Work reported in this document was sponsored by the Department of Energy under contract No. EX-76-A-01-2295. This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product or process disclosed or represents that its use would not infringe privately owned rights.

METHODOLOGY AND DEFINITION OF SOLAR PHOTOVOLTAIC PLANNING REGIONS

Richard D. Tabors and Paul R. Carpenter

Energy Laboratory Report No. MIT-EL-78-034

July 1978

PREPARED FOR THE UNITED STATES

DEPARTMENT OF ENERGY

Under Contract No. EX-76-A-01-2295
Task Order 37

METHODOLOGY AND DEFINITION OF SOLAR PHOTOVOLTAIC PLANNING REGIONS

Richard D. Tabors and Paul R. Carpenter

Energy Laboratory Report No. MIT-EL-78-034

July 1978

PREPARED FOR THE UNITED STATES

DEPARTMENT OF ENERGY

Under Contract No. EX-76-A-01-2295
Task Order 37

ABSTRACT

There are in use at the present time at least 10 differently defined sets of energy planning regions for the United States. This paper identifies and describes the existing energy planning regions and discusses their intended function. It then presents an argument for development of a set of solar regions within the United States and discusses the criterion (climate, economics and energy supply/demand) which are required for definition of such a set of regions. The final two sections of the paper discuss a methodology two stage factor analysis, for the definition of solar planning regions and the application of that methodology to the definition of a set of seven planning regions for the United States. Those regions so defined are: the Northeast (south as far as Virginia); the south (west as far as Oklahoma and Kansas); the southwest (including California and Nevada); the northwest (as far east as the Dakotas); the midwest (stretching from Minnesota to Ohio) and individually the state of Texas.

METHODOLOGY AND DEFINITION OF SOLAR PHOTOVOLTAIC PLANNING REGIONS

Richard D. Tabors

Paul R. Carpenter

> The domestic energy problem and its solution are national, regional, and local in nature. The problem extends beyond traditional RD&D efforts and includes, as an essential ingredient, the market penetration of near-term technologies. Success or failure in meeting the Nation's energy needs will depend as much on the ability to resolve complex economic, social, political and ecological issues at the regional and local levels as on the technical quality of the specific energy RD&D programs. The Federal Government must therefore be sensitive to local and regional needs. It must also reach public and private groups at these levels to provide information to them; to develop effective productive communication links with regional, state, local, university financial and industrial representatives; and to receive feedback from them on the problems, progress, public acceptability and overall effectiveness of ERDA's programs and the National Plan for Energy RD&D. To assist in achieving ERDA's overall energy mission and in carrying out its specific assigned energy program responsibilities, an enhanced regional capability may be desirable.[1]

I. INTRODUCTION: EXISTING ENERGY REGIONS

Choice or definition of planning regions for Department of Energy (DOE) activities has been given only slight attention to date. While the regional office structure of DOE exists, these centers are historical, inherited from the Atomic Energy Commission, rather than planned to meet the requirements of the new department. The purpose of the paper which follows is to define a set of solar photovoltaic planning regions for the United States. To accomplish this task this paper will present briefly a discussion of different classifications of regions and a review of the more common regions developed and utilized for energy analysis and planning. The second section of the paper will present a methodology, two-stage factor analysis, for regional definition specific to the requirements of the Photovoltaic Program at DOE. The third section of the paper presents the results of the first stage of the analysis and the data employed. The fourth section of the paper presents a discussion of the definition of composite regions and the fifth section compares the regions defined in section four with those currently under use by others, specifically by individuals within DOE.[2]

In general terms regions are defined for one of three reasons (or for a combination of the three). These are:

- Grouping of similar areas for policy or program implementation
- Grouping of areas for research or testing purposes i.e. similar climate
- Grouping of areas for administration i.e. administration of a census.

There are three types of regions to which regional economists or geographers refer. The first is a nodal region in which the intensity of the activity under study decreases as you move away from the node. This is analogous to an urban region where population density decreases as one moves out from the central city. Regional boundaries occur in the "countryside" between the nodes. The second regional type is the homogeneous region showing relative uniformity within its boundaries and relative dissimilarity across regional boundaries. The third category consists of political or administrative regions, those built up by administrative convenience such as states or groups of states (New England). The analysis which follows uses states as data units for the creation of a set of homogeneous solar planning regions.

There are at the present time more than nine regional groupings used for energy planning within the United States. These range from rather general multipurpose energy regions, such as those developed by the Bureau of Mines, to highly specific fuel planning regions such as the Petroleum Administration for Defense (PAD) regions. The section which follows presents in summary form the more significant of the energy planning regions currently in use.

Bureau of the Census Regions: The most frequently adopted U.S. regional structure is the Census regions shown in Figure 1. While these regions satisfy many administrative criteria, they are not well suited to energy planning. Because their purpose was the enumeration of the U.S. population, they represent -- or did -- relatively even census groupings but have little bearing upon economic structure or upon resource availability. The Census regions were those adopted by the old FEA for its planning, implementation and administrative structure. These are also the demand regions used in the FEA/PIES analysis.

Regional Electric Reliability Councils: Developed and used by the Edison Electric Institute, the regional Reliability Councils of the National Electric Reliability Council are defined to capture the interconnectivity of specific electric utilities. They are used for planning and reporting of customer service reliability. As shown in Figure 2, the boundaries do not follow state or county lines; they represent electricity utility service areas, are power sheds and electric power planning regions but are not of assistance in more general energy planning.

Edison Electric Institute Regions: The Edison Electric Institute Regions (Figure 3) are data-reporting regions which conform where possible with the Census regions. These are the units on which Edison Electric reports its weekly electric output. These represent a

FIGURE 1

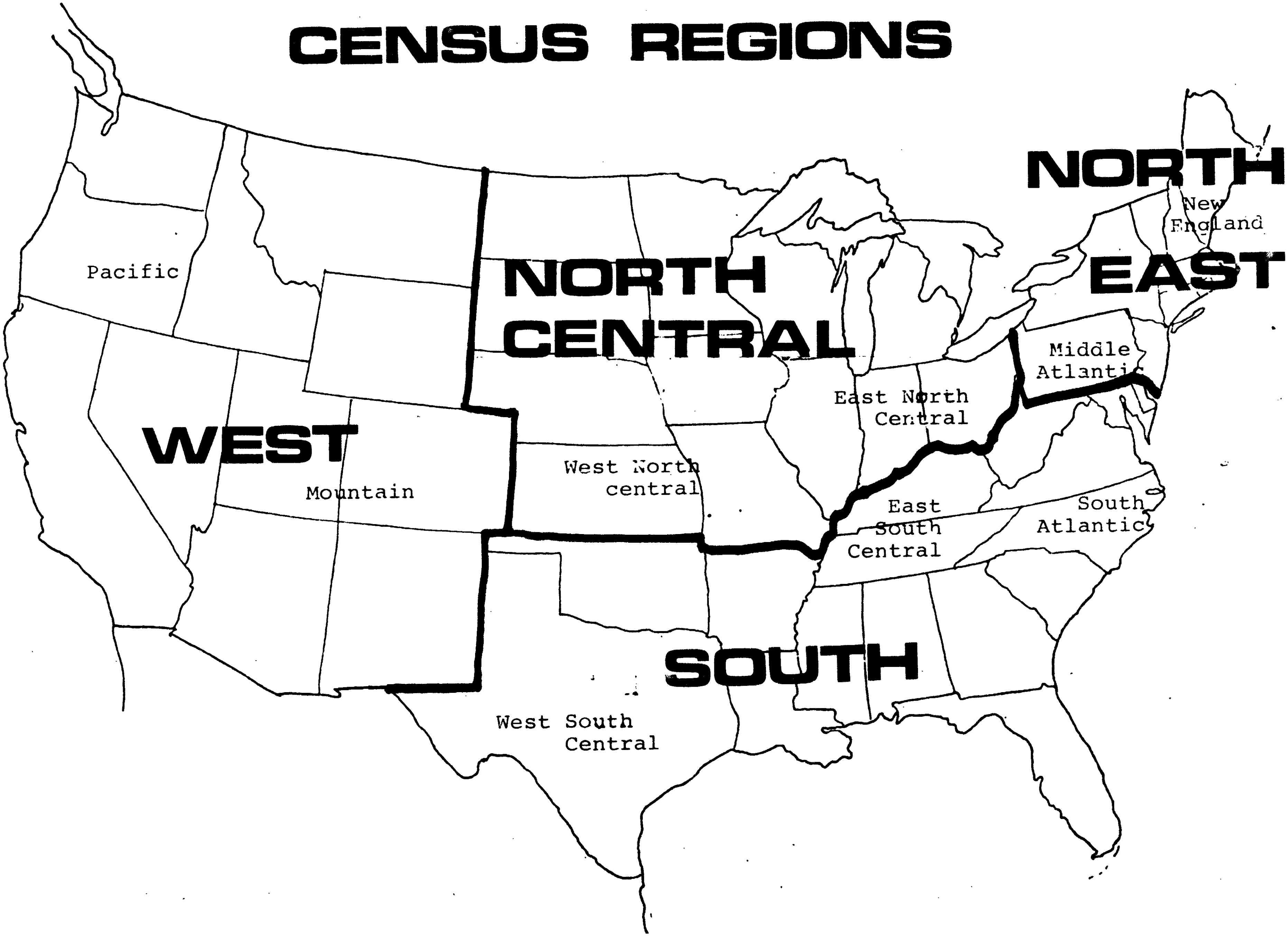
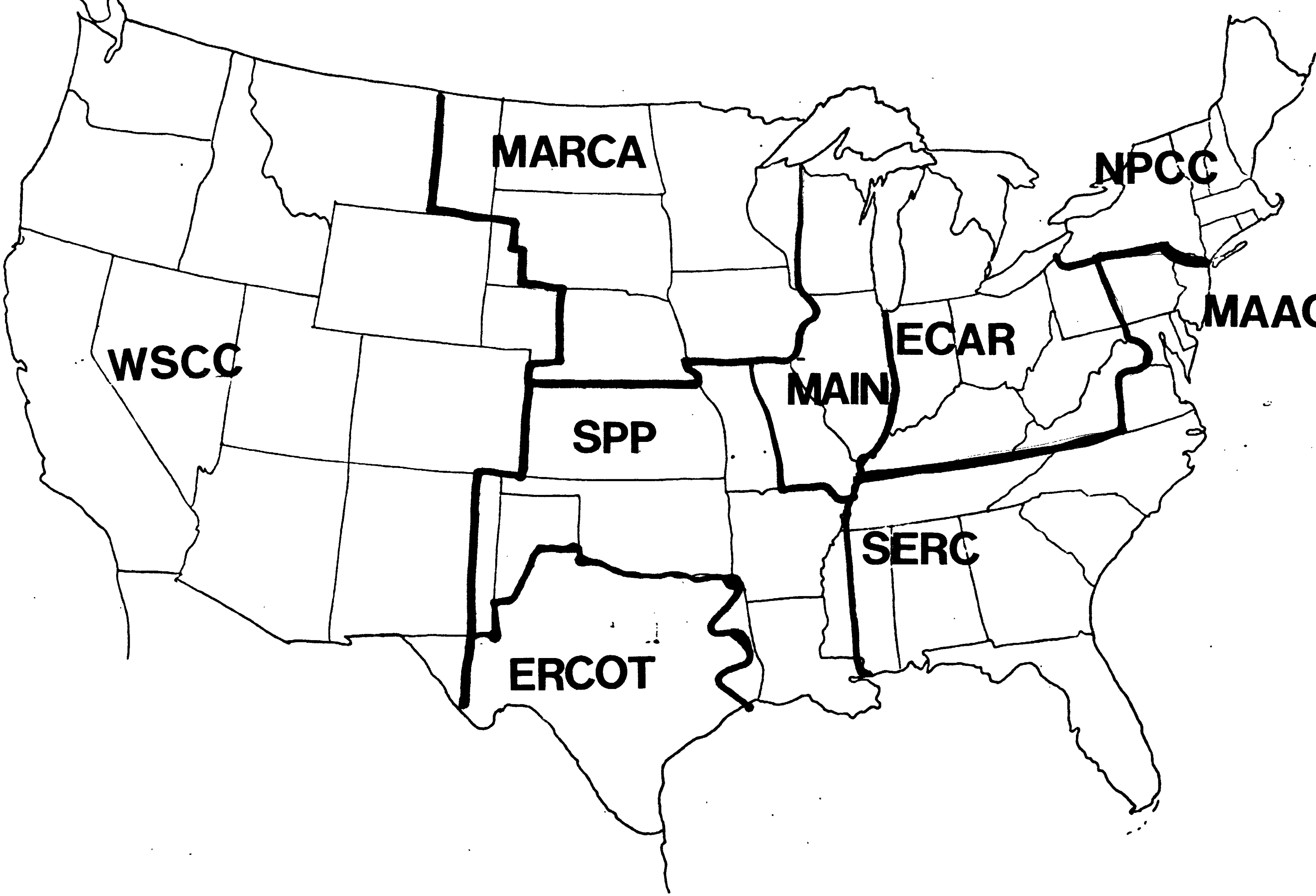

CENSUS REGIONS
NORTH EAST
New England
Middle Atlantic
NORTH CENTRAL
East North Central
West North central
WEST
Pacific
Mountain
SOUTH
South Atlantic
East South Central
West South Central

FIGURE 2

REGIONAL RELIABILITY COUNCILS OF THE NATIONAL ELECTRIC RELIABILITY COUNCIL



"compromise" between the information units of Figures 1 and 2.

<u>Petroleum Administration for Defense</u> (PAD): The PAD regions were developed to reflect oil supply and refining regions within the United States. As is shown in Figure 4 these regions are large and tend to be grouped by coastal location -- east, west or gulf and two central categories.

The following five regional groupings were used in the Project Independence Evaluation System (PIES) developed by FEA for planning and analysis of energy policy. In general they are not geographically consistent but rather capture the regional characteristics of individual energy supply types.

<u>PIES Coal</u>: The PIES coal regions reflect the major coal producing regions within the United States. It is most important to note in Figure 5 that not all of the United States is covered and that the regions do not correspond to state or other political boundaries.

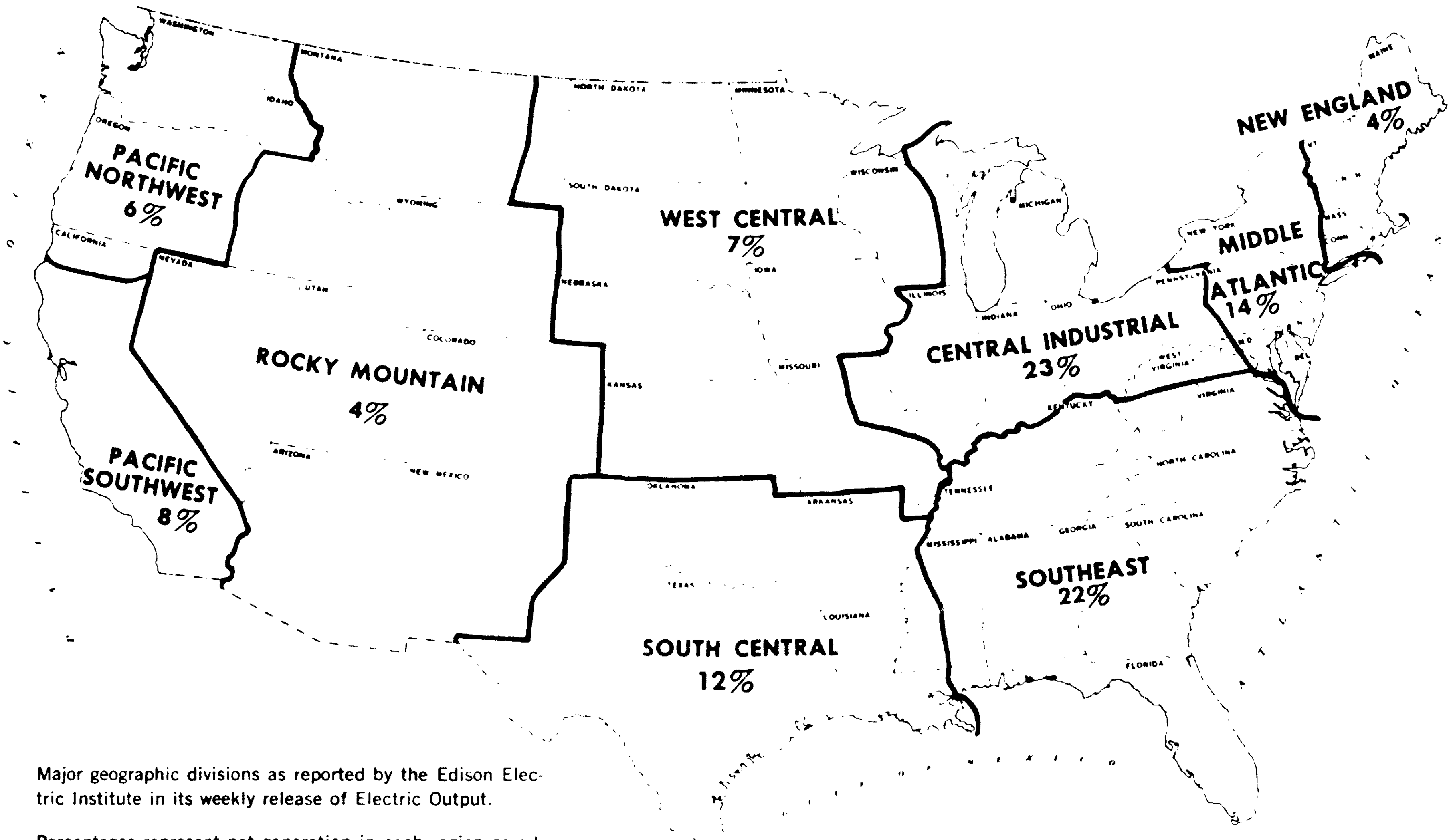
<u>PIES Natural Gas</u>: The PIES natural gas regions, Figure 6, correspond to production areas and again do not cover the United States but only the gas producing regions.

<u>PIES Refinery Regions</u>: The PIES refinery model (Figure 7) utilized the PAD regional definition and split PAD 1 into two regions and PAD 2 into two regions.

<u>PIES Electric Utility Regions</u>: The PIES electric utility regions (Figure 8) follow the nine Census region boundaries.

<u>PIES Demand Regions</u>: The PIES demand regions are the same set used by FEA for administrative purposes and for PIES utilities (Figure 8); they are also the same state boundaries used by the Bureau of the Census (Figure 1).

FIGURE 3



Major geographic divisions as reported by the Edison Electric Institute in its weekly release of Electric Output.

Percentages represent net generation in each region as adjusted for net imports.

FIGURE 4

PETROLEUM ADMINISTRATION FOR DEFENSE (PAD) DISTRICTS

FIGURE 5

Coal Supply Regions

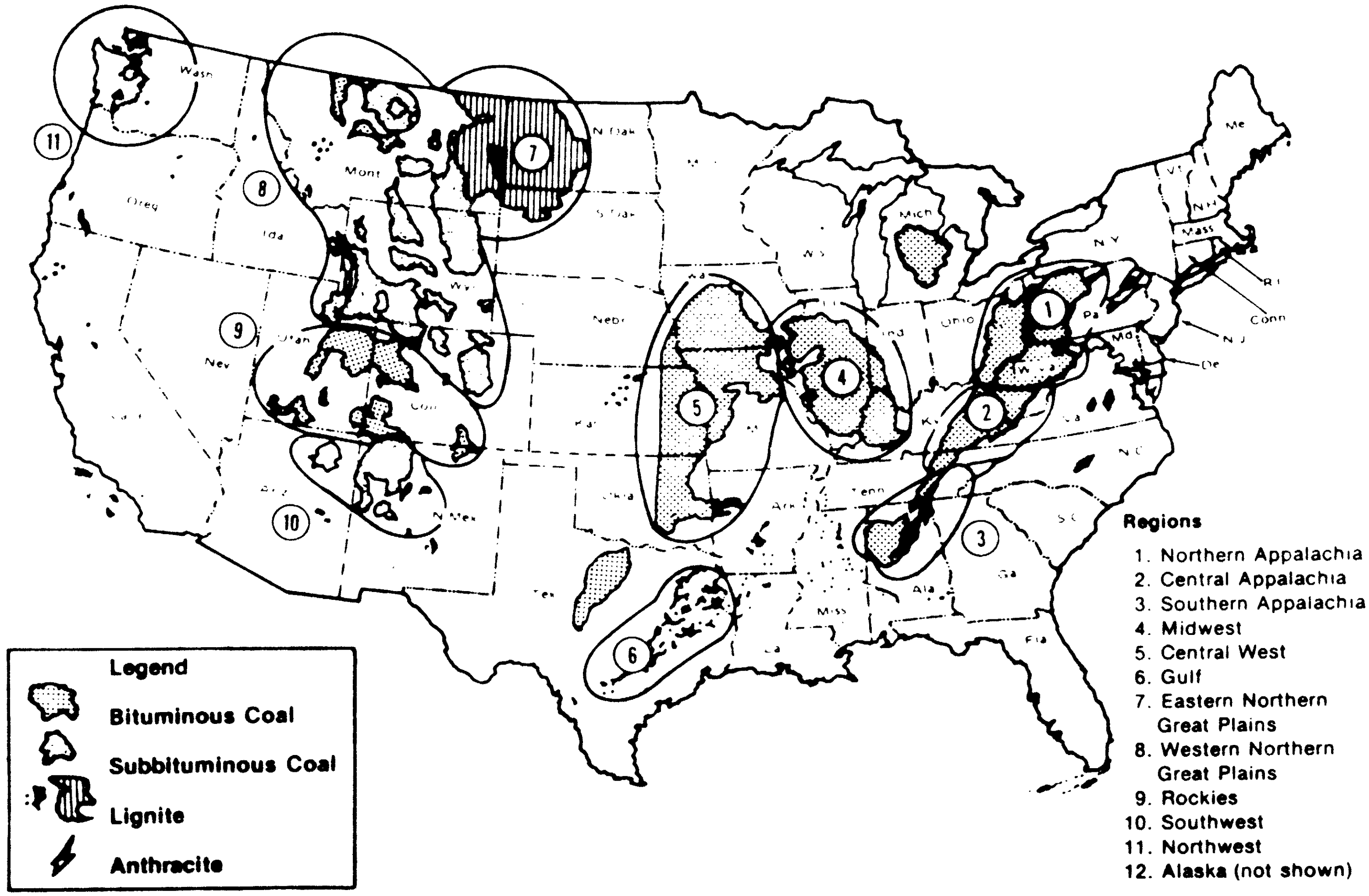


FIGURE 6

Major Natural Gas Producing Regions And Pipelines With Significant Curtailments

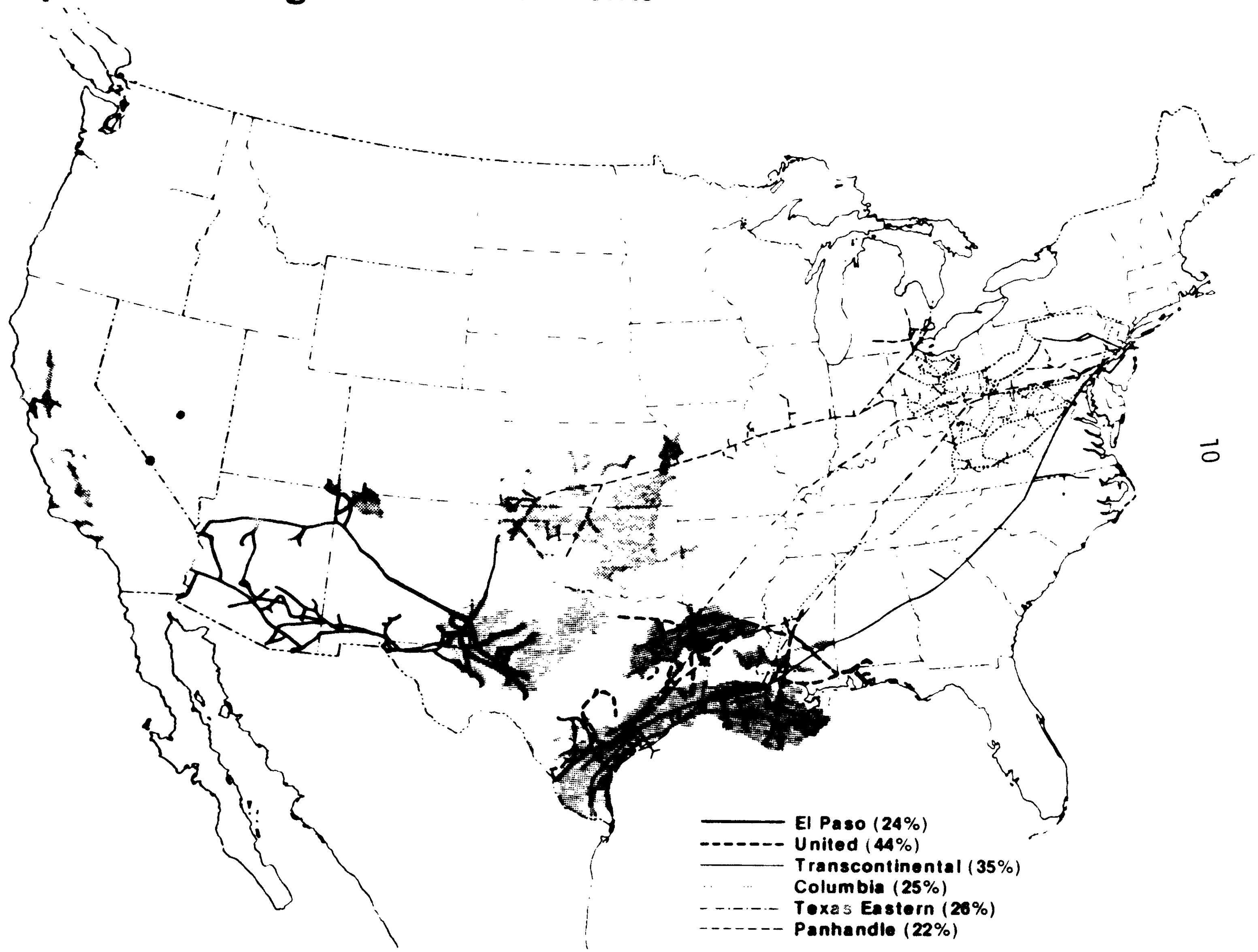


FIGURE 7

Refinery Regions

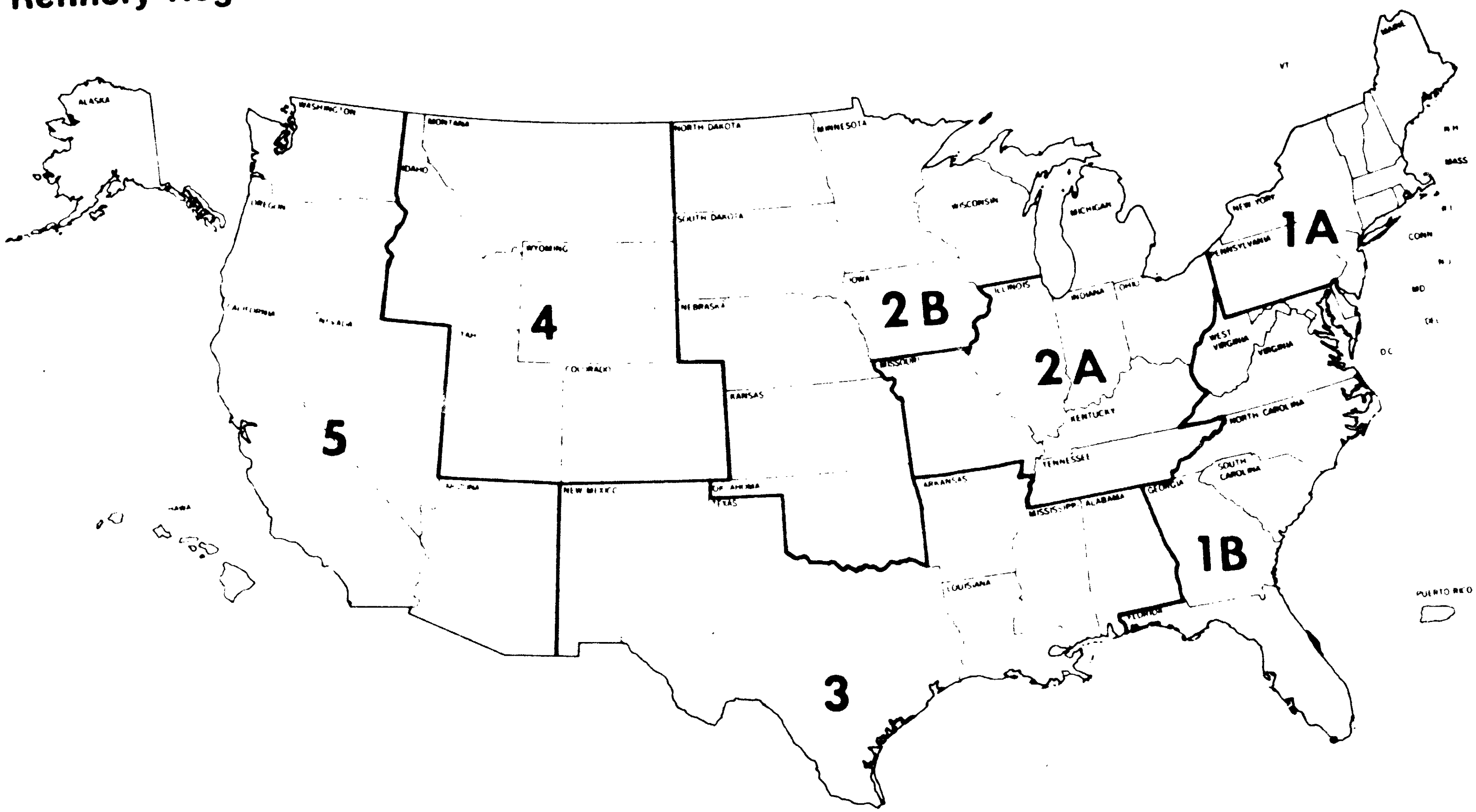
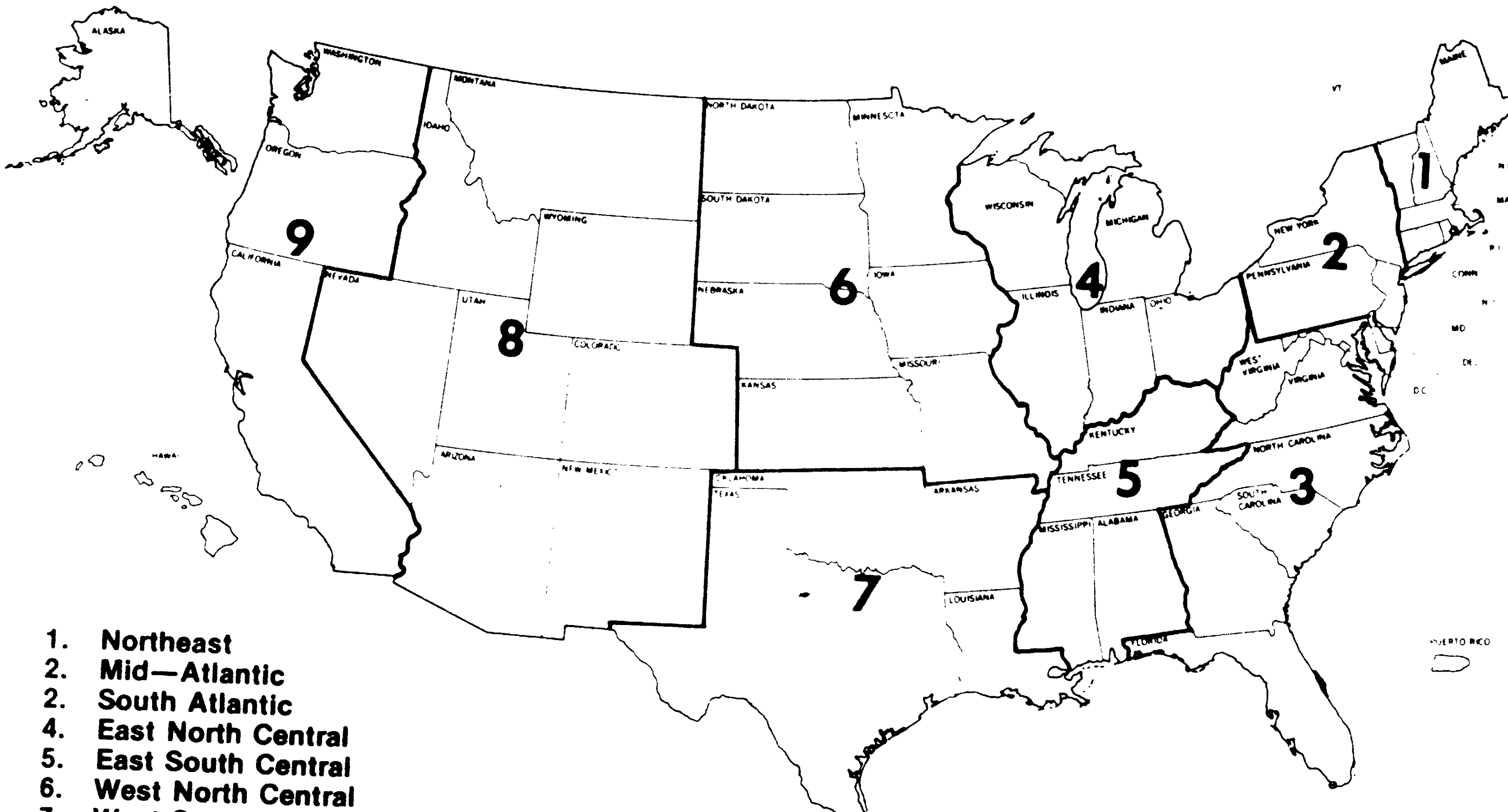


FIGURE 8

Electric Utility and Demand Regions



1. Northeast
2. Mid—Atlantic
2. South Atlantic
4. East North Central
5. East South Central
6. West North Central
7. West South Central
8. Mountain
9. Pacific

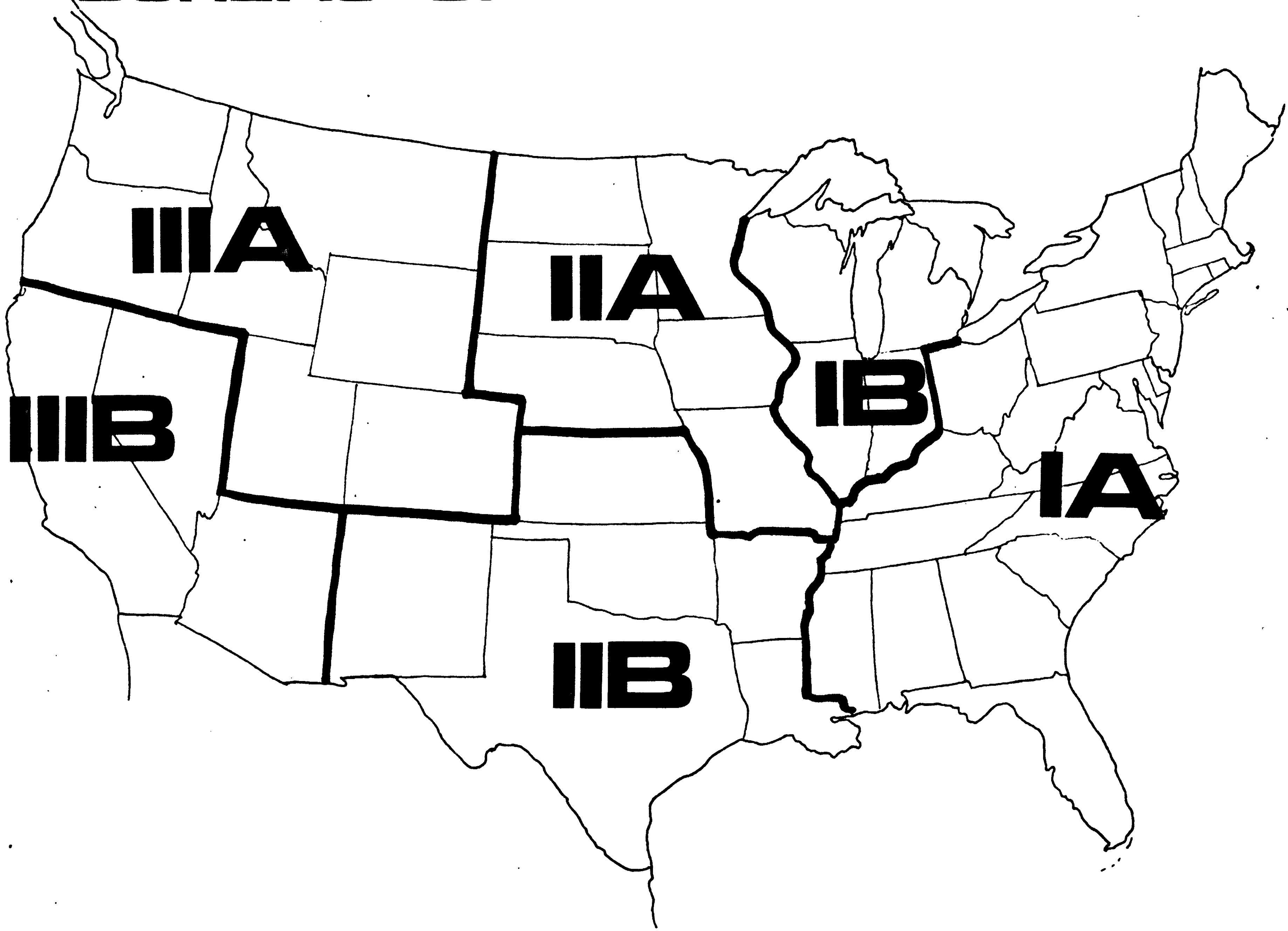
Bureau of Mines Regions: Figure 9 shows the regional groupings developed by the Bureau of Mines for the reporting of general regional energy production and consumption. This set of regions comes the closest to being multifunctional but still does not take into consideration any of the solar and climatic variables required for a set of regions to be used in an analysis of solar and solar-electric applications.

In summary, the regions utilized for energy planning in the United States are generally focused on energy supplies; where there is a concern for energy demand, the regions chosen are most commonly those defined by the Bureau of the Census. Several generalizations may be made about the energy regions discussed above. The first is that either five or nine regions cover the continental United States. The second is that there is significant geographic overlap in all of the definitions. New England or the Northeast appears in all of the regional groupings. In addition, one region covers the Pacific coast and another covers the southeastern portion of the United States. The central regions most frequently encompass Indiana and Tennessee with another region further west, centering on Nebraska. The state of Texas frequently appears alone or as the major block in a west south central regional area.

The material which follows will present both a rationale for development of solar planning regions for the United States and present the regions so defined. Additional references on regional energy planning may be found in the general reference section at the end of this paper.

FIGURE 9

BUREAU OF MINES REGIONS



II. METHODOLOGY

The two stage factor analysis methodology employed for definition of solar planning regions for the United States was originally developed by Tabors[1] and extended in Lake et al.[2] It followed the work of Berry and others.[3] Two stage factor analysis employs mathematical analysis and grouping procedures to identify groups of, in this instance, states whose characteristics across a predefined variable set are nearly homogeneous within groups and relatively heterogeneous between groups.

During the past decade and a half, considerable emphasis has been placed upon the use of principal components and factor analysis as tools in grouping and in the identification of multifunctional, uniform regions. These appear to be effective statistical techniques when used correctly, and as such, can be a rational, systematic approach to regional delimitation for energy planning.[4]

Factor analysis was developed for use in the field of psychology for the characterization of underlying psychological traits and a great deal of the literature with regard to the characteristics of this mathematical technique may be found in the psychometrics literature. The earliest reporting of the technique is that of Charles Spearman in 1904. The 1930s were, however, the period of greatest development.[5] Factor analysis was first used to define regions by Margaret Hagood in 1943 to regionalize agriculture and population information for the United States.[6] More work in the field appeared in Hagood's Statistics for Sociologists in 1952.[7]

In the last decade Brian J.L. Berry has been the primary developer of factor analysis as a regionalization tool. The Atlas of Economic Development edited by Norton Ginsburg contains a section by Berry on the use of multiple regression methods in regionalization.[8] In the Atlas, the authors present the patterns of economic development as expressed in terms of a series of demographic and economic characteristics. The chapter by Berry offers one of the earliest and most complete discussions of the value of factor analysis in regionalization for the identification, summation, and condensation of a large series of detailed variables into a relatively small group of components or factors. In this exercise, Berry ran a factor analysis on 43 indices for 95 countries. Isolated were four basic dimensions characterized by Berry as (1) technological (2) demographic, (3) income and external relations, and (4) large or small countries. Berry maintained at that time that the "basic similarities were such that 4 maps would, for certain purposes, do about the same job as 43!"[9]

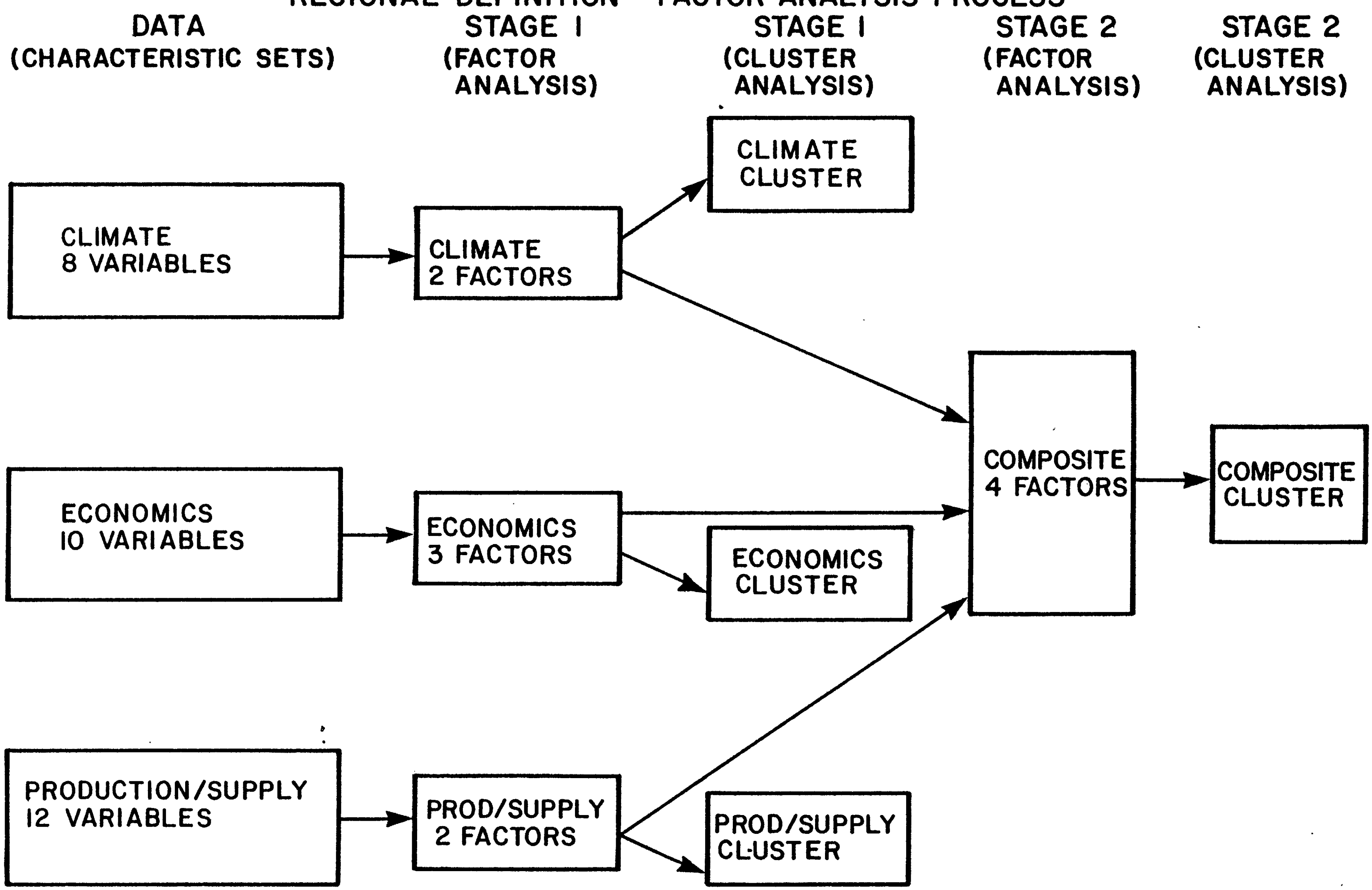
A long and detailed series of articles concerning the process of regionalization through the use of factor analysis have followed during the past ten years. At first, factor analysis was considered a panacea for the process of regionalization. Later many research scholars looked upon it as being unapplicable for use in the process of grouping and regionalization.[10] It would appear that neither of these reactions to the mathematical technique are correct. It is obvious that if one has no considered hypothesis or little theoretical basis for the regionalization being performed then, it is highly unlikely that one will meet with success in the result. The factor analysis technique is only as good as the information and thought used with it. By the same token, if the logic is sound, the results may be interpreted in the light of knowledge about, first, how factor analysis operates, and second, the expected results based on the hypothesis. Knowledge of both of these would appear necessary for the use of any analytical technique, including factor analysis.

Figure 10 shows in block form the analytic steps undertaken in this study to define solar energy planning regions for the United States. For each of the continental forty-eight states the analysis developed three sets of characteristic data; Climate, Economic and Energy Production/Supply.[11] Each of these three data sets was subjected to a factor analysis to identify those characteristics which best defined the variables chosen. A grouping or clustering algorithm was then used to group the forty-eight states into regions with a general criterion of minimizing the variance within groups while maintaining variance (and therefore distinction) between groups. As will be discussed in later sections there is no hard and fast rule which can be applied to such analyses. The choice of a precise cutoff point is dependent upon the purpose for which the groupings are to be used and the general number of regions required. The analysis of the first stage regions accomplishes two purposes. The first is that it identifies the underlying factors in each characteristic set. The factor scores developed become the input to the second stage analysis. The second purpose of the first stage analysis is that it allows for verification of regional patterns on the part of the analyst. Much of the early criticism of factor analysis was based on the lack of contact between the analyst and his results. The two stage analysis guarantees a level of 'accountability' between the hypothesized input variables and the regions defined in the first stage analysis[12].

The second stage of the analysis performs a factor analysis on the first stage factor scores, the first stage factors having been selected so as to meet the general criteria for selection of number of factors:

FIGURE 10

REGIONAL DEFINITION - FACTOR ANALYSIS PROCESS



interpretability and addition to cumulative variance explained in the original data set. A more complete discussion of the factors chosen from stage one will follow in a detailed discussion of the results of this analysis.

The short section which follows describes briefly the process and vocabulary of factor analysis, introduces briefly the mathematics and lists references for more detailed information. The reader familiar with the technique should proceed to section III which follows.

The purpose of factor analysis is to collapse a large set of data into a smaller and more simple set of orthogonal factors which summarize the correlations of the larger set of original variables.[1] Mathematically, the technique is quite similar to principal components analysis, with the exception that, in factor analysis, an error term is added to the basic form of the equation which allows for the inclusion of random variation. This variation may be stochastic, or may simply occur due to variation not measured in the sample. The factor analysis model is "concerned with finding a matrix (V) such that $(V) \times (V)^t = (R)$. The matrix (V) contains the factor loadings or correlations between variables and factors....The factor analysis model is

$$Z_j = \sum_{r=1}^{m} v_{jr} f_r + e_j \qquad \text{for } j = 1,2\ldots,\ p{:}m \leq p$$

[1] ~~For a complete~~ discussion of factor analysis see Harmon, Factor Analysis, or Lawley and Maxwell, Factor Analysis as a Statistical Method, or Kendall, A Course in Multivariate Analysis. For an excellent summary of the technique and its use in geography see Leslie J. King, Statistical Analysis in Georgraphy, chap. vii, pp. 165-193.

where:

Z_j = Standardized original variables

V_{jr} = Variances

f_r = The common factor

e_j = The error term

and the correlation matrix (R) is only partially reproduced by the product of $(V) \times (V)^t$."[1] The diagonal of the correlation matrix, unlike that for principal components, is not necessarily made up of ones. This principal diagonal contains the communalities of the factor analysis program, and hence, the most significant portion of the model.

"Under the assumed composition of variables, the communalities are the basic quantities to be analyzed. Herein lies the trouble, there is no *a priori* knowledge of the values of the communalities."[2] There are several channels of estimation of communalities open to the investigator. The simplest of these is to assume that all are equal to unity. As stated by Harmon, "as a saving grace, there is much evidence in the literature that for all but very small sets of variables the resulting factorial solutions are little effected by the use of 'communalities' or unities in the principal diagonal of the correlation matrix."[3]

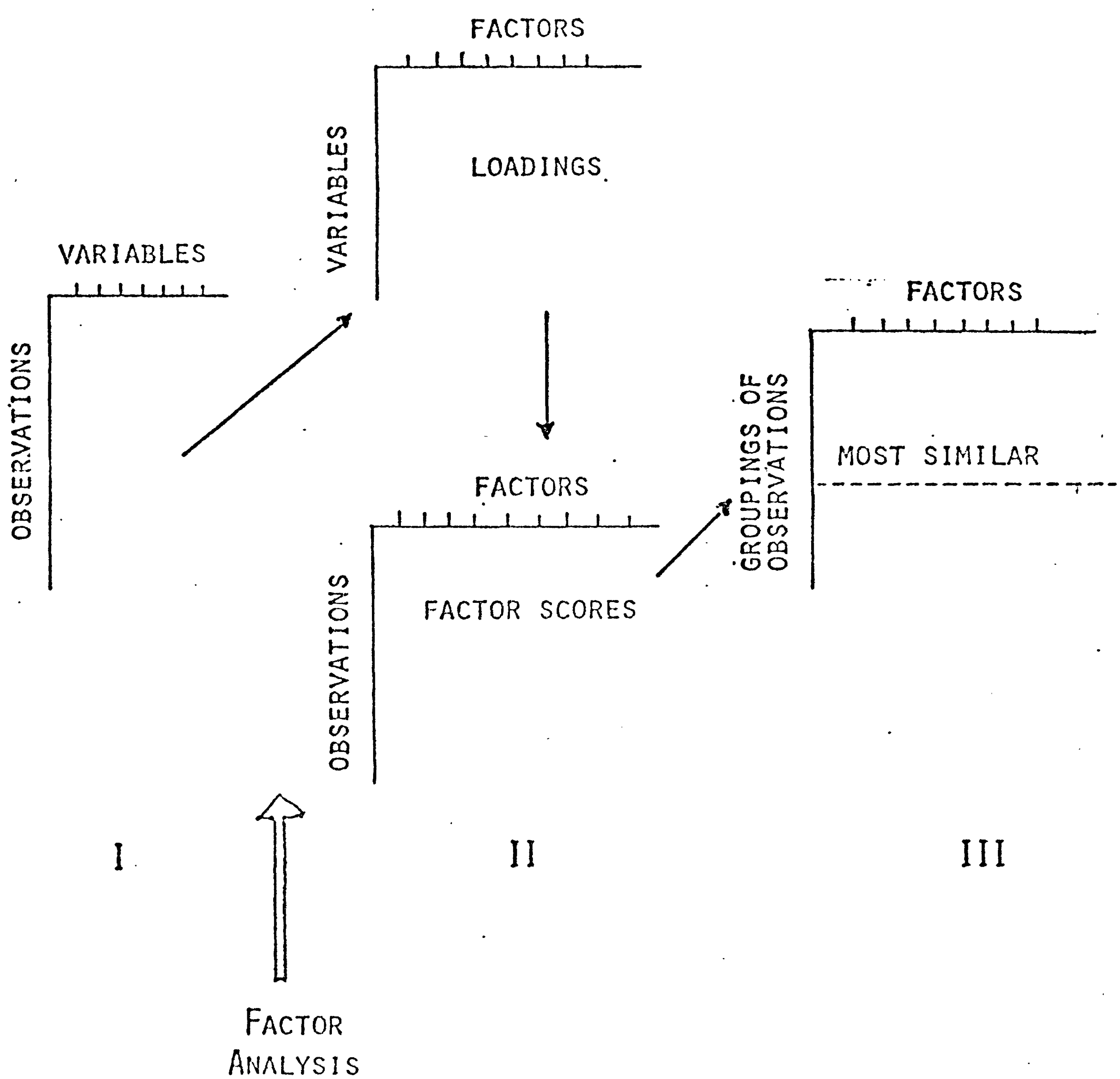
As shown in figure 11, the resultant matrix of factor loadings relates the original variables to the set of factors. Factor loadings represent correlations between the original set of variables and the factor defined in the analysis. As such, a high (close to 1.0) positive or negative

---

[1] King, *Statistical Analysis in Georgraphy,* p. 184.
[2] Harmon, *Factor Analysis,* p. 69.
[3] *Ibid.*, p. 86.

Figure 11

FACTOR ANALYSIS AND GROUPING DEFINITIONS



loading represents high positive (or negative) correlation between the variable and the factor. Low loadings indicate poor correlation. The factors can now be described in terms of the set of variables with which they are most highly correlated and hence in the section which follows we discuss, under the climate characteristic set, a "sunshine" factor which collapses the solar variables and other correlated weather variables into a single factor. The second matrix generated within the factor analysis is one of factor scores in which each observation point is "scored" for each factor. Factor scores are surrogate variables which are, by definition, orthogonal (independent). In summary the factor analysis collapses a set of variables (which may be highly correlated) into a set of underlying factors which are fewer in number and are independent. The procedure produces, in addition to factor scores and factor loadings, an additional statistic, the eigenvalue or VP for each factor identified. The eigenvalue is the sum of the squared factor loadings. Since a factor loading is in fact a correlation coefficient, the squared factor loading is analagous to an R^2 and plus describes the "fit" of the variable to the factor. When divided by the number of variables, the resultant value is the percentage of the total variance in the original data set explained by the individual factor. Eigenvalues are indicators of the significance of the factors and as such can be used to judgementally select the number of factors appropriate to the data set. There is no hard and fast rule as to the appropriate number of factors to select. The final mathematical step taken in any factor analysis used for regional definition is the grouping (or clustering) of like observations. The algorithm used for this purpose was developed by Tweetie and Meyers and modified by Tabors.[13] All observations are arrayed in "N" dimensional space according to factor scores on individual factors (if 3 factors then each observation can be placed in 3 dimensional space, if 4 then 4 dimensional and so on).

At this point the system has maximum variance. The two most similar observations are then combined and their scores "averaged" to a new centroid. The next two most similar are combined; the next etc until there is only one group. At each step the variance within groups is calculated and compared with the variance between groups. While no perfect rule exists for setting the cut off point in the grouping procedure it is important to be certain that at the point chosen the variances between groups is still greater than that within the group. The stepwise grouping procedure may be made more clear by referring to the "Tree" diagrams included in the appendix to this report.

III. STAGE ONE ANALYSIS

The discussion which follows summarizes the data set development and first stage factor analysis undertaken for the United States.

CLIMATE DATA SET

Solar Insolation: Since precise solar insolation data was unavailable for reporting stations in all states, analysis was performed to determine which of the widely available sunshine surrogates best approximated direct solar insolation on a tilted plate. A correlation analysis was performed to compare direct insolation on a 45^{0} tilted plate[1], available for 26 reporting stations, to the commonly collected data; number of cloudy, partly cloudy, and clear days per season; cloud cover in tenths; and percent possible sunshine. Percent possible sunshine correlated with direct insolation with a coefficient of .9 and was selected for inclusion in the data set. This confirms similar results obtained by the Aerospace Corporation.[2] To obtain percent possible sunshine values by state, the values for all typical reporting stations within the state were averaged. The variable names and definitions (and sources) are:

SUNF: percent possible sunshine - fall season (NOAA)

SUNSP: percent possible sunshine - spring season (NOAA)

SUNSM: percent possible sunshine - summer season (NOAA)

SUNW: percent possible sunshine - winter season (NOAA)

Temperature: Temperature data is based on the National Oceanographic and Atmospheric Administration's historical daily normals, 1941-1970, collected for the 379 U.S. weather reporting stations. The data is in the form of heating and cooling degree days, defined as deviations from a daily average temperature of 65° F. To illustrate, a day with an average temperature of 50° F will register 15 heating degree days, indicating that heating is required on that day. Degree day totals were summed to obtain winter and summer season totals. Winter is defined to include the months November through March, summer includes the months April through October. To obtain state averages the seasonal values for the typical state reporting stations were averaged. The variables and their definitions are:

SHDD: summer heating degree days (NOAA)

WHDD: winter heating degree days (NOAA)

SCDD: summer cooling degree days (NOAA)

WCDD: winter cooling degree days (NOAA)

ECONOMIC DATA SET

The base data for this and the following variable sets was compiled by Frank Drysdale and Charles Calef of the Brookhaven National Laboratory and is contained in the data set, Energetics of the United States of America, September 1976.[3] This data was supplemented by the MIT Energy Laboratory where indicated below. The variables in the economic set were chosen to indicate energy consumption characteristics, economic growth rates, and absolute income and industrial activity levels. The variables, their definitions and their sources are:

PCAUTOS: number of automobiles registered in SMSAs per capita (Brookhaven National Lab.)

PCTRUCKS: number of trucks registered in SMSAs per capita (Brookhaven National Lab)

PCCYCLES: number of motorcycles registered in SMSAs per capita (Brookhaven National Lab)

PCJETS: Total commercial jet aircraft departures from all airports per capita (Brookhaven National Lab)

CHPOP: percent change in population 1960-1970 (U.S. Census of Population)

PCPI: per capita personal income 1975 (Survey of Current Business, April 1977)

CHPI: percent change in personal income 1969-1975 (U.S. Census, Survey of Current Business)

PCVA: per capita value added 1972 (U.S.Census)

CHVA: percent change in value added 1967-1972 (U.S. Census)

CHSALES: percent change in retail sales 1967-1972 (U.S. Census)

ENERGY PRODUCTION/SUPPLY DATA SET

The selection of the variables in this set was designed to capture energy supply and production characteristics of the states. This set includes energy production variables by fuel type, refinery capacity by process type, and electricity prices by user type as a proxy for energy availability. The variables, their definitions and their sources are:

OIL: oil production per year (Brookhaven National Lab)

GAS: natural gas production per year (Brookhaven National Lab)

COAL: coal production per year (Brookhaven National Lab)

HYDRO: hydroelectric ouput, average hourly production (Brookhaven National Lab)

REFINE: refinery capacity (Brookhaven National Lab)

CRACK: catalytic cracking capacity (Brookhaven National Lab)

REFORM: catalytic reforming capacity (Brookhaven National Lab)

ALKYL: alkylation capacity (Brookhaven National Lab)

THERM: thermal cracking capacity (Brookhaven National Lab)

HYCRACK: hydrocracking capacity (Brookhaven National Lab)

INDPRICE: cost of electric energy to industrial consumers, 1974 (U.S. Census of Manufacturers, 1974)

RESPRICE: typical residential electric bill, 250 kWh service, 1976 (Federal Power Commission)

Having specified the data content of each characteristic set, factor analysis is performed on each set to collapse the variables into their underlying dimensions. The factor scores for each set are then used as input to a cluster analysis routine. In this manner it is possible to evaluate the underlying dimensions and regional grouping of each characteristic set. The results of this stage one analysis are presented below for each set. Detailed stage one computer output is presented in the Appendix.

CLIMATE: As might be expected, the eight climate variables were condensed by the factor analysis into two dimensions. Table 1 indicates the sorted rotated factor loadings for the climate set. Recall from earlier discussion that the factor loadings are the correlation coefficients of the variables with the factor.

The first factor can be interpreted as a sunshine factor. Apparently, the percent possible sunshine for a given location is highly correlated season to season. The VP eigenvalue or sum of the squared loadings, when

Table 1

FACTOR LOADING

CLIMATE SET

VARIABLE		FACTOR 1	FACTOR 2
SUNF		.939	-.269
SUNSP		.932	.0
SUNSM		.878	.0
SUNW		.837	.0
SHDD		.0	.909
WHDD		.0	.895
SCDD		.257	-.873
WCDD		.0	-.759
	VP	3.308	3.192
	AVE.VP	.413	.398

divided by the number of variables, indicates the percentage of the total data variance explained by the factor. The first factor explains over 40% of the total variance. The second factor describes cold climate conditions. Areas with many total heating degree days tend to have fewer total cooling degree days and vice versa. This factor also explains nearly 40% of the variance. Together they explain 81% of the total data variance.

The results of the intermediate cluster analysis for the two climate factors are shown in Figure 12.

The mean factor scores for each climate cluster are indicated in Table 2. The more detailed tree diagram and mean cluster values for each amalgamation step are presented in the Appendix.

Two states, Arizona and Florida, which did not group with any other states due to the magnitude of their factor scores, were included in the cluster of states whose mean factor scores were most similar. The individual factor scores for these two states are also indicated in Table 2. Notice that Arizona is 3.2 standard deviations above the mean state in sunshine, while Florida is 3.4 standard deviations below the mean state in cold climate.

ECONOMIC DATA SET: The stage one factor analysis on the ten economic variables identified three orthogonal factors. As indicated by the factor loadings in Table 3, each factor explained about 25 percent of the variance in the data.

Factor one can be interpreted as a consumption factor, as indicated by the high loadings on the automobile, truck and motorcycle registrations per capita variables. Factor two characterizes economic growth but low manufacturing output. Factor three consolidates

FIGURE 12

Climate

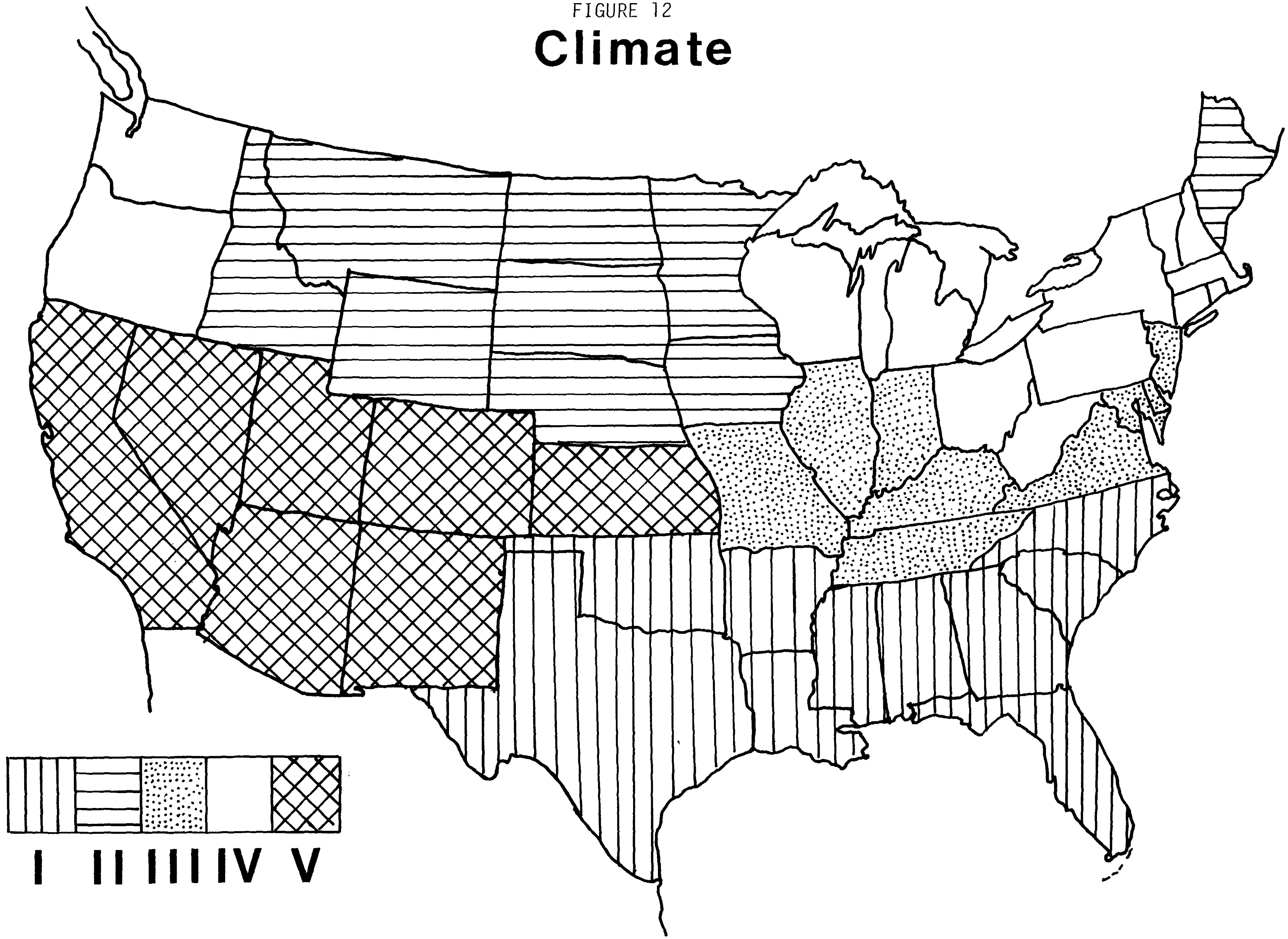


Table 2

MEAN CLUSTER VALUES

CLIMATE SET

GROUP	FACTOR 1 (SUNSHINE)	FACTOR 2 (COLD CLIMATE)	NUMBER OF CASES IN GROUP[1]
1	.001	-1.521	10
2	.289	1.069	9
3	-.345	-.203	9
4	-1.006	.468	13
5	1.713	.171	8
ARIZ	3.233	-.570	
FLOR	-.061	-3.360	

[1]does not add to 48 due to California being divided into two "states."

Table 3

FACTOR LOADINGS

ECONOMICS SET

VARIABLE		FACTOR 1	FACTOR 2	FACTOR 3
PCTRUCKS		.918	.0	.265
PCAUTOS		.916	.0	.0
PCCYCLES		.888	.0	.306
CHPI		.0	.915	.0
CHVA		.0	.886	.0
PCVA		.0	-.791	.0
CHPOP		.391	.0	.847
PCPI		.0	.0	.810
PCJETS		.350	.310	.662
CHSALES		.0	.367	.499
	VP	2.805	2.565	2.310
	AVE. VP	.280	.256	.231

those variables that represent affluence and population growth.

The result of the cluster analysis based on these three economic factors is indicated in Figure 13 and Table 4.

To be particularly noted in these results is the identification of the so called "sun-belt" indicated by Group 4. Once again there were a few states that did not cluster due to the magnitude of their scores on particular factors. Nevada, in particular, is over four standard deviations above the mean state in the income and population growth factor. These outlying states were grouped with the closest characteristic cluster.

ENERGY PRODUCTION SUPPLY DATA SET: The final characteristic set for stage one analysis consists of twelve energy production, processing and price variables. The factor analysis performed on this set isolated three underlying dimensions, the factor loadings for which are indicated in Table 5.

High correlation was found between the oil and gas production variables, all of which combined to form factor one. As indicated, factor one explains nearly 60 percent of the variance in the data by itself. Factor two can be interpreted to represent high electric energy prices (a surrogate for supply availability) and lack of hydroelectric power resources. Factor two explains approximately 17 percent of the variance in the data. Factor three is exclusively a coal factor and was eliminated from further analysis due to the small amount of variance it explained.[4] Figure 14 describes the results of the cluster analysis on the two production/supply factors. Table 6 indicates the mean factor scores for the four groups. Notice that oil and gas production and fuel processing are concentrated in basically four states. Several states again

FIGURE 13 **Economics**

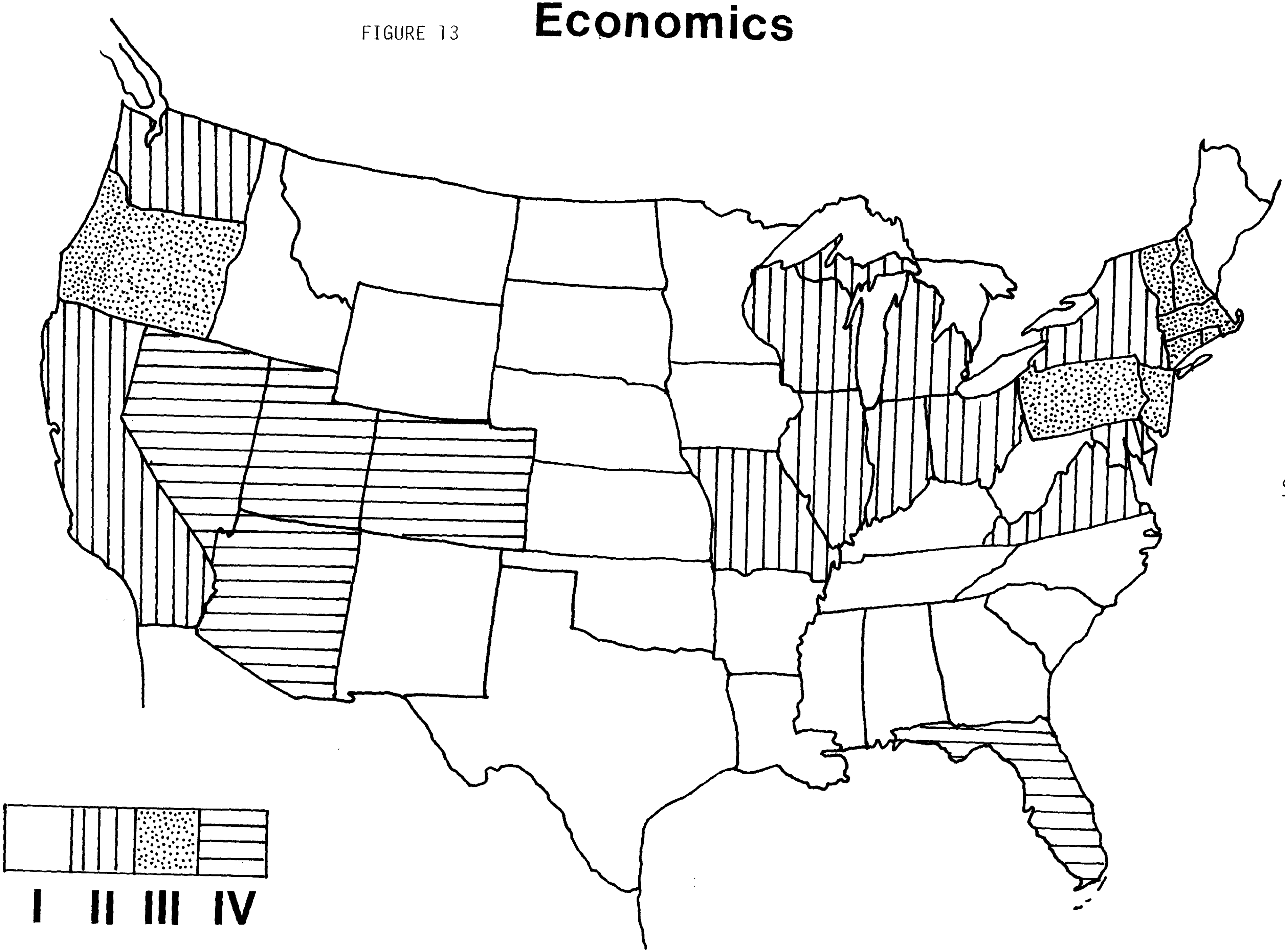


Table 4

MEAN CLUSTER VALUES

ECONOMICS SET

GROUP	FACTOR 1 (CONSUMPTION)	FACTOR 2 (GROWTH, LOW VA)	FACTOR 3 (HIGH INC-POP GROWTH)	NO. OF CASES IN GROUP
1	-.114	.617	-.612	23
2	.576	-.927	.050	13
3	-1.432	-.868	.689	8
4	1.319	.958	1.582	5
UTAH	2.232	.283	-.400	
WYOM	-1.687	2.062	.573	
NEVA	1.370	1.194	4.358	

Table 5

FACTOR LOADINGS

ENERGY PRODUCTION/SUPPLY SET

VARIABLE		FACTOR 1	FACTOR 2	FACTOR 3
REFINE		.991	.0	.0
CRACK		.983	.0	.0
REFORM		.983	.0	.0
ALKYL		.968	.0	.0
OIL		.951	.0	.0
THERM		.917	.0	.0
GAS		.897	.0	.0
HYCRACK		.729	.0	.0
INDPRICE		.0	.895	.0
RESPRICE		.0	.892	.0
HYDRO		.0	-.623	-.359
COAL		.0	.0	.946
	VP	6.951	2.007	1.084
	AVE. VP	.579	.167	.090

FIGURE 14

Production/Supply

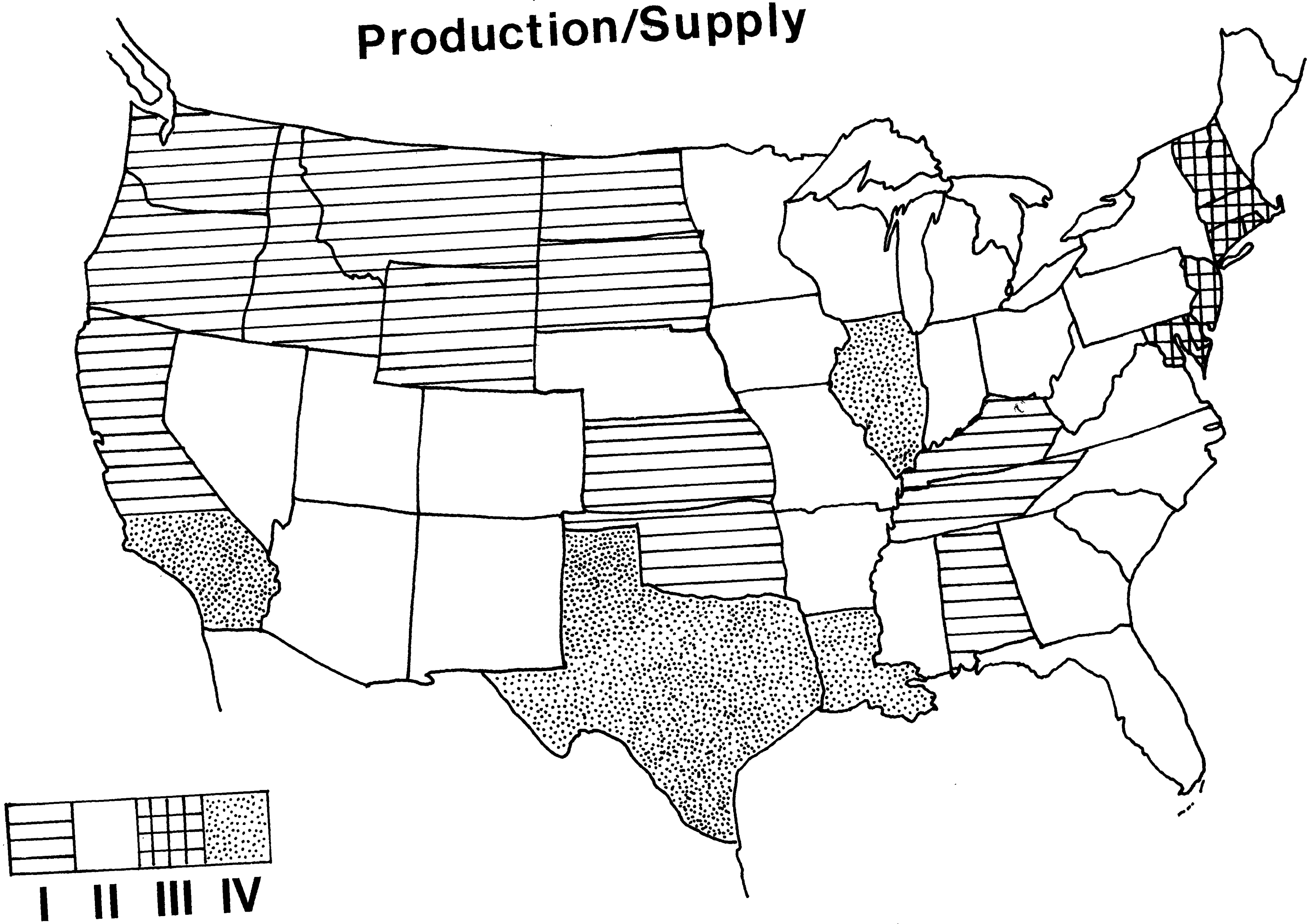


Table 6

MEAN CLUSTER VALUES

PRODUCTION/SUPPLY SET

GROUP	FACTOR 1 (OIL, GAS, PROCESSING)	FACTOR 2 (HIGH PRICE, NO HYDRO)	NO. OF CASES IN GROUP
1	-.172	-1.154	13
2	-.257	.130	24
3	-.273	1.462	8
4	2.485	.266	4
SCALIF	1.584	.330	
WASH	-.136	-3.752	
OREG	-.489	-2.224	
TEXA	5.796	.055	

did not cluster due to the magnitude of their factor scores. For the purposes of this intermediate presentation of results they were grouped with the clusters most similar to them. The factor scores for these states are also indicated in Table 6.

IV. STAGE TWO ANALYSIS

As was discussed in Section II, the methodology employed in this study to develop solar energy planning regions for the United States is two stage factor analysis. The first stage, described in Section III, constructed regions based on variable sets characterizing climatic, economic and energy production and supply conditions. The objective of the second stage analysis is to develop a set of composite characteristics which will describe the relationship between these three variable sets and the states within the U.S. The results of this analysis are a set of composite planning regions which constitute groupings of states that are most similar with respect to the composite characteristics.

Table 7 lists the factor characterizations from the three characteristic sets in the stage one analysis. As can be seen, the final analysis utilizes two factors from the climate set, three factors from the economic set and two factors from the production/supply set. The factor scores from the seven stage one factors now define new surrogate variables for analysis in stage two. Table 8 presents the factor loadings derived from the stage two factor analysis on the stage one factor scores. As indicated, the four composite factors are each essentially combinations of two correlated stage one factors. Each new factor explains about 20 percent of the variance in the stage one factor scores matrix. Given the nature of the factor loadings, verbal characterizations can be attached to the new factors.

Table 7

VARIABLES (FACTORS) FOR STAGE 2 ANALYSIS

	Variable (Factor)	VP
1)	Climate #1 - Sunshine	3.308
2)	Climate #2 - Cold Weather	3.192
3)	Economic #1 - Consumption (Cars, Trucks, Motorcycles)	2.805
4)	Economic #2 - Growth in Income and Value Added; Low Initial Value Added	2.565
5)	Economic #3 - High Income; Population Growth	2.310
6)	Production/Supply #1 - Oil and Gas Production and Fuel Processing	6.951
7)	Production/Supply #2 - High Electric Prices; No Hydro	2.007

Table 8

FACTOR LOADINGS

COMPOSITE FACTORS

VARIABLES	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
ECONOMIC 2	.973	.0	.0	.0
CLIMATE 1	.615	.567	.0	.378
ECONOMIC 1	.0	.922	.0	.0
CLIMATE 2	.0	.0	.886	.0
PROD/SUPPLY 1	.0	.377	-.700	.0
ECONOMIC 3	.0	.0	.0	.837
PROD/SUPPLY 2	-.447	-.289	-.262	.634
VP	1.569	1.417	1.369	1.264
AVG.VP	.224	.202	.196	.181

TOTAL VARIANCE EXPLAINED .803

Table 9 describes these composite factors. These four factors constitute new standardized variables distilled from the thirty original variables.

The factor scores for each of these four new factors are now inserted into the cluster analysis to determine the final regional groupings. Table 10 indicates the mean cluster values for the final composite regions. A certain amount of manipulation is necessary to smooth completely the regional groupings and it is useful to look at the mean cluster values in Table 10 to accomplish this task. As indicated in Table 10 there are a number of states with high-magnitude factor scores on the four final factors that have a tendency in the cluster analysis to break out individually, defining their own regions. It is possible, however, to compare them with the larger groups that were defined to uncover their underlying associations. For example, the basic southeast region is formed in Group I with a moderate mean factor score for each factor, ranging from -.693 to .377. Louisiana and Mississippi each have the same pattern as Group I although the factor score magnitudes in both cases are greater. It would appear expedient therefore to consider combining Mississippi and Louisiana with Group I given their similarities and patterns and despite their differences in magnitude. Using the same logic as that discussed above, one can begin to combine or discuss the combination of other states. While discussing the definition of factors in the composite analysis of stage two, it was pointed out that a number of states scored exceedingly highly on several of the factors. Such high scores on one or more factors tend to cause a state not to group with others that may have a similar pattern but a lower magnitude for each

Table 9

CHARACTERIZATION OF STAGE-2 FACTORS

FACTOR I

High Growth and Sunny Climate

(Climate Factor 1, Economic Factor 2)

FACTOR II

High Consumption and Sunny Climate

(Climate Factor 1, Economic Factor 2)

FACTOR III

Cold Climate, Low Oil, Gas Production and Processing

(Climate Factor 2, Production/Supply Factor 1 (-))

FACTOR IV

Affluent, High Energy Price

(Economic Factor 3, Production/Supply Factor 2)

TABLE 10

MEAN CLUSTER VALUES

COMPOSITE REGIONS

		Factor 1	Factor 2	Factor 3	Factor 4
SE	Group I	.377	-.187	-.424	-.693
NW	Group II	1.085	-.481	.951	-.511
NC	Group III	-.953	.423	.437	-.399
NE	Group IV	-.905	-.906	.020	.914
WCENT	Group V	.405	1.746	.629	.350
MIDSOUTH	Group VI	.685	-1.049	-1.954	-1.006
States with High-Magnitude Loadings					
	S. California	-.981	2.340	-.961	.477
	New Mexico	1.827	.136	-.069	.317
	Florida	1.058	-.569	-2.079	1.471
	Washington	-.381	1.182	1.364	-2.162
	Nevada	1.510	1.706	1.318	3.376
	Arizona	1.871	1.280	-.201	1.903
	Texas	-.319	1.979	-3.925	-.373
	Louisiana	.569	-.762	-2.470	-.939
	Mississippi	.801	-1.336	-1.438	-1.073
	Maine	-.398	-.559	.891	-.467

score. Nevada is an excellent example of this. Again, referring to Table 10, Nevada has exceedingly high positive scores on each of the four factors ranging from a low of 1.318 to a high of 3.376. The pattern for Nevada is, however, of the same sign throughout as that for Group 5, the West-Central, though the range in that case is from 1.350 to 1.746. As can be seen from Table 10, Arizona has a pattern in which, relative to Nevada, it scores quite low on factor 3 though it scores exceedingly highly on factors 1, 2 and 4. In much the same way, New Mexico parallels the patterns of Arizona and Nevada though with lower scores on 2, 3 and 4 than either of the other two. Thus, it is suggested that the combination of sun-belt states be put with the West-Central or Group 5 to form a South-West grouping that includes Northern California, Nevada, Wyoming, New Mexico and Arizona. Southern California offers a considerable challenge. While Northern California is within Group 5, scoring positively on all four factors, Southern California scores negatively on factor 1, very highly positive on factor 2, being a zone of high consumption and high sunshine, scores negatively on factor 3 and positively again on factor 4, thus making it unlikely that it would group with any of the regions available. Clearly, however, it is necessary that Southern California be grouped with the Southwestern region even though in this analysis we have shown it to be somewhat different in characteristics.

The state of Washington presents a particular problem in the analysis. As can be seen from Table 10, the pattern of coefficients for Washington is most similar to that for Group 3, the North Central. On the other hand, it is clearly geographically most close to the North-West group, Group 2. It, like the state of Maine, represents a non-contiguous grouping with the North-Central states. While it is not identical to the North-West group, looking within that group to Oregon and Idaho, it is possible to see that Washington state shows a not dissimilar pattern to that area, and must be grouped there to maintain geographic consistency. In much the same way the state of Maine is clearly much more similar in climate and economic conditions to much of the North-Central region. Its geographic location requires it be included as part of the North-East region, however.

The state of Florida, as can be seen from Table 10, shows a pattern similar to that of the rest of the South-East although is has higher factor scores than the South-East group. It is, however, fairly similar to Mississippi and Louisiana. Again, geographic continuity requires that Florida be included in the South-East region.

The state of Texas offers a unique challenge in any regional analysis. As can be seen from Table 10, the pattern of factor scores for Texas is unlike any other. It shows an exceptionally high positive value on factor 2 making it similar to Arizona and Nevada and Southern California, reflecting a high endowment in sunshine and per capita consumption. It shows an exceptionally strong negative factor score in factor 3, giving it a high score for oil and gas reserves, making it similar to Louisiana and Mississippi. It shows a moderate negative on factor 4 and a moderate negative on factor 1. Both because of its size and resource endowment, we have suggested that Texas remain a region unto itself.

In summary, Table 11 presents the final mean factor scores for the six solar energy planning regions. Figure 15 presents these regions pictorially. The solar energy regional analysis carried out here defined

Table 11

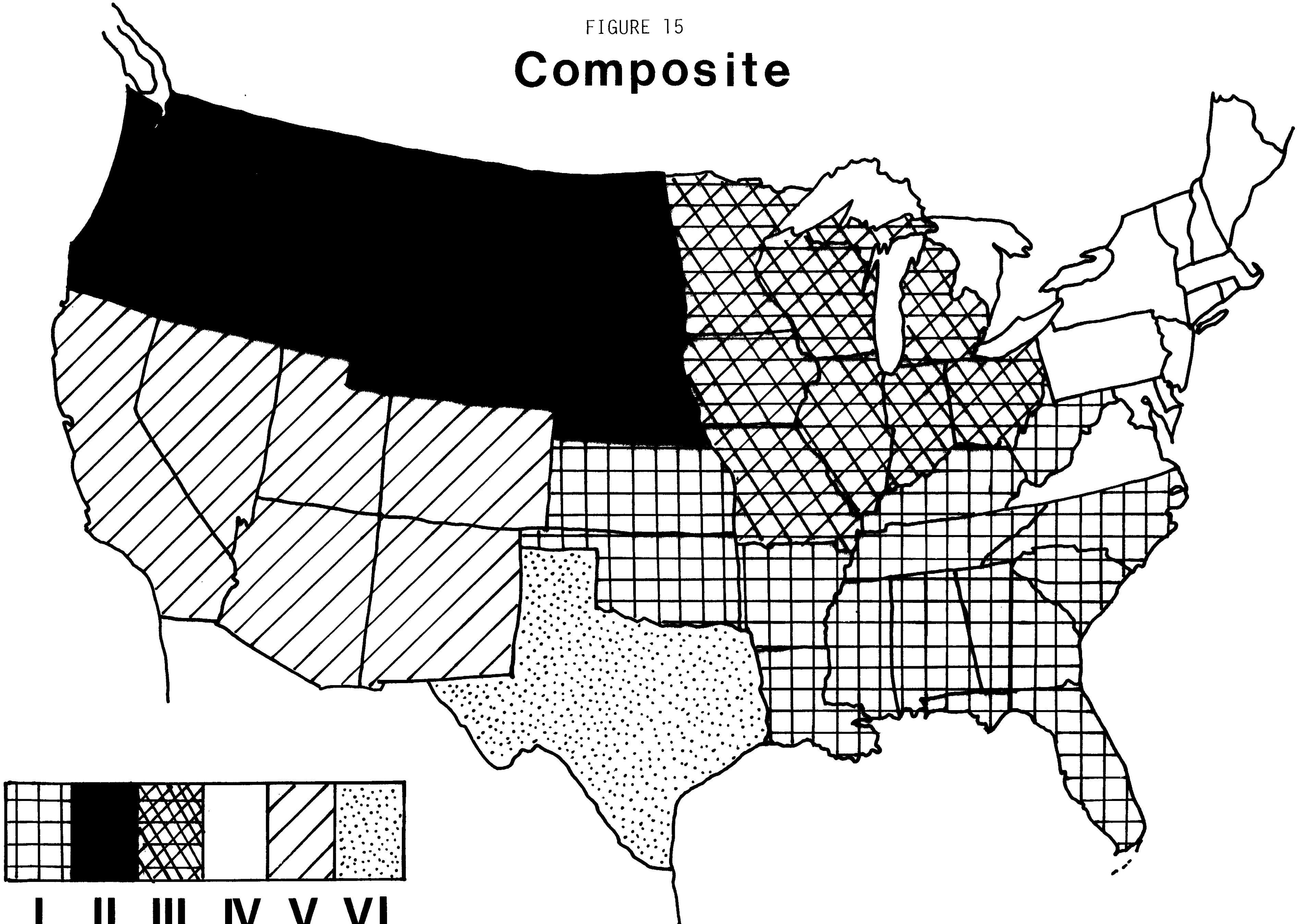
ADJUSTED MEAN CLUSTER VALUES

COMPOSITE REGIONS

GROUP	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	NO. OF CASES IN FINAL GROUP
1.	.477	-.344	-.787	-.575	13
2.	.902	-.273	1.003	-.717	8
3.	-1.022	.546	.380	-.391	8
4.	-.863	-.877	.093	.799	12
5.	.778	1.529	.282	1.018	7
Texas	-.319	1.979	3.925	-.373	1

FIGURE 15

Composite



five planning regions with an additional sixth region, the state of Texas. The North-East and the North-West groups contain uncharacteristic states, Maine and Washington, which fit less well within the region than the remainder of the states. In much the same way the South-East and the South-West have peripheral states whose characteristics would hold them apart from the region as a whole, both Southern California and Florida do not group using this particular methodology.

Having reviewed the regions used in energy planning in the United States at the present time, it is significant to compare the results of this analysis with that reported earlier. The regions so defined are, in general dissimilar to all of the unifunctional regional groupings discussed in Section I. The most striking difference is the division of the West into regions which more nearly reflect climatic differences North to South and which identify the Solar and Economic uniqueness of the "Sun Belt". The Midwest also shows the effect in the definitional process of both the climatic and the economic variables, though here the dissimilarity to both the Census and the Bureau of Mines regions is not as great. Breaking the Northeast and Southeast is less of a problem although, as with the Civil War, their are difficulties in dealing with boundary states.

The final test of the usefulness of the proposed solar electric regions will be in their adoption for planning and testing.

FOOTNOTES TO SECTION I

1. USERDA, 1976, p.

2. A preliminary draft of the background and methodology sections of this paper was presented by Richard Tabors at the AAAS meeting in Denver, Colorado, February 1977.

FOOTNOTES TO SECTION II

1. Tabors, Definition of Multifunctional Planning Regions: A Case Study of East Pakistan. (Cambridge, Mass. Harvard University Population Center,) 1971.

2. Lake, Blair, Hudson and Tabors Classification of American Cities for Case Study Analysis, Carried out for the Environmental Protection Agency by Urban Systems Research and Engineering, July 1976.

3. Brian J.L. Berry, "An Inductive Approach to the Regionalization of Economic Development," Essays on Geography and Economic Development, ed. by Norton Ginsburg (Chicago: University of Chicago, Department of Geography, Research Paper No. 62, 1960): pp.78-107; Brian J.L. Berry, "A Method for Deriving Multi-factor Uniform Regions," Przeglad Georaficzny, XXXIII, No. 2 (1961): 263-279; Brian J.L. Berry, "Grouping and Regionalizing: An Approach to the Problem Using Multivariate Analysis," in Quantitative Geography, Par I: Economic and Cultural Topics, ed, by W.L. Garrison and D.F. Marbel (Chicago: Northwestern University Studies in Geography, No. 13, 1967): 219-251. For a more complete bibliography on factor analysis and regionalization, see R.J. Johnson, "Grouping and Regionalizing: Some Methodological and Technical Observations," Economic Geography (June supplement, 1970):

4. Although this paper will deal with the use of factor analysis in regionalization, it is not its purpose to dwell upon the mathematical aspects of the technique. References to the technique of factor analysis may be found in the following: Harry H. Harmon, Modern Factor Analysis (Chicago: University of Chicago Press, 1960); N. Lawley and A.E. Maxwell, Factor Analysis as a statistical Method (London: Butterworths, 1963); and M.G. Kendall, A Course in Multivariate Analysis (New York: Griffins Statistical Monographs and Courses, Hafner Publishing Company, 1968).

5. Harmon, Modern Factor Analysis, pp. 3-10.

6. Margaret Jarman Hagood, "Statistical Methods for Delineation of Regions Applied to Data on Agriculture and Population," Social Forces, XXI (March 1943): 287-297.

7. Margaret Jarman Hagood and Daniel O. Price, Statistics for Sociologists (New York: Henry Hold and Co., 1952): Chapter xxvi.

8. Brian J.L. Berry, "Basic Patterns of Economic Development," in Atlas of Economic Development, ed. by Norton Ginsburg (Chicago: University of Chicago Press, 1961): 110-119.

9. IBID. p. 110

10. An overview and bibliography of the use of Factor Analysis may be found in Philip H. Rees, "Factorial Ecology: An Extended Definition, Survey, and Critique of the Field," in Comparative Factorial Ecology Supplement to Economic Geography, Vol. 47, No. 2 (Supplement), June 1971, pp. 220-234.

11. It should be noted that for this analysis the state of California was divided into Northern and Southern California on the San Bernadino county line.

12. For a more complete discussion of the dangers in "Blind" use of factor analysis for regional definitic see Tabors (1971) op.cit.

FOOTNOTES TO SECTION III

1. Supplied by Eldon Boes, Sandia Laboratories in Boes, et al., Availability of Direct, Total, and Diffuse Solar Radiation to Fixed and Tracking Collectors in the U.S.A., Draft Report SAND 77-0885, 1977.

2. Aerospace Corp., Mission Analysis of Photovoltaic Energy Systems, NSF/RANN #61-44099, El Segundo, California, Aerospace Report No. ATR-76 (7476-01)-1, 1 December 1975.

3. Drysdale, F.R., Calef, C.E. The Energetics of the U.S.A.: An Atlas, National Center for Analysis of Energy Systems, Brookhaven National Laboratory, Prepared for Division of Technology Overview, ERDA, September 1976,

4. This decision was not arbitrary in the sense that the automatic cutoff for factor generation is a VP value of 1.0. See the methodology section for further discussion of choice of number of factors.

GENERAL REFERENCES

1. Battelle for U.S.E.R.D.A., Regional Analysis of the U.S. Electric Power Industry (ERDA: Washington, July 1975)

2. Dixon, W.J., ed., BMDP Biomedical Computer Programs., University of California, Los Angeles, 1975.

3. Drysdale, Frank R., Calef, Charles E., The Energetics of the United States of America: An Atlas, National Center for Analysis of Energy Systems, Brookhaven National Laboratory, September 1976.

4. Federal Power Commission, Typical Electric Bills, 1976., Washington, D.C., 1977

5. W.K. Foell, The Wisconsin Energy Model: A Tool for Regional Energy Policy Analysis. (University of Wisconsin-Madison Institute for Environmental Studies = Madison, 1974).

6. William Loehr, "Public Policy and Energy Balance", Journal of Energy and Development, Autumn 1976 pp. 132-147.

7. William H. Miernyk, "Some Regional impacts of the Rising cost of Energy", Papers of the Regional Science Association, Volume 37, 1976 pp. 213-227.

8. William H. Miernyk, "Regional Economic Consequences of high Energy Prices in the United States", The Journal of Energy and Development, 1976 pp. 213-239

9. Midwest Research Institute, Program for Regional Energy Analysis

10. Nonnally, Jum C., Psychometric Theory., McGraw-Hill Book Company, New York, 1967.

11. Karen Polenske and Paul F. Levy for Department of Transportation, Office of Transportation Planning Analysis, Multiregional Economic Impacts of Energy and Transportation Policies (D.O.T: Washington, March 1975)

12. Rummel, R.J., Applied Factor Analysis, Northwestern University Press, Evanston, Ill., 1970.

13. R. Thomas Van Arsdall, Regional Patterns of Energy Consumption in the U.S., 1967. (University of Illinois, Champaign/Urbana: Urbana, 1975) under grant from the National Science Foundation NTIS PB 242 689.

14. In addition to those listed above the Policy Analysis Division, National Center for Analysis of Energy Systems, Brookhaven National Laboratory produced a series of reports under their Northeast Energy Perspectives Series which include both supply and demand analyses of the Northeast (New England plus Pennsylvania, New York and New Mexico, Maryland, Delaware and the District of Columbia).