I. INTRODUCTION

Coal provides more ton-miles of both rail and inland waterway traffic than does any other commodity. Yet the ratio of its value to weight is one of the lowest of any major commodity that has to be transported long distances. Because of this consumers in general have tended to locate near supply sources for saving in transportation. While a substantial portion of the energy resources in the United States have become increasingly scarce in recent years, coal is still in relatively abundant supply here. The energy shortage that came into clearer focus following the 1973 oil embargo has intensified the need to improve the efficiency of coal use and transportation. This and rising freight costs are playing increasingly important roles in the selection of appropriate locations of new coal-fired plants in the electric utility and other industries.

It is essential to understand geographic and economic space and most importantly the concept of a region before the implementation of regional transportation and other related policies can be formulated. There has been an increasing interest in recent years in the development of better methods for the delineation of regions and in explaining the size and shapes of these market areas [1, 3, 5, 11]. There are two broad classifications in defining regions [2]: uniform region and functional region. A uniform region is defined on the basis of attributes of the areas being grouped and the areas are contiguous. A functional region is composed of areas or locational entities which have more interaction or connection with each other than with outside areas.

Our approach belongs to the second definition. In his analysis of international commodity flows and common memberships in intergovernmental organizations, Russett applied factor analysis to a matrix containing standardized measures of the flow or association between places [13]. Brown and many others [2, 4, 8, 14] further uses hierarchical clustering procedures for interaction data in their studies for the definition of region. For our purpose the clustering process is used to establish functional collections of relatively strong intragroup ties. The tree diagram is then used in defining coal flow regions among states. This diagram is useful because the flows are unusually sensitive to interstate changes in the quantity of coal demanded and supplied.

* Virginia Commonwealth University, West Virginia University, and West Virginia University, respectively.
After regions are identified, smooth spatial market boundaries are then drawn on the basis of both mine prices and freight rates. The purpose of this paper is to focus attention on the delineation of markets in the coal industry so that wasteful transportation can be minimized.

II. PROCEDURES IN CLUSTER ANALYSIS

The methodology is concerned primarily with the development of aggregation procedures in the spatial interaction model. The data sets involved are typically coal movements which relate to interaction between a large number of basic spatial units defined for data collection purposes. A two-stage algorithm is used to generate final outcome [2, 8, 14].

In the first stage, an n x n table of recorded coal flows among states is employed to derive measures of functional linkages. The rows and columns of a transaction flow table are alternately scaled and adjusted by the iterative proportional fitting procedures (IPFP)—also called the biproportional or RAS method.

The iteration procedure is continued until all row and column sums are equal to the same number (e.g., 100). Zeros are posted on its diagonal. After this double standardization the entries of the flow table are statistical estimates of coal flow that would occur if all states had the same output capacity in coal production. There will be no size difference between the n entries in the total amounts of movement into and out of each of them. The functional linkage can therefore be measured by the entries of the adjusted table. If there is no coal flow from state i to state j, the adjusted ij-entry will be zero. On the other hand, if all coal trade involving states i and j is accounted for by a single flow, the adjusted ij entry is 100. These adjusted entries are useful to discover patterns of strong interaction among states [14].

In the second stage, a hierarchical clustering method is used. It is based on the concept of a strong component, which is a set of directed vertices between any two members, of which there exists a path of directed links. At the beginning each unit is a strong component and thus contains n strong components as no links are between units. Following the magnitude of values in the adjusted table, a series of directed graphs is obtained through the sequential insertion of links in decreasing order. If the link between i and j entries is the largest a connected line is drawn. Next goes the second largest link. These entries are partitioned into non-overlapping groups. Finally, all entries lie in one cluster or one strong component.

This clustering process can be used to determine functional collections of units with relatively strong intragroup ties. The significance of these groups can be measured by the magnitudes of the difference between the thresholds defined by the limits on the ranges of existence of groups. If these ranges are relatively large the groups are well-defined. A tree diagram or dendrogram that shows the clustering process is produced.
Figure 1

TREE DIAGRAM: Coal Movements Among States

ND, SD, MN
MT, WY, ID, WA, OR
MD
DE, DC
VA
NH, VT
PA
RI
IN
WI
OH
NY, NJ
MI
IL
KY
CT, ME
TN
MA
VA
NC, SC
AL
MS
GA
FL
IA
NB
KS, MO, AR, OK, TX
LA
CO
NM
CA
AZ
NV, UT
III. DATA AND IDENTIFICATION OF REGIONS

Data from Mineral Industry Surveys [9] were collected and extracted for analysis. The data show the distribution of bituminous coal and lignite in the United States during the calendar year of 1978 from coal production districts to destination states. To fit into square $n \times n$ table, data must be rearranged and a $33 \times 33$ flow table was obtained. States with little flow of coal were excluded since the row or column sum corresponding to it was zero and could not be adjusted by the IPFP.

The IPFP was then used to obtain an adjusted table with zeros on its diagonal. All row and column sums were adjusted to equal selected constant (e.g., 100). A diagram can then be constructed with 33 points in it. The largest value in the adjusted table is found. An arrow is then drawn between these two points. Following that, the second largest value is found. This process is represented hierarchically in Figure 1. Clusters of states as a region can be found through visual examination of the tree diagram together with the range of threshold. A threshold value of more than 69 would produce an extreme case of 33 regions and a value of less than 19 would give only one region. In light of the apparent five clusters and possible comparison with regions defined in our previous effort, a threshold of less than 25 is used as an indication of weakly connected regions.

Five regions are therefore identified and are indicated in Table 1. For example, region IV is formed with a threshold of 43 and is "better defined" than region I which is formed with a threshold of 31. Region I is, in turn, better defined than region V of 28, region III of 24 and region II of 22. Note that, because of data limitation, region IV is established through two spatial points only. An increase in the number of spatial points could have an adverse affect on the intragroup tie of the region.

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGIONS IN COAL INDUSTRY IDENTIFIED FROM CLUSTER ANALYSIS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regions</th>
<th>States</th>
<th>Regional Centers</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>MD, DE, DC, VA, NH, VT, PA, RI</td>
<td>Scranton, PA</td>
</tr>
<tr>
<td>II</td>
<td>KY, CT, ME, TN, MA, VA, NC, SC, AL, MS, GA, FL</td>
<td>Chattanooga, TN</td>
</tr>
<tr>
<td>III</td>
<td>IN, WI, OH, NY, NJ, MI, IL</td>
<td>Springfield, OH</td>
</tr>
<tr>
<td>IV</td>
<td>ND, SD, MN, MT, WY, ID, WA, OR</td>
<td>Sheridan, WY</td>
</tr>
<tr>
<td>V</td>
<td>IA, NB, KS, MO, AR, OK, TX, LA, CO, NM, CA, AZ, NV, UT</td>
<td>Amarillo, TX</td>
</tr>
</tbody>
</table>
A regional center can now be established from the supply center and the demand center in each region. Coal supply centers are estimated from output and location of coal mines and production districts. Demand centers are identified from the size and the location of coal utility power plants. These regional centers listed in Table 1 are needed before smooth market boundaries can be derived.

IV. SPATIAL MARKET AREAS DEFINED

A spatial market area model is formulated in this section to define optimal coal markets. Our purpose is to identify market centers from regions determined in section III and then derive market boundaries from the centers. The concept of intra-industry dispersion is employed as a starting point in developing the analytical framework and estimating the empirical data. We also assume, for simplicity, that the aggregate of firms in a market zone are located at a market center and are capable of supplying their own market. The market boundaries are not predetermined but should be drawn on the basis of both mine prices and freight rates as they are the major sources that influence final delivered prices.

To define market boundary, market centers must first be established from supply and demand centers in each region. Coal supply centers are estimated from output and location of coal mines and production districts. Demand centers are established from the size and location of coal utility power plants. Freight rate functions in these five regions must also be estimated and they are given in Table 2 along with mine prices.

The market boundary is determined by the same delivered prices where the difference between freight costs from two market centers is equal to the difference between mine prices. The delivered prices $P_1$ and $P_2$ for market I and market II are then given, respectively as:

\[
\begin{align*}
(1) \quad P_1 &= m_1 + f_1(d) \\
(2) \quad P_2 &= m_2 + f_2(d)
\end{align*}
\]

where $m$'s are mill prices and freight cost $f$ is the function of distance $d$. Following f.o.b. system, the boundary line defined by $P_1 = P_2$ in equations (1) and (2) leads to

\[
(3) \quad m_1 + f_1(d) = m_2 + f_2(d).
\]

It can be shown, when freight rates are the same $f_1 = f_2$ and mine prices are different $m_1 \neq m_2$, the market boundary is either an ellipse or a hyperbola; when freight rates are different and mine prices are the same, the market boundary is in circular form. However, when both freight rates and mine prices are different, the following relations are established.
where $x$ and $y$ are horizontal length and vertical height of a triangle as in Figure 2. $a_1, a_2, a_3, b_1, b_2, b_3$ are constants and $h$ is the distance between two firms I and II. Equation (4) is established from relation (3) with freight rate functions in quadratic form. Equation (5) is formulated from triangle relation determined by two market centers and boundary point.

Let distance between market center I and boundary point be $d_1 = \sqrt{x^2 + y^2}$ and $d_2 = \sqrt{(h-x)^2 + y^2}$ for distance between market center II and that boundary point. The following matrix form is established to be used for computer programming:

\[
\begin{bmatrix}
  d_1^{i+1} \\
  d_2^{i+1}
\end{bmatrix} =
\begin{bmatrix}
  d_1^i \\
  d_2^i
\end{bmatrix} - \begin{bmatrix}
  2a_1d_1 + a_2 & -2d_1d_2 - b_2 \\
  2d_1 & 2d_2 - 2h\cos\theta
\end{bmatrix}^{-1} \begin{bmatrix}
  F_1 \\
  F_2
\end{bmatrix}
\]

where the elements in the inverse matrix are obtained by taking the partial derivative of $F_1$ and $F_2$ with respect to $d_1$ and $d_2$. The initial values of $d_1^0$ and $d_2^0$ can be found when the direct distance between two market centers is known with $\theta = 0$. By adding another degree of angle, another set $(d_1^{i+1}, d_2^{i+1})$ can be found from relation (6). A computer program has been derived from...
TABLE 2

MINE PRICES AND TRANSPORT RATE FUNCTIONS*

<table>
<thead>
<tr>
<th>Regions</th>
<th>Average Mine Price (per ton)</th>
<th>Mine Price (Cents per million Btu)</th>
<th>Transport Rate Functions (cents per million Btu transported)</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>21.33</td>
<td>87.30</td>
<td>( TR = 13.14 - 0.0000297d^2 + 0.0891d ) (1.66) (−1.76) (2.60)</td>
<td>0.46</td>
</tr>
<tr>
<td>II</td>
<td>25.08</td>
<td>107.74</td>
<td>( TR = 6.81 - 0.0000159d^2 + 0.0737d ) (1.88) (−2.29) (2.73)</td>
<td>0.63</td>
</tr>
<tr>
<td>III</td>
<td>14.11</td>
<td>61.70</td>
<td>( TR = 2.93 - 0.0000159d^2 + 0.0704d ) (0.79) (−1.20) (2.01)</td>
<td>0.88</td>
</tr>
<tr>
<td>IV</td>
<td>11.41</td>
<td>58.85</td>
<td>( TR = 7.43 - 0.0000182d^2 + 0.0723d ) (1.42) (−3.52) (4.03)</td>
<td>0.55</td>
</tr>
<tr>
<td>V</td>
<td>17.59</td>
<td>97.04</td>
<td>( TR = 1.48 - 0.0000197d^2 + 0.0740d ) (0.51) (−1.67) (1.98)</td>
<td>0.41</td>
</tr>
</tbody>
</table>

*Data are obtained from *Mineral Year Book* [9] and Interstate Commerce Commission [6]. The figures in parentheses are the t statistics.

the relation (6), and the boundary points between the two market centers are generated.  

Our purpose is to define market boundaries from five market centers identified in the last section. The market boundaries are not predetermined but should be determined on the basis of both mine prices and freight rates as they are not only different in different regions but also are the major sources that influence final delivered prices.

In order to generate these boundary points, mill prices and freight rates in terms of cents per million Btu are estimated. Estimated mine prices and freight rates in these five regions are summarized in Table 2. The mine prices per ton in region II are the highest with $25.08 per ton and region I is the next highest. Regions V, III and IV follow. However, the heating value of coal in each region is different. The value of the coal therefore must be transformed into cents per million Btu to reflect its heating value. After the transformation, the mine prices in Btu terms in region V become more expensive than region I. The coal of region II continues to be the most expensive, but the difference becomes less significant as the heating value in this area is the highest.

Transport cost must receive special consideration as coal is a bulky commodity. The market areas of Figure 3 are determined from computer yields through use of mine price and freight rate data from Table 2. Given the market centers identified in Table 1, the market boundaries can be defined on the basis of the same delivered prices which, in turn, are determined by mill prices and freight rates in adjacent markets. The market boundary is not predetermined but is determined after the structure of the markets is ascertained and their parameters estimated. Some cities on the boundaries are identified, such
as Durham, North Carolina between regions I and II and Youngstown, Ohio between regions I and III as was indicated in Figure 3.

The relative advantage in transport rates can also be determined. When the curve of market boundary is bent toward a market center, it means that the transport costs in that region become relatively more expensive as distance increases. For example, transport rates in region I are greater than those of region III. This is consistent with modifications in railroad freight rates with larger portions of the coal moving in with trains each year. Unit train rates tend to be closely related to costs of the service. Furthermore, recent relaxation of public regulation of rail freight rates are causing the rates to be even more cost sensitive than in past years when coal freight rate structures were more rigid in nature. No undue capacity constraints are anticipated in the railway system.

Different transport modes such as barge and slurry pipeline could distort market boundaries. If barge transportation is available as an alternative mode for part of the market while railway is the only possible means for the rest of the market, market boundaries would take irregular shapes as barge transport can penetrate longer distances. The smooth boundary estimated in Figure 3 is based strictly on rail transport.

Market centers are no longer arbitrarily determined as was done by Project Independence Report of 1974 [12] shown here in Figure 4. Our markets are first identified objectively through cluster analysis. Market centers are then determined by both demand and supply centers. Because of stronger demand in the eastern portion of these regions, market centers as well as most markets shift towards the east.

The type and quality of coal varies from region to region and even within regions. Differences in heating value were taken into consideration in our model as the value of coal is estimated in terms of cents per million Btu. However, there are also variations in other characteristics, such as sulfur, ash, and moisture contents. These variations cause some types of coal to be useful for only certain purposes. The high-quality steam and coking coal of Pennsylvania and West Virginia in region I and II, especially the well-known low volatile Pocahontas coal of West Virginia, is low in ash and sulfur and is suitable for coke manufacture. Some coal in these regions, especially in Ohio and parts of Pennsylvania and West Virginia, is sufficiently near the surface to permit surface mining. Some of the seams in these regions are six feet or more in thickness. However, toward the southern portion of regions I and III, and region II, most of the seams are much thinner. Western deposits of region III are primarily medium to high-volatile coal. The coal of this area is relatively high in ash and sulfur and is generally suitable for heating and steam purposes only. Seams of coal in this region generally are so thin that underground mining is prohibitively expensive. Colorado and Utah of region IV, with deposits of relatively high-quality coking coal, have been the source of coke for the smelting of iron and other metals in the Rocky Mountain and West Coast regions. However, most of region IV and almost all of region V have coal that
is low in rank. And most of it is found close to the surface and is suitable for electric power generation purposes.

V. CONCLUSION

The central focus of the study is a definition of coal markets through the interaction between a large number of basic spatial units. The markets are determined objectively through iterative proportional fitting procedures and hierarchical clustering method. The market boundary specified in this study is drawn from both mine price and freight functions, both of which are main factors influencing delivered prices. Since coal is a bulky product that is heavy in weight relative to its value, a spatial feature (distance cost) is included in the identification and definition of coal markets. This provides more realistic and useful basis for study and regional analysis than the 23 old mining regions designated by the U.S. Bureau of Mines in 1932. The market centers are identified on the basis of both supply and demand forces, and differ from the markets defined in the Project Independence Report of 1974, which was on a supply basis only.

That coal is and will continue to be a prime source of fuel for electric generating stations is a virtual certainty. Electric utility and other industries that must have dependable supplies of boiler fuel look to coal for substantial portions of their energy requirements over the remaining years of this century. This means that greater coal tonnages will move increasingly longer distances. Due to the magnitude of these requirements, the scope of current and growing, coal markets and the clarification of minimum transportation cost assumes a role of progressively greater importance. In the absence of careful study and thorough analysis of coal markets, the cost could become unduly great for many coal consuming firms and their customers.

FOOTNOTES

1. For a review of regions defined by the Project Independence Report of 1974 and its revision by our previous effort see Campbell and Hwang [3].
2. A weighted average procedure is used to determine supply and demand centers and subsequently regional centers.
3. For a review of literature along this line see Greenhut [4].
4. For more details, see Campbell and Hwang [3].

REFERENCES