EPA-600/5-77-008a
May 1977

Socioeconomic Environmental Studies Series

CLASSIFICATION OF AMERICAN CITIES FOR CASE STUDY ANALYSIS: VOLUME I. Summary Report

Office of Monitoring and Technical Support
Office of Research and Development
U.S. Environmental Protection Agency
Washington, D.C. 20460

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development.
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the SOCIOECONOMIC ENVIRONMENTAL STUDIES series. This series includes research on environmental management, economic analysis, ecological impacts, comprehensive planning and forecasting, and analysis methodologies. Included are tools for determining varying impacts of alternative policies; analyses of environmental planning techniques at the regional, state, and local levels; and approaches to measuring environmental quality perceptions, as well as analysis of ecological and economic impacts of environmental protection measures. Such topics as urban form, industrial mix, growth policies, control, and organizational structure are discussed in terms of optimal environmental performance. These interdisciplinary studies and systems analyses are presented in forms varying from quantitative relational analyses to management and policy-oriented reports.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/5-77-008a
May 1977

CLASSIFICATION OF AMERICAN CITIES FOR CASE STUDY ANALYSIS

Volume I

Summary Report

by

Elizabeth Lake
Carol Blair
James Hudson
Richard Tabors
Urban Systems Research & Engineering Inc.
Cambridge, Massachusetts 02138

Contract No. 68-01-3299

Project Officer

Samuel Ratick
Office of Monitoring and Technical Support
Washington, DC 20460

OFFICE OF MONITORING AND TECHNICAL SUPPORT
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, DC 20460

DISCLAIMER

This report has been reviewed by the Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ABSTRACT

Attempts to analyze and evaluate the impacts of federal programs has led to the extensive use of case studies of program impacts at selected sites. This project has developed a methodology for the systematic selection of representative case study sites and for generalizing the study results. The methodology, involving two stage factor analysis and clustering, is applied to a specific program/policy problem, the selection of metropolitan areas for case studies in analyzing the impact of federal policies on general environmental quality.

The methodology begins with a data base on standard metropolitan statistical areas, SMSAs, including variables related to environmental quality, urban form, and household, industrial, and government activity. It analyzes these variables through a two-stage factor analysis technique which allows heuristic consideration of the significant characteristics. Finally, it develops city clusters which group areas with similar attributes. Modal (or representative) cities are selected for each group and suggested as case study sites. These groups may be used to generalize the study results and to analyze the transferrability of results between areas. The methodology is sufficiently flexible to consider a wide range of research hypotheses.

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

The research described in this report has developed two major products, one a direct output, and the other a methodology for further analysis. The first output has been a classification of Standard Metropolitan Statistical Areas (SMSAs) based on broad measures of environmental quality and other attributes. This classification depends on a large scale data base which includes 262 SMSAs, and has data on activity in the industrial, demographic, and government sectors, on the attributes of the urban form and the physical environment, and on the pollutant residuals and ambient environmental quality resulting from these activities and attributes.

The second product is a methodology for developing alternative classifications, oriented towards specific policy or research issues and the urban characteristics related to them. The methodology can be directed at issues such as the choice of sites for case studies, and demonstration projects or the transfer of results from one case study area to other cities. The data base and methodology are being maintained by EPA for further applications in case study analyses.

The results of this study are focused on analytic needs involving general environmental quality or other subject areas. Environmental quality is determined by actors in the urban socioeconomic system and the physical environment together. A simplification of the interrelationships is illustrated in Figure 1. Note that, at least in this crude model, polluting residuals and ambient environmental quality are entirely endogenous to the system. Because these two aspects of the system are closely related to the attributes of the four other actors and because available measures of environmental quality and residuals are less numerous and less reliable than most others, the major classification scheme of this research was developed from data on the public sector, households, industry, and the physical environment, and then evaluated by comparison with data on environmental quality and residuals. Due to the fact that these four aspects are comprehensive in terms of the urban system and are not biased toward environmental quality, they may have far-reaching applications.

Each box in Figure 1 is represented in the data base by a group of variables called a significant characteristic set (SCS). Altogether, then, there are six SCSs which together contain approximately 200 variables. The SCSs are:

1. Ambient Environmental Quality

2. Urban Form and the Physical Environment

3. Residuals

4. Demographic Characteristics

Figure 1

CAUSATION: ENVIRONMENTAL QUALITY

(First Order Effects)

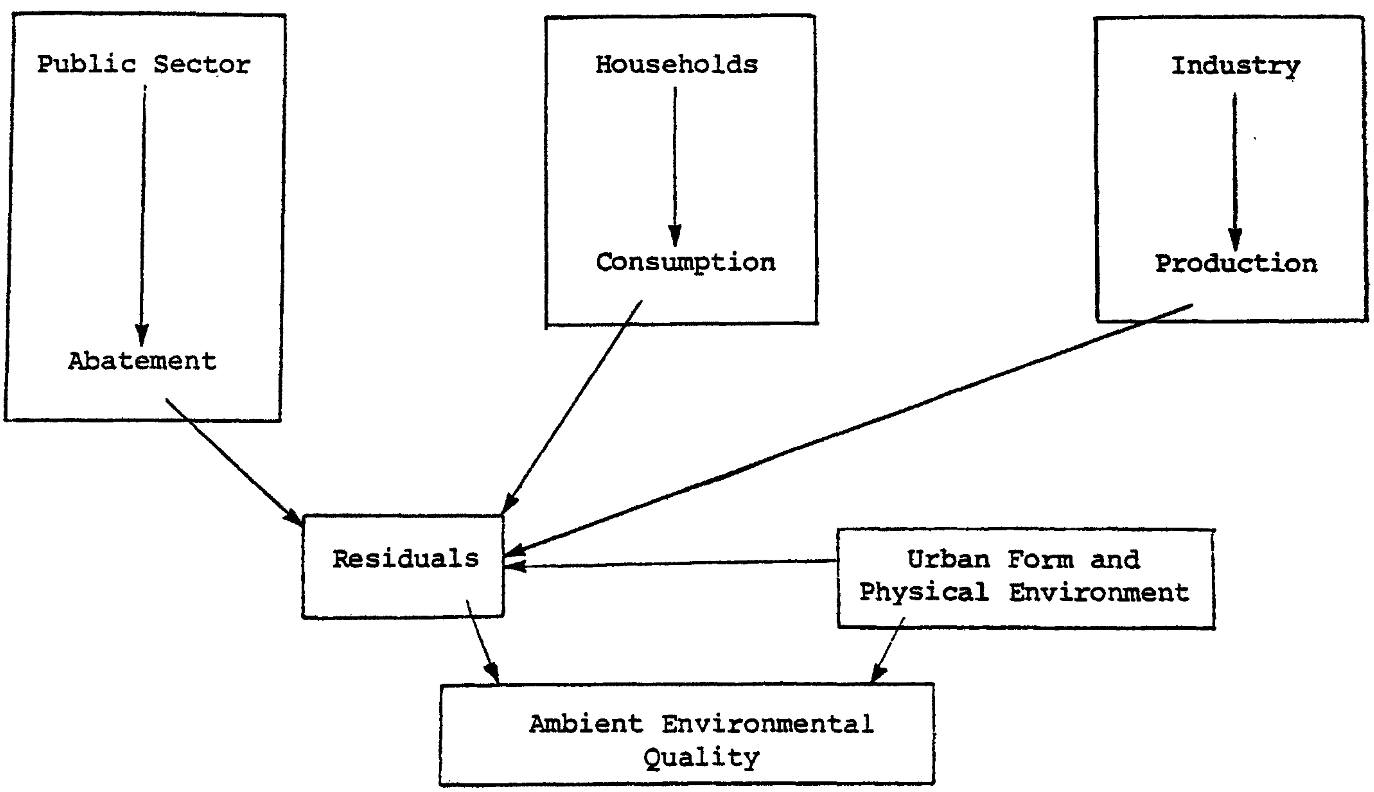


Table 1

MAJOR SMSA GROUPINGS FOR GENERAL ENVIRONMENTAL PURPOSES
(based on total SMSA Sample)

Group No.	No. of Cities in Group	Representative City	Cities Close to Modal City
1	36	Little Rock-North Little Rock, AR	Baton Rouge, LA Corpus Christi, TX Lafayette, LA Midland, TX Montgomery, AL Odessa, TX Tyler, TX
2	46	Lake Charles, LA	Spartanburg, NC Parkersburg-Marietta, WV-OH
3	27	Williamsport, PA	Davenport, IL Evansville, IN-KY Lawrence, MA Peoria, IL
4	48	Albany-Schenectady-Troy, NY	Appleton-Oshkosh, WI New Britain, CT Portland, OR
5	18	Dallas, TX	Charlotte, NC Richmond, VA

5. Government

6. Industry.

Correlation between variables within an SCS was anticipated as well as relationships between SCSs. The methodology used to classify SMSAs takes advantage of these correlations to reduce the vast amounts of data through factor analysis.

The factor analysis technique was applied to the data in two stages. In Stage I, a small number of factors was extracted from the data in each SCS. These factors summarized the basic dimensions of the data available to describe relevant attributes of urban areas. In the second stage, factors derived in Stage I were treated as variables. The factors derived in Stage II, then, reflect relationships both within the SCSs and between SCSs. The four factors* from Stage II which explain the greatest amount of variance in the data base were taken to characterize the SMSAs for purposes of classification. As the factors generated are linear combinations of the original variables, it is possible to estimate scores for each observation on each factor. These provide the location of each SMSA in the four-dimensional factor space derived in Stage II. (The 262 SMSAs have been classified by applying a simple "nearest neighbor" clustering technique to these factor scores.) Five major city groups were developed, including 175 of the 262 SMSAs. (See Table 1)

These groups were tested with respect to their ability to discriminate between cities with different levels of environmental quality. A series of t- and F-tests (statistical comparisons of means) were performed using a select set of environmental quality measures. Testing revealed the groups to be significantly different in an environmental prospective. The groups appear to be useful for environmental research and may be tested in a similar manner to determine their applicability to any given area of research.

New classifications may be developed by the same method used here by modifying the data base. The availability of new and relevant data will often justify such an effort. For instance, land use is a valuable measure of a number of influences on environmental quality, yet little data measuring land use is available except that which comes from dispersed sources in various forms. Should a new body of uniform data on land use in a large number of SMSAs become available, a more enlightened classification of SMSAs might be developed.

Other classifications might be developed to satisfy a more specific emphasis. This research included such an effort for the Energy Resources Development Administration, interested in potential energy savings through changes in transportation patterns. The basis of the classification was limited to variables related to auto use: auto ownership, per capita vehicle-miles travelled, household size, urban density, etc. Comparison of the resulting groups with those previously developed indicates a great deal is common to the two classifications. The data bank, the methodology, the classification and modal cities will be valuable in a variety of applications related to urban development and environmental quality. Specific classification of this sort is applicable to a wide range of environmental and urban policy research problems, wherever detailed case studies are performed.

---

*The number of second stage factors used for clustering was arbitrarily limited to four. A larger number of factors represents more dimensions along which cities may differ, fragmenting city groups into a large number of small clusters.

1.0 INTRODUCTION

1.1 Research Objectives

The officials of the Environmental Protection Agency and other Government agencies are frequently faced with the task of evaluating the effects of programs or policies at the local and regional levels. For example, EPA officials may be concerned with the effects of parking restrictions on urban air quality. To analyze this, they may monitor air quality in every city where parking restrictions are imposed, or they may restrict their monitoring activities to a more limited representative sample of cities. The second alternative is clearly more economical, however, it requires an appropriate urban classification scheme. The objective of this research project was the development of a flexible methodology for the classification of cities, which would be appropriate for the purposes of testing the effects of general environmental and other programs, and to aid in assessing the impacts of specific environmental policies. These typologies then group similar cities and identify modal, or representative urban areas for each group, facilitating the generalization of case study results. In this report the two terms, cities and SMSAs (Standard Metropolitan Statistical Areas), are used interchangeably. Cities normally make up parts of SMSAs, as demonstrated in Figure 1-1.

In the last few decades, a large number of city classification schemes have been developed. It may be appropriate then to ask why yet another methodology and typology was necessary? A brief review of past attempts to classify cities may answer this question.

Although considerable resources have been directed toward developing community typologies, few of the resulting classifications have been applied to further research or practical problems. One reason is that every potential use of a community typology has specific requirements in terms of community characteristics considered, and the universe of communities to be investigated. No classification, then, is useful in every case. The majority of the earlier classification schemes did not include environmental characteristics.

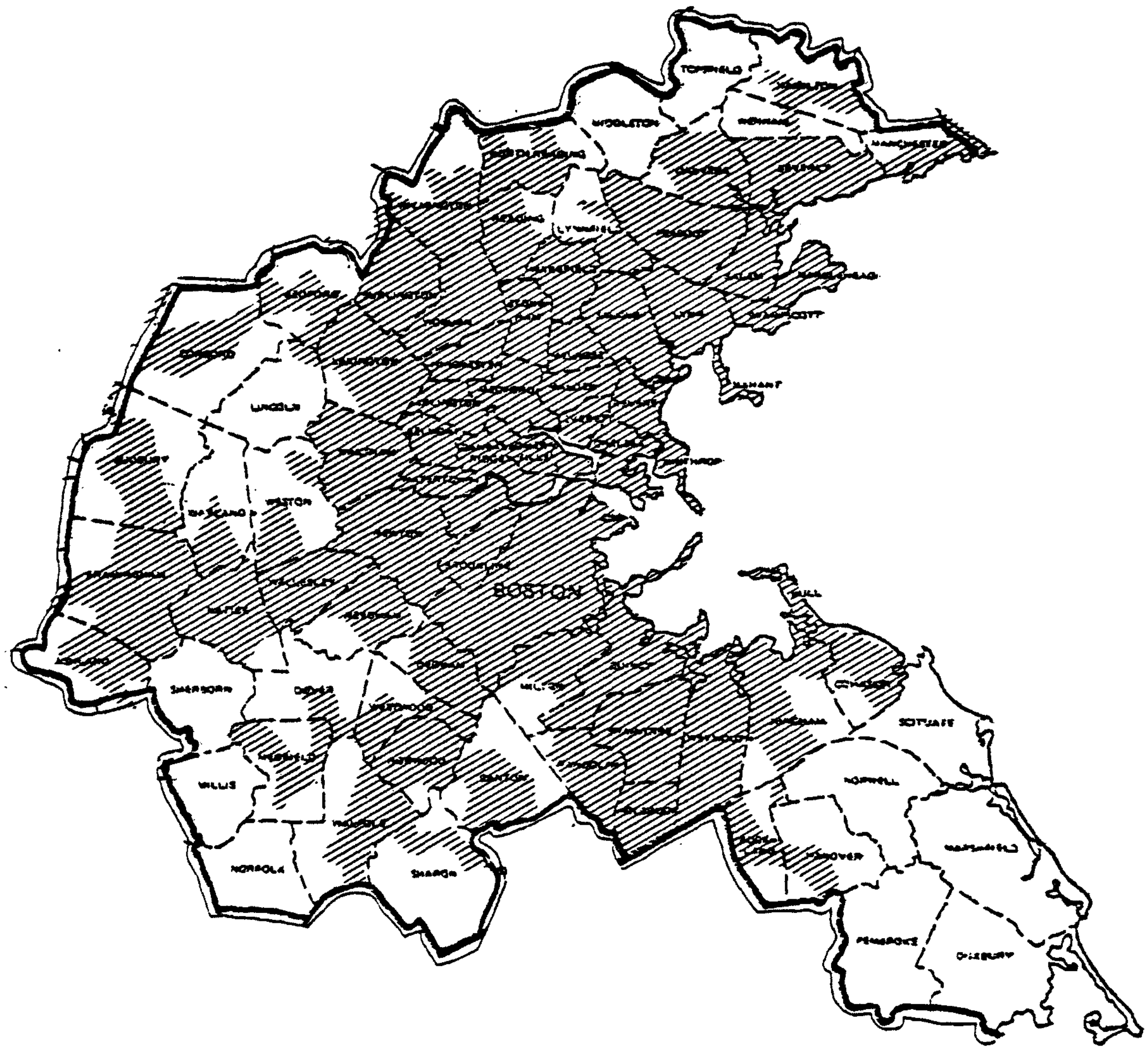
Recently there has been a great interest in environmental quality and quality of life. Coughlin* performed factor analysis for 101 metropolitan areas on sixty indicators of environmental quality and quality of life. The

---

*Robert E. Coughlin, Goal Attainment Levels in 101 Metropolitan Areas, RSRI Discussion Paper Series No. 41 (Philadelphia: Regional Science Research Institute, 1970).

Figure 1-1

Boston SMSA



Outer Boundary of Boston SMSA

Urbanized Area, 1970

analysis, however, was biased toward social and economic characteristics since data on physical conditions was sparse. The John Somers' study* performed for EPA represents another recent example. This study utilized 1960-61 Census data for the most part, therefore, its results are somewhat obsolete. More relevant is Berry's study,** Land Use, Urban Form and Environmental Quality, which provides a city classification based on social, economic, and environmental characteristics. Although there are weaknesses in both the data base and the methodology used for this research, Berry has made a beginning and provides inspiration, as well as a core data base, for future research.

A complete review of community classification studies may be found in Appendix B to Volume III. Most of these studies developed groupings for single research purposes (e.g., transportation analysis, environmental quality analysis, etcetera); in many, the groupings are limited to a subset of U.S. metropolitan areas; and some of the data sets used are incomplete or out of date. In contrast, the research described here did not identify a single urban typology, rather it developed a flexible methodology through which urban classifications may be developed for testing a variety of research hypotheses. Further, all 262 SMSAs are included in the analysis, which utilized an extensive data base, with much of the information recently becoming available.

1.2 Research Methodology

As the initial research objective specified case study site selection for environmental analysis, the data base was designed to include descriptors of ambient environmental quality as well as its causal variables. This data bank thus includes information on ambient air and water quality, on the types and quantities of residuals being discharged into the environment, on socio-economic parameters, on the activities of the local government which affect environmental quality, and on variables describing the urban form, including land use, density, and so on. The data sources used include STORET, SEAS, the Bureau of the Census, the Department of Transportation, and the Department of Agriculture. Specific policy analyses would only use the relevant portions of the data, of course.

Theoretically, it is possible to develop city groupings directly on the basis of the variables. However, the number of variables in the data bank represent too many "axes" along which cities may differ, making it impossible to develop consistent city groupings. We have used an iterative, two stage factor analysis procedure for data reduction purposes.

Factor analysis is an arithmetic means of reducing a complex and highly intercorrelated data set to a smaller number of underlying factors. For example, the research design may require information on the percent of families below the poverty level, or the prevalence of substandard housing

*John Somers, George B. Pidot, Jr., Modal Cities, prepared under Contract #EPA-600/5-74-027, for The Office of Research and Development, U.S. Department of Environmental Protection Agency.

**Brian J.L. Berry, et al, Land Use, Urban Form and Environmental Quality, Chicago: Department of Geography, University of Chicago, 1974.

units, and unemployment statistics: three highly correlated variables. Factor analysis offers one method of combining such variables into a single dimension for further statistical analysis or for grouping communities. This procedure is described in Chapter 2, as well as in Volume III.

The use of factor analysis for data reduction purposes and for the development of groupings has been widely critiqued. In a tongue-in-cheek study of the dangers of indiscriminate use of factor analysis, J. Scott Armstrong uses an example in which Tom Swift, the young analyst, must collect data of significance and then analyze a sample of metal blocks. Armstrong chose the variables in such a way that there are only five significant variables in the grouping of eleven, the other six being only combinations of the first five. The results are amusing with the characteristics of the metal block identified as "intensity, shortness and compactness."* The point made in the Armstrong article is that the investigator must have some prior knowledge of the sample under study, first, to frame the hypotheses, and second, to interpret the results in light of reality.

In order to frame the research hypotheses, and to assist in the interpretation of the results, an iterative, or two stage factor analysis procedure, was used. This approach also structured and facilitated the data collection efforts. The variables, on which information was to be collected, were separated into six categories, or significant characteristic sets (SCSs), each of which describe or affect ambient environmental quality. These sets are:

SCS 1. Ambient Environmental Quality

SCS 2. Urban Form and the Physical Environment

SCS 3. Residuals

SCS 4. Household Sector

SCS 5. Government Sector

SCS 6. Industrial Activity

Simple factor analyses were performed for each of these SCSs, with the resultant factors being inputs into the second stage factor analysis. City groupings were then developed on the basis of the second stage factors obtained. The objective here was to minimize the within group variance, and to maximize the between group variance in terms of the dimensions defined by the second stage factor analysis. In other words, the objective was to form groups of cities similar to one another, but different from the cities of other groups.

---

*J. Scott Armstrong, "Derivation of a Theory by Means of Factor Analysis or Tom Swift and His Electric Factor Analysis Machine," The American Statistician (December 1967): 17-21.

Modal, or most representative, cities were then selected for each of the groups by simply identifying the city in each group which lies the closest in the multidimensional space to the geometric centroid (center) of the group.

1.3 Results

This research project developed an urban typology, and identified representative SMSAs appropriate for general environmental analysis. In addition, it developed a flexible capability for developing similar typologies and identifying representative SMSAs for testing alternative research hypotheses. Each of these will be described in turn.

1.3.1 SMSA Groupings for General Environmental Research--

Five major city groups were identified for general environmental purposes; these five groups include 175 of the 262 SMSAs considered. The remaining cities are either outliers, single cities significantly different from the cities in the five major groups, or they are in "minor" groups comprised of a smaller number of cities, having different characteristics.

Table 1-1 describes the five major city groupings. The largest group contains 48 SMSAs, with the modal city being Albany-Schenectady-Troy, New York. Other cities in this group include Appleton-Oshkosh, WI; New Britain, CT; Portland, OR; etcetera. The second largest group contains 46 SMSAs, with Lake Charles, LA, being the modal city, while the smallest group contains 18 cities, with Dallas, TX, being its representative.

Table 1-1

Major SMSA Groupings for General Environmental Purposes

Group No.	No. of Cities	Representative City	Cities Close to Modal City
1	36	Little Rock-North Little Rock, AR	Baton Rouge, LA Corpus Christi, TX Lafayette, LA Midland, TX Montgomery, AL Odessa, TX Tyler, TX
2	46	Lake Charles, LA	Spartanburg, NC Parkersburg-Marietta, WV-OH
3	27	Williamsport, PA	Davenport, IL Evansville, IN-KY Lawrence, MA Peoria, IL
4	48	Albany-Schenectady-Troy, NY	Appleton-Oshkosh, WI New Britain, CT Portland, OR
5	18	Dallas, TX	Charlotte, NC Richmond, VA

From this classification the single most representative city for environmental analysis in the United States is Louisville, Kentucky. If one were limited to a single case study, or demonstration project, the results would suggest that it should be located in Louisville, KY. The study results then could be allowed for a greater number of case studies/demonstration projects, say five, these should be located at Little Rock, AR; Lake Charles, LA; Williamsport, PA; Albany-Schenectady-Troy, NY; and Dallas, TX; with the results being appropriate for the other cities in each of the five groups.

To assess the effects of city size on city characteristics, the set of 262 cities was divided into small (less than 200,000) population, medium (between 200,000 and 500,000 population), and large (greater than 500,000 population) SMSAs. Two analyses were performed: first, the second stage factor analysis was repeated for each of the city size groups. Second, a separate clustering, similar to that described for the entire sample, was performed within each of these strata.

Second stage factors remained stable for the three size groups. For example, factor 1 (largest explanatory power) from the all city analysis, indicating low income, low expenditures for sewerage and low levels of manufacturing activity, showed up as factor 1 in the analysis for each group of cities. These clusters were based on a single set of stage 2 factors, the set used for the general classification.

Separate classifications may be very useful where city size is of great importance. The classifications appear to provide similar results. Note from Table 2-7 that small, medium, and large SMSAs are distributed throughout the general classification. Clusters within city size strata were found to be similar to the general SMSA groups, and to clusters in the other size groups. For example, Group 1 of the small SMSAs and Group 1 of the medium SMSAs show similar characteristics, as do Group 3 of the medium size SMSAs and Group 4 of the large SMSAs. In other words, city size did not significantly affect the classification scheme.

1.3.2 Application to Alternative Research Hypotheses --

The data collected in this research project, and the methodology developed may be used to assist other environmental and urban research in three major ways: through the direct use of the data, through the identification of appropriate case study sites, and through facilitating the generalization of study results. Each of these will be described in turn.

The use of the data collected during this project for other research purposes is an obvious function. Although our data collection efforts were limited to secondary sources, some of the information contained in the data base was not easily accessible to the public. Information on ambient water and air quality, obtained from STORET and from the SEAS model, are two such examples. Some of the descriptors of land use represent another case in point: the urbanized proportions of SMSAs, and the land area devoted to outdoor recreation were obtained from USDA, and the Bureau of Outdoor Recreation, respectively. This data collection effort should not be duplicated by other researchers; a comprehensive description of our data collection efforts, as well as a complete listing of data may be found in Volume III of this series of reports.

As described above, a broad data base containing some 200 variables was collected containing descriptors of ambient environmental quality and a diversity of other phenomena believed to affect environmental quality. Alternative research hypotheses may be described in terms of the variables contained in the data base, on the basis of which representative cities appropriate for case study/demonstration project siting may be selected. For example, a program analyst interested in the effects of the bottle bill on resource recovery may be interested in funding a limited number of demonstration projects. The optimal sites for these may be identified by first specifying the variables believed to affect the outcome, then performing factor analysis, groupings, and the selection of representative cities as described above. If the number of variables is limited, the research hypothesis is well defined, a simple one-stage factor analysis may be appropriate. Additional variables of interest available from secondary sources may also be added. In addition, the universe of cities may be limited to fit the requirements of the particular research project; this may be limited to cities in a certain size range, in certain geographical region, or to cities possessing certain attributes, such as high unemployment.

Environmental deterioration and other problems frequently occur in SMSAs which are "outliers," which do not fit into any of the groups. Although the studies analyzing these effects cannot be easily generalized to other cases, a limited extrapolation may be possible. An examination of the data for the outlying SMSAs will identify the factor axes or variables with extreme observations, which are in fact responsible for the outlier position of the city. If these variables are not crucial to the analysis, the city may be grouped with others in terms of the remaining variables. The study results then can be generalized to this group, although at a lower level of confidence.

This capability was tested during the course of the project in connection with siting potential demonstration projects by the Energy Research Development Agency (ERDA). This agency is interested in a limited number of demonstration projects for electric cars; potential sites for these demonstration projects were identified by USR&E. The results of this application are described in more detail in Chapter 4 of this volume, as well as in Volume II.

Case studies or demonstration projects may not always be performed at their ideal sites; data limitations or other constraints may prevent this. Alternatively, a researcher may be interested in generalizing the results of a previously performed study. The methodology developed in this project may facilitate this process as well. Variables are selected, factor analysis is performed, and groupings are developed in the same manner as described above. The cities of interest are then located within or outside the groups, indicating the degree the study results may be generalized. The development of the general city typology, the data bases, factor analytic techniques, and clustering methods used are described in Chapter 2. The applicability of the groups and their modal cities to general environmental research is discussed in Chapter 3; the development of alternative urban typologies is summarized in Chapter 4.

2.0 GENERAL CLASSIFICATION OF SMSAs

2.1 Research Design and Data Base

Essential to the design of any classification methodology are (1) definition of the entities to be classified, (2) identification of attributes to be considered in the classification and the formulation of hypotheses concerning those attributes, and (3) selection of appropriate techniques by which the data can be used together to differentiate between observations. Each of these will be discussed below.

2.1.1 The Set of Localities to be Classified --

For this research, the definition of entities to be classified was a simple manner. Standard Metropolitan Statistical Areas (SMSAs) have been defined by the Office of Management and Budget (OMB) to represent the areas in and around one or more cities that act as a center in which the activities form an integrated economic and social system. Although other definitions of U.S. metropolitan areas are available, none have been utilized as extensively as the SMSA for the collection of data. The use of SMSAs achieved the greatest possible amount of data consistent between metropolitan areas of the U.S.

As cities are constantly changing and OMB revises the definition of an SMSA periodically, it was necessary to perform the classification from the perspective of a single point in time. The set of SMSAs defined as of January 7, 1972 was arbitrarily selected as the set of SMSAs to be classified and data utilized in the classification was that which most closely represented each SMSA at that point in time.

2.1.2 The Data Base --

Figure 1 in "Executive Summary" presents a general diagram intended to identify the major determinants of environmental quality. Beginning at the bottom of the diagram, ambient environmental quality is shown to be a function of both residuals and the capacity of the environment to dilute and/or neutralize pollutants (urban form and physical environment). Urban form also influences residuals generation, particularly in the transportation sector. Residuals are also the net result of pollution generated by households and industry and abatement efforts by the public sector. The public sector also influences consumption through investment in public facilities. However, the three areas at the top are reduced to simple form by associating abatement with the public sector, consumption with households, and production with industry. Second order effects, such as the influence of environmental quality on the three actors at the top of the causal path, are not considered here.

The primary areas of interest, then, are:

1. ambient environmental quality
2. urban form and the physical environment
3. residuals
4. demographic characteristics
5. government
6. industry.

The variables in each of these categories comprise a significant characteristic set (SCS). Initial sets of variables used to represent the six SCSs are listed in Appendix A to this volume. These variables represent the data which are presently available; many items are indicators of relevant activity or surrogates for otherwise unquantified attributes.

For the general environmental classification of cities, only four SCSs were utilized. These are: urban form and the physical environment; demographic characteristics; government; and industry. These SCSs include the variables appropriate for a general urban classification scheme. Further, these variables describe some of the factors which determine ambient environmental quality, as well as the sources and levels of pollutants. The resulting typology should then be appropriate for urban research, as well as for general environmental policy analysis.

SCS2 (urban form) contains variables describing the distribution of activities within the SMSA, the density of the SMSA and its urbanized portions, the assimilative attributes of the city, as well as some transportation related measures. SCS4 is comprised of household descriptors, providing information on the demographic characteristics of the population, on housing quality and living conditions, on economic welfare, on population and income changes, and on the modal split in transportation. The public sector, SCS5, contains information on Government expenditures for improving environmental quality, as well as general community concern and involvement. SCS6 is comprised of industry variables indicating the importance of industries critical to environmental quality,* the importance of manufacturing and describing the industry mix in terms of 20 manufacturing categories, wholesale trade, retail trade, and selected services.

Because previous studies have found size and regional location to dominate groupings, the variables were "standardized" where appropriate: in that they were expressed in per capita or normalized terms. Regional biases were also eliminated where possible; variables that describe only regional location (e.g., latitude) were excluded.

2.1.3 Methodology --

As described in Chapter 1, the data set contains too many variables to perform groupings directly so a data reduction technique is necessary. An iterative, two stage factor analysis was chosen in order to facilitate framing the research hypotheses, and to aid in the interpretations of the resulting factors. Selected factors derived from simple factor analyses of the four individual SCSs served as the second stage factor analysis, the second stage factors then formed variables, the basis for the clustering.

The basic function of factor analysis is to ascertain and measure the fundamental dimensions or interrelationships of a set of variates. The transformation is made possible by moderate or high correlation between variables which compose the original data base. When two or more variables are very highly correlated, a single "factor" may describe this related variable set. In geometric terms, correlations between the original variates indicate nonorthogonal

*Industries critical to environmental quality have been defined as the heaviest polluters, and industries where abatement is the most difficult.

dimensions. In most cases, resolution into orthogonal (independent) dimension: factors will simplify the vector space.*

Many alternative techniques have been devised in the development of factor analysis. Although new methods often improve on the last for some analyses, no single method is regarded as better than all others. In fact, none is applicable to all research efforts. The methods differ in basic assumptions about the data base as well as in solution techniques. As a result, the value of any given application of factor analysis rests in the ability of the researcher to choose the method most appropriate to his research. Alternative methods of performing factor analysis including principal components analysis and common factor analysis are described in Appendix A to Volume II.

For purposes of this research, principal components analysis was selected as the preferred method; with common factor analysis being the principal alternative considered. The result of the two techniques did not differ substantially, with principal components analysis being the less expensive alternative, in terms of computer costs. The choice of this technique is described in greater detail in Section 4.5.3 of Volume II. Other aspects of refining the methodology include selecting the method of rotating the factors and the choice of the clustering technique. These are discussed in Chapters 5 and 6 of Volume II, respectively.

2.2 Stage I Factor Analysis

For each SCS, the general approach to developing Stage I factors was to perform initial statistical descriptions and then repeat factor analyses to:

1. clear the data set of unrealistic and erroneous observations;

2. modify hypotheses and the variables in each SCS to reflect knowledge gained from initial analysis;

3. estimate values to replace missing observations;

4. complete the Stage I factor analysis.

The sections of this chapter will describe, separately for each SCS, the course of analysis followed and the outcome of Stage I analysis.

2.2.1 SCS2--Urban Form and the Physical Environment --

The analysis of SCS2 included a large number of factor analysis runs--each yielding new information about the measures involved. From the 30 variables originally included in this SCS, 15 were chosen for the final Stage I factors which best reflected the urban form concept of the original research hypothesis.

*Note, however, that the final (preferred) solution may include non-orthogonal factors.

The remaining set of 15 variables was factor analyzed and yielded the five factors shown in Table 2-1, which explain 65.7 percent of the variance in the variable set. Factor 1, for example, loads on all five central tendency measures, and thus presents a measure of the proportion of activity which takes place in the central city(s) of the SMSA. It is not surprising this appears as the primary indicator of urban form.

Factor 3 has a high positive loading on the number of primary radial facilities and a positive loading on the number of major circumferential facilities (roads) in the urban area. The factor loads negatively on square miles per person in the SMSA. Together, these indicate a heavily urbanized area.

Factor 5 loads heavily on the number of population centers in the area (PCR) and the number of square miles per person in the central city(s) (SQMC). This indicates a dispersed pattern of urbanization involving more than one community in the core of the area.

Table 2-1

SCS2 - Stage I - Factor Analysis for 15 Variables

Variable	FACTOR 1	2	3	4	5
Portion of Employment in Central City	++				
Portion of Total Population in Central City	++				
Portion of Manufacturing Employment in the Central City	++				
Portion of Retail Sales in Central City	++				
Portion of Manufacturing in Central City	+				
Portion of Land Which is Urbanized		++			
Portion of Workers Working Outside SMSA		++			
Portion of Land Devoted to Outdoor Recreation		+			
Number of Radial Roadways			++		
Number of Major Circumferential Roadways			+		
Square Miles Per Person			-		
Total Miles of Roadway per Capita				++	
Portion of Principal Arterials				--	
Number of Population Centers					++
Square Miles per Person - Central City					+

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained

65% of variance explained.

2.2.2 SCS4--The Household Sector --

Analysis of SCS4 data yielded factors which describe the socioeconomic character of an urban area. In this sense, the analysis of this SCS is more similar than any other to previous efforts in city classification. The variables used in this factor analysis, and their factor loadings are presented in Table 2-2. The interpretation of the resulting seven factors is summarized in Table 2-3.

Table 2-2

SCS4 - Stage I - Factors

Variable	Factors 1	2	3	4	5	6	7
Movers into the SMSA	++						
Individuals Residing in the Same Dwelling for Five Years	—						
Population Change - SMSA	+						
Female/Male in 20-64 Age Group	-						
Population Change - Central City	+						
Employmtn/Population - Total		++					
Employment/Population - Male		++					
Employment/Population - Black		++					
Employment/Population - Female		+					
Income - Gini Coefficient			++				
Median Family Income			-				
Married Couples without Own Household			+				
Crowded Housing Units			+			++	
Per Capita Income			-		+		
Work Trip - Drivers				++			
Work Trip - Passengers				-			
Single Family Dwellings				+			
Households with one or more Automobiles				+			
Owner-Occupied Housing				+			
Units in Structure -- More than Five				-	+		
Portion of Population which is Urban					+		
Units in Structure -- More than Fifty					+		
Portion of Population which is of Foreign Stock					+		
Average Household Size						++	
Median Age						-	
Fertility Rate						+	
Portion of Population which is Black							++
Infant Death Rate							+
Relative Death Rate							+

growth; employment; standard of living; transportation; housing; cosmopolitan; family structure; health standard

78.3% of variance explained

Table 2-3

SCS4 - Stage I - Interpretation of Factors

Factor	Positive Score Indicates some Combination of--
1. Growth	Immigration to the SMSA Increasing population in SMSA and central city
2. Employment	High employment participation rates in all sectors of the population
3. Standard of Living	Relatively low level of income as indicated by median and per capita measures Unequal distribution of income over the population Crowded housing
4. Location-Density	High proportion of single unit and owner-occupied housing Heavy dependence on the automobile
5. Cosmopolitan	Relatively high level of income as indicated by a per capita measure Dense housing Urban and foreign elements of the population relatively large
6. Family Structure	Crowded housing Large proportion of young families
7. Health Standard	High infant death rate Relatively high overall death rate given the age distribution of the population

2.2.3 SCS5--The Public Sector --

As indicated in Table 2-4, eight variables have been used to describe the public sector. These yielded four factors.

Table 2-4

SCS5 - Stage I - Final Factors

Variable	Factors 1	2	3	4
Local Government				
General Revenues	+++			
General Expenditures	++ large government			
Employment	+++			
Severage Expenditures — Total		+++		
Severage Expenditures — other than capital		+++ severage		
Expenditures on Water Supply			+++ water	
Expenditures on Parks & Recreation			+ & parks	
Expenditures on Sanitation other than Severage				sanitation +++

85.5% of variance explained

2.2.4 SCS6: The Industrial Sector --

Results of the first stage factor analysis for this SCS are shown in Table 2-5. The eight factors described here explain 59 percent of the variance indicated by the 24 input variables. The 8 factors appear to describe well the basic dimensions of industry mixes and could be easily titled as follows:

1. overall level of industrial activity
2. services and trade
3. textiles and apparel
4. miscellaneous manufacturing and instruments
5. fuel and chemicals
6. paper and allied products
7. leather and leather products
8. lumber and wood products.

The factors are listed in order according to the amount of variance explained by each; the latter factors, therefore, are the least valuable to the factor description of SMSAs.

Table 2-5

SCS6 - Stage I - Final Factors

Factor	Variables	
1	VAM	Value added in Manufacturing
	EM	Employment in Manufacturing
	S34	Value added in Fabricated metal products
	S35	Value added in machinery, except electrical
2	WHOL	Wholesale Sales
	RETT	Retail Sales
	SS	Selected Services
	S27	Value added in printing and publishing
3	S22	Value added in textile mill products
	S23	Value added in apparel and other textile products
4	S39	Value added in miscellaneous manufacturing industries
	S38	Value added in instruments and related products
5	S28	Value added in chemicals and allied products
	S29	Value added in petroleum and coal products
6	S26	Value added in paper and allied products
7	S31	Value added in leather and leather products
8	S24	Value added in lumber and wood products

59% of variance explained.

NOTE: These factors did not load on the remaining 7 variables: S20--value added in food and kindred products; S25--value added in furniture and fixtures; S30--value added in rubber and plastic products; S32--value added in stone, clay and glass products; S33--value added in primary metal industries; S36--value added in electrical equipment and supplies; S37==value added in transportation equipment.

2.3 Stage II Factor Analysis

The two-stage application of the factor analytic technique was utilized to insure the proper evaluation of research hypotheses. Stage I analysis involved separate factor analysis for each SCS to reveal the hypothesized underlying dimensions within each group of variables. The output from individual Stage I analyses was then combined and used as the input to Stage II. Stage II thus identifies the relationships between SCSs and the basic underlying attributes of U.S. metropolitan areas.

For a variety of reasons--both conceptual and statistical, a limited number of Stage I factors were chosen for input to Stage II. This set included the first four factors from four SCSs--2, 4, 5, and 6--with the exception of Factor 3 from SCS2.

Stage I analysis yielded a different number of factors for each of the four SCSs to be pursued in Stage II. For SCSs 2, 4, 5, and 6, the number of factors was 4, 7, 4, and 8, respectively as indicated in the previous section. To use this set of factors in Stage II would create a significant bias toward SCS4 (the household sector) and SCS6 (the industry sector). In addition, since Stage I factors from each SCS are mutually independent, excess factors in SCS4 and SCS6 will lead to the formation of additional separate factors beyond the primary factors indicated by interactions between the four SCSs. With these considerations, the number of factors from Stage I to be included in Stage II was limited to four per SCS.

The loss of explanatory power from the exclusion of these factors is significant but tolerable. The loss will be 28.3 percent in SCS4 and 21.0 percent in SCS5. Percent of variance explained by the first four factors in each SCS is as follows:

SCS2*	Urban Form	70.4 percent
SCS4	Household Sector	50.0 percent
SCS5	Public Sector	85.5 percent
SCS6	Industry Sector	38.0 percent

Although several alternative combinations of Stage I factors were tested, the even distribution of factors between SCSs yielded the most meaningful factors, therefore, the reduced set was accepted as the best input to Stage II.

Factor analysis of the 15 Stage I factors shown in Table 2-6 yielded 6 Stage II factors which, together, explain 63.1 percent of the variance in Stage I factors. These factors are intuitively satisfying as well as statistically valid in that each factor represents a set of characteristics which are likely to be encountered together in an urban area.

Factor 1 indicates a low standard of living, low government expenditures for sewerage, and a low level of total manufacturing activity; this factor would characterize an economically depressed area on this factor scale.

*Factor 3 was dropped from SCS2 because of data problems.

Factor 2 indicates a low level of total manufacturing activity, heavy growth in recent years and a high concentration of population and economic activity in the core (central city) of the SMSA.

Factor 3 indicates manufacturing activity in the miscellaneous category, a compact core, and heavy expenditures for sanitation other than sewerage.

Factor 4 indicates high employment and service trade activities.

Factor 5 indicates low residential densities, high auto dependence, little manufacturing in industries such as textiles and apparel, and heavy expenditures on water supply and recreation.

Finally, factor 6 indicates a highly urbanized SMSA with a relatively large local government.

Table 2-6

Stage II Factors

SCS	Stage 1 Factor No.	Stage I Factor Name	Stage II Factor 1	2	3	4	5	6
4	3	Low Standard of Living	++					
5	2	Expenditure on Sewerage	-					
6	1	Overall Level of Manufacturing	-	-				
4	1	Growth		++				
2	1	Central Tendency		+				
6	4	Miscellaneous Manufacturing & Instruments			++			
2	4	Sprawling Core			-			
5	4	Expenditure on Non-Sewerage Sanitation			+			
4	2	Employment				++		
6	2	Services & Trade				++		
4	4	Location-Density (many single-family homes, high auto dependence)					+	
6	3	Textiles & Apparel					-	
5	3	Expenditure on Water Supply & Recreation					+	
5	1	Large Government						++
2	2	Highly Urbanized						-

Before proceeding to the next stage of the analysis, which identified groups of similar cities, a test was performed to determine the stability of Stage II factors between different size cities. The factor analysis was repeated for each of three groups of cities:

Group	Population
small	less than 200,000
medium	200,000-500,000
large	more than 500,000

The proportion of variance explained by primary factors is stable, varying only from 61.7 percent to 65.4 percent.

2.4 SMSA Groupings and Representative Cities

Groups of similar cities were identified through a simple geometrical clustering technique; in which each SMSA is initially considered a separate point. The two groups separated by the smallest geometric distance are then located and combined to form a new group with its centroid (center) midway between the two points. Then, the two groups with nearest centroids of the new set are combined; and a new centroid located; the process can continue until only one groups remains. The centroids are weighted by the number of SMSA's already in the group.

Criteria for choosing a stopping point in the process include the size and number of groups, and the relationship between within-group variance and between-group variance.

Once the set of groups is selected, modal cities are identified by simply determining which city in each group lies closest in the multidimensional space to the geometric center of the group.

It would have been possible to develop SMSA groups directly from the initial variables using the geometrical clustering routine. In practical terms, however, the variables of the data base provide too many dimensions along which cities may differ--the additional descriptive information provided by the variables stresses the uniqueness of each city rather than underlying basic characteristics which the cities have in common. Because the use of too many dimensions creates an unmanageable set of groups, input to the cluster analysis was arbitrarily limited to the first four Stage II factors.

Table 2-7 shows how the 262 SMSAs grouped together form basic classes of SMSAs. Five major groups of cities were identified; these include 175 of the 262 cities. The remaining cities grouped together as follows:

36 SMSAs in groups of five or more
34 SMSAs in groups of two to five

---

*Computer program written by Howard Gilbert and Steve Chasen, Health Sciences Computing Facility, University of California, Los Angeles, California. Reference: R.R. Sokal and P.H.A. Sneath (1973) Numerical Taxonomy: the Principles and Practice of Numerical Classification (San Francisco: W.H. Freeman and Co.).

17 SMSAs in groups of one.

In Table 2-7 the double lines indicate division between groups and the dotted lines delineate subgroups. Subgroups within the major groups exhibit minor dissililarities; the major groups are dramatically different.

Group I shows a low level of manufacturing activity, low income and low expenditures on sewerage. This is tempered by moderate loadings on Stage II factors 2 and 4 which signify a growing economy oriented more than the average toward services and trade. Little Rock, the modal city for this group, is very close to the group's centroid. In 1970, Little Rock was less wealthy than the average SMSA as is indicated by the number of families below the low income level: 13.5 percent as compared to the national level of 8.5 percent. But the area is growing and, in 1970, enjoyed a high rate of employment of 3.3 percent (average for all SMSAs was 4.3 percent), and 34.7 percent of all housing units were built since 1960 (average for all SMSAs was 25.5 percent). Employment in manufacturing was below the national level: 20.1 percent as compared to 25.8 percent for all SMSAs in 1970. Other cities found near the center of this group include Baton Rouge, LA; Corpus Christi, TX; and Montgomery, AL.

Group II includes cities which are closer to the centroid of all cities than Group I. Factor scores indicate these cities have high unemployment and are not active in services and trade. In addition, they generally have slightly lower than average income levels and economic activity and they may have experienced less than average growth. Lake Charles, LA, the modal city for this group, is still different from a hypothetical city at the centroid of the group, loading more heavily on Factor 1 and not at all on Factor 3. A high Factor 1 score is the result of a lower than average standard of living. Of all families in Lake Charles, 16.6 percent have incomes below the low income level and 4.7 percent of all housing units lack some or all plumbing facilities (national averages for SMSAs are 8.5 and 2.9 respectively). And, a large negative score on Factor 4 reflects the combined effect of slightly more than average activity in services and trade accompanied by very high unemployment (5.7 percent as opposed to the average 4.3 percent). Other cities near the centroid of the group are Spartanburg, NC; and Parkersburg, WV. The group also includes the overall modal city of Louisville, KY.

Group III has high negative scores on Factors 2 and 3. Cities in this group, then are expected to be small SMSAs (in area) with an industrial base and little recent growth. Williamsport, PA, the modal city for this group, includes only one county of moderate size, 42.6 percent of its employment is in manufacturing and population growth in the decade ending 1970 was only 3.6 percent, as compared with the national average of 16.6 percent for SMSAs. Other cities near the centroid of this group include Davenport-Rock Island-Moline, IA-IL; Evansville, IN-KY; and Lawrence-Haverhill, MA-NH.

Group IV is even closer to the overall centroid than Group II. Scores are moderately negative for Factors 1 and 2, moderately positive for Factor 3 and almost zero for Factor 4. Thus, these cities are expected to rank about average on the dimensions defined by the factor analysis. The modal city, Albany-Schenectady-Troy, NY, has a more negative score than the centroid

Table 2-7
Groups of Similar SMSAs*

GROUP 1

- Abilene, TX
- Lafayette, LA
- San Angelo, TX
- Midland, TX
- Lubbock, TX
- Tuscon, AZ
- Albany, GA

- Knoxville, TN
- West Palm Beach-Boca Raton, FL
- Tampa, FL
- Monroe, LA
- Orlando FL

- Sherman-Denison, TX

- Corpus Christi, TX
- Shreveport, LA
- Macon, GA
- Texarkana, TX-AR
- Wilmington, NC
- Savannah, GA
- Tyler, TX
- Montgomery, AL
- Jackson, MS

- New Orleans, LA
- Portland, ME
- Waco, TX
- Little Rock-N. Little Rock, AR**
- Odessa, TX
- Tulsa, OK
- Baton Rouge, LA

- St. Joseph, MO
- Sioux City, IA-NB
- Billings, MT
- Boise City, ID
- Huntsville, AL
- Springfield, MO
- Amarillo, TX

GROUP 2

- Huntington-Ashland, WV-KN-OH
- Stockton, CA
- Modesto, CA

- Augusta, GA-SC
- Pueblo, CO
- Fresno, CA

- Killene-Temple, TX
- Lewiston-Auburn, ME
- Pine Bluff, AR
- Spokane, WA
- Owensboro, KN
- Fort Myers, FL
- Yakima, WA

- Florence, AL
- Santa Rosa, CA
- Pensacola, FL
- Riverside-San Bernadino-Ontario, CA
- Provo-Orem, UT
- Charleston, SC
- Salem, OR
- Duluth-Superior, WI-MN

- Altoona, PA
- Gadsden, AL
- Lake Charles, LA**
- Mobile, AL
- Bakersfield, CA
- Chattanooga, TN-GA
- Lakeland-Winter Haven, FL
- Birmingham, AL

- Erie, PA
- Wheeling, WV-OH
- Poughkeepsie, NY

- Louisville, KN-IN
- Richland-Kennewick, WA
- Springfield, OH
- Parkersburg-Marietta, WV-OH
- Sacramento, CA

- Gastonia, NC
- Salt Lake City, UT
- Petersburg-Colonial Heights-Hopewell, VA
- Charleston, WV
- Wilkes Barre-Hazelton, PA
- Beaumont, TX
- Spartanburg, SC
- New Bedford, MA
- Alexandria, LA

MINOR GROUP

- El Paso, TX
- Tuscaloosa, AL
- San Antonio, TX
- Columbus, GA-AL

MINOR GROUP

- Galveston-Texas City, TX
- Manchester, NH
- Santa Barbara, CA
- Columbia, SC

GROUP 3

- Allentown-Bethlehem-Easton, PA-NJ
- Harrisburg, PA
- St. Louis, MO-IL
- Greenville, SC
- Reading, PA
- York, PA
- Lancaster, PA

- Cedar Rapids, IA
- Waterloo-Cedar Falls, ID
- Fort Wayne, IN
- Toledo, OH-MI
- Rockford, IL
- Grand Rapids, MI

- Davenport-Moline-Rock Island, IA-IL
- Williamsport, PA**
- Elmira, TX
- Mansfield, OH
- Lima, OH
- Peoria, IL
- Lawrence-Haverhill, MA-NH
- Evansville, IN-KN

- Indianapolis, IN
- Springfield, IL
- Wichita, KS
- Decatur, IL
- Terre Haute, IN
- Anderson, IN

*This table is to be used in conjunction with Table 6-5 as discussed below pertaining to Group V.
**Modal SMSA

Table 2-7 (continued)

MINOR GROUP

Hartford, CT
Minneapolis-St. Paul, MN-WI
San Jose, CA
Milwaukee, WI
Washington, DC-MD-VA
Norwalk, CT
Rochester, NY

GROUP 4

Bristol, CT
Meriden, CT
New London-Norwich, CT-RI
Fitchburg-Leominster, MA

Baltimore, MD
Los Angeles, CA
Melborune-Titusville-Cocoa Beach, FL
Daytona Beach, FL
Newark, NJ
Philadelphia, PA
Buffalo, NY

Appleton-Oshkosh, WI
Syracuse, NY
Racine, WI
Scranton, PA
Loraine-Elyria, OH
Worcester, MA
South Bend, IN
Binghamton, NY-PA

New Brunswich-Perth-Amboy-Sayreville, NJ
Wilmington, DE-NJ-MD
Cleveland, OH
Detroit, MI
Trenton, NJ
Dayton, OH
Cincinnati, OH

Brockton, MA
Pittsfield, MA
Lowell, MA-NH
Kenosha, WI

Anaheim-Santa Ana-Garden Grove, CA
New Haven-West Haven, CT
Chicago, IL
Jersey City, NJ
Seattle-Everett, WA
San Francisco-Oakland, CA
Portland, OR-WA
New Britain, CT
Oxnard-Simi Valley-Ventura, CA
Nashville, TN
Danbury, CT

Green Bay, WI
La Crosse, WI
Dubuque, IA
Ogden, UT
Santa Cruz, CA
Hamilton-Middletown, OH
Albany-Schenectady-Troy, NY**

**Modal SMSA

GROUP 5

Charlotte, NC
Dallas, TX**
Oklahoma City, OK
Raleigh, NC
Fort Worth, TX

Lexington, KY
Tallahassee, FL
Jacksonville, FL
Durham, NC
Memphis, TN-AR-MI

Des Moines, IA
Kansas City, KS-MO
Stamford, CT
Richmond, VA
Omaha, NB
Denver, CO
Columbus, OH
Houston, TX

MINOR GROUPS

Fort Lauderdale-Hollywood, FL
Phoenix, AZ
Roanoke, VA
Sarasota, FL
Fall River, MA-RI
Albuquerque, NM

Eugene-Springfield, OR

Fargo-Moorehead, ND-MN
Lincoln, NB
Lafayette-West Lafayette, IN
Rochester, MN
Topeka, KS
Bloomington-Normal, IL

Canton, OH
Youngstown-Warren, OH
Pittsburgh, PA
Paterson-Clifton-Passaic, NJ
Utica-Rome, NY
Steubenville-Weirton, OH-WV
Johnstown, PA

Long Branch-Asbury Park, NJ
Atlantic City, NJ

Vineland-Millville-Bridgeton, NJ

Fort Smith, AR-OK
Lynchburg, VA
Asheville, NC

Las Vegas, NV

New York, NY

Madison, WI

Bryan-College Station, TX
Gainesville, FL
Columbia, MO
Austin, TX

MINOR GROUPS

Greensboro-Winston-Salem-Highpoint, NC

Nashville-Davidson, TN

Sioux Falls, SD
Reno, NV

Atlanta, GA

Baltimore, MD

Jackson, MI
Muskegon-Muskegon Heights, MI

Gary-Hammond, IN
Saginaw, MI
Muncie, IN
Bay City, MI

Flint, MI

Lansing-East Lansing, MI
Kalamazoo-Portage, MI

Ann Arbor, MI

Fayetteville, NC
Lawton, OK

Newport News-Hampton, VA
Norfolk-Virginia Beach-Portsmouth, NC
San Diego, CA

Champaign-Urbana-Rantoul, IL
Colorado Springs, CO

Tacoma, WA
Salinas-Seaside-Monterey, CA
Vallejo-Fairfield-Napa, CA
Great Falls, MT
Biloxi-Gulfport, MS

Miama, FL

Providence-Warwick-Pawtucket, RI-MA
Waterbury, CT

Springfield-Chicopee-Holyoke, MA-CT

Boston, MA

Bridgeport, CT
Akron, OH

Laredo, TX

McAllen-Pharr-Edinburg, TX

Brownsville-Harlingen-San Benito, TX

on Factor 1, perhaps the result of higher incomes and manufacturing activity. Cities very similar to the modal city include Appleton-Oshkosh, WI, and New Britain, CT.

Group V has high scores on Factors 2 and 4, describing large SMSAs which are prosperous, as indicated by growth and high employment, and which are active in services and trade rather than manufacturing. Dallas, TX, has been designated as the modal city for this group and appears to fit the factor description.

A great deal of caution should be exercised in dealing with the modal city and groups of cities since the group includes a large number of cities for which several of the values were estimated. Most of these estimated values are for descriptors of the industry mix, which is important to the grouping fof these cities.

3.0 APPLICABILITY TO GENERAL ENVIRONMENTAL RESEARCH PROBLEMS

As described in the previous chapter, the general city classification scheme excluded the SCSs describing ambient environmental quality, and the residuals discharged into the environment. There were several reasons for this; a classification scheme based on the other four SCSs would result in city groupings useful for general urban research; and the data describing environmental quality had some limitations. Further, the general city groupings should reflect differences in the generation of residuals, and in ambient environmental quality if the causal relationships hypothesized in our research design are true (see Figure 1).

In this chapter, first the data base contained in the ambient environmental quality SCS and in the residuals SCS are described. Second, differentials in environmental quality are analyzed between the general city groupings.

3.1 Environmental Data Base

This data base consists of eleven ambient water quality indicators, two measures of air quality, a single subjective measure of perceived water quality, and eleven drinking water quality variables.

The water quality variables were obtained from STORET. For each SMSA, up to eleven longitude-latitude points were identified along the boundaries; information was retrieved from all STORET stations within the polygon defined by these longitude/latitude points. A simple average of the readings was then calculated for each variable and used to indicate water quality differences across SMSAs. These measures represented approximations at best because of the uneven distribution of sampling over time and over space. Because the motive for sampling varies, the parameters measured, the sampling methods and the location of the STORET stations also vary. Missing values also represented a significant problem.

Suspended particulates and sulfur dioxide were the only two air quality parameters included in our data base; for other descriptors of air quality (oxidants, carbon monoxide, nitrogen dioxide, for example) information was available for a limited number of SMSAs only. The SEAS data file was the source of the air quality information.

The PDI index is a subjective measure of the prevalence, duration, and intensity of water pollution, calculated by the Office of Water Programs in EPA. Drinking water quality data has been obtained from the Water Supply Division of EPA. The ten drinking water quality parameters include information on the chemical content of the water supply, its alkalinity, hardness, and acidity.

Information on the quantities of residual pollutants discharged into the environment has been obtained from the SEAS data base. This data base includes information on the quantities of residuals discharged into the air: particulates, sulfur oxides, etcetera into the water; BOD, suspended solids, etcetera, and on the generation of solid wastes. Data on residuals from the SEAS data bank is computed rather than directly measured. Industry coefficients are developed for approximately 400 pollution producing economic sectors and subsectors. These coefficients relate the generation of a specific pollutant by the particular industry to the output of that industry at the national level. The coefficient times the output of the industry in the given SMSA equals the total gross residual. A second coefficient estimating abatement by sector is applied to the gross residual at the SMSA level to obtain the net residual. The use of national coefficients for most sectors ignores regional differentials in the production of residuals generation process.

Industry output at the SMSA level is measured by total economic value of production. The 1975 data used here is actually forecast by the SEAS model rather than measured. The national forecast is shared out between SMSAs based on disaggregate forecasts prepared by the Bureau of Economic Analysis (OBERS), the Economic Information System (EIS) tapes and other appropriate sources.

3.2 Applicability to Environmental Research

The simplest method of testing the applicability of the general SMSA groups to environmental policy analysis was to look at the variation in environmental measures between groups of SMSAs. Three approaches to comparison have been followed: regression analysis, factor analysis, and t- and F-tests (comparison of means).

Regression analysis was used to test the hypothesis that there is a significant relationship between environmental variables and Stage I factors derived from nonenvironmental data. The statistics support this hypothesis. For example, dissolved oxygen (DO) was found to rise with sewer expenditures (S5F2)*, although it is negatively related to other sanitation expenditures (S5F4). Growing cities have, on the average, lower DO than older centralized cities (S2F1, S2F2, S4F1), and low income also is correlated with low stream DO. A portion of these effects may be related to the fact that northern cities naturally have higher DO because of lower temperatures. However, the general relationship is that sewers, higher incomes, and slow growth all improve DO.

Similarly, significant relationships were found between the other environmental quality variables, and the Stage I factors. For a comprehensive description of these results, see Chapter 6 of Volume II.

The second approach to testing the relationship between environmental characteristics and more basic urban attributes was to include Stage I factors and variables from SCS1 and SCS3 in the Stage II factor analysis. The second stage analysis was repeated with each of the eight variables indicating ambient environmental quality. A close relationship between environmental and general attributes would be expected to cause the added measures to join Stage I factors from other SCSS to form factors similar to those which resulted from the basic fifteen factor set. If environmental attributes were not closely aligned with other attributes, the added measures would cause the factors to be restructured to some extent—perhaps forming an entirely new Stage II factor.

*S5F2 indicates factor number 2 from SCS5.

When SCS1 indicator variables were added, the result was as expected--the indicator variables appended themselves directly to factors derived from the basic fifteen factors. In no case was an additional factor developed.

The final approach to testing the suitability of the groups to environmental analysis involved t- and F-tests, comparisons of means to test whether the groups were significantly different in terms of the environmental quality variables.

In the simplest case--testing whether Group A and Group B have significantly different values for a single variable--a t-test is used with the null hypothesis indicating equal means for the two groups. For each variable, then, each pair of groups was compared to generate t-statistics which indicate the magnitude of any difference in the means relative to the variance of the given variable. Significant differences between groups were found for all but one variable--the PDI index.

Groups can also be compared in terms of general environmental quality by performing F-tests between the groups using all eight indicator variables together. The null hypothesis is that no group has its own characteristics, that there is a high probability that the eight environmental quality variables do not show significant differences between the groups. If this was true for a pair of groups, they could be combined to form a single, unique group in terms of environmental quality.

In every case, however, F-statistics indicate with at least 60 percent probability that the groups were significantly different given the eight environmental indicators.

4.0 OTHER RESEARCH PURPOSES

As discussed in Chapter 1, the data base and the methodology developed during this research project may be useful for other research purposes. In particular, the data may be used directly, city groupings and representative cities may be selected for testing alternative research hypotheses, and the results of studies performed in particular localities may be extrapolated to other areas. During the project, this site selection capacity was tested for ERDA; so that potential sites for electric car demonstration projects were identified. The example described in the following section is followed by other potential applications described in Section 4.2.

4.1 Transportation Demonstration Project

In the context of this project, a specific city classification scheme was developed in response to a problem proposed by the Energy Research Development Agency. ERDA is concerned with the potential for energy conservation which may be achieved through alternative transportation policies, in urban areas, in particular, through the use of electric cars. The development of an appropriate urban classification scheme and the identification of SMSAs for performing case studies and/or siting demonstration projects are the objectives for this task.

Given the more limited scope of this classification, the factor analysis was performed in a single stage. A set of thirteen variables was chosen from the assembled data base to reflect attributes of an urban area which are important in urban transportation analysis (see Table 4-1).

The transportation analysis is particularly interesting because it has been performed for three city size strata (see page 6) as well as for the full sample of 262 SMSAs.

Table 4-1

Variables Selected for the Transportation Classification

Variable	Description
1. H1U	Percent of families in single-unit housing
2. TD	Percent of workers commuting as auto drivers
3. AUTO	Percent of households with one or more cars
4. CEME	Percent of SMSA employment in central city
5. PCS	Percent population change, 1960-1970
6. EPF	Percent of women 18 or older who are employed
7. YM	Median household income
8. PBL	Percent of population which is Black
9. PPH	Persons/housing unit
10. SQMU	Square miles/person, urbanized area
11. RAD	Count of radial highways
12. CIR	Count of circumferential highways
13. VMTP	Vehicle miles travelled/capita-day

Four separate single stage factor analyses were performed; one for the 262 SMSAs and one for each of the three city size strata. Although a total of twelve factors were generated, three factors describing auto use, income/racial characteristics, and highways dominate all the runs. In other words, city size appears to have a limited effect on the factors describing urban transportation characteristics.

Four separate clustering procedures were performed; one for all cities, and one for each of the three size strata. The four most important factors were used for clustering in each case; the factors selected varied somewhat between the city size groups.

The 65 large cities formed two major groups with Providence, RI, and Louisville, KY-IN, being their representatives. About a quarter of the large cities are outliers, indicating the wide divergence in characteristics shown by the large cities.

The medium size cities also formed two major groups, with Little Rock, AR, and Tacoma, WA, being their modal SMSAs. Of the 87 cities in this group, only 9 were outliers.

Within the 107 small cities, there are five major groups, and about 10 percent unclassified (outlier) cities. Modal cities were Sarasota, FL; Lincoln, NB; St. Joseph, MO; Parkersburg, WV; and Spartanburg, NC.

The results of this classification may be used for a variety of purposes. The modal cities suggest natural case study sites, or locations for demonstration projects; the results of these may then be generalized to other cities in their groups. Further considerations may lead to other choices, data availability represents a case in point. The results of these studies may also be generalized to a larger set of cities. In addition, factor scores may be used to compare cities along the urban transportation related dimensions defined by the factors.

Large city case studies, for example, should be located in Providence, RI, and Louisville, KY-IN, with Providence results being applicable to Akron, OH; Rochester, Pittsburgh, and so on, and the Louisville results being relevant for Atlanta, Baltimore, Omaha, and so on. Medium city case studies should ideally be located in Little Rock and in Tacoma. Should study results be available for "outlier" cities such as Nashville, Hartford, or San Antonio, the factor scores for these cities should be examined individually. They may indicate that the SMSA--San Antonio, for example--is like no other city, in this case the results cannot be generalized. For other outliers, similarities may be discovered at least along some of the axes, for example, Nashville resembles groups 5 and 6 (of the large city groupings)--but does not fall into them because of an extreme value on Factor 1. Thus, its results have at least limited relevance to Toledo, Norfolk, and other cities in these groups.

4.2 Potential Applications

The data base and the methodology developed during this research project may be applied for alternative research purposes in three main uses. These are: the direct use of the data base, case study/demonstration project site selection, and the extrapolation of results of existing case studies to other sites.

The direct use of the data base does not merit extensive discussion. Clearly, researchers requiring the information available for our data base should not duplicate our data collection efforts, particularly since some of the information has been obtained from unpublished secondary sources. A complete list of the variables included in the data base is included in Appendix A to this volume; a comprehensive description of the data sources, strengths and weaknesses, as well as a data listing of all data may be found in Volume III.

Case study or demonstration project sites may be selected for a variety of research purposes through the use of the data base and methodology developed in this project. The major constraint to developing an appropriate classification scheme is that the research hypotheses to be tested must be capable of being framed in terms of the variables included in the data base. Indeed, some subset of the data base may be adequate for that purpose. For example, if a program analyst was interested in studying the effects of an antipoverty program, environmental quality and urban form descriptors would be of peripheral interest for his research purposes. Alternatively, if air quality maintenance programs represented the focal point of the inquiry, then water quality descriptors, and some of the income variables may not be relevant. The first step in developing appropriate city groupings is the specification of the relevant data set to be used for factor analysis and clustering. Secondly, the universe of SMSAs may be limited to fit the requirements of the research project. The program being tested may apply to a limited geographical area such as the South, or may be relevant for cities in a certain size category only. It is possible to identify cities with certain attributes to be excluded/included from the analysis. Once the data set and the universe of SMSAs is delimited, factor analysis is performed. If the data set is composed of a limited number of variables, and if the research hypotheses are relatively simple and well-defined, a one stage factor analysis may be adequate. A more complicated research design may call for a two stage factor analysis. Clustering is then performed on the basis of the first and second stage factors obtained, and representative cities are selected for each of the city groups. The case studies or demonstration projects, should be sited at these representative cities to best ensure that their results can be generalized to the other SMSAs in the group.

In some cases, it may not be possible to perform a case study at an ideal location in the appropriate representative city because of costs, data limitations, or simply due to lack of cooperation. Alternative sites may then be chosen. Further, the site selection process may be random, based on less rational criteria than the ones described here. It is appropriate to ask whether the results of such studies may be extrapolated to other sites.

The data base and methodology developed here may be used for this purpose as well. The data base and the universe of SMSAs must be delimited first, and second, the factor analysis and clustering are performed as described above. The sites of the case studies/demonstration projects are identified relative to the city groups; with the results being capable of generalization to the other cities in the groups. If the sites do not fall into any of the groups, then the study results may not easily be generalized to other cities. A large number of case studies have been performed in outlier cities, such as

New York, not because of a random site selection process but because the problem areas are frequently outliers. It is possible to analyze the data for such SMSAs, and to determine the axes or variables with extreme observation, which are in fact responsible for the outlying position of the SMSA. If these variables are not crucial to the analysis, the SMSA may be grouped with other cities in terms of the remaining variables. Study results may be generalized to this group--although at a lower level of confidence.

APPENDIX A

LISTING OF DATA VARIABLES

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
1	BOD	*Biochemical Oxygen Demand (5-day, 20° C)
1	FCOL	*Fecal Coliforms, measured by membrane filter method
1	N	*Total Nitrogen (mg/l)
1	C	*Total Organic Carbon (mg/l)
1	TDSS	*Dissolved Solids (mg/l)
1	TSS	*Suspended Solids (mg/l)
1	TURB	*Turbidity (Jackson Candle Units)
1	AALK	*Alkalinity (mg/l as $CaCO_3$)
1	PHS	*Acidity (standard units)
1	OAG	*Oil and Grease, soxhlet extraction (mg/l)
1	MBAS	*Methylene Blue Active Substance (mg/l)
1	SUS	Suspended Particulates (micro-g/cu.m.)
1	SO2	Sulfur Dioxide (micro-g/cu.m.)
1	PDI	PDI Index
1	CL	+Chloride (mg/l)
1	FL	+Fluoride (mg/l)
1	FE	+Iron (mg/l)
1	MG	+Manganese (mg/l)
1	NO3	+Nitrate (mg/l)
1	SO3	+Sulfate (mg/l)
1	ALK	+Alkalinity (mg/l as $CaCO_3$)
1	HD	+Hardness (mg/l as $CaCO_3$)
1	PH	+Acidity (pH standard units)

*Ambient Water Quality   +Drinking Water Quality

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
1	TDS	+Total Dissolved Solids (mg/l)
2	SQMS	Square Miles Per Person
2	SQMC	Square Miles Per Person - Central City
2	SQMU	Square Miles Per Person - Urban Places
2	CENP	Portion of Total Population in Central City
2	CENM	Portion of Manufacturing in Central City (by value added) (percent)
2	CENS	Portion of Retail Sales in Central City (percent)
2	CENE	Portion of Employment in Central City (percent)
2	CENME	Portion of Manufacturing Employment in the Central City (percent)
2	ARC	Arc of SMSA around the Center (quadrants)
2	RAD	Number of Major Radial Roadways
2	CIR	Number of Major Circumferential Roadways
2	PCR	Number of Population Centers
2	LU	Portion of Land Which is Urbanized (percent)
2	REC	Portion of Land Devoted to Outdoor Recreation (percent of total land area)
2	LA	Land in Farms (percent)
2	TCOM	Portion of Workers Working Outside SMSA (percent)
2	LAT	Latitude of SMSA (degrees)
2	LONG	Longitude of SMSA (degrees)
2	ALT	Altitude (feet)
2	PREC	Mean Annual Precipitation (inches)

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
2	SUN	Mean Annual Possible Sunshine (percent)
2	WIND	Mean Annual Wind Velocity (miles per hour)
2	HWPC	Total Miles of Roadway per Capita
2	HWPR	Portion of Principal Arterials (percent)
2	INV	Inversions (mean annual frequency)
2	WTMP	Water Temperature (ambient) (^{o}C)
2	DO	Dissolved Oxygen in Water (ambient) (mg/l)
2	HARD	Hardness of Water (ambient) (mg/l as $CaCO_3$)
2	WTR	Large Water Bodies (number)
3	RPAR	Particulates (tons per year per capita)
3	RSO	Sulfur Oxides (tons per year per capita)
3	RNOX	Nitrogen Oxides (tons per year per capita)
3	RHC	Hydrocarbons (tons per year per capita)
3	RCO	Carbon Monoxide (tons per year per capita)
3	RBOD	Biochemical Oxygen Demand (tons per year per capita)
3	RSS	Suspended Solids (tons per year per capita)
3	RDS	Dissolved Solids (tons per year per capita)
3	RNUT	Nutrients (tons per year per capita)
3	RWW	Wastewater (million gallons per year per capita)
3	RNSW	Noncombustible Solid Waste (tons per year per capita)
3	RIS	Industrial Sludges (tons per year per capita)

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
4	AGE	Median Age (years)
4	PF	Portion of Population which is of Foreign Stock (percent)
4	FR	Fertility Rate (children ever born per thousand women ever married)
4	NOH	Married Couples Without Own Household (percent)
4	PU	Portion of Population Which is Urban (percent)
4	NMOV	Individuals Residing in the Same Dwelling for Five Years (percent)
4	PPH	Average Household Size (persons per household)
4	PBL	Portion of Population Which is Black (percent)
4	MIGP	Movers into the SMSA (percent of individuals over five years of age)
4	FTM	Female/Male in 20-64 Age Group
4	PCC	Population Change -- Central City (percent)
4	PCS	Population Change -- SMSA (percent)
4	YMC	Change in Median Income (percent)
4	H1U	Single Family Dwellings (percent)
4	HOCC	Owner-Occupied Housing (percent)
4	H5U	Units in Structure -- More than Five Units (percent)
4	H50U	Units in Structure -- More than Fifty Units (percent
4	GINI	Income -- Gini Coefficient
4	YM	Median Family Income (dollars)
4	YP	Per Capita Income (dollars)

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
5	GX	Local Government -- General Expenditures (dollars per capita)
5	GREV	Local Government -- General Revenues (dollars per capita)
5	GEMP	Local Government -- Employment (full time equivalent per capita)
6	EM	Total Employment in Manufacturing (percent)
6	VAM	Value Added by All Manufacturing (dollars per capita)
6	IFLT	Total Value of Production in Meat Animals and other Livestock (dollars per capita)
6	IOIL	Total Value of Production in Crude Petroleum, Natural Gas (dollars per capita)
6	IMET	Total Value of Production in Meat Products (dollars per capita)
6	ICTH	Total Value of Production in Broad and Narrow Fabrics (dollars per capita)
6	I45	Total Value of Production in Household Furniture (dollars per capita)
6	IPLP	Total Value of Production in Pulp Mills (dollars per capita)
6	IPPR	Total Value of Production in Paper and Paperboard Mills
6	ICHM	Total Value of Production in Industrial Chemicals (dollars per capita)
6	IPRT	Total Value of Production in Commercial Printing (dollars per capita)
6	IFRT	Total Value of Production in Fertilizers (dollars per capita)
6	IMCM	Total Value of Production in Miscellaneous Chemical Products (dollars per capita)

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
6	IPLA	Total Value of Production in Plastic Materials and Resin (dollars per capita)
6	IPNT	Total Value of Production in Paints (dollars per capita)
6	IFUL	Total Value of Production in Petroleum Refining (dollars per capita)
6	IASP	Total Value of Production in Paving and Asphalt (dollars per capita)
6	IGLS	Total Value of Production in Glass (dollars per capita)
6	ICLY	Total Value of Production in Structural Clay Products (dollars per capita)
6	ICMT	Total Value of Production in Cement, Concrete, Gypsum (dollars per capita)
6	ISTL	Total Value of Production in Steel (dollars per capita)
6	IALM	Total Value of Production in Aluminum (dollars per capita)
6	IAPL	Total Value of Production in Household Applicances (dollars per capita)
6	ICAR	Total Value of Production in Motor Vehicles (dollars per capita)
6	IELC	Total Value of Production in Electric Utilities (dollars per capita)
6	ICOL	Total Value of Production in Coal Mining (dollars per capita)
6	IVEG	Total Value of Production in Canned and Frozen Foods (dollars per capita)
6	IFBR	Total Value of Production in Cellulous Fibers (dollars per capita)

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
6	ITAN	Total Value of Production in Leather and Industrial Leather Products (dollars per capita)
6	IABS	Total Value of Production in Other Stone and Clay Products (dollars per capita)
6	ICU	Total Value of Production in Copper (dollars per capita)
6	IPB	Total Value of Production in Lead (dollars per capita)
6	IZN	Total Value of Production in Zinc (dollars per capita)
6	IMTL	Total Value of Production in Other Fabricated Metal Products (dollars per capita)
6	IWST	Total Value of Production in Wholesale Trade (dollars per capita)
6	RETT	Total Retail Sales ($000 per capita)
6	WHOL	Total Wholesale Sales ($000 per capita)
6	S20	Total Value Added in Food and Kindred Products (SIC 20) ($ millions per capita)
6	S22	Total Value Added in Textile Mill Products (SIC 22) ($ millions per capita)
6	S23	Total Value Added in Apparel and Other Textile Mill Products (SIC 23) ($ millions per capita)
6	SS	Selected Services ($000 per capita)
6	S24	Total Value Added in Lumber and Wood Products (SIC 24) ($ millions per capita)
6	S25	Total Value Added in Furniture and Fixtures (S 25) ($ millions per capita)
6	S26	Total Value Added in Paper and Allied Products (SIC 26) ($ millions per capita)

ASSIGNED TO SCS	CODE	STATISTICAL UNITS
6	S27	Total Value Added in Printing and Publishing (SIC 27) ($ millions per capita)
6	S28	Total Value Added in Chemicals and Allied Products (SIC 28) ($ millions per capita)
6	S29	Total Value Added in Petroleum and Coal Products (SIC 29) ($ millions per capita)
6	S30	Total Value Added in Rubber and Plastic Products (SIC 30) ($ millions per capita)
6	S31	Total Value Added in Leather and Leather Products (SIC 31) ($ millions per capita)
6	S32	Total Value Added in Stone, Clay and Glass Products (SIC 32) ($ millions per capita)
6	S33	Total Value Added in Primary Metal Industries (SIC 33) ($ millions per capita)
6	S34	Total Value Added in Fabricated Metal Products (SIC 34) ($ millions per capita)
6	S35	Total Value Added in Machinery, Except Electrical (SIC 35) ($ millions per capita)
6	S36	Total Value Added in Electrical Equipment and Supplies (SIC 36) ($ millions per capita)
6	S37	Total Value Added in Transportation Equipment (SIC 37) ($ millions per capita)
6	S38	Total Value Added in Instruments and Related Products (SIC 38) ($ millions per capita)
6	S39	Total Value Added in Miscellaneous Manufacturing Industries (SIC 39) ($ millions per capita)

	CODE	STATISTICAL UNITS
control data*	P	Population (in thousands)
control data*	HTOT	Total Housing Units (in thousands)
control data*	TALL	Total Commuters (hundreds)
control data*	LS	Total Land Area -- SMSA (square miles)
control data*	LURB	Total Land Area -- Urbanized Portion (square miles)

*Control data were used for computing normalized variables.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/5-77-008a	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Classification of American Cities For Case Study Analysis-Volume I Summary Report		5. REPORT DATE May 1977 (Issuing date)
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Elizabeth Lake, Carol Blair, James Hudson, Richard Tabors		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Urban Systems Research & Engineering Inc. 1218 Massachusetts Avenue Cambridge, Massachusetts 02138		10. PROGRAM ELEMENT NO. 1HA091
		11. CONTRACT/GRANT NO. 68-01-3299
12. SPONSORING AGENCY NAME AND ADDRESS Office of Monitoring & Technical Support - Wash., DC Office of Research and Development U.S. Environmental Protection Agency Washington, D.C. 20460		13. TYPE OF REPORT AND PERIOD COVERED Final
		14. SPONSORING AGENCY CODE EPA/600/19

15. SUPPLEMENTARY NOTES

Volumes II - Detailed Report and III - Documentation of Study are available from the National Technical Information Service.

16. ABSTRACT

Attempts to analyze and evaluate the impacts of federal programs has led to the extensive use of case studies of program impacts at selected sites. This project has developed a methodology for the systematic selection of representative case study sites and for generalizing the study results. The methodology, involving two stage factor analysis and clustering, is applied to a specific program/policy problem, the selection of metropolitan areas for case studies in analyzing the impact of federal policies on general environmental quality.

The methodology begins with a data base on standard metropolitan statistical areas, SMSAs, including variables related to environmental quality, urban form, and household, industrial, and government activity. It analyzes these variables through a two-stage factor analysis technique which allows heuristic consideration of the significant characteristics. Finally, it develops city clusters which group areas with similar attributes. Modal (or representative) cities are selected for each group and suggested as case study sites. These groups may be used to generalize the study results and to analyze the trans-ferrability of results between areas. The methodology is sufficiently flexible to consider a wide range of research hypotheses.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS		b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Economic Factors	Socioeconomic Stans	Modal Cities	08F
Economic Surveys	Urban Areas	City Classification	05C
Economic Geography	Urban Sociology		05K
Census	Urban Geography		13B
Central City	Factor Analysis		
Demographic Surveys			
Populations			

18. DISTRIBUTION STATEMENT Unlimited	19. SECURITY CLASS *(This Report)* Unlimited	21. NO. OF PAGES 48
	20. SECURITY CLASS *(This page)* Unlimited	22. PRICE

EPA Form 2220-1 (9-73)

☆ U. S. GOVERNMENT PRINTING OFFICE: 1977-757-056/6426 Region No. 5-11