Classification of Coastal Communities Reporting Commercial Fish Landings in the U.S. Northeast Region: Developing and Testing a Methodology

SARAH L. SMITH, RICHARD B. POLLNAC, LISA L. COLBURN, and JULIA OLSON

Introduction

This paper introduces a method for classifying coastal communities for either sampling purposes or further analysis. Along the coastline from North Carolina to the Canadian border we find nearly 2,000 communities associated with commercial and/or recreational fishing. When NOAA's National Marine Fisheries Service (NMFS) plans to implement fishery management plans, it is necessary to conduct (among other analyses) a social impact assessment (SIA). These SIA's can be quite complex and time consuming (e.g. Pollnac et al., 2006); nevertheless, they are often required to be submitted in a very short time period. In an attempt to be prepared to conduct SIA on short notice, all NMFS Regions have prepared profiles of a subset of the numerous coastal communities with fishing activity. These are called Community Profiles. This raises the question of how one selects the communities to be profiled.

One hundred seventy-seven community profiles were created and have been posted on the web site "Community Profiles for the Northeast U.S. Fisheries" (http://www.nefsc.noaa.gov/read/socialsci/community_profiles/). The profiles were developed as part of a nationwide initiative to develop community profiles for each of the NMFS regions for use in Environmental Impact Statements (EIS). The profiles provide basic descriptive information, including a historic, demographic, cultural, and economic context, for understanding a community's involvement in fishing and also furnishes a baseline from which to measure future change.

Thus far, communities to be profiled have been selected on the basis of size and importance of fishery, types of fishing present, and overall knowledge possessed by experts working in the region. We posit that this technique is too unsystematic for this important endeavor, as important fishing communities could possibly be overlooked. SIA's describe important implications of potential impacts of management actions on fishermen and the communities in which they live. If SIA's are based on the limited information available in community profiles, and if the communities profiled are not representative of the communities involved in the target fishery, then the SIA's produced may not reflect an understanding of the potential impact of fishery management plans (FMP's). Inaccurate SIA's can result in decreased fishing activity, which may affect household and community well-being and lead to social dysfunction within communities reliant on fishing, exacerbating the resistance to fisheries management that is evident in the Northeast Region and elsewhere (Pollnac et al., 2006).

If we could classify the large number of coastal communities into smaller, meaningful groupings, SIA data from a sample of communities within relevant subgroups would provide more accurate data for management decision making. Relevant subgroups would be those characterized by varying degrees of nonfishery and fishery attributes associated with participation in the target fishery or fisheries. Hence, the subgroups should be based on multivariate criteria—an analytic task for some form of numerical taxonomy.

Sarah L. Smith is with the Department of Marine Affairs, Washburn Hall, University of Rhode Island, Lower College Road, Kingston, RI 02881 (email: sarahlindleysmith@gmail.com). Richard B. Pollnac is with the Coastal Resources Center, University of Rhode Island, South Ferry Road, Narragansett, RI 02882. Lisa L.Colburn is with the Narragansett Laboratory, Northeast Fisheries Science Center, National Marine Fisheries Service, NOAA, 28 Tarzwell Dr., Narragansett, RI 02882-1152. Julia Olson is with the Northeast Fisheries Science Center, National Marine Fisheries Service, NOAA, 166 Water Street, Woods Hole, MA 02543.

ABSTRACT—The National Marine Fisheries Service is required by law to conduct social impact assessments of communities impacted by fishery management plans. To facilitate this process, we developed a technique for grouping communities based on common sociocultural attributes. Multivariate data reduction techniques (e.g. principal component analyses, cluster analyses) were used to classify Northeast U.S. fishing communities based on census and fisheries data. The comparisons indicate that the clusters represent real groupings that can be verified with the profiles. We then selected communities representative of different values on these multivariate dimensions for in-depth analysis. The derived clusters are then compared based on more detailed data from fishing community profiles. Ground-truthing (e.g. visiting the communities and collecting primary information) a sample of communities from three clusters (two overlapping geographically) indicates that the more remote techniques are sufficient for typing the communities for further in-depth analyses. The in-depth analyses provide additional important information which we contend is representative of all communities within the cluster.

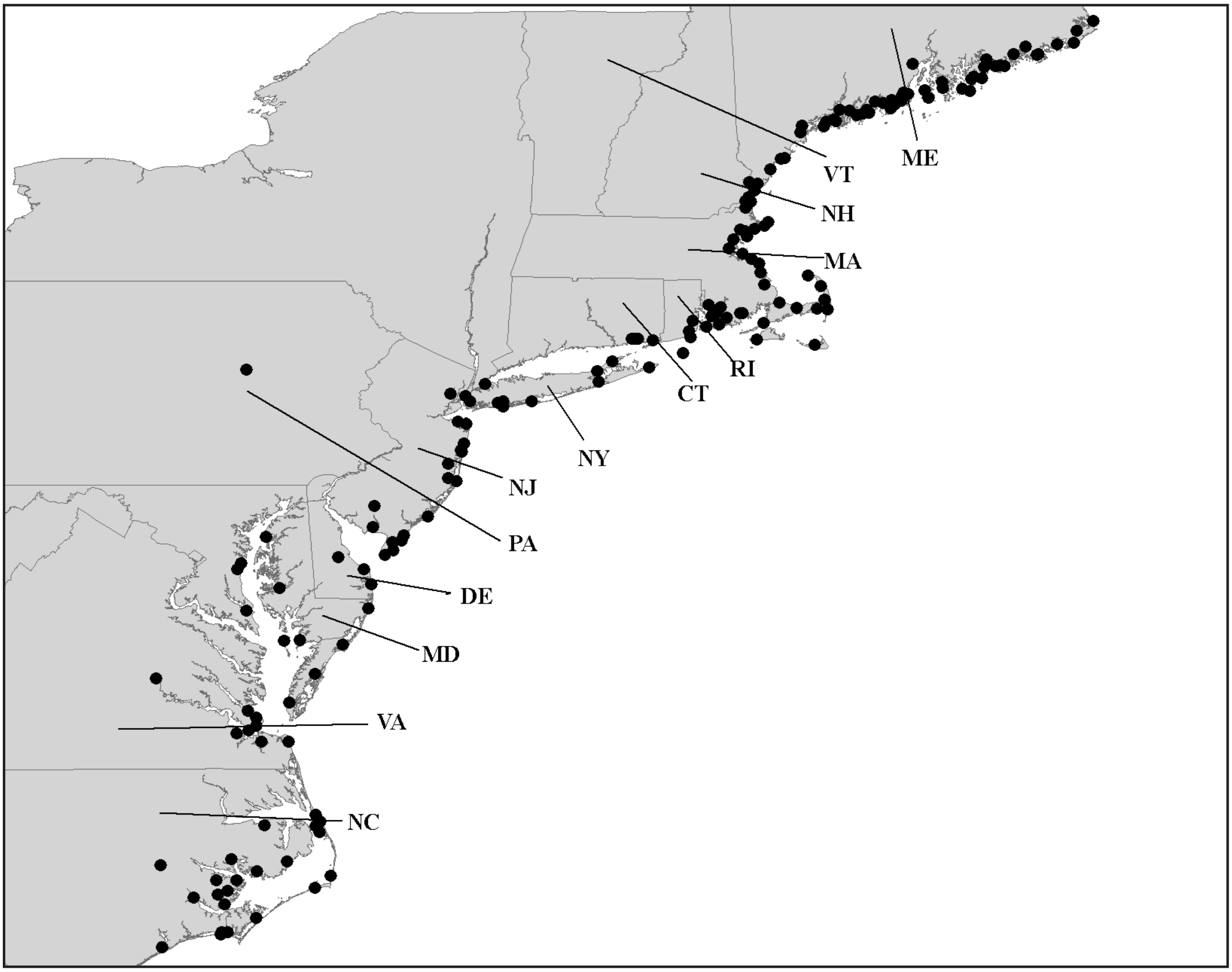


Figure 1.—Communities profiled for the U.S. northeast community profiling project.

Many disciplines use multivariate analyses for the purposes of classification. For example, modern biology uses numerical-based systematics to classify organisms—tools such as multiple discriminant analysis and cluster analysis. These techniques are not foolproof. First, unless all attributes of the "thing" to be classified are used, human decision making is significantly involved in the process. Second, a variety of techniques are used in numerical taxonomy (Sokal and Sneath, 1963), and the method selected can influence the results (Frey and Duek, 2007; Brusco and Kohn, 2008). For this reason, we felt it essential to test our results against several independent data sets, a process we refer to as "ground-truthing."

Methods

Sample

The attributes selected for the numerical taxonomy are derived from the NMFS "Social Science Data Base" (NMFS-SSDB) which includes commercial fisheries and U.S. Census data for 1,835 "ports" from North Carolina to the Canadian border. Those ports selected for community profiling are depicted in Figure 1 to demonstrate the geographic range of communities. By "ports" we mean coastal communities that report commercial fish landings, are the vessel owner port of residence, or homeport for permitted vessels, or are sites of processing, seafood/shellfish dealers, or recreational fishing activity. From the NMFS-SSDB, we selected 43 "fishery" and 25 "social" variables for analysis—a total of 68 variables (Tables 1, 2). The fishery variables selected were drawn from a number of variables characterizing fishing activity over a ten-year period, and included data relevant to quantifying fishing activity, such as landings by species, numbers of vessels, and numbers of vessel owners. The social variables used were those data from the 2000 United States Census that could most accurately reflect changes in port communities that may result from or result in changes in fishing activity, such as the numbers of people employed in fishing related activities, the number of people who are self-employed, median household and per capita income, and other relevant factors.

Data Reduction Techniques

Principal component analysis was selected as the most appropriate technique for accomplishing a reduction in variables because it creates a smaller number of new variables, grouping them into factors based on shared covariance. The 43 fishery and 25 social variables were reduced to fewer variables with the use of principal component analyses. The scree test (Cattell, 1966) was used to determine the number of components, resulting in four components which account for a total of 70.4 percent of the variance in the data set. Components were rotated using the varimax technique. The results of this analysis are found in Table 1. Items loading highest on the first component (large landings, large vessels, sea scallops, *Placopecten magellanicus;* large groundfish, skates, *Raja* spp.; red crab, *Geryon quinquedens*; and monkfish, *Lophius americanus*, decreasing landings) reflect a fishery characterized by large vessels and large, but decreasing, landings of sea scallops, large groundfish, skates, red crab, and monkfish. Items loading highest on the second component (small vessels, many vessels, lobster, *Homarus americanus*; herring, *Clupea harengus*; and many species) indicate a fishery characterized by many small vessels, landing various species including lobster and herring. The third component reflects a fishery characterized by medium-sized vessels with landings composed principally of bluefish, *Pomatomus saltatrix;* tilefish, *Lopholatilus chamaeleonticeps;* butterfish, *Peprilus triacanthus;* mackerel, *Scomber scombrus;* squid, *Loligo pealeii, Illex illecebrosus;* summer flounder, *Paralichthys dentatus;* scup, *Stenotomus chrysops;* and black sea bass, *Centropristis striata*. The final component reflects ports with changing numbers and sizes of vessels.

Table 1.—Principal component analysis of fishery data. Items in boldface type indicate highest loadings on those factors.

Variable	Component 1	2	3	4
Value of scallops, 2003	**0.932**	0.024	0.068	0.201
Landings value for home-ported vessels, 2004	**0.930**	0.196	0.215	0.110
Number of large vessels (>70ft), 2004	**0.932**	0.184	0.219	0.084
Average value of home-ported vessels, 1997–2003	**0.907**	0.243	0.265	0.059
Value of landings at dealer reported port, 2004	**0.867**	0.282	0.187	0.140
Number of large vessels by owner city, 2003	**0.881**	0.220	0.122	0.182
Total gross tonnage for home-ported vessels	**0.852**	0.345	0.326	0.154
Value of large-mesh groundfish, 2003	**0.832**	0.407	0.007	0.023
Value of skates, 2003	**0.821**	0.175	0.220	0.071
Average landed value, 1997–2003	**0.816**	0.290	0.248	0.071
Total gross tonnage for city owner vessels, 2004	**0.789**	0.376	0.185	0.302
Value of red crab, 2003	**0.730**	0.014	–0.042	0.132
Value of monkfish, 2003	**0.668**	0.435	0.216	–0.058
Number of small vessels (<50ft) by owner city, 2003	–0.027	**0.901**	0.054	0.307
Number of small vessels by homeport, 2003	0.041	**0.904**	0.272	0.162
Average number of vessels by owner city, 1997–2003	0.322	**0.843**	0.128	0.282
Number of vessels by owner city, 2004	0.350	**0.825**	0.105	0.346
Average number of home-ported vessels, 1997–2003	0.393	**0.798**	0.381	0.097
Number of home-ported vessels, 2004	0.416	**0.793**	0.344	0.169
Number of active owner city vessels, 2004	0.507	**0.691**	0.193	0.308
Number of federal dealers, 2004	0.487	**0.657**	0.071	0.020
Number of active home-ported vessels, 2004	0.535	**0.646**	0.451	0.145
Average number of dealers, 1997–2003	0.484	**0.688**	0.084	–0.020
Value of lobster, 2003	0.087	**0.575**	0.012	0.091
Value of herring, 2003	0.516	**0.555**	–0.023	–0.106
Number of medium vessels (50–70ft) by owner city, 2003	0.502	**0.525**	0.281	0.282
Species diversity (number of species landed), 2003	0.147	**0.502**	0.452	–0.026
Value of summer flounder, scup, black sea bass, 2003	0.243	0.087	**0.780**	0.002
Value of butterfish, mackerel, squid, 2003	0.193	0.103	**0.710**	–0.080
Value of smallmesh multispecies, 2003	0.440	0.119	**0.683**	0.201
Value of tilefish, 2003	–0.093	0.070	**0.648**	0.431
Number of medium (50–70ft) vessels by home-port 03	0.518	0.489	**0.557**	–0.003
Value of bluefish, 2003	–0.007	0.107	**0.488**	0.147
Difference in HP gross tons from 1997/98 to 2003/04	–0.230	0.021	–0.199	**–0.776**
Difference in city owner gross tons from 1997/98 to 2003/04	–0.306	–0.094	–0.339	**–0.656**
Difference in HP vessels from 1997/98 to 2003/04	–0.130	–0.304	0.097	**–0.641**
Difference in number of city owner vessels from 1997/98 to 2003/04	–0.200	–0.274	0.019	**–0.622**
Value of dogfish, 2003	–0.059	0.398	0.055	0.028
Value of surf clam, ocean quahog, 2003	0.357	0.013	0.116	–0.006
Difference in dealers from 1997/98 to 2003/04	0.144	0.363	0.024	–0.156
Value of other species, 2003	0.091	0.065	0.166	–0.025
Difference in landings values for 1997/98 to 2003/04	**–0.857**	–0.247	–0.052	–0.227
Difference in sum landings for HP vessels 1997/98 to 2003/04	**–0.928**	–0.101	–0.085	–0.242
Percent total variance	**32.5**	**21.1**	**9.7**	**7.1**

Table 2 presents a principal component analysis of a set of variables from the 2000 Census. Variables selected can be seen in Table 2. Once again, the scree test was used to select number of components and components were rotated using the varimax technique. This resulted in three components which explain a total of 52.9 percent of the total variance in the data set.

Component scores representing the position of each port on each component were created for each port. The component scores are the sum of the component coefficients times the sample standardized variables. These coefficients are proportional to the component loadings. Hence, items with high positive loadings contribute more strongly to a positive component score than those with low or negative loadings. Nevertheless, all items contribute (or subtract) from the score; hence, items with moderately high loadings on more than one component (e.g. percent black and percent white in Table 2) will contribute at a moderate level, although differently, to the component scores associated with each of the components. This type of component score provides the best representation of the data.

Cluster Analysis

Cluster analysis was then used to systematically group like communities based on these newly-created component scores. As a means of combining the communities into relevant subgroups to be used for efficiently obtaining data for management decision making, we used K-means cluster analysis (Hartigan and Wong, 1979). The K-means procedure split the fishing communities into a selected number of groups by simultaneously maximizing between group (or cluster) variation and minimizing within group variation. Component scores, which were used as input to the cluster analysis, are standardized, hence providing equal weight for each of the nine components used. Only cases that had no missing data on any of the variables used in the principal component analyses are used in the cluster analysis (*n*=446). This eliminated any ports that did not have associated census data, which occurred when the port name did not correspond to either a geopolitically defined entity or a census designated place, bringing the number of ports used in the analysis from 1,835 down to 446. The procedure first selects the same number of "seeds" as the number of groups desired. The "seeds" selected are as far as possible from the center of all the cases. Then all cases are assigned to the nearest "seed," and cases are reassigned to other clusters, as needed, to reduce within-groups sum of squares.

Number of clusters selected was based on an iterative procedure wherein we started at a relatively low number, examined the output, then increased the number if it was felt that, based on our knowledge of the ports, similar ports were combined. This iterative procedure resulted in a decision to use 40 clusters as the requested number. The results of the analysis are in Appendix I, and an example of selected clusters is provided in Table 3.

The F-ratios across the 40 groups are impressive, but one must remember that they are an artifact of the clustering technique which maximizes these values. Twelve of the clusters contain only one port, as illustrated by Montauk, N.Y., in Table 3. We believe that this is a valid clustering since our knowledge of ports included in these single port clusters suggests that they are unique, and any grouping of them with other

Table 2.—Principal component analysis of Census data. Items in boldface type indicate highest loadings on those factors

	Component		
Variable	1	2	3
Median household income	**–0.793**	0.395	0.018
High school (%)	**–0.766**	0.172	–0.413
High school males (%)	**–0.745**	0.243	–0.359
Poverty rate	**0.735**	–0.209	0.309
High school female (%)	**–0.732**	0.088	–0.444
Unemployed (%)	**0.727**	0.279	0.038
Unemployed males (%)	**0.659**	0.277	0.029
Unemployed females (%)	**0.657**	0.229	0.044
Household income >200K (%)	**–0.624**	0.302	0.104
Share of HH income >200k	**–0.579**	0.296	0.118
Share of HH income retired	**0.526**	–0.291	–0.247
Black (%)	**0.520**	0.121	**0.447**
Males in fishing related job (%)	0.080	**–0.846**	0.002
Fishing related employment (%)	0.054	**–0.845**	0.018
Population in urban area (%)	–0.156	**0.599**	0.271
Females in fishing related job (%)	0.035	**–0.549**	0.001
Tourist housing (%)	0.016	**–0.475**	–0.256
Hispanic (%)	0.216	0.174	**0.766**
Other ethnic group (%)	0.276	0.135	**0.745**
White (%)	**–0.455**	–0.200	**–0.690**
Two or more ethnicities	0.187	0.155	**0.612**
Population	–0.078	–0.095	**0.570**
Aggregate household income	–0.111	–0.091	**0.566**
Asian (%)	–0.230	0.280	**0.451**
Male population (%)	–0.179	–0.186	0.083
Percent of Total Variance	**24.072**	**13.358**	**15.469**

ports would be questionable. Each of these single-port clusters represents a community with either an exceptionally large fishery (e.g. New Bedford, Mass.; Cape May, N.J.), or is a large city and thus the census data factors are very different from the other clusters (e.g. New York, N.Y.; Boston, Mass.). That these ports appear in their own individual clusters indicate that they are unique enough to be studied on their own and should not be grouped with other ports.

Note the distance for Montauk. This is a measure of the distance of a port from the center of all the cases in the cluster, and since there is only one, the distance is zero. In cluster 8, Portsmouth, N.H., is closest to the center of all eight cases in the cluster for all seven component scores. Hence, this distance measure can be used in selecting cases from clusters for more intensive analysis.

For example, one may only desire ports close to the center or want a representative sample from the cluster and select ports across the range of distances. Numbers of ports in each cluster range from 1 to 123. As can be seen in Appendix I, many of the clusters (12) contain only a single case, followed by 7 clusters containing 2–9 cases, 2 clusters containing 22 cases, 3 clusters containing 32–38 cases, and 1 cluster containing 57 cases (not all clusters are shown in Appendix I).

Those clusters plotted in multidimensional space allow us to view similarities and differences on more than one component at a time. Figure 2 illustrates relative positioning of the 12 single-port clusters on one social component (population, percent in fishing related jobs and tourist housing) and two fishery components (component 2: small vessels, landing many species including lobster and herring and component 4: ports with decreasing numbers and sizes of vessels). A high number on fishery component 4 reflects rising numbers and sizes of vessels; hence, the name for the dimension—Rising.

Figure 3 illustrates relative positioning of seven multiport clusters in the same three-dimensional space. In this figure, the number following the name indicates cluster number as indicated in Appendix I. Where there are only a few states involved (MAME32), the states are abbreviated (e.g. MAME32 is cluster 32 which includes six cases from Maine and Massachusetts). MIXED refers to too many states to abbreviate in a brief title. GROUNDT refers to clusters that are "ground-truthed" (see below). You can see the ports included in cluster 8 in Table 3. Ports included in cluster 40 are mainly in Massachusetts with some from Maine, New Hampshire, and Rhode Island.

Plots of clusters, such as those illustrated in Figures 2 and 3, can be rotated to identify groups of communities that cluster in various selected component spaces, such as clusters numbered 8, 32, and 40. Clusters can then be examined by mean scores on all components, as in Figure 4. While communities in these three clusters overlap geographically and are quite similar on most of the fishery and social components, clusters 8 and 40 are on opposite sides of the component mean (zero for a standardized variable) with regard to growth trends. The type of analysis presented here allows one to identify differences between any subset of clusters in the data set, but to illustrate the process we will focus on these two clusters (8 and 40) which are used in further analyses below.

Table 3.—Segment of K–Means cluster analysis output.

Summary statistics for all cases

Variable	Between SS	df	Within SS	df	F–ratio
FAC1FSH9 (fishery component 1)	1796.537	39	21.907	406	853.716
FAC2FSH9 (fishery component 2)	1421.244	39	69.732	406	212.176
FAC3FSH9 (fishery component 3)	1074.305	39	36.679	406	304.911
FAC4FSH9 (fishery component 4)	1491.341	39	100.399	406	154.635
SOCFA1 (social component 1)	281.368	39	119.428	406	24.526
SOCFA2 (social component 2)	760.698	39	88.195	406	89.790
SOCFA3 (social component 3)	435.648	39	86.124	406	52.659
TOTAL	7261.142	273	522.465	2842	

Cluster 7 of 40 contains 1 cases

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NY, Montauk	0.00	FAC1FSH9	–4.46	–4.46	–4.46	–
		FAC2FSH9	2.60	2.60	2.60	–
		FAC3FSH9	22.73	22.73	22.73	–
		FAC4FSH9	18.56	18.56	18.56	–
		SOCFA1	0.91	0.91	0.91	–
		SOCFA2	–0.94	–0.94	–0.94	–
		SOCFA3	0.37	0.37	0.37	–

Cluster 8 of 40 contains 8 cases

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Harwich	0.51	FAC1FSH9	–1.25	–0.64	0.62	0.61
MA, Rockport	0.34	FAC2FSH9	3.03	4.08	5.46	0.85
MA, Plymouth	0.52	FAC3FSH9	–0.90	–0.14	0.82	0.57
MA, Scituate	0.88	FAC4FSH9	–0.47	0.96	1.85	0.74
ME, Kittery	0.43	SOCFA1	–0.90	–0.28	0.38	0.40
NH, Hampton	0.47	SOCFA2	–0.30	0.07	0.31	0.22
NH, Portsmouth	0.29	SOCFA3	–0.92	–0.47	–0.20	0.24
RI, Narragansett	0.58					

Cluster 9 of 40 contains 3 cases

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
ME, Stonington	0.33	FAC1FSH9	–0.85	–0.77	–0.65	0.10
ME, Vinalhaven	0.47	FAC2FSH9	5.41	6.29	6.76	0.76
ME, Jonesport	0.47	FAC3FSH9	–2.82	–2.47	–1.78	0.59
		FAC4FSH9	2.49	2.98	3.46	0.48
		SOCFA1	–0.33	0.33	0.73	0.57
		SOCFA2	–4.95	–4.25	–3.89	0.61
		SOCFA3	–0.10	0.13	0.36	0.23

Testing the Usefulness of the Cluster Analysis

If the cluster analysis actually does group communities which differ on sociocultural and fishery variables, we would expect these differences to be manifest in other aspects of the community which were not measured as part of the original data set. To test this hypothesis we coded a select set of sociocultural variables found in the existing 177 community profiles, which were compiled from a wide range of available data. Eleven variables not used in the cluster analysis were coded, and percent distribution across clusters 8 and 40 can be found in Figure 5. Despite the fact that there are some large differences between clusters 8 and 40, for example, presence of a fishermen's memorial (50% versus 11%, respectively; Fisher's Exact Test $p>0.05$) the small number of communities in each cluster (8 and 9, respectively) necessitates a relatively large difference to achieve statistical significance.

It would be more revealing to examine combinations of the sociocultural variables than individual items. Once again, we used principal component analysis with varimax rotation to develop scales from the profile-derived, sociocultural data set. Number of components was selected on the basis of the scree test. The results of the analysis are in Table 4.

Table 4.—Principal component analysis of cultural and recreational fishing information from profiles

Item	Fishing Culture	Fishing Recreation
Fishermen's festival	0.667	0.258
Blessing of fleet	0.657	–0.001
Fishermen's memorial	0.619	–0.257
Fishermen's assistance	0.597	–0.314
Fishermen's competition	0.553	0.107
Fishermen's association	0.539	0.081
Recreational fishing pier	–0.090	0.718
Fishing tournament	–0.010	0.713
Fishing education	0.361	0.487
Percent variance	26.109	16.777

Table 4 indicates that the two components account for 43% of the variance in the data set. Items loading highest on the first component are related to aspects of a commercial fishing culture, such as presence of a commercial fishermen's memorial, a fishermen's museum, blessing of the commercial fleet, etc. Items loading highest on the

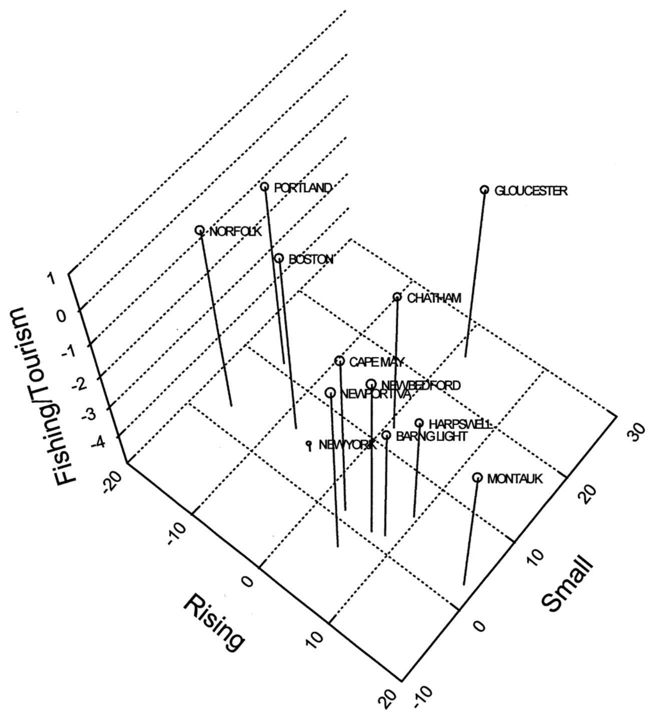


Figure 2.—Plot of single port clusters on one social and two fishery components.

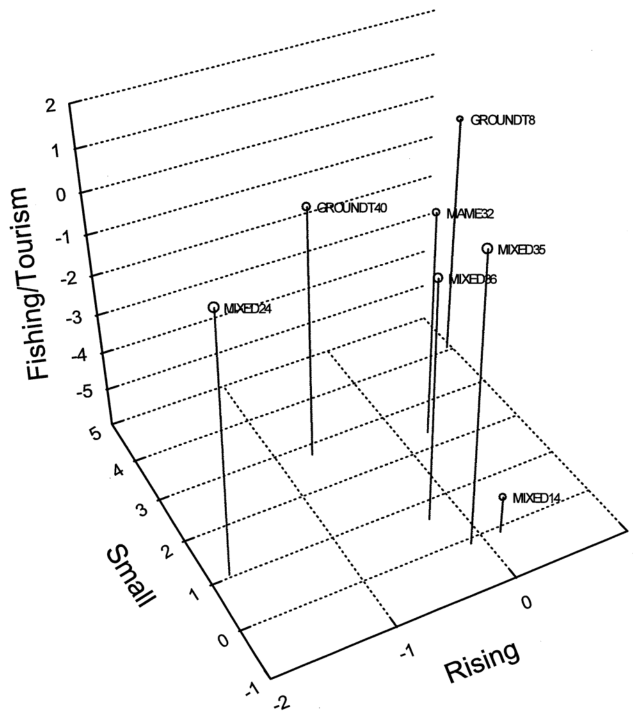


Figure 3.—Plot of multiple port clusters on one social and two fishery components.

second component are more related to recreational fishing, including presence of a recreational fishing tournament and a recreational fishing pier. Presence of fishermen's educational programs loads about the same on both components. Component scores, as described above, were calculated for each port in the profile data set.

Since sample size within clusters 8 and 40 are relatively small for statistical analyses, we decided to cluster the clusters to allow comparison between larger groupings of ports that include both clusters 8 and 40. Data input were mean values for each of the 40 clusters (Appendix I) on the four fishery and three social component scores described above, and a hierarchical cluster analysis using median linkage and Euclidean distances was performed (Appendix II). A segment of the hierarchical tree which will be analyzed further is in Figure 6. All of the clusters found in Figure 6 can be found within cluster 1 of a K-means cluster analysis of the same data set (Appendix III).

We will now compare two clusters depicted in Figure 6 on the two scales developed from the profile data. We will refer to the four clusters represented by MASS/ME32 through MIXED38 depicted at the bottom part of Figure 6 as Group A (n=25), and MIXED1 through MIXED12 as Group B (n=31). Mean scores for Group A and Group B on the Fishing Culture Component are 0.297 and –0.594, respectively (t = 4.393, df = 54, $p<0.001$), and on the Recreational Component they are –0.247 and 0.192, respectively (t = 1.581, df = 54, $p>0.05$). This analysis indicates that the cluster analysis identified clusters that differ on sociocultural variables not included in the initial data set used for the clustering, providing a measure of external validity for the analysis.

A final test of the usefulness of the clusters derived from the K-means cluster analysis was to "ground-truth" the various clusters. In contrast to the preceding analyses, which are based on secondary data (the initial database) and more detailed community profiles, which were also based on secondary data from publications, websites, and telephone inquiries as needed (see the community profiles), the ground-truthing is based on actual visits to the communities and interviews with community members.

The ground-truthing method used the following techniques:

1) A photo-survey that included infrastructure (dock areas, fish processing and marketing facilities), fishing related cultural items (fishermen's memorials, statues), and general snapshots that would provide an overall picture of the ambience of the community;
2) Interviews with key informants concerning infrastructure and other points included in the profiles to provide field validity checks;
3) A brief survey that included the following six questions: 1) If you were to list five things that characterize [community name],

what would they be? 2) Would you say that [community name] is a fishing community (if not included in the response to the first question)? 3) What are three important issues facing [community name] today? 4) Has [community name] changed over the past 5–10 years? How? 5) Would you advise a young person to live in [community name]? Why? 6) If the person interviewed is a fisherman, he or she will be asked "What's it like fishing out of [community name]?"

To provide a rigorous test of the clustering technique we selected clusters 8 and 40 as the first two clusters to be compared. These two clusters overlap geographically and are composed of relatively small ports in Rhode Island, Massachusetts, and New Hampshire (Appendix I). Ground-truthed ports from Cluster 8 are Plymouth, Harwich, and Scituate, Mass., as well as Portsmouth, N.H. (Sample size of surveys (n=89). The ports from Cluster 40 that were ground-truthed are Seabrook, N.H., and Westport, Barnstable, and Marshfield, Mass. (n=81).

When ground-truthing was completed for the eight communities, we noted that communities from Cluster 8 were somehow "nicer." The people in the communities seemed to be friendlier, speaking of their community in a manner that made it seem more cohesive. These qualitative observations are supported by a content analysis of responses made by community members during the ground-truthing exercise. While 11% of those interviewed in Cluster 40 said their communities were "spread out" and "composed of different parts" only 2% of respondents from Cluster 8 made this observation (χ^2 = 5.505, p<0.05).

Additionally, a common issue in coastal communities is that of "gentrification"—a change from being a fishing port to that of a desired residential and recreational location. This was manifested by respondents' complaints concerning the development of "condos," "million dollar homes," and an increase in "yuppies" as well as a "loss of character" in the port. Once again, Clusters 8 and 40 differed with respect to these responses. Forty-six percent of respondents from Cluster 8 voiced these complaints in contrast to only 20% from Cluster 40 (χ^2 = 13.175, p<0.001). These findings provide more external validity to the results of the classification methods used.

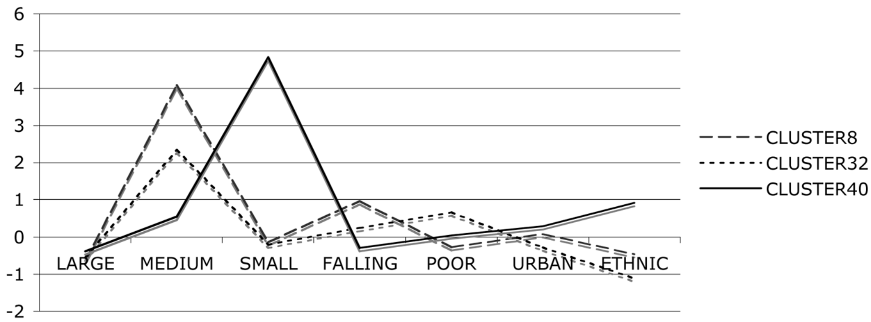


Figure 4.—Mean component values plotted for three similar clusters.

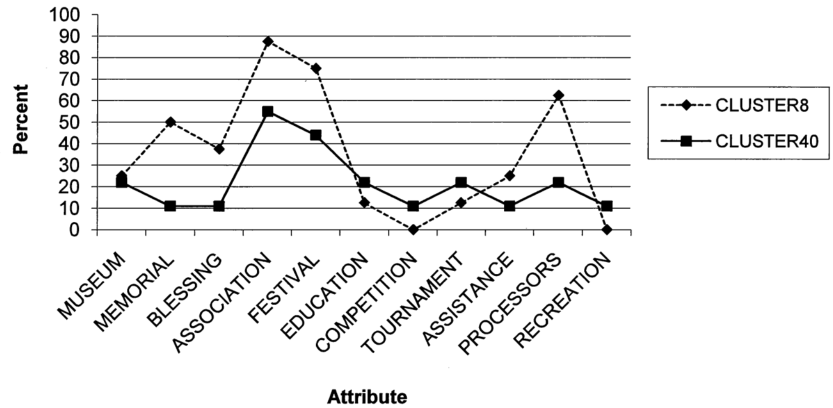


Figure 5.—Selected profile attributes compared across clusters 8 and 40.

Conclusions

In sum, the tests of external validity for the cluster analyses provide support for the claim that the analysis actually did cluster communities into groupings that are different—different on the items used in the initial clustering as well as other variables identified by the analysis of the data from the community profiles and the ground-truthing exercise.

We argue here that this type of classification of coastal communities is a necessary first step in providing representative information to be used in SIA. Community Profiles form an important part of the information used in developing SIA's, and communities to be profiled have thus far been selected on the basis of size and importance of

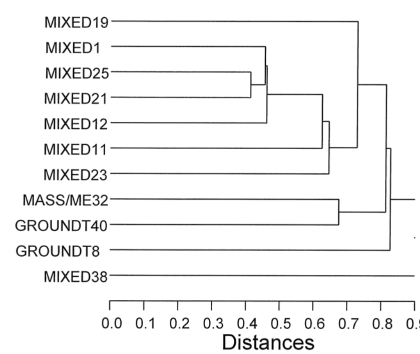


Figure 6.—Segment of hierarchial cluster analysis of 40 clusters from K-means cluster analysis

fishery, types of fishing present, and overall knowledge possessed by experts working in the region. This technique is too unsystematic for such an important endeavor. SIA's detail important implications with regard to the impacts of management on fishermen and the communities in which they live. As noted in the introduction, the lack of a statistically representative range of communities that may be impacted by proposed regulations can result in inadequate SIA's, resulting in undesirable effects on household and community well-being. All of these can exacerbate the types of resistance to fisheries management that are evident in most, if not all, fisheries. Using the methodology described here to first select the communities to be profiled, as a way of improving the sampling process, would result in more representative and useful community profiles and, ultimately, improve SIA's.

The type of classification of coastal communities presented here should be done on a regular basis to reflect the rapid changes that are taking place in our fisheries. One of the principal components of the analysis of the fishery data reflected these changes. If regularly conducted, such analyses would allow those responsible for SIA's to observe the changes in fishing communities in terms of their similarities and differences, determine the factors influencing these changes, and use this information to craft more reliable and timely SIA's related to specific, proposed management measures.

Acknowledgments

The authors would like to thank NOAA for providing support for this project, Azure D. Westwood for her work on the community profiles and ground-truthing exercises, and the three anonymous reviewers for their insightful comments on this manuscript.

Literature Cited

Brusco, M. J., and H. F. Kohn. 2008. Comment on "Clustering by passing messages between data points". Science 319:726

Cattell, R. B. 1966. The scree test for the number of factors. Multivariate Behavioral Res. 1:245–276.

Frey, B. J., and D. Duek. 2007. Clustering by passing messages between data points. Science 315:972–976.

Hartigan, J. A., and M. A. Wong. 1979. A K-means clustering algorithm. Applied Statistics 28:100–108.

Pollnac, R. B., S. Abbott-Jamieson, C. Smith, M. L. Miller, P. M. Clay, and B. Oles. 2006. Toward a model for fisheries social impact assessment. Mar. Fish. Rev. 68(1–4): 1–18.

Sokal, R. R., and P. H. A. Sneath. 1963. Principles of numerical taxonomy. San Francisco: W.H. Freeman & Co., 359 p.

Appendix I.—K-Means Cluster Analysis.Distance metric is Euclidean distance, K-means splitting cases into 40 groups. Data for the following results were selected according to: HBOATS04> 0) AND (SOCMISDA= 0).

Summary statistics for all cases

Variable	Between SS	df	Within SS	df	F-ratio
FAC1FSH9	1796.537	39	21.907	406	853.716
FAC2FSH9	1421.244	39	69.732	406	212.176
FAC3FSH9	1074.305	39	36.679	406	304.911
FAC4FSH9	1491.341	39	100.399	406	154.635
SOCFA1	281.368	39	119.428	406	24.526
SOCFA2	760.698	39	88.195	406	89.790
SOCFA3	435.648	39	86.124	406	52.659
TOTAL	7261.142	273	522.4652	842	

Cluster 1 of 40 contains 57 cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
CT, Greenwich	0.58	FAC1FSH9	–0.27	–0.06	0.09	0.07
CT, Guilford	0.21	FAC2FSH9	–0.31	–0.06	0.80	0.22
CT, Madison	0.17	FAC3FSH9	–0.41	–0.10	0.62	0.14
CT, North Branford	0.22	FAC4FSH9	–0.45	0.15	1.31	0.31
MA, Aquinnah	0.52	SOCFA1	–2.98	–1.17	–0.40	0.57
MA, West Tisbury	0.39	SOCFA2	–0.75	0.38	1.44	0.38
MA, Georgetown	0.20	SOCFA3	–0.67	0.00	1.05	0.42
MA, Manchester	0.41					
MA, Middleton	0.26					
MA, West Newbury	0.39					
MA, Bedford	0.17					
MA, Hopkinton	0.23					
MA, Cohasset	0.39					
MA, Dover	0.68					
MA, Norfolk	0.33					
MA, Norwood	0.26					
MA, Marion	0.46					
MA, Southborough	0.39					
MA, Sutton	0.27					
ME, Yarmouth	0.22					
NC, Ocean Island Beach	0.46					
NH, Hollis	0.25					
NH, Greenland	0.21					
NH, Hampton	0.33					
NH, New Castle	0.38					
NH, Windham	0.16					
NJ, Medford	0.12					
NJ, Avalon	0.37					
NJ, East Brunswick	0.41					
NJ, Sewaren	0.40					
NJ, Manasquan	0.27					
NJ, Monmouth	0.20					
NJ, Rumson	0.56					
NJ, Sea Bright	0.25					
NJ, Wall	0.26					
NJ, Wayne	0.24					
NY, Atlantic Beach	0.11					
NY, East Rockaway	0.23					
NY, Lido Beach	0.20					
NY, Massapequa	0.15					
NY, Seaford	0.26					
NY, Wantagh	0.20					
NY, Babylon	0.20					
NY, East Islip	0.25					
NY, Huntington Bay	0.19					
NY, Islip	0.44					
NY, Mount Sinai	0.16					
NY, Northport	0.17					
NY, Oakdale	0.23					
NY, Port Jefferson	0.20					
NY, Sayville	0.16					
NY, Southampton	0.43					
NY, Stony Brook	0.10					
NY, Armonk	0.60					
NY, Bronxville	0.89					
RI, Barrington	0.19					
RI, East Greenwich	0.30					

continued

Appendix I.—(Continued).

Cluster 3 of 40 contains 1 case

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, New Bedford	0.00	FAC1FSH9	39.65	39.65	39.65	—
		FAC2FSH9	1.12	1.12	1.12	—
		FAC3FSH9	–0.68	–0.68	–0.68	—
		FAC4FSH9	6.91	6.91	6.91	—
		SOCFA1	1.94	1.94	1.94	—
		SOCFA2	0.10	0.10	0.10	—
		SOCFA3	1.50	1.50	1.50	—

Cluster 5 of 40 contains 1 case

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
VA, Norfolk	0.00	FAC1FSH9	2.87	2.87	2.87	—
		FAC2FSH9	2.89	2.89	2.89	—
		FAC3FSH9	3.51	3.51	3.51	—
		FAC4FSH9	–15.71	–15.71	–15.71	—
		SOCFA1	1.12	1.12	1.12	—
		SOCFA2	0.41	0.41	0.41	—
		SOCFA3	1.10	1.10	1.10	—

Cluster 6 of 40 contains 1 case

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NJ, Barnegat Light	0.00	FAC1FSH9	1.27	1.27	1.27	—
		FAC2FSH9	1.67	1.67	1.67	—
		FAC3FSH9	6.20	6.20	6.20	—
		FAC4FSH9	8.33	8.33	8.33	—
		SOCFA1	–0.58	–0.58	–0.58	—
		SOCFA2	–1.31	–1.31	–1.31	—
		SOCFA3	–0.74	–0.74	–0.74	—

Cluster 7 of 40 contains 1 case

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NY, Montauk	0.00	FAC1FSH9	–4.46	–4.46	–4.46	—
		FAC2FSH9	2.60	2.60	2.60	—
		FAC3FSH9	22.73	22.73	22.73	—
		FAC4FSH9	18.56	18.56	18.56	—
		SOCFA1	0.91	0.91	0.91	—
		SOCFA2	–0.94	–0.94	–0.94	—
		SOCFA3	0.37	0.37	0.37	—

Cluster 8 of 40 contains 8 cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Harwich	0.51	FAC1FSH9	–1.25	–0.64	0.62	0.61
MA, Rockport	0.34	FAC2FSH9	3.03	4.08	5.46	0.85
MA, Plymouth	0.52	FAC3FSH9	–0.90	–0.14	0.82	0.57
MA, Scituate	0.88	FAC4FSH9	–0.47	0.96	1.85	0.74
ME, Kittery	0.43	SOCFA1	–0.90	–0.28	0.38	0.40
NH, Hampton	0.47	SOCFA2	–0.30	0.07	0.31	0.22
NH, Portsmouth	0.29	SOCFA3	–0.92	–0.47	–0.20	0.24
RI, Narragansett	0.58					

continued

Appendix I.—(Continued).

Cluster 10 of 40 contains 1 case

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NY, New York	0.00	FAC1FSH9	0.61	0.61	0.61	—
		FAC2FSH9	4.31	4.31	4.31	—
		FAC3FSH9	–0.83	–0.83	–0.83	—
		FAC4FSH9	–4.77	–4.77	–4.77	—
		SOCFA1	–4.16	–4.16	–4.16	—
		SOCFA2	–4.73	–4.73	–4.73	—
		SOCFA3	15.44	15.44	15.44	—

Cluster 11 of 40 contains 32 cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
DE, Frederica	0.29	FAC1FSH9	–0.13	0.01	0.26	0.11
DE, Milford	0.61	FAC2FSH9	–0.28	–0.04	0.81	0.27
DE, Millsboro	0.31	FAC3FSH9	–0.37	–0.05	0.50	0.18
MA, Onset	0.34	FAC4FSH9	–1.14	0.06	1.11	0.46
MD, Cambridge	0.24	SOCFA1	0.76	1.57	3.78	0.68
MD, Crisfield	0.54	SOCFA2	–1.82	–0.10	0.74	0.63
MD, Willards	0.38	SOCFA3	–1.08	0.18	1.05	0.49
MD, Berlin	0.42					
MD, Snow Hill	0.13					
ME, Eastport	0.58					
NC, Aurora	0.38					
NC, Belhaven	0.61					
NC, Gloucester	0.44					
NC, Marshallberg	0.56					
NC, Morehead City	0.26					
NC, Newport	0.32					
NC, Swan Quarter	0.75					
NC, Wilmington Beach	0.36					
NC, Bayboro	0.48					
NC, Vandemere	0.45					
NJ, Millville	0.32					
NJ, Keansburg	0.27					
NJ, Neptune City	0.44					
NY, Mastic Beach	0.42					
RI, East Providence	0.31					
RI, Woonsocket	0.34					
VA, Melfa	0.42					
VA, Onancock	0.45					
VA, Hallwood	0.61					
VA, Exmore	0.20					
VA, Nassawadox	0.85					
VA, Portsmouth	0.25					

continued

Appendix I.—(Continued).

Cluster 12 of 40 contains 38 cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
DE, Lewes	0.34	FAC1FSH9	–0.61	–0.24	0.55	0.24
MA, Brewster	0.34	FAC2FSH9	–0.29	0.66	1.58	0.50
MA, Dennis	0.39	FAC3FSH9	–1.16	–0.29	0.39	0.33
MA, Eastham	0.31	FAC4FSH9	0.20	0.97	2.31	0.49
MA, South Dennis	0.23	SOCFA1	–1.17	–0.21	0.64	0.44
MA, Yarmouth	0.42	SOCFA2	–0.93	–0.09	0.82	0.41
MA, Vineyard Haven	0.32	SOCFA3	–1.13	–0.53	0.04	0.33
MA, Essex	0.25					
MA, Newburyport	0.71					
MA, Salisbury	0.40					
MA, Swampscott	0.38					
MA, Nantucket	0.38					
MA, Kingston	0.24					
MA, Middleboro	0.26					
MA, Ocean Bluff	0.33					
ME, Falmouth	0.37					
ME, Scarborough	0.47					
ME, South Portland	0.37					
ME, Hancock	0.43					
ME, Buxton	0.34					
ME, Kittery	0.34					
ME, Ogunquit	0.46					
ME, Saco	0.32					
ME, Wells	0.34					
NH, Newington	0.62					
NH, Dover	0.22					
NJ, Middletown	0.53					
NJ, Beach Haven	0.35					
NJ, Forked River	0.24					
NJ, Manahawkin	0.51					
NJ, Point Pleasant	0.31					
NJ, Toms River	0.26					
NJ, Tuckerton	0.41					
NJ, Waretown	0.37					
NY, Oceanside	0.64					
RI, Charlestown	0.43					
VA, Wachapreague	0.40					
VA, Poquoson	0.35					

Cluster 14 of 40 contains 9 cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Gosnold	0.70	FAC1FSH9	–0.16	–0.07	–0.01	0.05
MD, Smith Island	0.49	FAC2FSH9	–0.26	–0.09	0.20	0.14
ME, Cranberry Isles	0.43	FAC3FSH9	–0.27	–0.09	–0.00	0.08
ME, Matinicus	0.43	FAC4FSH9	–0.14	0.15	0.68	0.25
ME, North Haven	0.51	SOCFA1	–1.05	0.09	1.89	0.96
ME, Roque Bluffs	0.64	SOCFA2	–6.51	–5.05	–3.37	1.06
NC, Smyrna	0.43	SOCFA3	–0.04	0.62	1.84	0.63
VA, Saxis	0.82					
VA, Tangier	0.51					

Cluster 15 of 40 contains 1 case

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Gloucester	0.00	FAC1FSH9	1.53	1.53	1.53	—
		FAC2FSH9	25.40	25.40	25.40	—
		FAC3FSH9	–2.45	–2.45	–2.45	—
		FAC4FSH9	1.57	1.57	1.57	—
		SOCFA1	–0.03	–0.03	–0.03	—
		SOCFA2	–0.01	–0.01	–0.01	—
		SOCFA3	–0.30	–0.30	–0.30	—

continued

Appendix I.—(Continued).

Cluster 16 of 40 contains 1 case

Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NJ, Cape May	0.00	FAC1FSH9	8.09	8.09	8.09	—
		FAC2FSH9	1.28	1.28	1.28	—
		FAC3FSH9	6.75	6.75	6.75	—
		FAC4FSH9	3.01	3.01	3.01	—
		SOCFA1	0.40	0.40	0.40	—
		SOCFA2	0.07	0.07	0.07	—
		SOCFA3	–0.44	–0.44	–0.44	—

Cluster 17 of 40 contains 1 case

Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Chatham	0.00	FAC1FSH9	–2.58	–2.58	–2.58	—
		FAC2FSH9	12.65	12.65	12.65	—
		FAC3FSH9	1.28	1.28	1.28	—
		FAC4FSH9	0.82	0.82	0.82	—
		SOCFA1	–0.17	–0.17	–0.17	—
		SOCFA2	–0.73	–0.73	–0.73	—
		SOCFA3	–0.37	–0.37	–0.37	—

Cluster 18 of 40 contains 1 case

Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
ME, Portland	0.00	FAC1FSH9	5.16	5.16	5.16	—
		FAC2FSH9	11.07	11.07	11.07	—
		FAC3FSH9	–0.70	–0.70	–0.70	—
		FAC4FSH9	–14.62	–14.62	–14.62	—
		SOCFA1	0.09	0.09	0.09	—
		SOCFA2	0.27	0.27	0.27	—
		SOCFA3	–0.12	–0.12	–0.12	—

Cluster 19 of 40 contains 22 cases

Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
CT, Bridgeport	0.66	FAC1FSH9	–0.15	0.01	0.36	0.11
CT, Norwalk	0.40	FAC2FSH9	–0.27	–0.02	0.48	0.20
CT, Stamford	0.45	FAC3FSH9	–0.23	–0.02	0.30	0.15
CT, New Haven	0.62	FAC4FSH9	–1.18	–0.31	0.33	0.44
DE, Wilmington	0.60	SOCFA1	–0.48	0.67	2.13	0.80
MA, Lynn	0.34	SOCFA2	0.04	0.52	1.19	0.26
MA, Framingham	0.49	SOCFA3	0.73	1.72	3.67	0.81
MA, Randolph	0.47					
MA, Revere	0.49					
MA, Worcester	0.17					
NJ, Ventnor City	0.30					
NJ, Jersey City	0.77					
NJ, Long Branch	0.24					
NJ, Clifton	0.23					
NY, Baldwin	0.44					
NY, Glen Cove	0.27					
NY, Inwood	0.36					
NY, Staten Island	0.39					
NY, Bay Shore	0.26					
PA, Philadelphia	0.60					
RI, Providence	0.73					
VA, Richmond	0.47					

Cluster 20 of 40 contains 1 case

Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
ME, Harpswell	0.00	FAC1FSH9	–1.75	–1.75	–1.75	—
		FAC2FSH9	5.45	5.45	5.45	—
		FAC3FSH9	–2.57	–2.57	–2.57	—
		FAC4FSH9	9.44	9.44	9.44	—
		SOCFA1	–0.57	–0.57	–0.57	—
		SOCFA2	–1.58	–1.58	–1.58	—
		SOCFA3	–0.21	–0.21	–0.21	—

continued

Appendix I.—(Continued).

Cluster 23 of 40 contains 22 cases

Members			Statistics				
Case	Distance		Variable	Minimum	Mean	Maximum	St.Dev.
DE, Bowers	0.15		FAC1FSH9	–0.33	–0.10	0.04	0.11
MA, Chilmark	0.58		FAC2FSH9	–0.28	0.13	1.38	0.48
ME, Brookline	0.17		FAC3FSH9	–0.43	–0.09	0.09	0.11
ME, Brooksville	0.18		FAC4FSH9	–0.36	0.06	1.07	0.37
ME, Castine	0.35		SOCFA1	–0.79	–0.05	0.87	0.46
ME, Franklin	0.39		SOCFA2	–2.75	–1.70	–0.87	0.56
ME, Sorrento	0.37		SOCFA3	–1.07	–0.53	0.11	0.30
ME, Sullivan	0.28						
ME, Tremont	0.44						
ME, Isle au Haut	0.24						
ME, St. George	0.34						
ME, Bremen	0.49						
ME, Bristol	0.42						
ME, Southport	0.33						
ME, Georgetown	0.45						
ME, Columbia	0.36						
ME, Jonesboro	0.17						
NC, Harkers Island	0.52						
NC, Ocracoke	0.22						
NC, Sneads Ferry	0.49						
NY, Orient	0.42						
VA, Onley	0.36						

Cluster 25 of 40 contains 35 cases

Members			Statistics				
Case	Distance		Variable	Minimum	Mean	Maximum	St.Dev.
CT, Branford	0.26		FAC1FSH9	–0.47	–0.13	0.14	0.14
CT, East Lyme	0.29		FAC2FSH9	–0.15	0.55	1.45	0.35
CT, Groton	0.45		FAC3FSH9	–0.47	0.07	0.66	0.29
CT, Mystic	0.25		FAC4FSH9	–1.58	–0.56	0.18	0.43
CT, Noank	0.27		SOCFA1	–1.75	–0.47	0.33	0.49
MA, Danvers	0.34		SOCFA2	–0.28	0.28	0.82	0.27
MA, Ipswich	0.27		SOCFA3	–0.67	–0.18	0.70	0.38
MA, Methuen	0.42						
MA, Nahant	0.31						
MA, Salem	0.44						
MA, Saugus	0.16						
MA, Quincy	0.36						
MA, Weymouth	0.21						
MA, Duxbury	0.59						
MA, Hingham	0.46						
MA, Hull	0.40						
MA, Pembroke	0.17						
ME, Cape Elizabeth	0.35						
ME, Bath	0.34						
ME, Eliot	0.28						
ME, Kennebunk	0.25						
ME, York	0.41						
ME, York Harbor	0.20						
NJ, Atlantic City	0.36						
NJ, Belmar	0.26						
NJ, Brielle	0.25						
NY, Island Park	0.49						
NY, Point Lookout	0.46						
NY, East Hampton	0.46						
NY, East Quogue	0.33						
NY, West Islip	0.28						
RI, Warwick	0.24						
RI, Jamestown	0.26						
RI, Cranston	0.40						
RI, Westerly	0.35						

continued

Appendix I.—(Continued).

Cluster 28 of 40 contains 1 case

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Boston	0.00	FAC1FSH9	1.84	1.84	1.84	—
		FAC2FSH9	5.47	5.47	5.47	—
		FAC3FSH9	0.25	0.25	0.25	—
		FAC4FSH9	–7.88	–7.88	–7.88	—
		SOCFA1	0.32	0.32	0.32	—
		SOCFA2	0.35	0.35	0.35	—
		SOCFA3	2.75	2.75	2.75	—

Cluster 31 of 40 contains 1 case

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
VA, Newport	0.00	FAC1FSH9	5.30	5.30	5.30	—
		FAC2FSH9	–2.58	–2.58	–2.58	—
		FAC3FSH9	1.52	1.52	1.52	—
		FAC4FSH9	5.13	5.13	5.13	—
		SOCFA1	0.61	0.61	0.61	—
		SOCFA2	0.34	0.34	0.34	—
		SOCFA3	0.83	0.83	0.83	—

Cluster 32 of 40 contains 6 cases

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Orleans	0.52	FAC1FSH9	–1.09	–0.55	–0.18	0.32
MA, Truro	0.65	FAC2FSH9	1.59	2.35	3.01	0.55
MA, Wellfleet	0.53	FAC3FSH9	–0.70	–0.21	0.29	0.41
ME, Bar Harbor	0.32	FAC4FSH9	–0.71	0.24	1.27	0.71
ME, Southwest Harbor	0.42	SOCFA1	–0.24	0.66	1.49	0.62
ME, Boothbay Harbor	0.44	SOCFA2	–1.02	–0.29	0.48	0.65
		SOCFA3	–1.64	–1.11	–0.68	0.38

Cluster 35 of 40 contains 8 cases

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
DE, Leipsic	0.26	FAC1FSH9	–0.11	–0.04	0.00	0.04
MA, Buzzards Bay	0.33	FAC2FSH9	–0.27	–0.09	0.16	0.15
ME, Gorham	0.29	FAC3FSH9	–0.22	–0.07	0.17	0.11
ME, Machias	0.43	FAC4FSH9	–0.77	–0.11	0.16	0.30
NC, Elizabeth City	0.56	SOCFA1	1.27	2.00	2.96	0.52
NH, Durham	0.12	SOCFA2	0.18	1.21	2.84	0.76
NJ, Wildwood	0.21	SOCFA3	–1.48	–1.13	–0.01	0.50
RI, Kingston	0.64					

Cluster 36 of 40 contains 6 cases

Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
CT, Stonington	0.68	FAC1FSH9	–0.38	–0.01	0.87	0.51
MA, Falmouth	0.40	FAC2FSH9	0.26	0.65	1.16	0.35
NJ, Sea Isle City	0.50	FAC3FSH9	0.83	1.48	2.46	0.72
NY, Mattituck	0.35	FAC4FSH9	–1.41	–0.25	0.74	0.79
RI, Little Compton	0.63	SOCFA1	–0.69	0.11	0.72	0.57
RI, Tiverton	0.40	SOCFA2	–0.71	0.06	0.81	0.52
		SOCFA3	–1.06	–0.77	–0.56	0.22

continued

Appendix I.—(Continued).

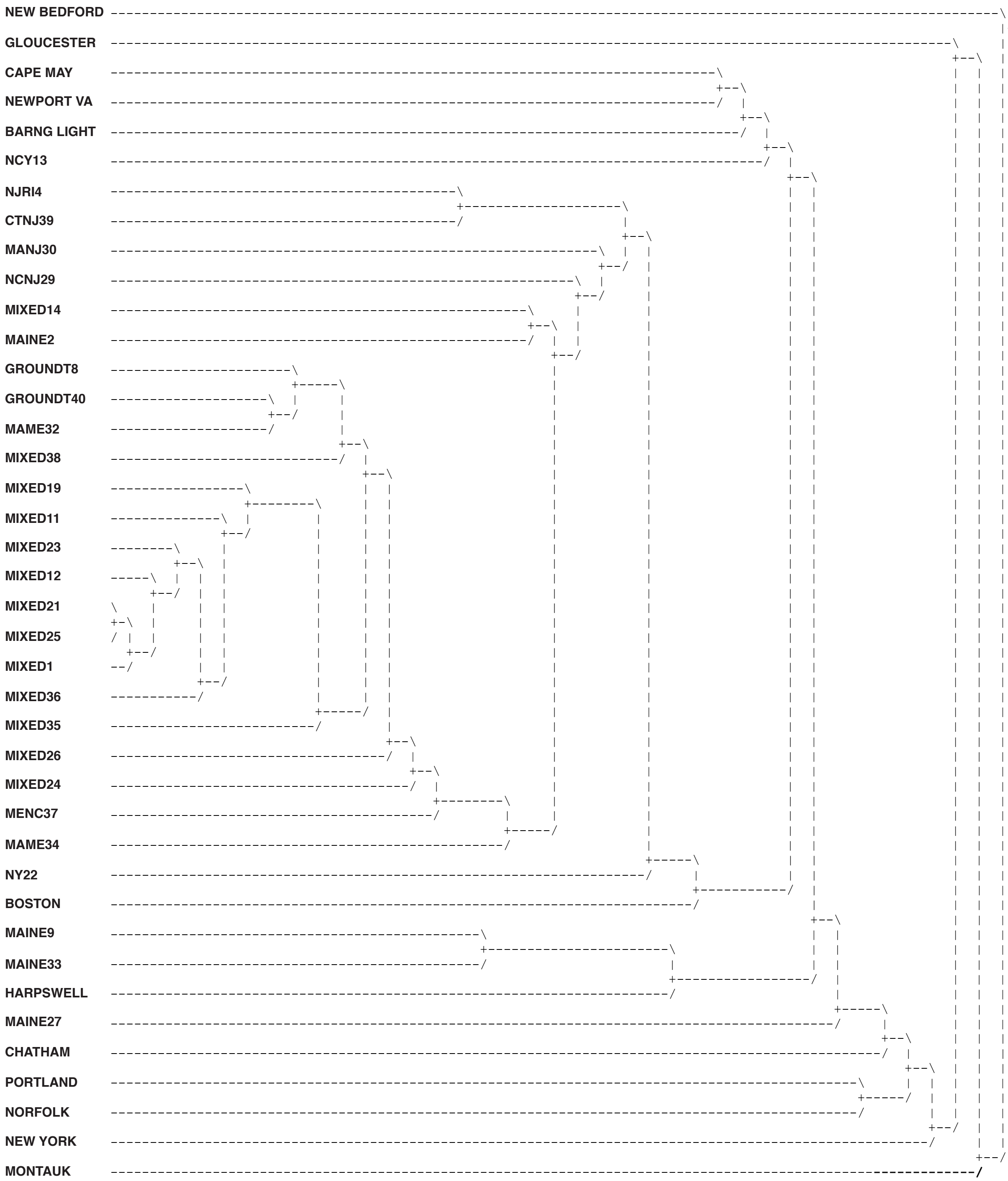
Cluster 38 of 40 contains 4 cases							
Members		Statistics					
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Sandwich	0.48	FAC1FSH9	–0.63	–0.29	0.33	0.42
NC, Beaufort	0.81	FAC2FSH9	0.56	1.84	2.94	1.00
VA, Chincote	0.50	FAC3FSH9	0.85	1.62	2.35	0.61
VA, Virginia	0.72	FAC4FSH9	0.68	1.20	2.41	0.82
		SOCFA1	–0.51	0.33	1.04	0.81
		SOCFA2	–1.20	–0.28	0.14	0.62
		SOCFA3	–0.66	–0.05	1.11	0.79

Cluster 40 of 40 contains 9 cases						
Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MA, Barnstable	0.29	FAC1FSH9	–0.84	–0.56	–0.40	0.14
MA, Westport	0.45	FAC2FSH9	2.05	2.78	3.44	0.48
MA, Beverly	0.32	FAC3FSH9	–0.01	0.34	0.96	0.37
MA, Marblehead	0.52	FAC4FSH9	–1.52	–0.71	0.17	0.62
MA, Newburyport	0.35	SOCFA1	–1.40	–0.34	0.24	0.49
MA, Marshfield	0.37	SOCFA2	–0.26	0.32	0.76	0.37
ME, Kennebunkport	0.46	SOCFA3	–1.08	–0.46	–0.19	0.27
NH, Rye	0.35					
NH, Seabrook	0.36					

Appendix II.—Hierarchial Cluster Analysis of K-means 40 Clusters. Distance metric is Euclidean distance. Median linkage method. Single port clusters use port name rather than cluster number.

Cluster containing	and	Cluster containing	Were joined at distance	No. of members in new cluster
MIXED25		MIXED21	0.411	2
MIXED1		MIXED25	0.378	3
MIXED1		MIXED12	0.437	4
MIXED1		MIXED23	0.599	5
MIXED36		MIXED1	0.624	6
MIXED36		MIXED11	0.631	7
MIXED36		MIXED19	0.639	8
GROUNDT40		MAME32	0.674	2
GROUNDT8		GROUNDT40	0.665	3
MIXED35		MIXED36	0.773	9
GROUNDT8		MIXED38	0.797	4
GROUNDT8		MIXED35	0.779	13
GROUNDT8		MIXED26	0.850	14
GROUNDT8		MIXED24	0.773	15
GROUNDT8		MENC37	0.945	16
CTNJ39		NJRI4	1.023	2
MAINE33		MAINE9	1.196	2
GROUNDT8		MAME34	1.256	17
MIXED14		MAINE2	1.353	2
MIXED14		GROUNDT8	1.294	19
MIXED14		NCNJ29	1.482	20
MIXED14		MANJ30	1.516	21
MIXED14		CTNJ39	1.772	23
NY22		MIXED14	1.593	24
HARPSWELL		MAINE33	2.329	3
NY22		BOSTON	2.351	25
NEWPORT VA		CAPE MAY	2.835	2
BARNG LIGHT		NEWPORT VA	2.578	3
NCNY13		BARNG LIGHT	2.541	4
NCNY13		NY22	2.918	29
NCNY13		HARPSWELL	2.660	32
NCNY13		MAINE27	2.725	33
PORTLAND		NORFOLK	3.657	2
NCNY13		CHATHAM	3.703	34
NCNY13		PORTLAND	4.775	36
NEW YORK		NCNY13	4.930	37
NEW YORK		GLOUCESTER	7.041	38
NEW YORK		MONTAUK	11.146	39
NEW BEDFORD		NEW YORK	13.814	40

Appendix II.—(Continued).



Appendix III.—K-means clustering of K-means 40 clusters. K-means splitting 40 cases into 10 groups (single port clusters use port name rather than cluster number).

Summary statistics for all cases

Variable	Between SS	df	Within SS	df	F-ratio
LARGE	1590.686	9	86.966	30	60.970
SMALL	784.370	9	88.921	30	29.403
MEDIUM	674.282	9	61.448	30	36.577
RISING	1039.884	9	178.613	30	19.407
POVERTY	28.208	9	19.093	30	4.925
URBAN	141.997	9	40.185	30	11.779
ETHNIC	232.583	9	36.724	30	21.111
TOTAL	4492.010	63	511.949	210	

Cluster 1 of 10 contains 21 cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MIXED1	0.80	LARGE	–0.64	0.13	3.30	0.88
MAINE2	1.34	SMALL	–1.33	1.09	4.13	1.42
GROUNDT8	1.30	MEDIUM	–0.99	0.20	2.11	0.78
MIXED11	0.62	RISING	–2.92	–0.19	3.78	1.56
MIXED12	0.62	POVERTY	–1.17	0.44	2.06	0.84
MIXED19	0.81	URBAN	–2.74	–0.19	1.21	0.96
MIXED21	0.52	ETHNIC	–1.13	0.06	4.37	1.22
NY22	2.03					
MIXED23	0.75					
MIXED24	1.14					
MIXED25	0.48					
MIXED26	0.91					
NCNJ29	1.58					
MANJ30	1.77					
MAME32	0.74					
MAME34	1.66					
MIXED35	1.02					
MIXED36	0.62					
MENC37	0.99					
MIXED38	0.82					
GROUNDT40	0.83					

Cluster 2 of 10 contains 3 cases

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NORFOLK	2.02	LARGE	1.84	3.29	5.16	1.70
PORTLAND	2.18	SMALL	2.89	6.48	11.07	4.18
BOSTON	2.06	MEDIUM	–0.70	1.02	3.51	2.21
		RISING	–15.71	–12.74	–7.88	4.24
		POVERTY	0.09	0.51	1.12	0.54
		URBAN	0.27	0.34	0.41	0.07
		ETHNIC	–0.12	1.24	2.75	1.44

Cluster 3 of 10 contains 1 case

Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NEW BEDFORD	0.00	LARGE	39.65	39.65	39.65	—
		SMALL	1.12	1.12	1.12	—
		MEDIUM	–0.68	–0.68	–0.68	—
		RISING	6.91	6.91	6.91	—
		POVERTY	1.94	1.94	1.94	—
		URBAN	0.10	0.10	0.10	—
		ETHNIC	1.50	1.50	1.50	—

continued

Appendix III.—(Continued).

Cluster 4 of 10 contains 6 cases							
Members		Statistics					
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NJRI4	1.97	LARGE	–1.16	2.31	8.09	3.61
BARNG LIGHT	2.29	SMALL	–2.58	0.42	1.67	1.54
NCNY13	2.11	MEDIUM	1.52	5.88	10.15	2.80
CAPE MAY	2.25	RISING	–2.18	2.69	8.33	3.76
NEWPORT VA	2.51	POVERTY	–0.58	0.07	0.61	0.41
CTNJ39	1.61	URBAN	–1.31	–0.21	0.34	0.74
		ETHNIC	–0.74	0.08	0.91	0.67

Cluster 5 of 10 contains 1 case						
Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
CHATHAM	0.00	LARGE	–2.58	–2.58	–2.58	—
		SMALL	12.65	12.65	12.65	—
		MEDIUM	1.28	1.28	1.28	—
		RISING	0.82	0.82	0.82	—
		POVERTY	–0.17	–0.17	–0.17	—
		URBAN	–0.73	–0.73	–0.73	—
		ETHNIC	–0.37	–0.37	–0.37	—

Cluster 6 of 10 contains 1 case						
Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MONTAUK	0.00	LARGE	–4.46	–4.46	–4.46	—
		SMALL	2.60	2.60	2.60	—
		MEDIUM	22.73	22.73	22.73	—
		RISING	18.56	18.56	18.56	—
		POVERTY	0.91	0.91	0.91	—
		URBAN	–0.94	–0.94	–0.94	—
		ETHNIC	0.37	0.37	0.37	—

Cluster 7 of 10 contains 1 case						
Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
NEW YORK	0.00	LARGE	0.61	0.61	0.61	—
		SMALL	4.31	4.31	4.31	—
		MEDIUM	–0.83	–0.83	–0.83	—
		RISING	–4.77	–4.77	–4.77	—
		POVERTY	–4.16	–4.16	–4.16	—
		URBAN	–4.73	–4.73	–4.73	—
		ETHNIC	15.44	15.44	15.44	—

Cluster 8 of 10 contains 3 cases						
Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
MAINE9	1.11	LARGE	–1.75	–1.36	–0.77	0.52
HARPSWELL	1.71	SMALL	4.50	5.41	6.29	0.90
MAINE33	0.97	MEDIUM	–2.57	–2.26	–1.73	0.46
		RISING	2.98	5.64	9.44	3.38
		POVERTY	–0.57	–0.17	0.33	0.46
		URBAN	–5.89	–3.91	–1.58	2.18
		ETHNIC	–0.21	0.19	0.64	0.43

continued

Appendix III.—(Continued).

Cluster 9 of 10 contains 2 cases							
Members		Statistics					
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.	
MIXED14	0.97	LARGE	–0.07	–0.05	–0.03	0.03	
MAINE27	0.97	SMALL	–0.09	0.02	0.14	0.16	
		MEDIUM	–0.09	–0.04	0.02	0.08	
		RISING	–0.37	–0.11	0.15	0.37	
		POVERTY	–2.35	–1.13	0.09	1.73	
		URBAN	–9.43	–7.24	–5.05	3.10	
		ETHNIC	0.62	1.04	1.45	0.59	

Cluster 10 of 10 contains 1 case						
Members		Statistics				
Case	Distance	Variable	Minimum	Mean	Maximum	St.Dev.
GLOUCESTER	0.00	LARGE	1.53	1.53	1.53	—
		SMALL	25.40	25.40	25.40	—
		MEDIUM	–2.45	–2.45	–2.45	—
		RISING	1.57	1.57	1.57	—
		POVERTY	–0.03	–0.03	–0.03	—
		URBAN	–0.01	–0.01	–0.01	—
		ETHNIC	–0.30	–0.30	–0.30	—