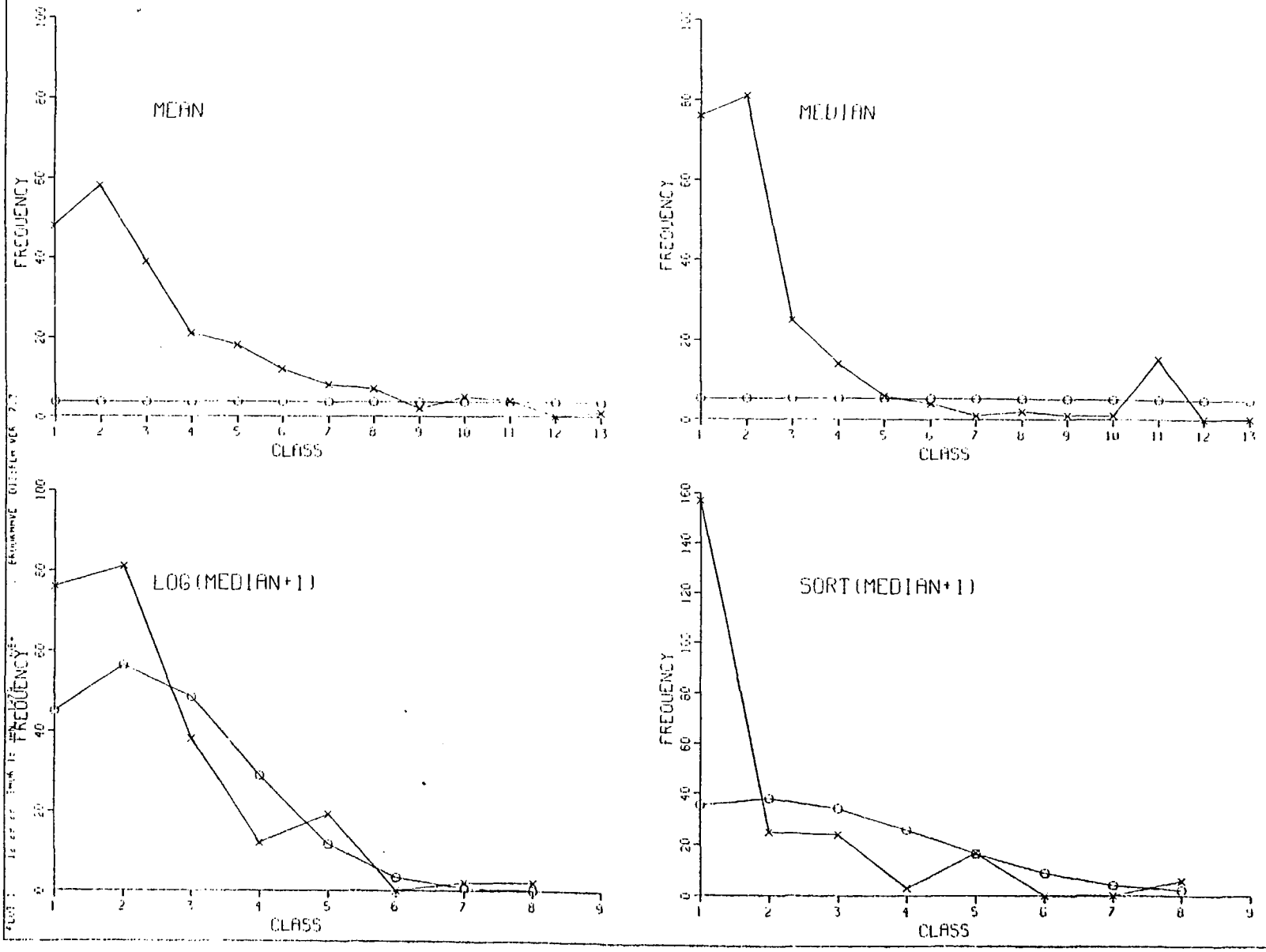
```
PGM=MSP,PURP=104/EPA,RMT=48,PRT=NO,
HEAD=TOXIC SUBSTANCE ANALYSIS--NEW ENGLAND;NY;NJ;PA;MD;DEL;OHIO,
EXTRACT=(1075 OR 1077 OR 720 OR 1090 OR 1092 OR
950 OR 951 OR 50060 OR 50064 OR 50066 OR
745 OR 746 OR 71875) AND TYPA AND AMBNT AND (LAKE OR STREAM),
AGGR=COUNTY,STC=9,STC=10,STC=13,STC=24,STC=25,STC=33,STC=34,
STC=36,STC=39,STC=42,STC=44,STC=50,
R= Z,
P=1075,PM=1077,RR,
P=1090,PM=1092,RR,
P=720,RR,
P=950,PM=951,RR,
P=50060,PM=50064,PM=50066,RR,
P=745,P=746,P=71875,RR,
MSPARMS=MODELL,%TILE15T=1,%TILE50 =5,%TILE85T=8,MEAN=M,MUL=*,
M+1S=+,MAX=H,MIN=L,
MODELR,%TILE15T=A,%TILE50T=B,%TILE85T=C,MAX=D,MIN=E,MUL=*,
M+1S=F,MEAN=G,
PLOT,LEFT1,P1>-1,AND,P1<5000,RIGHT2,P2>-1,AND,P2<10,
PLOT,LEFT3,P3>-1,AND,P3<10000,RIGHT4,P4>-1,AND,P4<15,
PLOT,LEFT5,P5>-1,AND,P5<5,RIGHT6,P6>-1,AND,P6<1500,
SAVEOUT,
STOPMSP,
./EPAEKK   JOB (BECE,DIS,,STORET,,),MSGLEVEL=(1,1),MSGCLASS=P TIME=2
**JOBPARM LINES=60
./S EXEC PGM=WOROOT,REGION=650K
./SOMSP DD DISP=(,KEEP),UNIT=(2400-3,,DEFER),
./ DSN=FILE3,DCB=(RECFM=VB,LRECL=110,BLKSIZE=5900),
./ LABEL=RETPD=365
./FCF DD UNIT=SYSDA,SPACE=(CYL,(95,95)),
./       DCB=(RECFM=VB,LRECL=309,BLKSIZE=6180,BUFNO=1)
./CF DD UNIT=SYSDA,SPACE=(TRK,(200,100))
./CSNRF DD DISP=(,PASS),UNIT=SYSDA,DSN=&RCSN
./  SPACE=(TRK,(50,50)),
./  DCB=(RECFM=FB,BLKSIZE=12000,LRECL=4,BUFNO=1)
./CDFTR DD DISP=(,PASS),UNIT=SYSDA,DSN=&FTR,
./  SPACE=(TRK,(50,50)),
./  DCB=(RECFM=FB,BLKSIZE=6400,LRECL=80,BUFNO=1)

# ALL KEYWORDS BETWEEN MSPARMS AND STOPMSP WILL BE PRINTED
BELOW ALONG WITH ANY SYNTAX ERRORS #
```

FIGURE 2

Figure 3 — ARSENIC FREQUENCY PLOTS

X - OBSERVED FREQ
O - EXPECTED FREQ



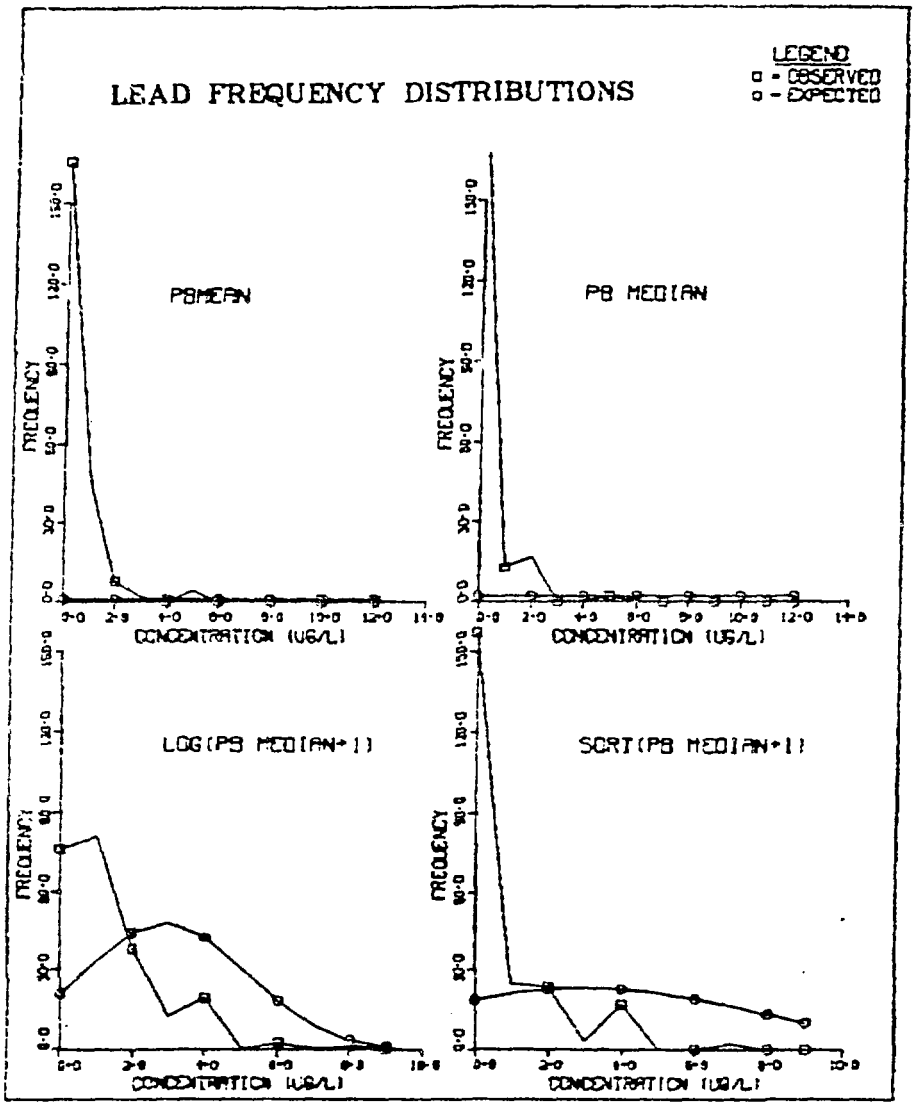

LEAD FREQUENCY DISTRIBUTIONS
LEGEND
□ - OBSERVED
○ - EXPECTED
PBMEAN
PB MEDIAN
LOG(PB MEDIAN+1)
SQRT(PB MEDIAN+1)
FREQUENCY
CONCENTRATION (UG/L)

FIGURE 4

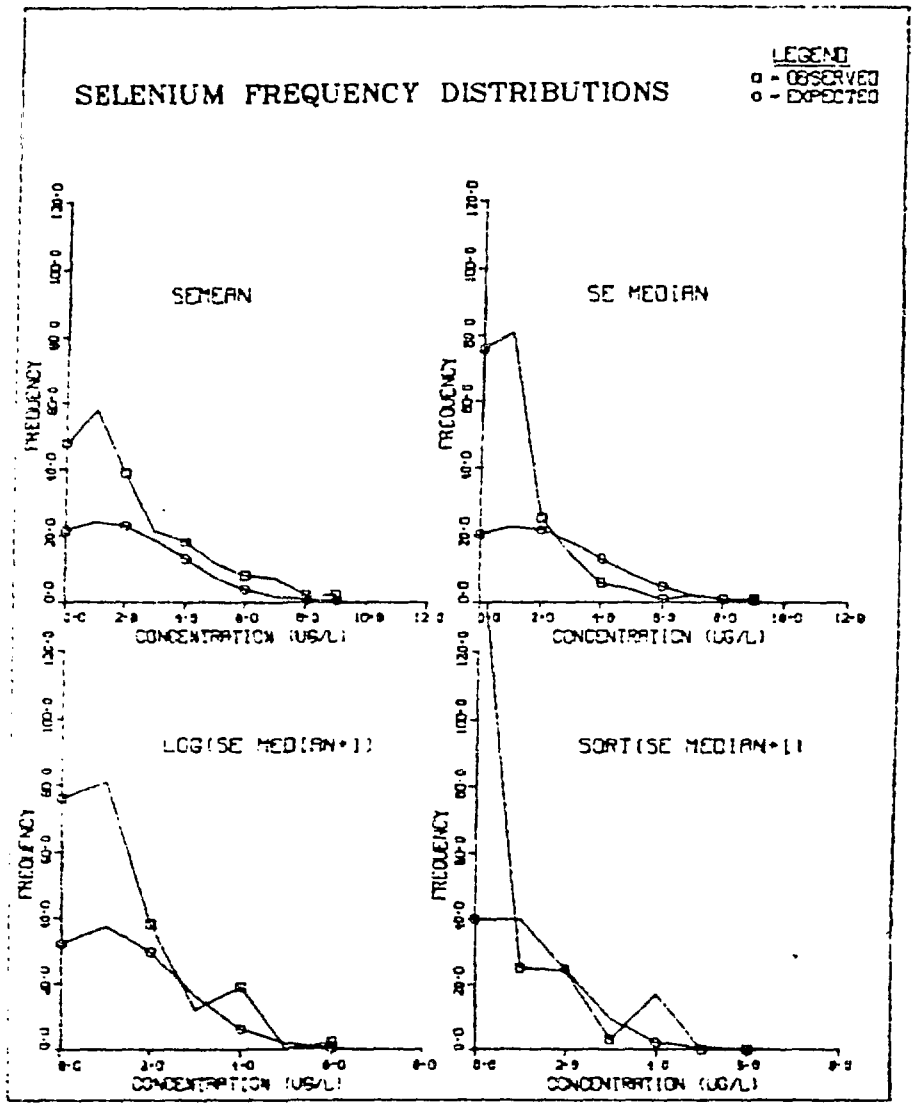
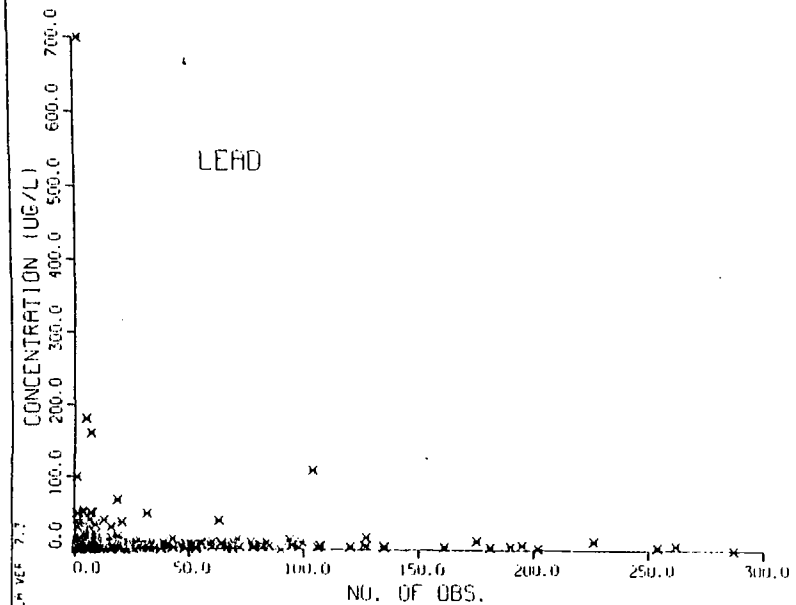
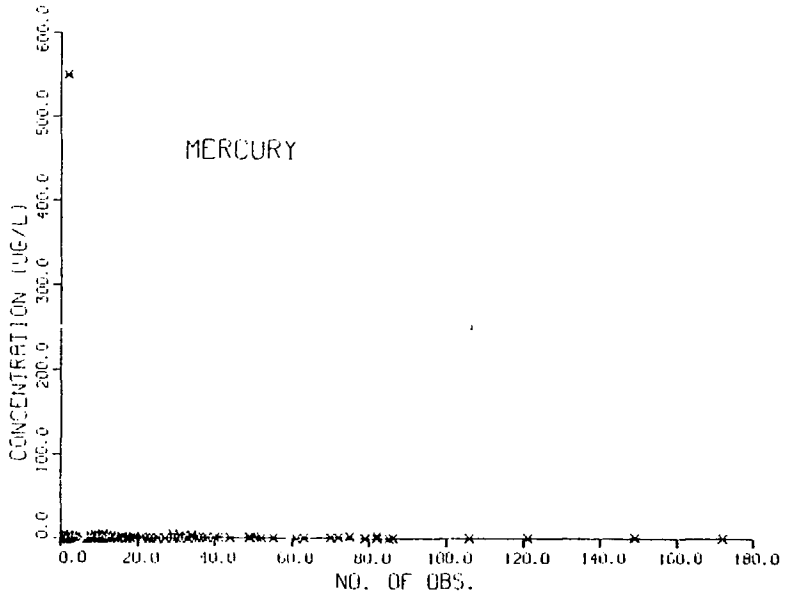
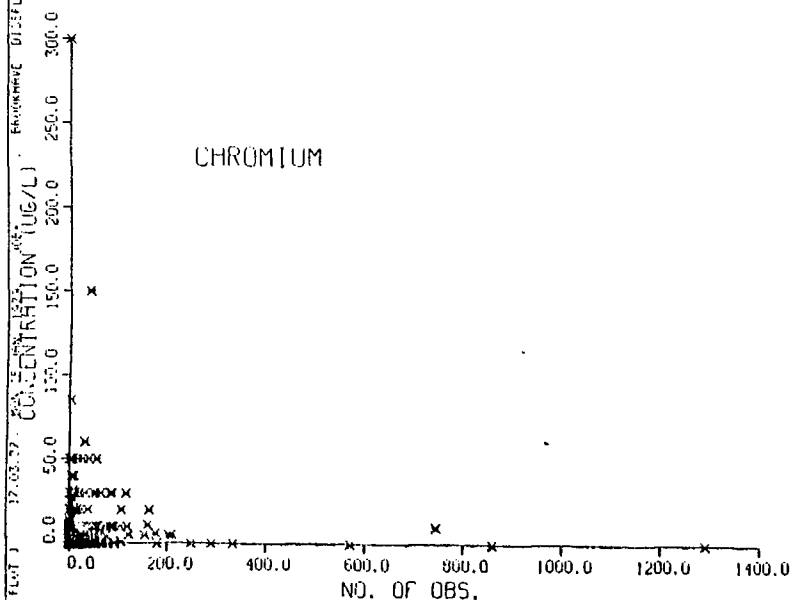

SELENIUM FREQUENCY DISTRIBUTIONS
LEGEND
□ - OBSERVED
○ - EXPECTED
SEMEAN
SE MEDIAN
LOG(SE MEDIAN+1)
SQRT(SE MEDIAN+1)
FREQUENCY
CONCENTRATION (UG/L)

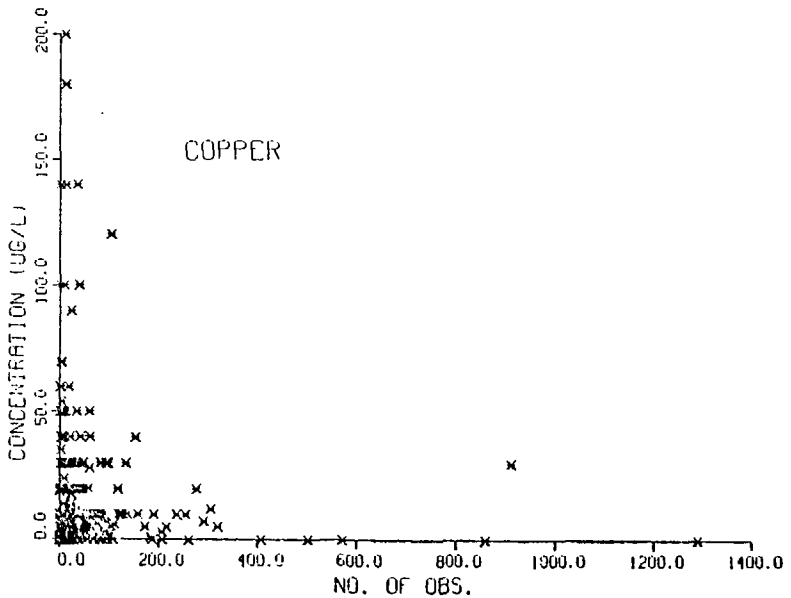
FIGURE 5

figure 6 - MEDIANS vs. NUMBER OF OBSERVATIONS









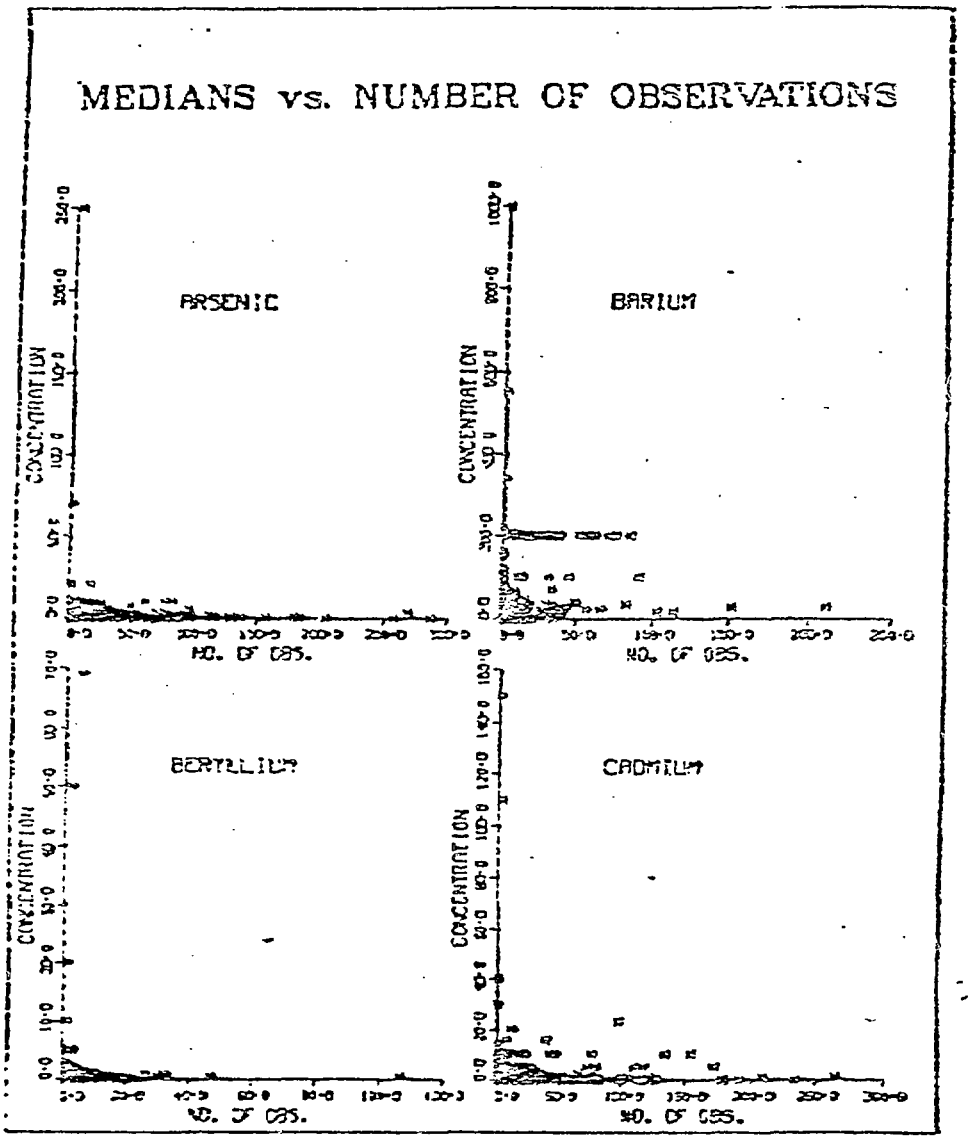
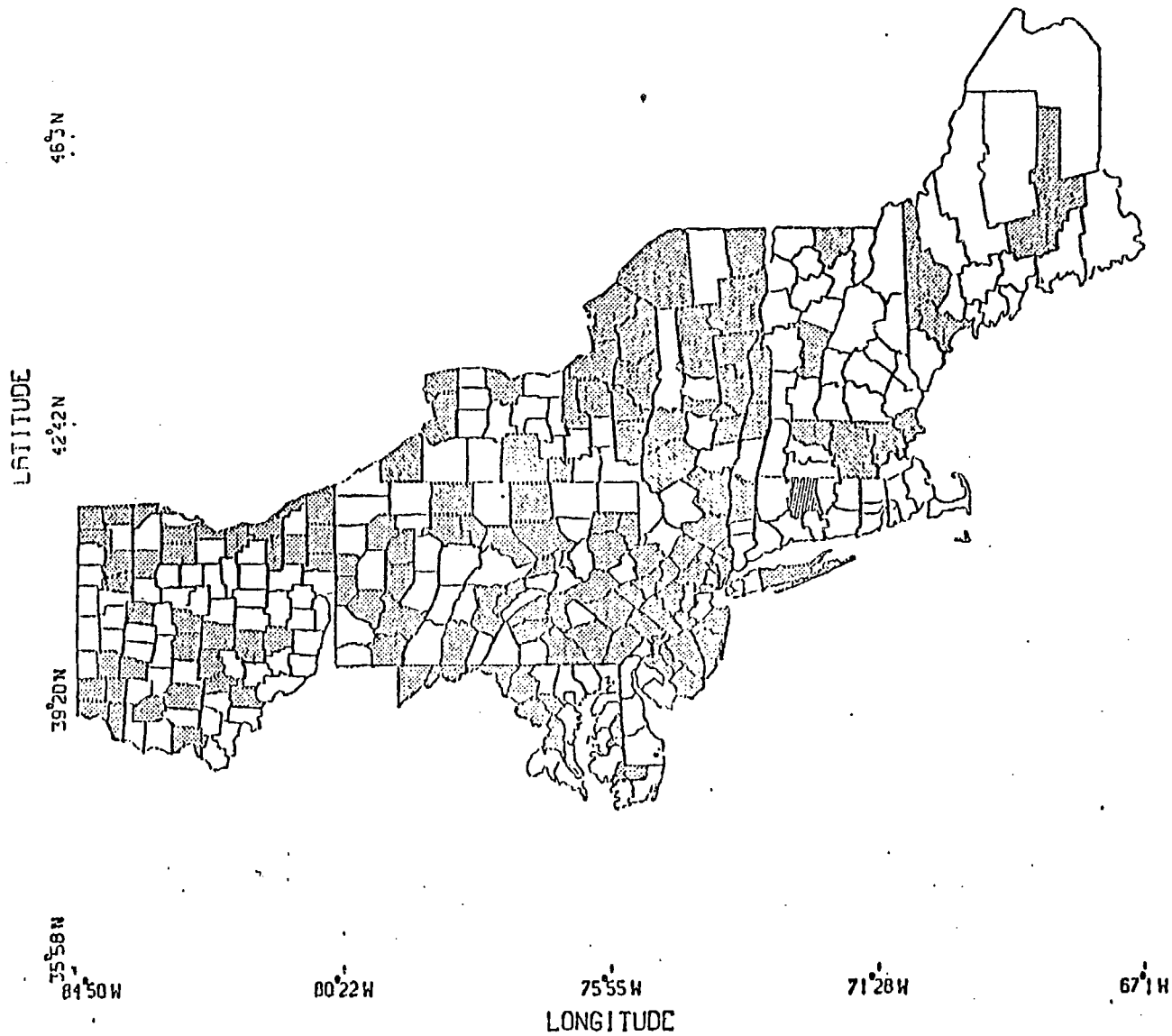

MEDIANS vs. NUMBER OF OBSERVATIONS
ARSENIC
BARIUM
BERYLLIUM
CADMIUM
CONCENTRATION
NO. OF OBS.

FIGURE 7

FIGURE 8 138 COUNTY SUBSET WITH 19 NON-MISSING WATER PARAMETERS



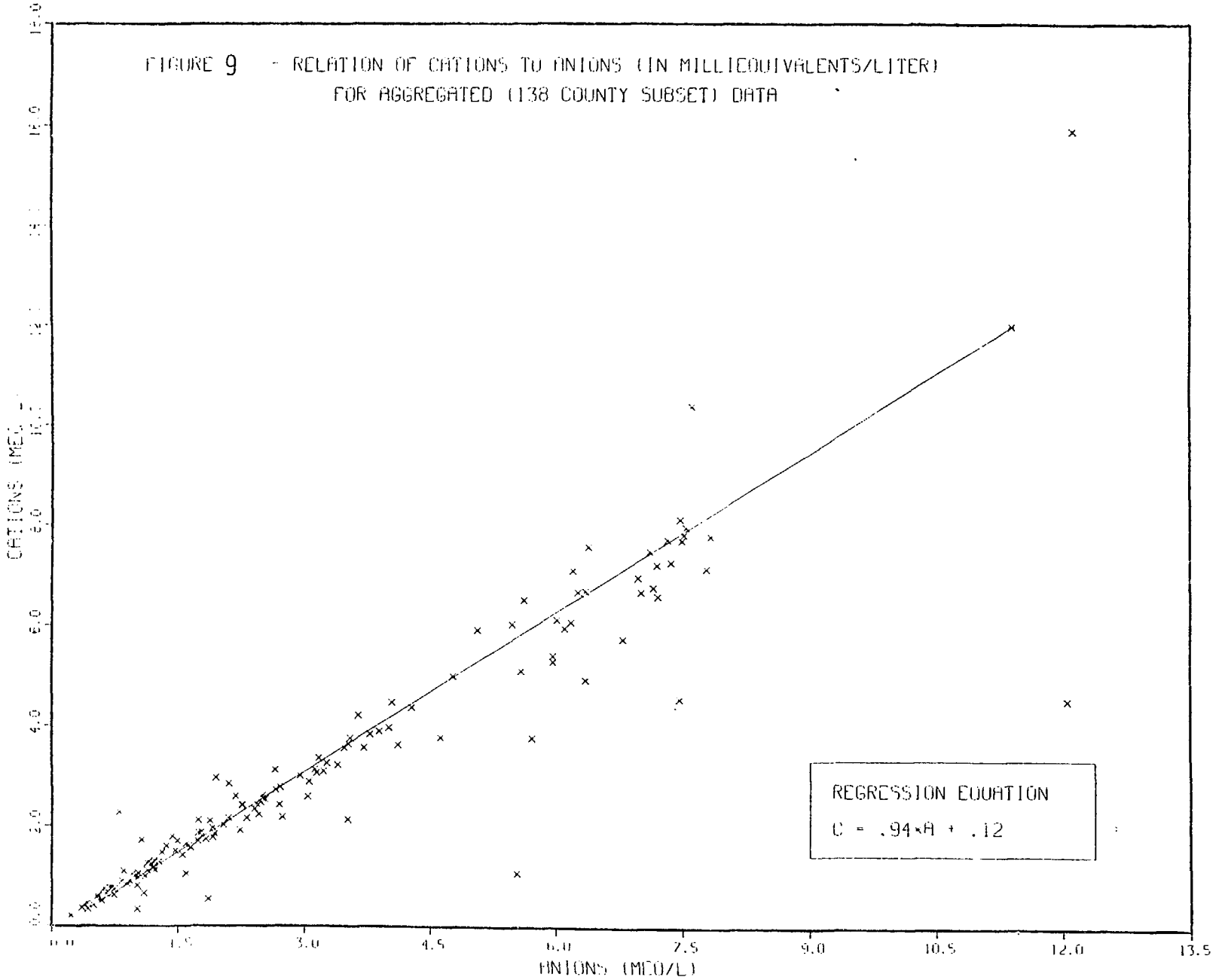
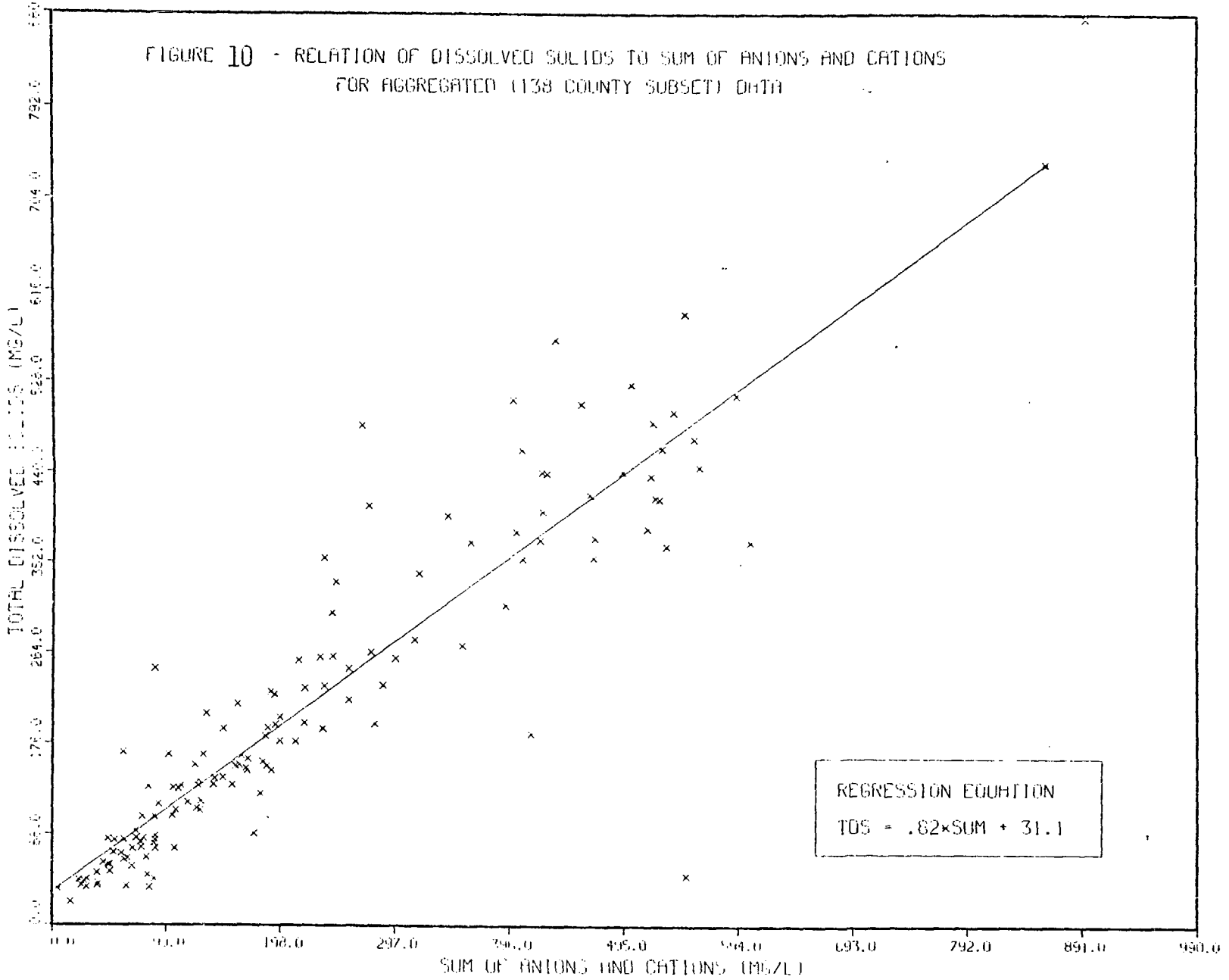


FIGURE 9 - RELATION OF CATIONS TO ANIONS (IN MILLIEQUIVALENTS/LITER)
FOR AGGREGATED (138 COUNTY SUBSET) DATA
CATIONS (MEQ/L)
ANIONS (MEQ/L)
0.0
1.5
3.0
4.5
6.0
7.5
9.0
10.5
12.0
13.5
REGRESSION EQUATION
C = .94*A + .12

FIGURE 10 - RELATION OF DISSOLVED SOLIDS TO SUM OF ANIONS AND CATIONS FOR AGGREGATED (138 COUNTY SUBSET) DATA



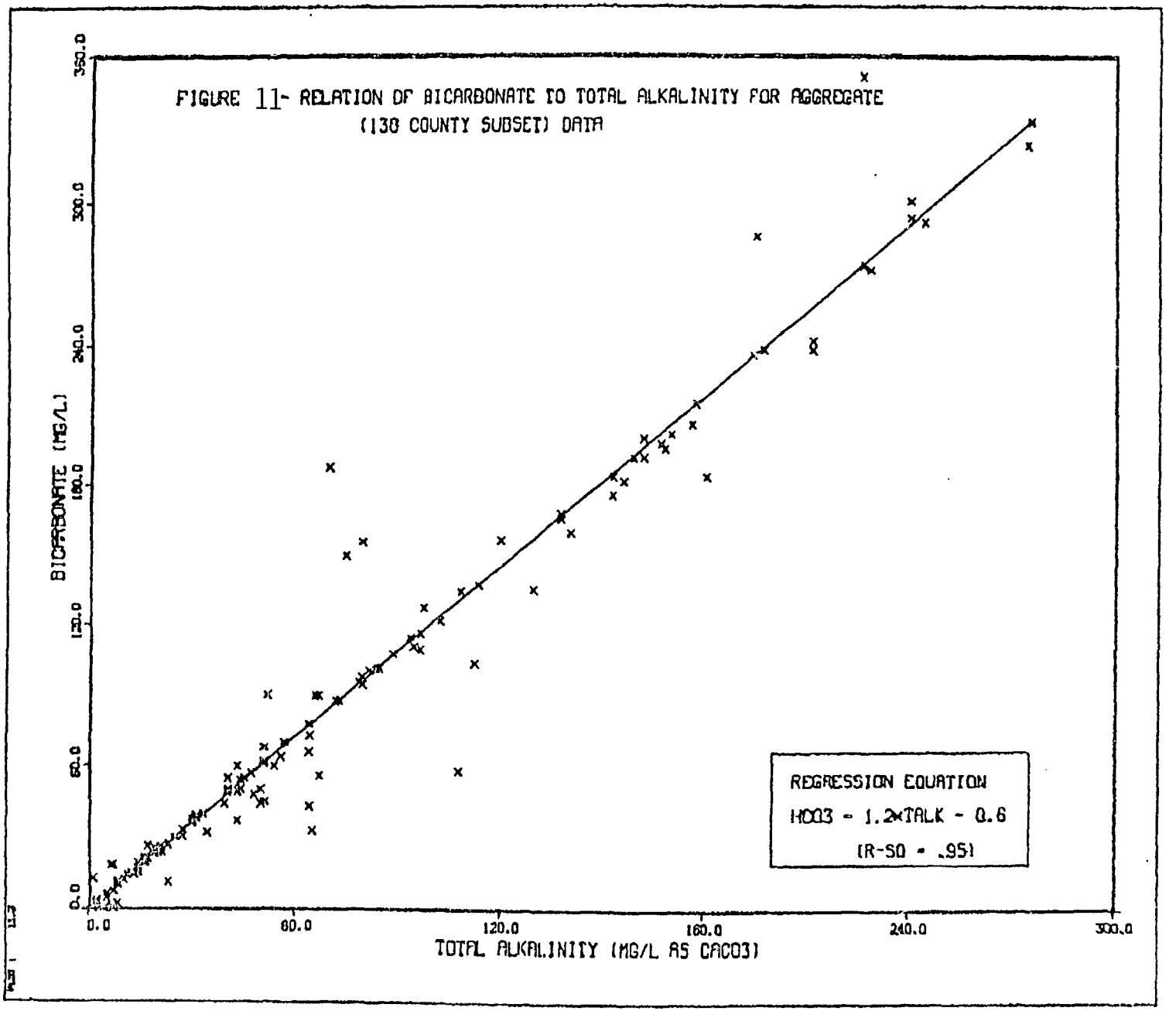


FIGURE 11- RELATION OF BICARBONATE TO TOTAL ALKALINITY FOR AGGREGATE (130 COUNTY SUBSET) DATA

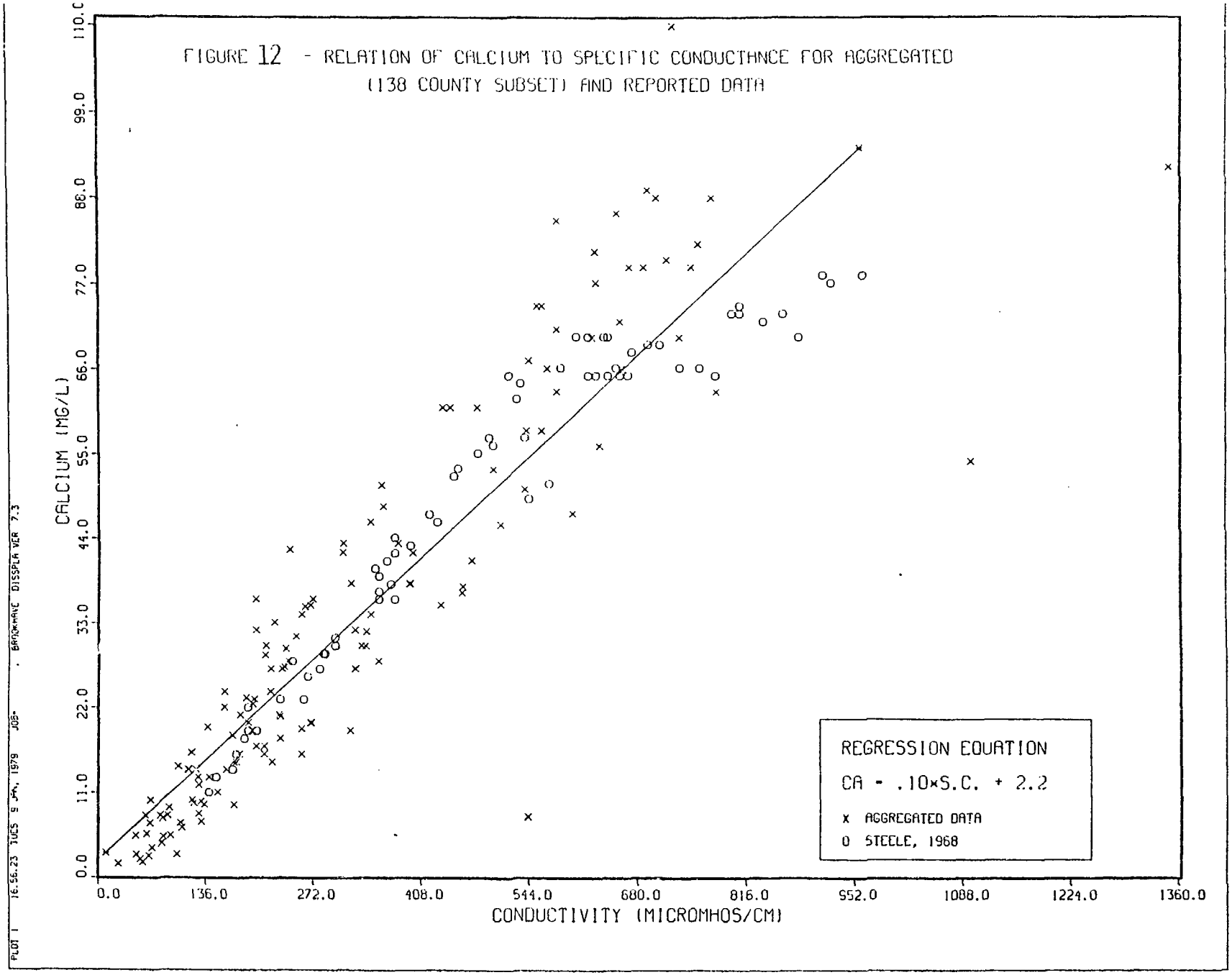


FIGURE 12 - RELATION OF CALCIUM TO SPECIFIC CONDUCTANCE FOR AGGREGATED (138 COUNTY SUBSET) AND REPORTED DATA
CALCIUM (MG/L)
0.0
11.0
22.0
33.0
44.0
55.0
66.0
77.0
88.0
99.0
110.0
CONDUCTIVITY (MICROMHOS/CM)
0.0
136.0
272.0
408.0
544.0
680.0
816.0
952.0
1088.0
1224.0
1360.0
REGRESSION EQUATION
CA = .10*S.C. + 2.2
x AGGREGATED DATA
O STEELE, 1968

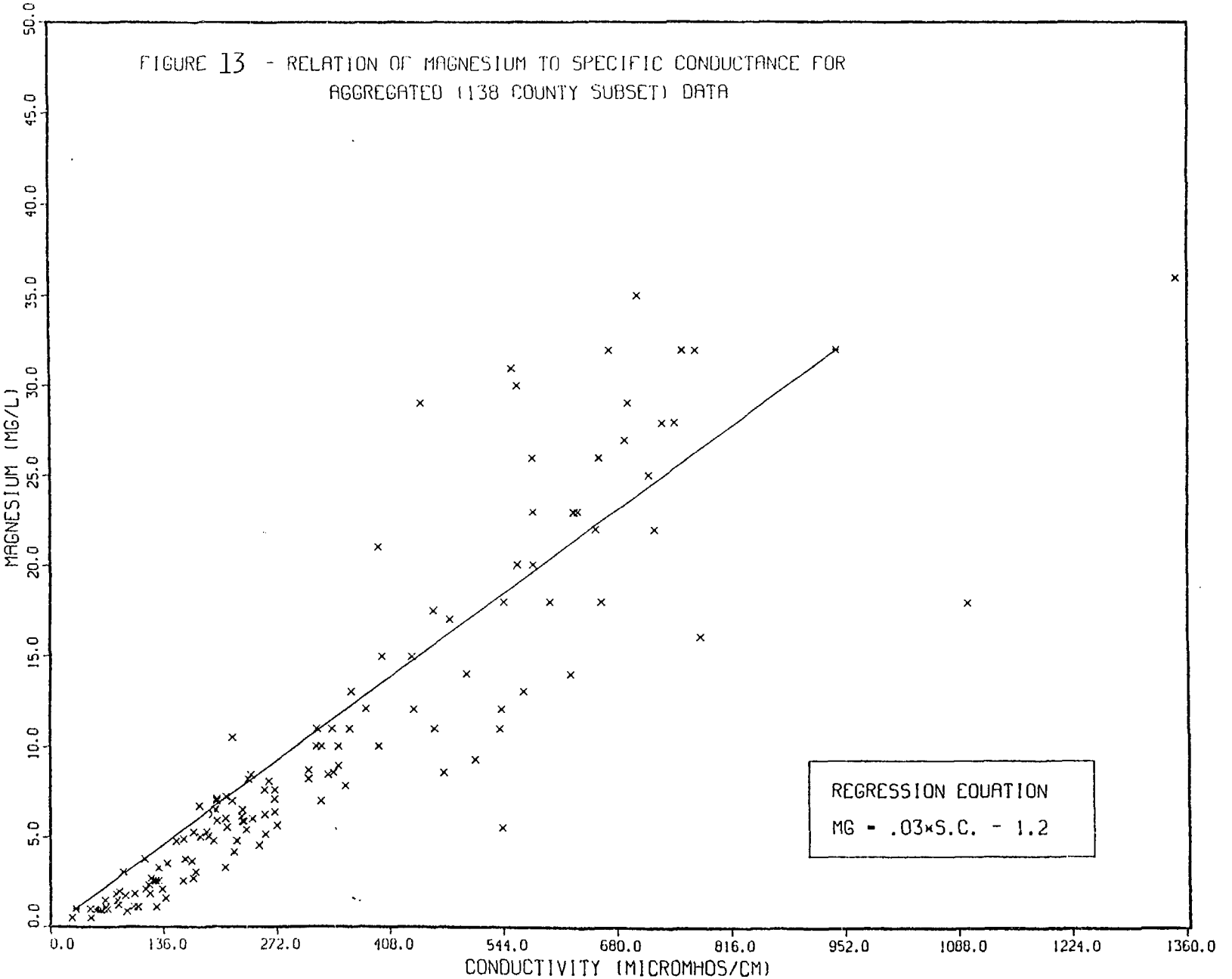


FIGURE 13 - RELATION OF MAGNESIUM TO SPECIFIC CONDUCTANCE FOR AGGREGATED (138 COUNTY SUBSET) DATA

PLOT 3 17.02.52 TUES 9 JAN, 1979 JOB- . BROOKHAVE DISSPLA VER 7.3

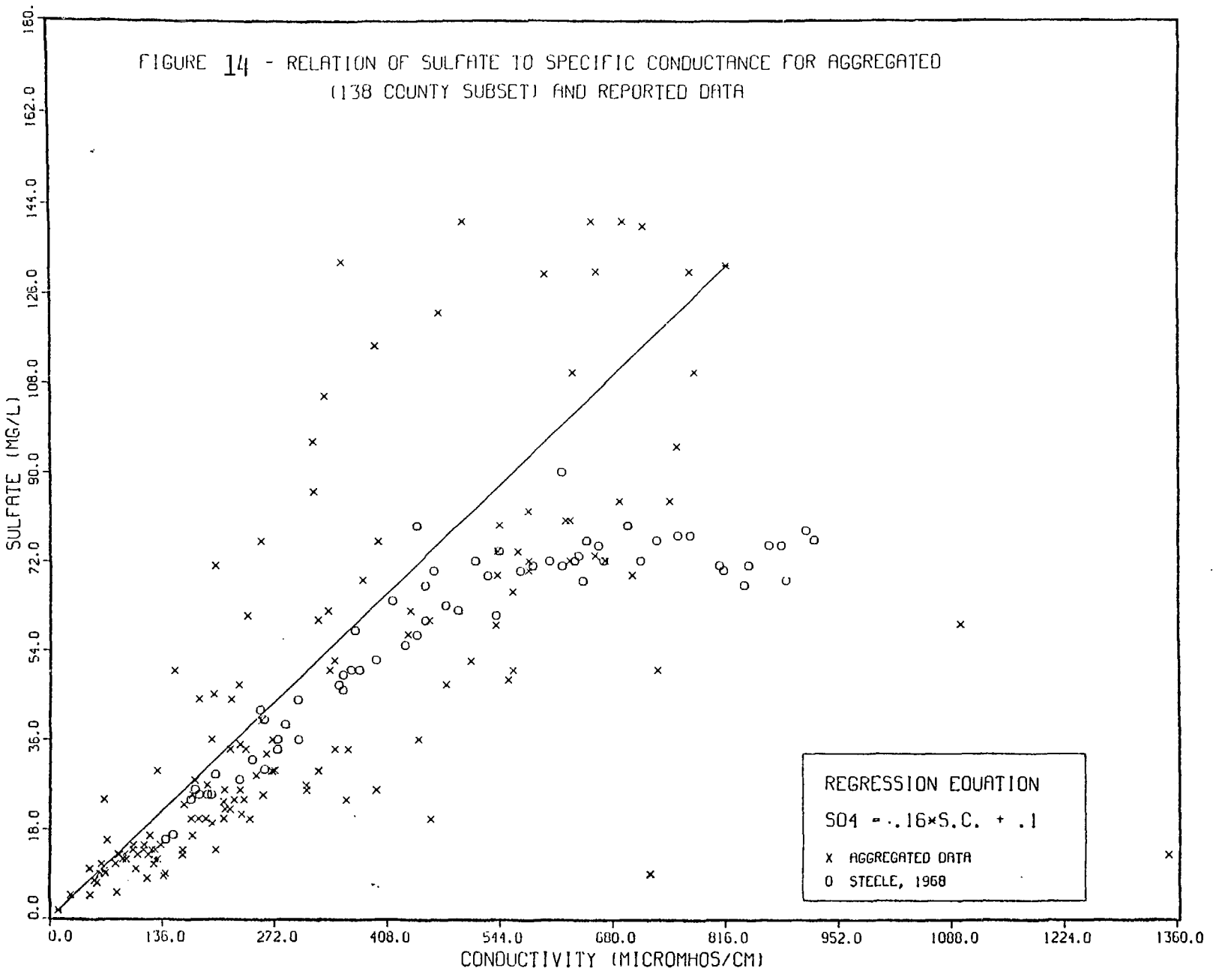


FIGURE 14 - RELATION OF SULFATE TO SPECIFIC CONDUCTANCE FOR AGGREGATED (138 COUNTY SUBSET) AND REPORTED DATA

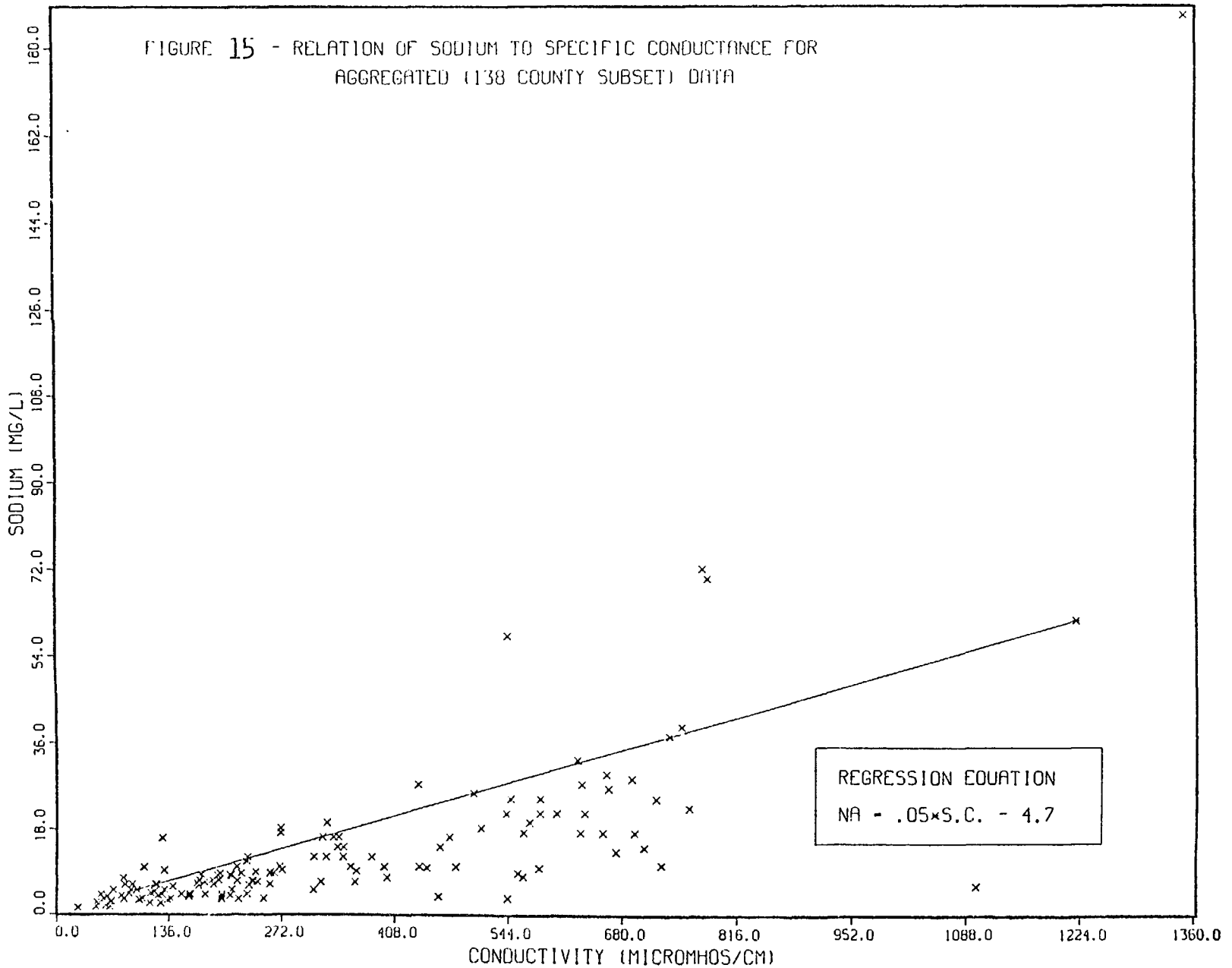


FIGURE 15 - RELATION OF SODIUM TO SPECIFIC CONDUCTANCE FOR AGGREGATED (138 COUNTY SUBSET) DATA

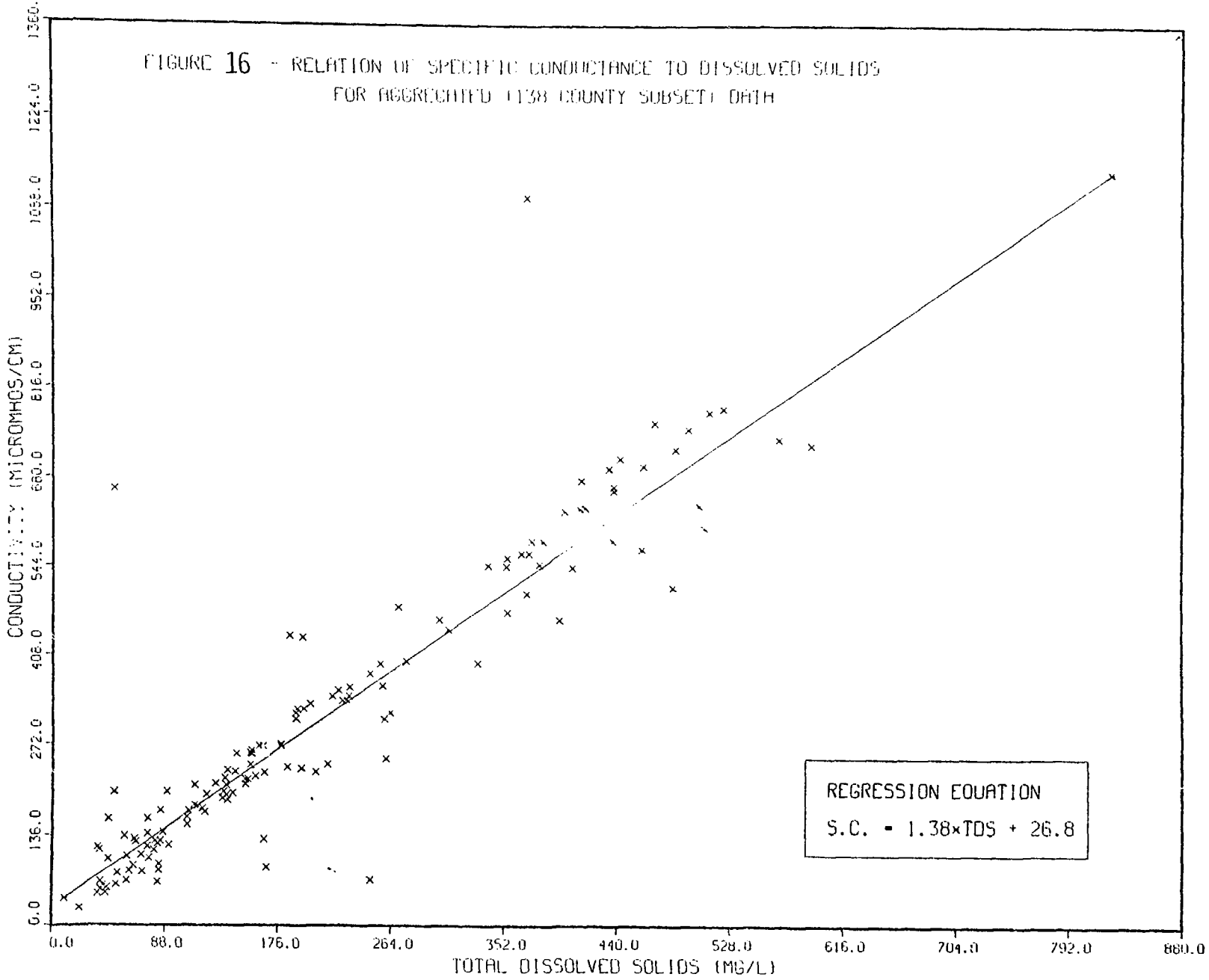
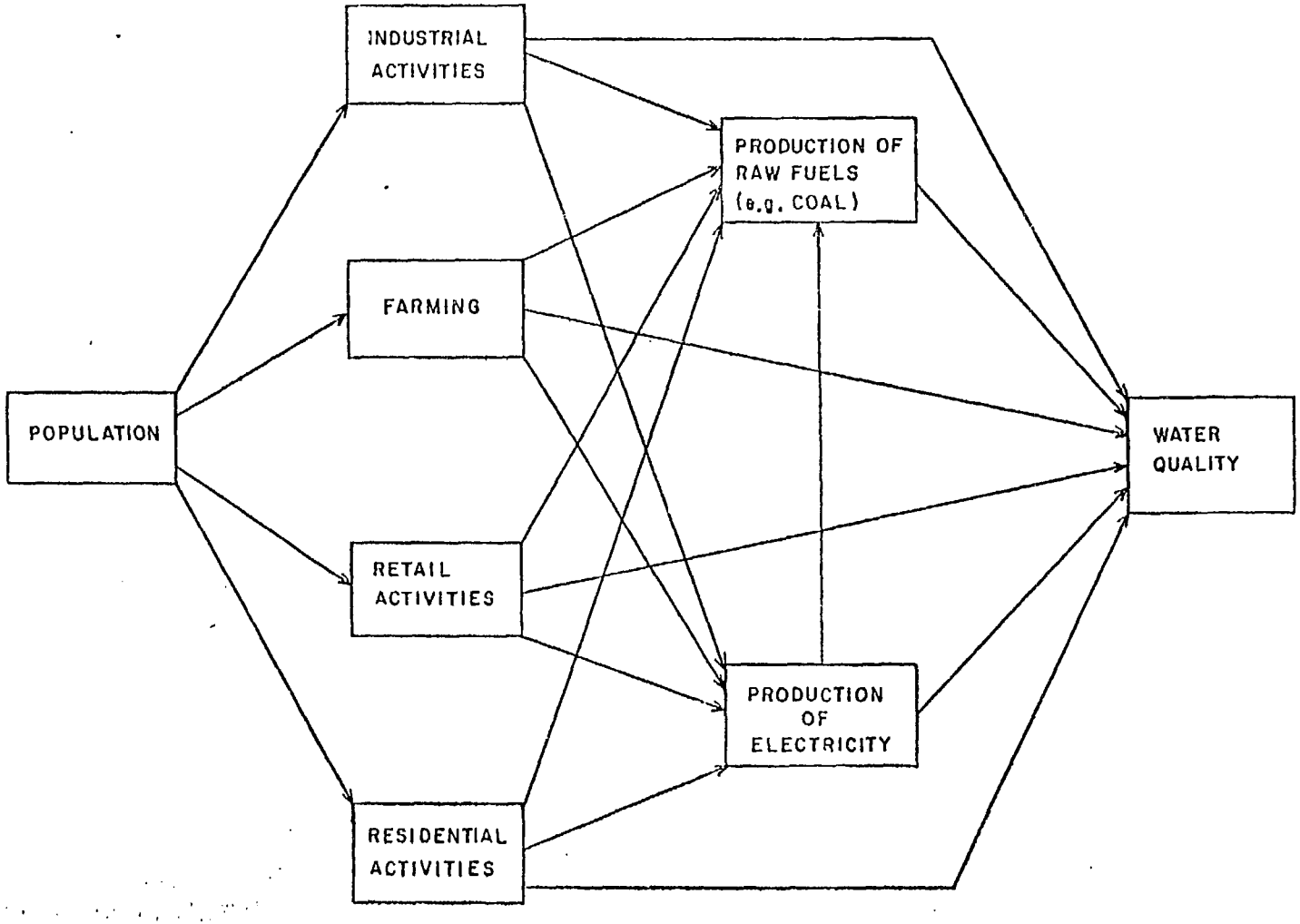


FIGURE 16 - RELATION OF SPECIFIC CONDUCTANCE TO DISSOLVED SOLIDS
FOR AGGREGATED (138 COUNTY SUBSET) DATA
CONDUCTIVITY (MICROMHOS/CM)
0.0
136.0
272.0
408.0
544.0
680.0
816.0
952.0
1088.0
1224.0
1360.
TOTAL DISSOLVED SOLIDS (MG/L)
0.0
88.0
176.0
264.0
352.0
440.0
528.0
616.0
704.0
792.0
880.0
REGRESSION EQUATION
S.C. = 1.38×TDS + 26.8



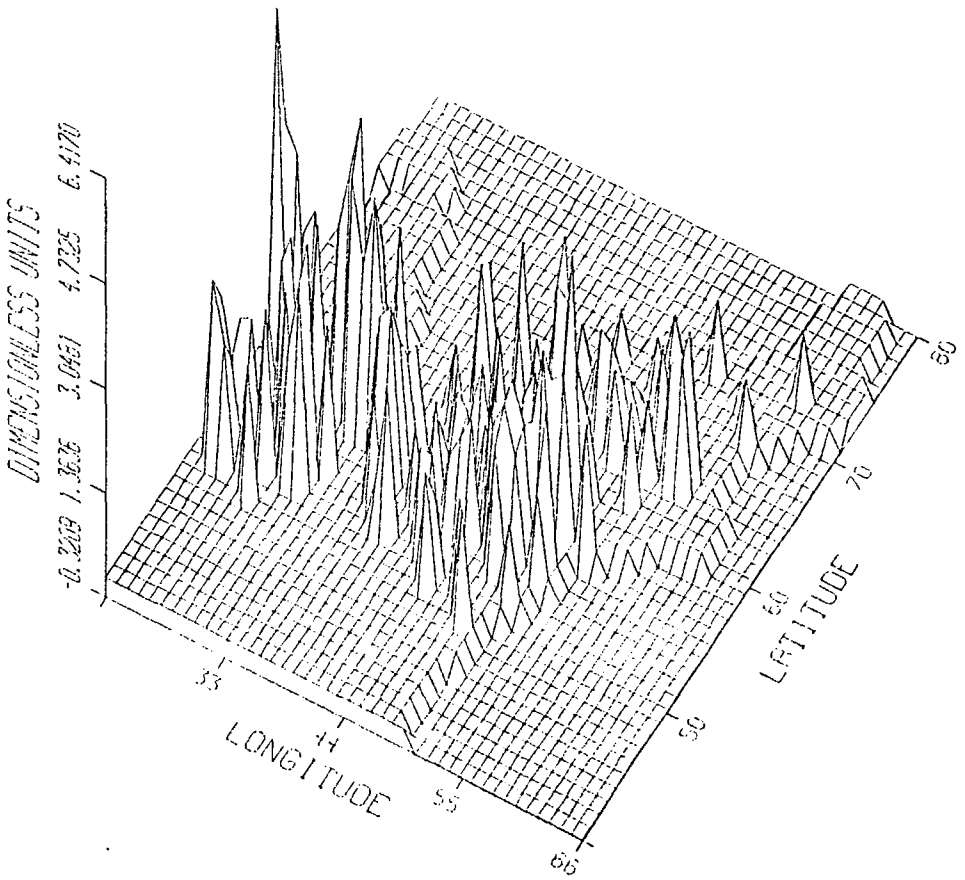
ENERGY DEMAND ENERGY SUPPLY

FIGURE 17 CONCEPTUALIZED PATH DIAGRAM

FIGURE 18 - WATER QUALITY FOR NE U.S.

FACTOR 1= F(K+NH+F+CL+TDS+COND.+SO4) + F(12 LESSER VARIABLES)

(VARIABLES ARE STANDARDIZED, FACTORS ARE DIMENSIONLESS)

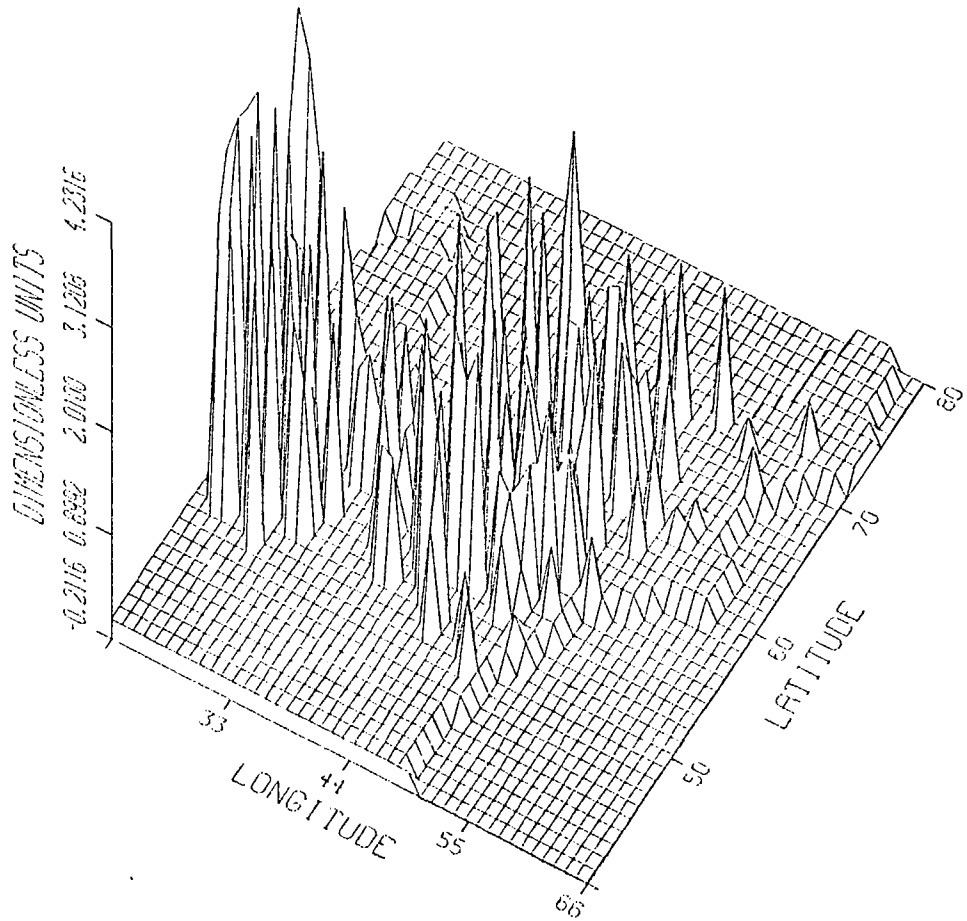


DR. ED KAPLAN, BIOMEDICAL AND ENVIRONMENTAL ASSESSMENT DIVISION
BROOKHAVEN NATIONAL LABORATORY 1-11-79

FIGURE 19 WATER QUALITY FOR NE U.S.

FACTOR 2= F(HCO3+ALKALINITY+CA) + F(16 LESSER VARIABLES)

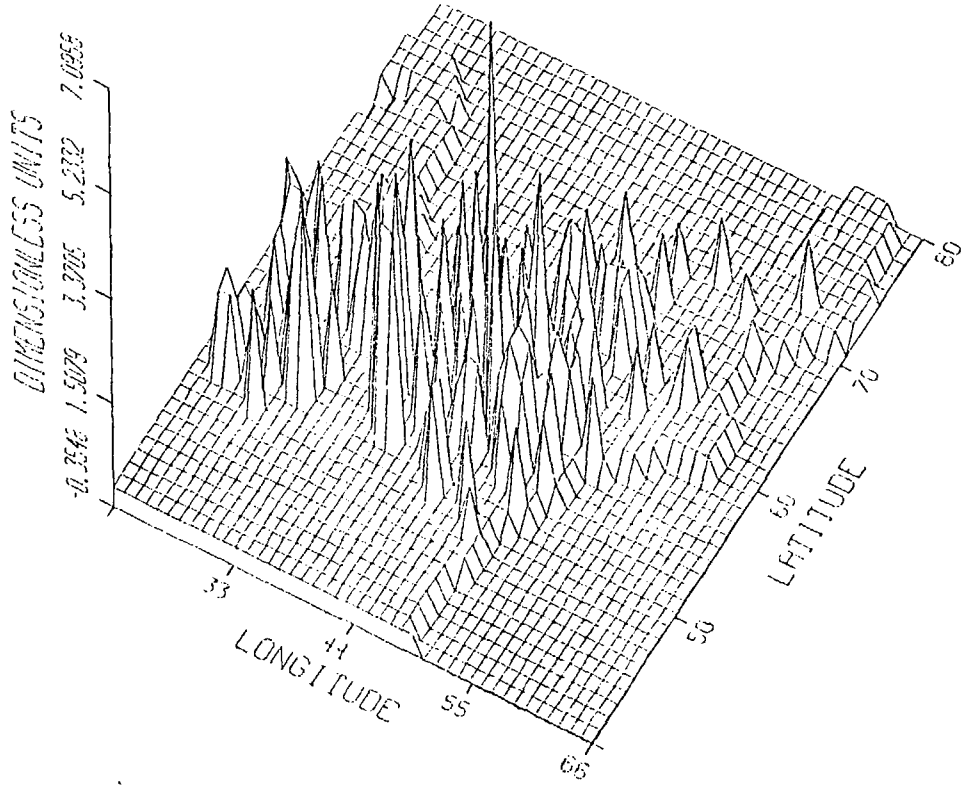
(VARIABLES ARE STANDARDIZED, FACTORS ARE DIMENSIONLESS)



DR. ED KAPLAN, BIOMEDICAL AND ENVIRONMENTAL ASSESSMENT DIVISION
BROOKHAVEN NATIONAL LABORATORY 1-11-79

FACTOR 3= F(MN+FE+SO4) + F(16 LESSER VARIABLES)

(VARIABLES ARE STANDARDIZED, FACTORS ARE DIMENSIONLESS)



DR. ED KAPLAN, BIOMEDICAL AND ENVIRONMENTAL ASSESSMENT DIVISION
BROOKHAVEN NATIONAL LABORATORY 1-11-79

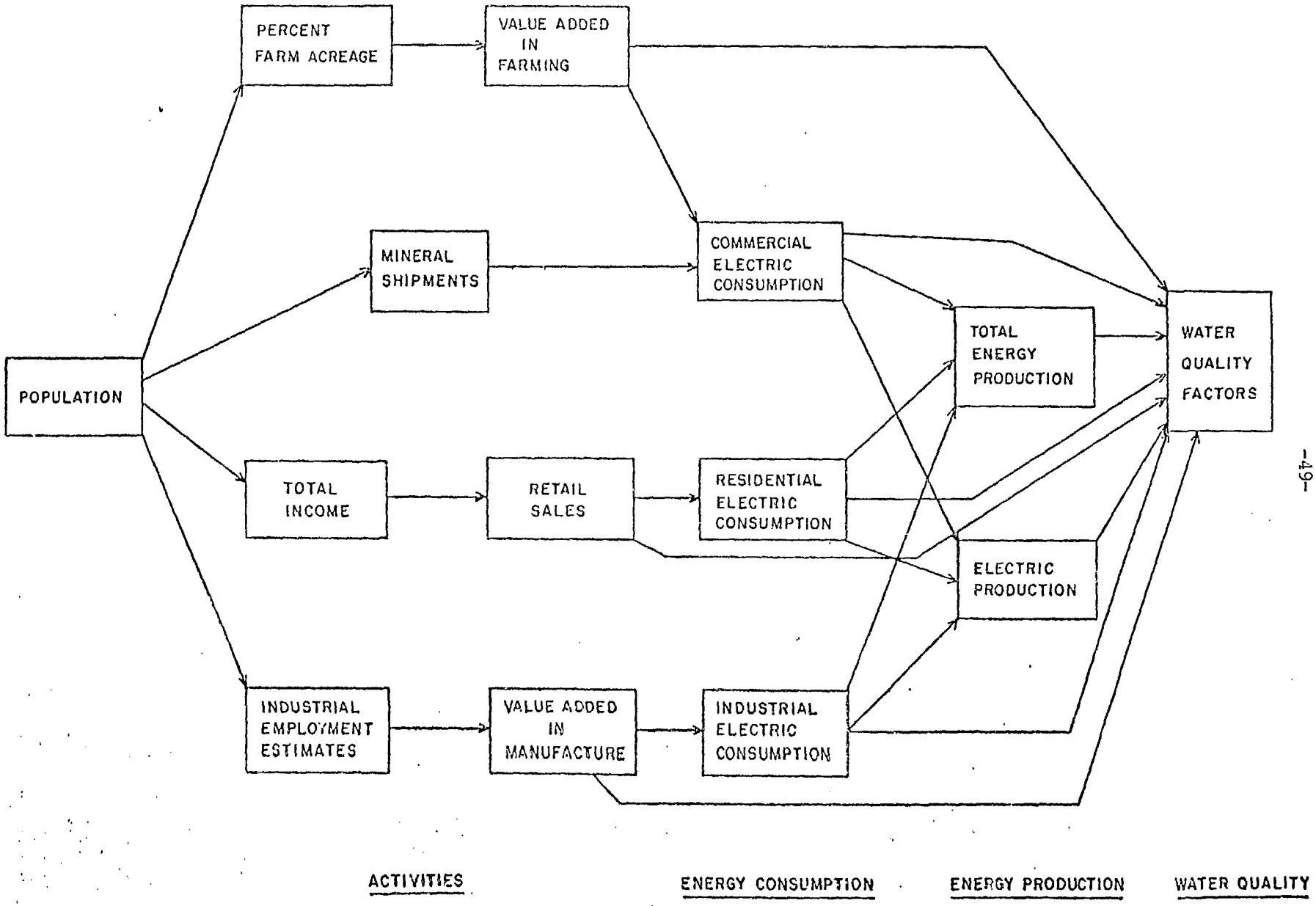


FIGURE 21 - PATH DIAGRAM FOR WATER QUALITY FACTORS AND AVAILABLE SOCIOECONOMIC/ENERGY DATA

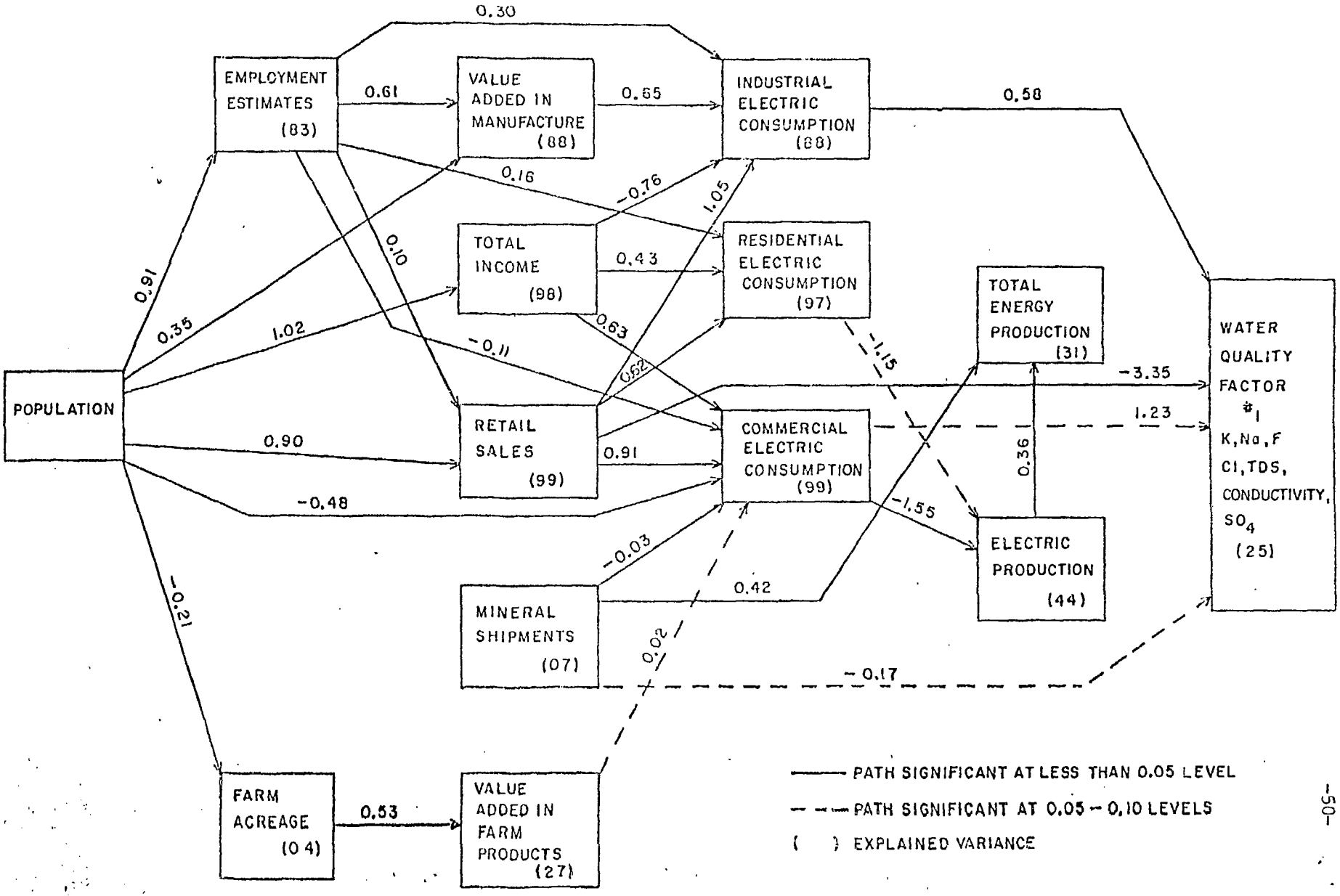


FIGURE 22 CAUSAL DIAGRAM FOR ENERGY, SOCIOECONOMIC AND WATER QUALITY

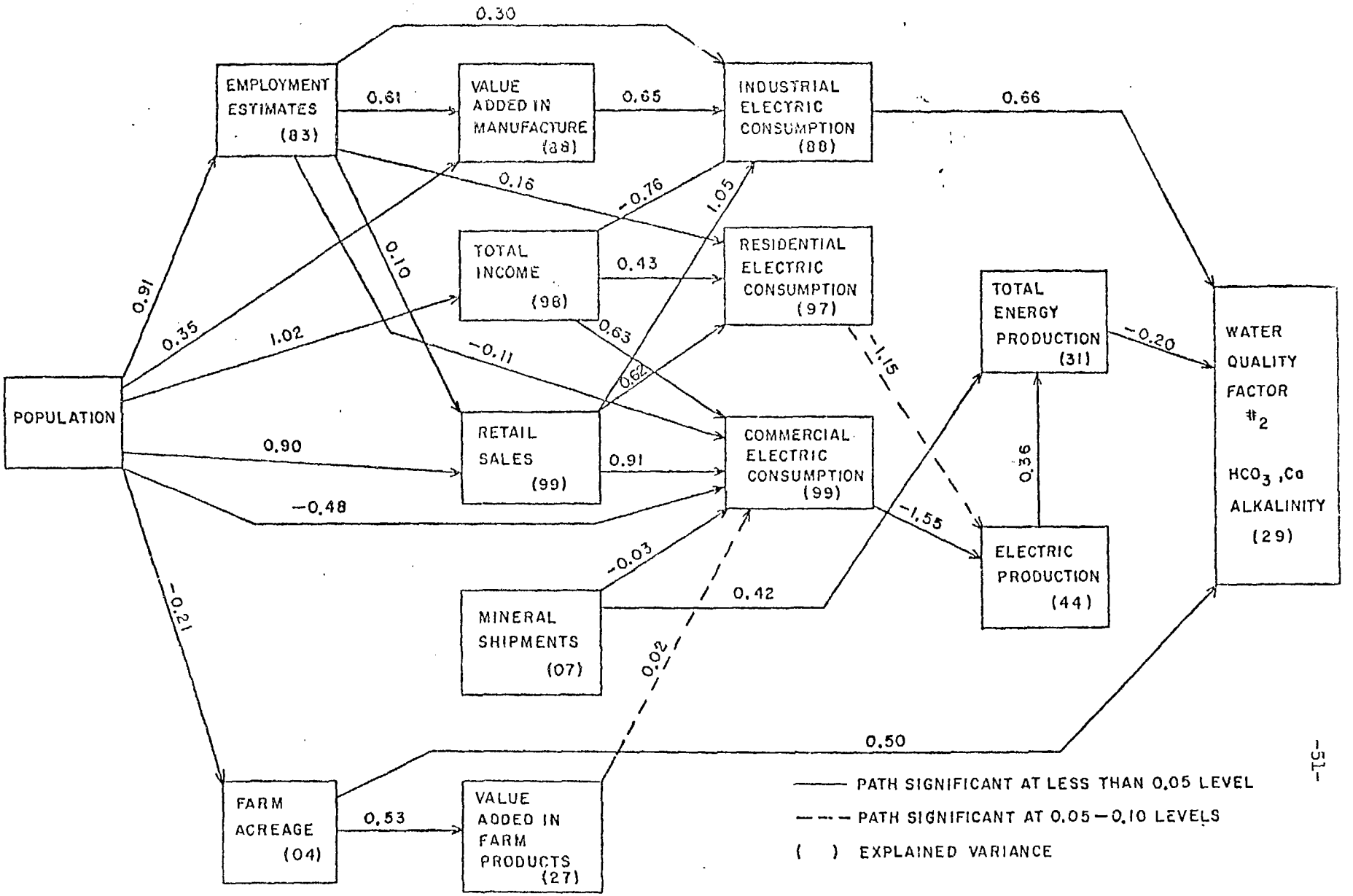


FIGURE 23 CAUSAL DIAGRAM FOR ENERGY, SOCIOECONOMIC AND WATER QUALITY

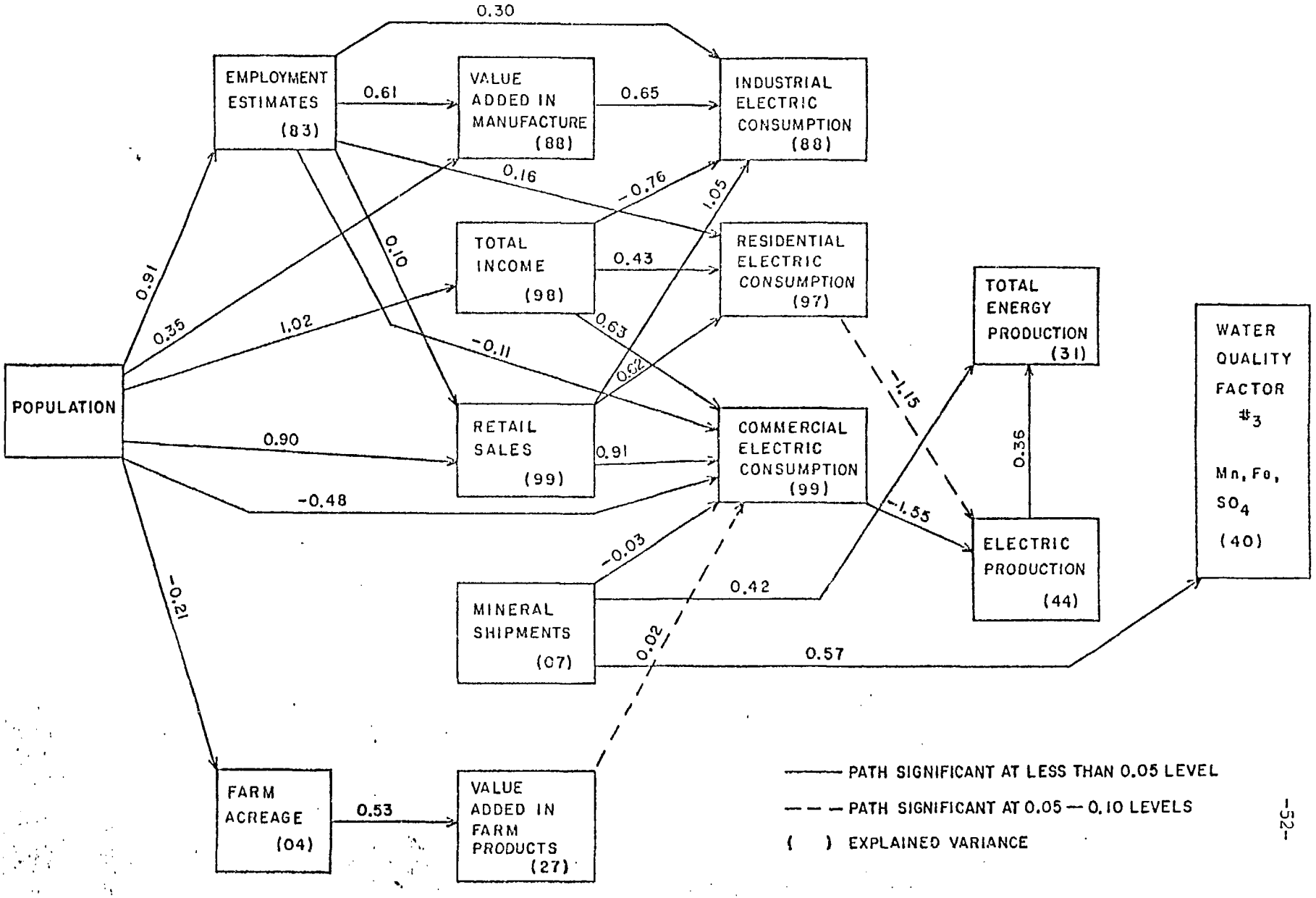


FIGURE 24 CAUSAL DIAGRAM FOR ENERGY, SOCIOECONOMIC AND WATER QUALITY