# TREATMENT CATEGORIES FOR COAL MINE DRAINAGE

Hittman Associates, Inc.

# TREATMENT CATEGORIES FOR COAL MINE DRAINAGE

by

H. Lee Schultz
Donald Koch
Carolyn Thompson
Dr. Kathleen Hereford
Hittman Associates, Inc.
Columbia, Maryland 21045

Contract No. 68-03-2566

Project Officer

Roger C. Wilmoth
Resource, Extraction, and Handling Division
Industrial Environmental Research Laboratory
Cincinnati, Ohio 45268

OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

## DISCLAIMER

# FOREWORD

When energy and material resources are extracted, processed, converted, and used, the related pollutional impacts on our environment and even on our health often require that new and increasingly more efficient pollution control methods be used. The Industrial Environmental Research Laboratory - Cincinnati (IERL-Ci) assists in developing and demonstrating new and improved methodologies that will meet these needs both efficiently and economically.

This report details the results of an effort to determine if a suitable basis existed to establish a separate Effluent Guidelines subcategory for coal mining point source discharges to differentiate between eastern and western coal mining. Unfortunately, problems with the existing data base prevented resolution of the issue. The information and sources contained herein would be of primary interest to the EPA regulatory offices. For further information, contact the Resource Extraction and Handling Division.

David G. Stephan
Director
Industrial Environmental
Research Laboratory
Cincinnati

## ABSTRACT

This effort involved the organization and statistical analysis of a large amount of data characterizing over 300 surface and underground coal mines and quantification of the pre- and post-treatment quality of their wastewaters. Only existing data, supplied to Hittman Associates by EPA, was utilized in this evaluation. The study objective was to determine whether the data supported the development of a separate Effluent Guidelines subcategory for coal-mining point source discharges to differentiate between eastern and western coal mining activities. An extensive effort was required to convert the data into a computerized format which would allow for expeditious statistical analysis. Following computerization, a variety of statistical analyses were conducted to determine if substantiation existed within the data base for development of treatment subcategories based on any or all of the following factors: precipitation, effluent flow, effluent source, influent concentration of pollutant parameters, treatment type, and acidity/alkalinity. The results of the statistical analyses illustrated that the data base provided to Hittman Associates was not adequate to provide a verifiable basis for subcategorization of the coal mining point source category. This inadequacy stemmed from numerous problems which existed within the data base concerning the non-uniformity population and the non-comparability of much of the data.

CONTENTS

# SECTION 1

# INTRODUCTION

The purpose of this study was to conduct computerized statistical analyses to answer the following interrelated questions:

- Does a given coal mine wastewater treatment technology provide significantly different results in the West than it does in the East?
- If so, what are the differences?
- Should different, more stringent treatment criteria (and hence different treatment technologies) be required for mines in different parts of the country?

The analyses conducted in addressing these questions were carried out based on existing data that were supplied to Hittman Associates, Inc. (HAI) by the U.S. Environmental Protection Agency (EPA). These data quantitatively characterized wastewaters originating at approximately 300 surface and underground coal mines located in the continental United States. The data also contained information characterizing the mining operation conducted at each site. Statistical analyses were conducted to determine if the data supplied to HAI were adequate to serve as a basis for addressing the study questions. The approach taken in evaluating these data, the results of the analyses, and conclusions drawn therefrom are discussed in the following sections.

# SECTION 2

## SUMMARY OF CONCLUSIONS

The data supplied for this analysis were not adequate to support a decision regarding the need for subcategorization of the coal mining point source category. This conclusion is based on the results of the analyses described here, and the funding and time limitations placed on the project. Results obtained in this analysis were inconclusive and support neither the need nor the lack of need for such subcategorization.

Data obtained through a statistically designed sampling program, or through such a program in concert with utilization of existing National Pollutant Discharge Elimination System (NPDES) data, could eliminate most of the problems identified in the data supplied to HAI and could, therefore, provide conclusive results.

## SECTION 3

## TECHNICAL APPROACH

### DATA OBTAINED

Data were obtained by HAI from four sources:

- EPA-IERL/Cincinnati: Mathematica report (unpublished at time of acquisition)
- EPA-EGD/Washington: Mine effluent and production data in computer cards (with corresponding coding sheets)
- EPA/Dallas: National Pollution Discharge Elimination System (NPDES) files
- EPA/Denver: National Pollutant Discharge Elimination System (NPDES) files.

The data contained in each of these sources are described below.

#### Mathematica Report

The Mathematica data, An Inventory of Western Mines, are in computer printout form. The data are primarily operational in nature, although ancillary information such as precipitation levels, surface rights data, land use information, coal characteristics, etc. are also included. The report covers 37 active and 7 planned surface coal mines in the West that had more than 100,000 tons of production in 1975 (which accounts for about 99.5 percent of the western surface coal tonnage produced in 1975).

#### EPA/Washington Computer Cards

EPA Effluent Guidelines Division (EGD) supplied some 20,000 computer cards containing data on mine identification, mine characteristics, and 60 water quality parameters. This data base covers 225 mines in all. Coverage is mainly for eastern mines, although a signficant number of western mines are also included. These data constitute the most comprehensive coverage of eastern mines available. Because the data are already in computerized form, significant time and effort are saved.

Versar Data

These unpublished data are in paper form (tabular) and contain water quality data for 23 mines located in the East and the West. The data primarily cover effluent concentrations of trace metals, although a few traditional water quality parameters are also included.

EPA Regions VI and VIII NPDES/DMR Files

Files from the National Pollutant Discharge Elimination System/Discharge Monitoring Report (NPDES/DMR) were made available to HAI in Dallas, Texas, and Denver, Colorado, so that the necessary data for mines located in EPA Regions VI and VIII could be obtained. Accessing these data involved identifying those files that contained pertinent data, copying the data, and coding and keypunching the data for computerization. The files for Regions VI and VIII make up the largest, most comprehensive data source for western mines available at this time.

## PROBLEMS IN DATA UTILIZATION

It should be noted that the coverage of the data supplied to HAI was rather variable. That is, data from different sources did not cover the same mine characteristics, water quality parameters, etc., in all instances. Thus, the ability to compare the data from different mines was significantly impaired.

Major problems were encountered as computerized data processing activities were initiated. Specifically, a number of problems of varying severity were identified in the EPA-EGD/Washington computerized data base. First, data appeared to be coded in a somewhat haphazard fashion, and data fields were found to be variable among the cards. Data items such as mine name, location, etc., were not coded in the same columns on all cards. This means that when the computer looks for information at a given location, it will in many instances find part of one data entry and part of another (instead of an entire entry). This would cause those data entries to which the problem applies to be useless and therefore be eliminated from evaluation.

An additional problem relating to coding of information in this data base was that all numerical data were not recorded in the same units. For example, in the mine size category, tonnages were usually recorded in terms of "ton per year"; however, in some instances such data were recorded in terms of "tons per day" or "tons per hour." While a relatively simple subroutine could be written to convert all such data entries to equal terms, the assumptions necessary to do so (such as the number of hours per day worked at the mine, the number of workdays per week, the number of work-weeks per year, etc.) were not provided. Therefore, any attempt at conversion would result in questionable outputs since without mine-specific operational data, any conversion would be arbitrary. However, neither this nor the previously defined problem were insurmountable. While they did require additional time and effort, and cut down on the sample of useable data in certain circumstances, they did not in themselves render the initial evaluation impossible

A third problem relative to the EGD cards uncovered later in the project did effect a serious impediment to the evaluation. The cards had been converted to tape form, and the tape read into the computer system. The taped data was printed out via terminal link prior to conducting analytical runs in order to assure the compatibility (in terms of format, etc.) with the other data to be used. At this time it was determined that this data bank was not useable in its initial form, as the data it contained were not logical. When the data bank was printed out, it was impossible to correlate the various data entries (that is, one could not ascertain which data went with which mine, etc.). It was determined that this problem was probably due to one of two causes:

- Blank data cards were not included in the data bank when a blank entry was encountered on the coding sheet. Unfortunately the computer does not know if some entries have been left out. It can only read the cards in the order in which they are put in and extract data from the cards based on this ordering. Thus, if some cards were deleted from the deck, all following cards would be out of order.

- All the necessary cards may have been present, but out of order. This could have occurred if the cards had been dropped or otherwise mixed up. Unfortunately, no sequential numbering system was used when keypunching the cards, so that there was no quick way to determine if all cards were both present and in order.

In order to resolve these problems, it was decided that the EGD computerized data base had to be completely recoded, then keypunched and entered into the computerized data management system. The original contract was modified to allow for this recoding. HAI was directed to conduct the recoding effort using the original coding sheets (from which the EGD-supplied computer cards were prepared) as the data source. During this effort data fields and units were standardized, and any other problems not yet identified relative to the data contained in this data base were to be resolved as far as was possible. A major portion of the time and funding allocation for this project was subsequently dedicated to the resolution of the EGD data base problems.

## METHODOLOGY

In order to address this project in a comprehensive fashion, any major mine characteristic which would be likely to affect effluent quality were to be considered as a likely candidate basis for a treatment category. For example, mine grouping could be based upon average annual precipitation levels generally considered to be indicative of various ecological classifications, such as the four categories presented below*:

- 0 to 10 inches of precipitation per year - desert
- 11 to 30 inches of precipitation per year - grassland, savanna, or open woodland
- 31 to 50 inches of precipitation per year - dry forest
- 51 inches of precipitation per year or more - wet forest

Alternately, such categories could be based upon the natural bisection of the United States by the 20-inch average annual precipitation isopleth which trends vertically through eastern North and South Dakota, central Nebraska, and western Kansas, Oklahoma, and Texas. This delineation would yield two categories**: 0 to 20 inches of precipitation per year and 21 inches of precipitation per year or more.

---

* *Odum, Eugene P., Fundamentals of Ecology, Second Edition. W.B. Saunders Company, Philadelphia, Pennsylvania, 1959, p. 112.*

** *Geraghty, Miller, Van Der Leeden, and Troise, Water Atlas of the United States, Water Information Center, Port Washington, New York, 1973.*

Effluent flow is another mine characteristic which could serve as a logical basis for categorization. While at first glance it might be assumed that this characteristic would be directly proportional to rainfall, this is not necessarily the case. Coal mines often intercept aquifers (water bearing strata), especially underground mines. When this occurs, water generally must be pumped from the area of active mining. Therefore, it is conceivable that a mine located in an arid area could have significant flows of water requiring treatment (flows greater than that expected based on rainfall alone).

Effluent source obviously plays a significant role in the characteristics of the wastewaters to be treated, and thus in the expected treatment efficiency. This consideration provides three potential categories for evaluation:

- Surface mines
- Underground mines
- Preparation plants.

With the above considerations in mind, evaluation of the supplied data indicated the appropriateness of subcategorization based on the following six characteristics:

- Precipitation
- Effluent flow
- Effluent source
- Influent concentration of pollution parameters
- Type of treatment
- Acidity/alkalinity.

These categories served as the basis for all subsequent evaluation. In addition, EPA chose to concentrate on nine of the water quality parameters contained in the overall data base during the evaluation process. The nine parameters included in analytical efforts were the following:

- pH
- Total suspended solids
- Total dissolved solids
- Zinc

- Aluminum
- Total iron
- Total manganese
- Acidity
- Alkalinity.

Once the data to be utilized in this evaluation had been entered into the computerized data management system and debugged, final computer runs were conducted and the results evaluated to arrive at a determination as to the appropriateness of one or more subcategorizations for the coal mining effluent limitation point source category. The output from the data management system consisted initially of mean and standard deviation values for the water quality parameters under study for all mines in the six predetermined categories. This information was then evaluated, with the aid of computerized techniques, to determine its statistical significance.

Multiple regression analysis was used to examine differences in water quality among the mines in the data base. An analysis of covariance approach was used in that some of the explanatory variables included in the regression equation were defined categorically (e.g., soil characteristics, effluent source, type of treatment) while other variables were numerically measured (e.g., precipitation, effluent flow, mine production). Standard statistical tests based on the t-statistic were applied to the estimated regression coefficients to determine which variables are statistically significant in explaining variations in water quality among mines.

The regression analysis was performed not only for the sample as a whole but also for selected subsamples defined by certain of the category variables. (For example, the data observed for surface mines defined one subsample, with the data for underground mines and for preparation plants defining other subsamples.) This approach permitted an assessment of interaction effects between the numerical variables and the categorical variables used to define the subsamples. A statistical test based on the F-statistic was used to determine if the relationship between water quality levels and the explanatory variables could be regarded as the same for different subsamples.

# SECTION 4

# RESULTS AND DISCUSSION

## RESULTS

Appendix A presents a listing of all computerized statistical analysis runs conducted. This section presents a discussion of the results obtained from these computer runs.

### t-test

The first type of statistical analysis run was a test of differences between groups, a t-test (at a 95% + 99% significance level). Groupings were constructed by dividing the data into the following categories:

- Mean average annual precipitation:
    - $<15$ in. vs. $\geq 15$ in.
    - $<20$ in. vs. $\geq 20$ in.
    - $<30$ in. vs. $\geq 30$ in.
- Effluent source:
    - Underground mine vs. surface mine
    - Underground mine vs. preparation plant
    - Surface mine vs. preparation plant
- Acidity vs. alkalinity of sample
- Effluent flow: $<1$ mgd vs. $\geq 1$ mgd.

The results of these analyses are discussed below.

#### Precipitation--

Significant differences were found to exist among means of all water quality parameters tested based on rainfall for $<20$ in. vs. $\geq 20$ in. Similar results were found to exist, however, when the test was conducted for the 15-in. and 30-in. demarcation categories as well (see Appendix B). In fact, the mean values changed little among the three cases tested.

Effluent Source--

Significant differences among means were found to exist for surface vs. underground mines for Al, Zn, Fe, Mn, and alkalinity, and also for log Al, log Fe, and log alkalinity.

Acidity vs. Alkalinity--

Significant differences among means were found to exist for all parameters based on the alkalinity or acidity of the sample.

Effluent Flow--

Significant differences among means found for only pH and log SS values when flows were grouped at <1 mgd and ≥1 mgd.

Influent Concentration of Pollutant Parameters vs. Effluent Concentration--

The data base contains only 74 influent cases, 70 of which are located in the East. Also, the water quality information corresponding to these cases was found to be variable (not all parameters were reported in all cases). Therefore, the influent sample data were found to be inadequate to provide any statistically significant results.

Type of Treatment--

Within the data base there exist 25 distinct treatment categories. These categories are made up of single treatment methods as well as various combinations of such methods. In addition, in many cases the type of treatment utilized, as well as data quantifying pollutant loadings before and after treatment, were not reported for all mines. This large number of potential treatment categories, combined with the non-uniformity in data coverage concerning pollutant output and method of treatment, results in a very small data sample for evaluation of each potential treatment category. Thus, the level of statistical significance which could be drawn from an analysis based on treatment type would yield results of minimal significance. It was determined, therefore, that the results of such an analysis would not constitute substantiation for a decision to develop an effluent treatment category based on type of treatment. Due to this problem, no computer runs were conducted based on treatment type.

## Regression Analysis

The second type of statistical analysis run was a test of correlation, a regression analysis. A set of independent variables was chosen and tested to see if these independent variables had any predictive value, hopefully explaining the group differences. The independent variables tested were average annual precipitation, effluent flow, effluent source, and acidity/alkalinity. Although a number of statistically significant relationships were found, none of these had much predictive value. For example, the simple regression of Al (dependent) on average annual precipitation (independent) had an $r^2$ of .012, meaning that 1.2 percent of the variance in Al is explained by rainfall. The signficance of this relationship is indicated by the F score, if one assumes that both samples are from normal populations. The F score was 10.2, indicating a highly significant relationship, i.e. there was a less than 0.01 percent probability that the indicated relationship occurred by chance.

Outlier Exclusion

Performance of a rigorous outlier exclusion exercise was not possible due to time and funding constraints. However, in order to illustrate the effects which such exclusion would have on analytical results, outliers were excluded from a number of computer runs based on best engineering judgement. Upper concentration limits for the water quality parameters under study were defined based on experience. Values in excess of the upper limits cited below were then eliminated:

- Al: 5 mg/l
- Zn: 5 mg/l
- Fe: 100 mg/l
- Mn: 100 mg/l
- Acidity: 5,000 mg/l
- Alkalinity: 5,000 mg/l
- TDS: 10,000 mg/l
- SS: 1,000 mg/l

Comparison of mean values of those cases with outliers included versus those with outliers excluded resulted in two observations:

- Excluding data points in excess of the above-cited limits significantly reduced the magnitude of the mean values of pollutant parameters.
- In each case a very small number of data points were excluded.

Thus it can be concluded that a small number of suspiciously large sample values were responsible for a significantly large increase in mean values. It is decided, therefore, that this best engineering judgement outlier exclusion exercise was reasonable. The majority of computer runs were then conducted incorporating this best engineering judgement outlier exclusion.

## DISCUSSION

The tests conducted here show that basic differences exist among different groups of data, but the reasons for the differences cannot be explained by the data. Specifically, the following conclusions were drawn from the statistical analysis for each category tested.

### Precipitation

Although significant differences were found to exist among means for all parameters based on average annual precipitation (at a 95% + 99% level of significance), the mean values obtained were numerically similar in all three cases tested, i.e., 15-in., 20-in., and 30-in. precipitation demarcation (see Appendix B). Therefore, although higher precipitation seems to be associated with lower pH and higher concentrations of pollutants (and although the opposite is true for lower precipitation), no obvious demarcation point was discovered. Also, the regression results indicated a lack of predictive ability.

### Effluent Source

Although significant differences were found among means for a number of parameters, the regression indicated that no predictive value could be identified. Therefore, the results while significant, are inconclusive.

### Alkalinity vs. Acidity

Although significant differences were consistently found among means of the water quality parameters tested at a 95% + 99% level of significance, the regression again indicated a lack of predictive ability as to the cause of these differences.

### Effluent Flow

The general lack of significant differences among means obtained based on effluent flow (at a 1-mgd demarcation) would seem to indicate that this category would not serve as an adequate basis for subcategorization. However, a number of general considerations relative to the adequacy of this data base which are presented in following discussions indicate that even this conclusion cannot be supported by the data.

General Considerations

In addition to the above-mentioned considerations, a number of general inadequacies and limitations have also been identified in relation to the data base.

The first problem regards the validity of the sample, which should incorporate a certain degree of uniformity. For example, data should be taken for all mines at approximately the same time of year so as to minimize seasonal variation in the effluent data population obtained. This was not the case with the data supplied for utilization in this study which was obtained during different seasons over a number of years. In addition, valid statistical analyses should be run on a random sample from the uniform population (all pollutant concentrations from all coal mines sampled during a given season). Our sample consisted of all NPDES data from Region VI, almost all the NPDES data from Region VIII, and a large amount of data from EGD of unknown origin (assumed to be a compilation of both previously recorded data and collected data). A more valid sample could be obtained by randomly selecting several coal mines in each Region, recording all NPDES data from these mines, and then weighting the data according to the total number of coal mines in each Region. Even this design is somewhat flawed because NPDES data may not be a timely representative sample (i.e., many mines may never bother to send in data, and these may be some of the worst in terms of pollution). Also, the scope of data reported by mine operators is variable (i.e., the same parameters are not always reported by all mines). Therefore, an ideal data acquisition effort would involve either obtaining additional water quality data so as to supplement NPDES data, or obtaining all data independent of the NPDES, thereby assuring both consistency in the data base and adequacy of the type of information obtained relative to the specific needs of the analysis to be conducted.

A second consideration regards the distribution of the variables. The t-test procedure and the F-test of the regression procedure are predicated on having a normal distribution in all variables. Water quality data does not come from a normal distribution. A normal distribution has zero skew. Water quality data is almost always skewed because the mean is often close to zero (and a figure of less than zero suspended solids is inaccurate). A Kolmogorov-Smirnov test of goodness-of-fit against a normal distribution indicated that none of the parameters under consideration were anywhere near a normal distribution.

The problem of normality can often be alleviated by transforming the data. The transformations should be based on experience or some knowledge about the behavior of the variable. Average annual precipitation should be from a normal population according to the central limit theorem which states that the sampling distribution of means of random samples of a population will be approximately normal if the sample size is sufficiently large. Since the average annual precipitation is a mean, and we had a reasonable sample size, we assumed in this analysis that precipitation is normally distributed. This also assumes that rainfall throughout the United States is a homogeneous population. Flow and the water quality parameters are probably not normally distributed. Thus, a transformation is needed to make these parameters approximate a normal distribution. It is often assumed in hydrologic work that flow is lognormally distributed, therefore taking the logarithm of flow would yield a normal distribution. The same assumption was made for all the water quality parameters in this evaluation except pH (which is already the log of the hydrogen ion concentration) for which a logarithmic transformation was conducted. The validity of these transformations, however, is open to question. A Kolmogorov-Smirnov test on the transformed variables indicated that although they were closer to a normal distribution than were the untransformed variables, they were still far enough from a normal distribution to put the absolute validity of the t-test, F-test, and regression analyses in question.

A third problem is not one of statistical theory but of the data. There are many values in the data that are questionable (and some that are highly unlikely). In addition, the data base contains a large number of missing values. The data base is also unsuited to the objective of proving a significant difference between eastern and western mines. Data on influents to treatment works are needed to prove this hypothesis. It has been determined that there are only 74 influent values in the data base and these are almost all from the East. Comparison of treatment technologies within the data base has also proven to be of little value. For instance, the treatment technology of settling includes a variety of undersized and poorly maintained ponds as well as highly effective installations. Thus, in order to identify and quantify the role which utilization of specific treatment technologies plays in determining effluent quality, that portion of the data base which contains data identifying the type of technology used should ideally also contain data specifically characterizing each of the treatment systems at each mine (design factors, for example). Also, of the samples which were not identified as influent, some were identified as effluent, and the remainder had no designation at all. This latter case existed for more than half of the entire data base. If only those samples for which the source is identified were to be included in the analysis the sample would be severely limited, thus reducing the statistical validity of any results obtained. If data with no designation as to source were included in the effluent category there would be no assurance that a large amount of influent data was not thereby being mixed with effluent data. If this were the case it would negate the validity of any results obtained. In conducting the analyses reported, runs were made based on:

- Cases identified as influent only
- Cases identified as effluent only
- Cases identified as effluent only plus those cases which are unidentified as the source.

It was determined that too few data points are contained in the influent-only sample to provide any significant results. In addition, essentially identical results were obtained when the effluent-only run results were compared with those from the effluent-only plus undesignated cases (this comparison was conducted for precipitation only, as the statistical results obtained based on precipitation, inclusive as they may be, were among the best obtained). The fact that little difference occurred in this comparison is probably a result of the non-uniformity of the data sample.

# APPENDICES

## APPENDIX A. STATISTICAL COMPUTER RUNS COMPLETED

### 1. Regression: (All Data):

| | | (a) | (b) | (c) |
|---|---|---|---|---|
| pH | vs | precipitation ; | flow ; | Alkalinity-Acidity |
| Zn | vs | precipitation | flow | Alkalinity-Acidity |
| Al | vs | precipitation | flow | Alkalinity-Acidity |
| Fe | vs | precipitation | flow | Alkalinity-Acidity |
| Mn | vs | precipitation | flow | Alkalinity-Acidity |
| Acidity | vs | precipitation | flow | Alkalinity-Acidity |
| Alkalinity | vs | precipitation | flow | Alkalinity-Acidity |
| TDS | vs | precipitation | flow | Alkalinity-Acidity |
| SS | vs | precipitation | flow | Alkalinity-Acidity |

| | | (d) |
|---|---|---|
| pH | vs | Underground, Surface, Preparation Plants |
| Zn | vs | Underground, Surface, Preparation Plants |
| Al | vs | Underground, Surface, Preparation Plants |
| Fe | vs | Underground, Surface, Preparation Plants |
| Mn | vs | Underground, Surface, Preparation Plants |
| Acidity | vs | Underground, Surface, Preparation Plants |
| Alkalinity | vs | Underground, Surface, Preparation Plants |
| TDS | vs | Underground, Surface, Preparation Plants |
| SS | vs | Underground, Surface, Preparation Plants |

| | | (e) | (f) |
|---|---|---|---|
| pH | vs | Alkalinity, Precip., & Flow ; | Log Alkalinity, Precip., Log Flow |
| Log Zn | vs | Alkalinity, Precip., & Flow | Log Alkalinity, Precip., Log Flow |
| Log Al | vs | Alkalinity, Precip., & Flow | Log Alkalinity, Precip., Log Flow |
| Log Fe | vs | Alkalinity, Precip., & Flow | Log Alkalinity, Precip., Log Flow |
| Log Mn | vs | Alkalinity, Precip., & Flow | Log Alkalinity, Precip., Log Flow |
| Log TDS | vs | Alkalinity, Precip., & Flow | Log Alkalinity, Precip., Log Flow |
| Log SS | vs | Alkalinity, Precip., & Flow | Log Alkalinity, Precip., Log Flow |

### 2. Regression: (outliers excluded):

| | | (a) |
|---|---|---|
| ph | vs | Log Flow, Precip. & Log Alkalinity |
| Log Zn | vs | Log Flow, Precip. & Log Alkalinity |

| | | |
|---|---|---|
| Log Al | vs | Log Flow, Precip. & Log Alkalinity |
| Log Fe | vs | Log Flow, Precip. & Log Alkalinity |
| Log Mn | vs | Log Flow, Precip. & Log Alkalinity |
| Log TDS | vs | Log Flow, Precip. & Log Alkalinity |
| Log SS | vs | Log Flow, Precip. & Log Alkalinity |
| Log Acid | vs | Log Flow, Precip. & Log Alkalinity |
| Log Alk. | vs | Log Flow, Precip. & Log Alkalinity |

3. <u>Scatterplot</u>

(a)

| | | |
|---|---|---|
| ph | vs | Precipitation |
| Log Zn | vs | Precipitation |
| Log Al | vs | Precipitation |
| Log Fe | vs | Precipitation |
| Log Mn | vs | Precipitation |
| Log Acidity | vs | Precipitation |
| Log Alkalinity | vs | Precipitation |
| Log TDS | vs | Precipitation |
| Log SS | vs | Precipitation |

4. <u>Kolmogorov-Smirnov Goodness-of-Fit Test</u>:

(a)

| | |
|---|---|
| Base Data For | pH |
| Base Data For | Al |
| Base Data For | Zn |
| Base Data For | Fe |
| Base Data For | Mn |
| Base Data For | Acidity |
| Base Data For | Alkalinity |
| Base Data For | TDS |
| Base Data For | SS |
| Base Data For | Precipitation |
| Base Data For | Flow |

(b)

| | |
|---|---|
| Log Transformed Data For | - |
| Log Transformed Data For | Al |
| Log Transformed Data For | Zn |
| Log Transformed Data For | Fe |
| Log Transformed Data For | Mn |
| Log Transformed Data For | Acidity |
| Log Transformed Data For | Alkalinity |
| Log Transformed Data For | TDS |
| Log Transformed Data For | SS |
| Log Transformed Data For | Flow |

5. <u>t-Test (excluding outliers)</u>:

| | | (a) | (b) |
|---|---|---|---|
| pH | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Zn | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Al | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Fe | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Mn | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Acidity | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Alkalinity | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| TDS | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| SS | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |

| | | (c) | (d) |
|---|---|---|---|
| Log Zn | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Log Fe | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Log Al | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Log Mn | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Log Acidity | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Log Alkalinity | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Log TDS | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |
| Log SS | vs | ≥20 in. Precip/≤20 in. Precip; | ≥30 in. Precip/≤30 in. Precip |

| | | (e) | | (f) |
|---|---|---|---|---|
| pH | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| Zn | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| Al | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| Fe | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| Mn | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| Acidity | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| Alkalinity | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| TDS | vs | Alkalinity/Acidity; | vs | Surface/Underground; |
| SS | vs | Alkalinity/Acidity; | vs | Surface/Underground; |

| | | (g) |
|---|---|---|
| pH | vs | ≥15 in. Precip/≤15 in. Precip |
| Zn | vs | ≥15 in. Precip/≤15 in. Precip |
| Al | vs | ≥15 in. Precip/≤15 in. Precip |
| Fe | vs | ≥15 in. Precip/≤15 in. Precip |
| Mn | vs | ≥15 in. Precip/≤15 in. Precip |
| Acidity | vs | ≥15 in. Precip/≤15 in. Precip |
| Alkalinity | vs | ≥15 in. Precip/≤15 in. Precip |
| TDS | vs | ≥15 in. Precip/≤15 in. Precip |
| SS | vs | ≥15 in. Precip/≤15 in. Precip |

| | | (h) | | (i) | |
|---|---|---|---|---|---|
| Log Zn | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |
| Log Al | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |
| Log Fe | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |
| Log Mn | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |
| Log Acidity | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |
| Log Alkalinity | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |
| Log TDS | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |
| Log SS | vs | Alkalinity/Acidity | ; | Surface/Underground | ; |

| | | (j) |
|---|---|---|
| Log Zn | vs | ≥15 in. Precip/≤15 in. Precip |
| Log Al | vs | ≥15 in. Precip/≤15 in. Precip |
| Log Fe | vs | ≥15 in. Precip/≤15 in. Precip |
| Log Mn | vs | ≥15 in. Precip/≤15 in. Precip |
| Log Acidity | vs | ≥15 in. Precip/≤15 in. Precip |
| Log Alkalinity | vs | ≥15 in. Precip/≤15 in. Precip |
| Log TDS | vs | >15 in. Precip/≤15 in. Precip |
| Log SS | vs | ≥15 in. Precip/≤15 in. Precip |

6. <u>t-Test (outliers not excluded)</u>:

| | | (a) | | (b) | | (e) |
|---|---|---|---|---|---|---|
| pH | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| Al | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| Zn | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| Fe | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| Mn | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| Acidity | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| Alk. | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| TDS | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |
| SS | vs | Acidity/Alkalinity | ; | Surface/Prep Plant | ; | Underground/Prep Plant |

| | | (d) | | (e) | |
|---|---|---|---|---|---|
| pH | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| Al | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| Zn | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| Fe | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| Mn | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| Acidity | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| Alkalinity | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| TDS | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |
| SS | vs | Surface/Underground | ; | Flow <1 MGD/Flow ≥1 MGD | ; |

(f)

| | |
|---|---|
| pH | ≤15 in. Precip/≥15 in. Precip |
| Al | ≤15 in. Precip/≥15 in. Precip |
| Zn | ≤15 in. Precip/≥15 in. Precip |
| Fe | ≤15 in. Precip/≥15 in. Precip |
| Mn | ≤15 in. Precip/≥15 in. Precip |
| Acidity | ≤15 in. Precip/≥15 in. Precip |
| Alkalinity | ≤15 in. Precip/≥15 in. Precip |
| TDS | ≤15 in. Precip/≥15 in. Precip |
| SS | ≤15 in. Precip/≥15 in. Precip |

| | | (g) | | (h) |
|---|---|---|---|---|
| pH | vs | ≤20 in. Precip/≥20 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |
| Al | vs | ≤20 in. Precip/≥20 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |
| Zn | vs | ≤20 in. Precip/≥20 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |
| Mn | vs | ≤20 in. Precip/≥20 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |
| Acidity | vs | ≤20 in. Precip/≥20 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |
| Alkalinity | vs | ≤20 in. Precip/≥20 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |
| TDS | vs | ≤20 in. Precip/≥30 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |
| SS | vs | ≤20 in. Precip/≥30 in. Precip | ; | ≤30 in. Precip/≥30 in. Precip |

| | | (i) | | (j) | |
|---|---|---|---|---|---|
| Log Al | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |
| Log Zn | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |
| Log Fe | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |
| Log Mn | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |
| Log Acidity | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |
| Log Alkalinity | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |
| Log TDS | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |
| Log SS | vs | Underground/Prep. Plant | ; | Surface/Prep. Plant | ; |

| | | (k) |
|---|---|---|
| Log Al | vs | Acidity/Alkalinity |
| Log ZN | vs | Acidity/Alkalinity |
| Log Fe | vs | Acidity/Alkalinity |
| Log Mn | vs | Acidity/Alkalinity |
| Log Acidity | vs | Acidity/Alkalinity |
| Log Alkalinity | vs | Acidity/Alkalinity |
| Log TDS | vs | Acidity/Alkalinity |
| Log SS | vs | Acidity/Alkalinity |

| | | (l) | (m) |
|---|---|---|---|
| Log Al | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |
| Log Zn | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |
| Log Fe | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |
| Log Mn | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |
| Log Acidity | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |
| Log Alkalinity | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |
| Log TDS | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |
| Log SS | vs | <15 in. Precip/≥15 in. Precip; | <20 in. Precip/≥20 in. Precip |

| | | (n) | (o) |
|---|---|---|---|
| Log Al | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |
| Log Zn | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |
| Log Fe | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |
| Log Mn | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |
| Log Acidity | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |
| Log Alkalinity | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |
| Log TDS | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |
| Log SS | vs | <30 in precip/≥30 in precip ; | Flow <1 MGD/Flow ≥1 MGD |

## APPENDIX B. ILLUSTRATIVE t-TEST RESULTS

TABLE B-1. t-TEST: PRECIPITATION
EXCLUDING OUTLIERS* (95% + 99% LEVEL OF SIGNIFICANCE)

| | <15 in | ≥15 in | ≤20 in | ≥20 in | <30 in | ≥30 in |
|---|---|---|---|---|---|---|
| pH | 8.0 | 6.7 | 8.0 | 6.6 | 7.9 | 6.5 |
| Zn(µg/1) | 69 | 406 | 71 | 409 | 70 | 421 |
| Al(µg/1) | 479 | 812 | 449 | 819 | 573 | 817 |
| Fe(νg/1) | 591 | 5884 | 834 | 6326 | 871 | 6591 |
| Mn(µg/1) | 246 | 11230 | 233 | 11309 | 217 | 11552 |
| Acid(µg/1) | 16 | 328 | 16 | 328 | 49 | 335 |
| Alk(mg/1) | 495 | 190 | 512 | 170 | 461 | 167 |
| TDS(mg/1) | 2724 | 2143 | 2511 | 2176 | 2462 | 2171 |
| SS(mg/1) | 47 | 66 | 40 | 70 | 42 | 71 |
| | | | | | | |
| Log ZN(µg/1) | 1.51 | 2.03 | 1.54 | 2.03 | 1.57 | 2.04 |
| Log Al(µg/1) | 2.20 | 2.58 | 2.17 | 2.59 | 2.38 | 2.57 |
| Log Fe(µg/1) | 2.42 | 2.90 | 2.34 | 2.95 | 2.39 | 2.97 |
| Log Mn(µg/1) | 1.96 | 3.0 | 1.93 | 3.0 | 1.94 | 3.02 |
| Log Acid (mg/1) | 0.89 | 1.92 | 0.89 | 1.92 | - | - |
| Log Alk (mg/1) | 2.60 | 1.98 | 2.62 | 1.94 | 2.55 | 1.93 |
| Log TDS (mg/1) | 3.31 | 3.11 | 3.27 | 3.11 | 3.27 | 3.10 |
| Log SS (mg/1) | 1.21 | 1.33 | 1.18 | 1.36 | 1.20 | 1.36 |

*Mean values (rounded off) are presented in Table 1 for all cases wherein a statistically significant difference in means was identified by the t-test.

TABLE B-2. t-TEST SCORES: PRECIPITATION EXCLUDING OUTLIERS* (95% + 99% LEVEL OF SIGNIFICANCE)

| Parameter | 15-in. case | 20-in. case | 30-in. case |
|---|---|---|---|
| pH | -17.4 | -24.2 | -22.7 |
| Zn | 10.0 | 10.1 | 10.7 |
| Al | 2.5 | 3.4 | 2.8 |
| Fe | 13.2 | 10.0 | 11.0 |
| Mn | 12.5 | 12.5 | 12.6 |
| Acid | 11.2 | 11.2 | 9.9 |
| Alk | - 9.0 | -10.6 | -10.2 |
| TDS | - 3.4 | - 2.1 | - 3.3 |
| SS | 2.4 | 4.6 | 4.5 |
| | | | |
| Log Zn | 8.4 | 8.1 | 8.8 |
| Log Al | 4.2 | 4.9 | 2.7 |
| Log Fe | 8.7 | 14.6 | 14.8 |
| Log Mn | 12.0 | 13.8 | 15.3 |
| Log Acid | 3.2 | 3.36 | 2.4 |
| Log Alk | 3.1 | -23.4 | -20.8 |
| Log TDS | - 6.0 | - 4.8 | - 6.1 |
| Log SS | 2.9 | 5.4 | 4.9 |

*t-scores are presented in Table 2 for all cases in which a statistically significant difference in means was identified by the t-test.

## TECHNICAL REPORT DATA

*(Please read Instructions on the reverse before completing)*

| | | |
|---|---|---|
| 1. REPORT NO. | 2. | 3. RECIPIENT'S ACCESSION NO. |
| 4. TITLE AND SUBTITLE<br>TREATMENT CATEGORIES FOR COAL MINE DRAINAGE | | 5. REPORT DATE<br>April 17, 1979<br>6. PERFORMING ORGANIZATION CODE |
| 7. AUTHOR(S)<br>H. Lee Schultz, Carolyn Thompson,<br>Donald Koch, Dr. Kathleen Hereford | | 8. PERFORMING ORGANIZATION REPORT NO. |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Hittman Associates, Inc.<br>9190 Red Branch Road<br>Columbia, Maryland 21045 | | 10. PROGRAM ELEMENT NO.<br>1NE623<br>11. CONTRACT/GRANT NO.<br>68-03-2566 |
| 12. SPONSORING AGENCY NAME AND ADDRESS<br>U.S. Environmental Protection Agency<br>Resource Extraction and Handling Division<br>Industrial Environmental Research Laboratory<br>Cincinnati, Ohio 45268 | | 13. TYPE OF REPORT AND PERIOD COVERED<br>Final (1/78 - 1/79)<br>14. SPONSORING AGENCY CODE<br>EPA 600/12 |
| 15. SUPPLEMENTARY NOTES | | |

16. ABSTRACT


This effort involved the organization and statistical analysis of a large amount of data characterizing over 300 surface and underground coal mines and quantification of the pre- and post-treatment quality of their wastewaters. Only existing data, supplied to Hittman Associates by EPA, was utilized in this evaluation. The study objective was to determine whether the data supported the development of a separate Effluent Guidelines subcategory for coal-mining point source discharges to differentiate between eastern and western coal mining activities. An extensive effort was required to convert the data into a computerized format which would allow for expeditious statistical analysis. Following computerization, a variety of statistical analyses were conducted to determine if substantiation existed within the data base for development of treatment subcategories based on any or all of the following factors: precipitation, effluent flow, effluent source, influent concentration of pollutant parameters, treatment type, and acidity/alkalinity. The results of the statistical analyses illustrated that the data base provided to Hittman Associates was not adequate to provide a verifiable basis for subcategorization of the coal mining point source category. This inadequacy stemmed from numerous problems which existed within the data base concerning the non-uniformity population and the non-comparability of much of the data.


17. KEY WORDS AND DOCUMENT ANALYSIS

| a. DESCRIPTORS | b. IDENTIFIERS/OPEN ENDED TERMS | c. COSATI Field/Group |
|---|---|---|
| Data Analysis, Mine Water, Field Data, Statistical Tests, Coal Mining, Water Quality, Water Treatment, Water Pollution | Effluent Guidelines | 13B |

| | | |
|---|---|---|
| 18. DISTRIBUTION STATEMENT<br>RELEASE TO PUBLIC | 19. SECURITY CLASS (This Report)<br>UNCLASSIFIED | 21. NO. OF PAGES<br>29 |
| | 20. SECURITY CLASS (This page)<br>UNCLASSIFIED | 22. PRICE |

EPA Form 2220-1 (9-73)