San Diego
ASSOCIATION OF GOVERNMENTS

# Understanding Cluster Analysis

**This section provides an overview of the San Diego Association of Government's methodology for defining and analyzing industrial clusters.**

Cluster Identification and Definition Methodology

Clusters are groups of inter-related industries that drive wealth creation in a region, primarily through export of goods and services. The use of clusters as a descriptive tool for regional economic relationships provides a richer, more meaningful representation of local industry drivers and regional dynamics than do traditional methods. The San Diego Association of Governments (SANDAG) and the San Diego Regional Technology Alliance (SDRTA) believe that in order for the region's policy makers to commit to using cluster-based data for planning and decision-making purposes, there must be an understandable, acceptable and replicable process or system in place. Thus, SANDAG has created a process that rigorously examines the composition of the driving clusters in the region.

Although there are neither official guidelines nor standardized approaches to cluster identification and definition, there are some accepted procedures. The methods that SANDAG has utilized are similar to those presently being employed in other regions, although SANDAG has placed more emphasis on creating a replicable, methodological approach to the cluster definition than many of the other areas.

SANDAG's approach builds on the regional clusters produced by Collaborative Economics, a private contractor hired by the SDRTA in 1995. SANDAG's process is intended to set a standard for identifying and defining employment clusters, and to use the information for economic development planning and analysis. Three technical factors are important in determining which industries constitute a cluster. The three factors are based on characteristics common to all clusters: interdependent; export-oriented; wealth generating. More specifically, the three factors are the Employment Concentration Factor (ECF), the Cluster Dependency Factor (CDF), and the Economic Prosperity Factor (EPF). The factors provide the quantifiable information necessary to standardize the identification of clusters and their components.

Employment Concentration Factor

Ideally, when attempting to identify regional clusters one would like to know which industries export goods and services out of the region and bring wealth back into it. Unfortunately, many regions lack accurate data about shipments out of the area. In order to overcome this hurdle, a proxy calculation is necessary. The standard approach is to use a "location quotient" which identifies the industries that employ more workers in the region than the national average for that same industry. The theory is that by employing more workers than the national average the industry is producing more goods and services than the region alone can consume; thus, the industries export the excess product out of the region. There are a few different titles for location quotients, SANDAG refers to the location quotient as the Employment Concentration Factor (ECF).

Employment Concentration Factors are determined by calculating the percentage of employment in a four-digit SIC code industry within a specified region to total regional employment. This ratio is then compared to the percentage of total national employment in that same industry divided by total national employment. If an industry's ECF is greater than 1.0, the national average, it can be assumed that some portion of its production is exported out of the region. For example, an Employment Concentration Factor of 3.0 would mean that employment in this particular industry is three times more concentrated in the region than for the nation as a whole. Essentially, ECFs are a type of export-strength measurement.

*Emerging Clusters and Employment Concentration Factors*

Quantitative calculation of economic concentrations has its limitations and may sometimes fail to identify certain types of clusters, like small and emerging clusters. Emerging cluster industries, such as environmental technology, may have a relatively lower current economic concentration when compared to the nation but have the potential to become more economically significant in the future. These are younger industries in their early stages of development. Emerging clusters may also be branches from older,

more established industries that have chosen to pursue a new direction. They may be more sensitive to market conditions and policy decisions due to their smaller size and lack of entrenchment in the regional economy. Emerging clusters are generally identified through analyzing growth and employment trends, interviews with local business people or by some source of local industry knowledge.

In addition to emerging clusters, relatively small cluster industries that belong to a large two or three-digit SIC grouping may also be overlooked during the initial calculations. These industries may need to be examined using other regions' cluster definitions (or a standardized definition in the future) as guidelines for analysis. Overall, Employment Concentration Factors are the most broad and accessible tool in the initial clustering process; while industry knowledge, data, research and dialogue with representatives from the private sector serve as important supplemental tools.

Input-Output Models and Inter-Industry Relationships

Using the Employment Concentration Factor has only produced a group of export-oriented industries in the region without any information about how they are related. The next step in the cluster definition process is to group the key exporting industry sectors into preliminary clusters. This step represents the second of the technical factors and is a measure of the strength of inter-relationships between the industries in a cluster. However, before we begin the discussion of the Cluster Dependency Factors (CDF) it is necessary to discuss input-output models and their relevant applications.

In order to determine the relationships between the region's industries, an economic input-output model was used. Specifically, SANDAG chose the Minnesota IMPLAN Group's input-output (IO) model for the San Diego region. When input-output models are used at a regional level, there may be differences between one region's cluster relationships versus another's, especially at the four-digit SIC code level. For example, electronic instruments might be considered part of an aerospace cluster in one region of the country, whereas in another region they might be part of an automotive cluster. Constructed from recent data for the San Diego region, the IO model represents the flow

of goods and services between industries, thus showing one industry's relationship with another. It enables the analyst to form clusters based upon the production process, grouping the suppliers and demanders of goods, services, and raw materials.

The input-output model is also a useful tool in determining whether or not a certain cluster needs to be further broken down into segments. Many of the Collaborative Economics clusters were initially too broad and all-encompassing. IO analysis was used to streamline the cluster definitions, improving the level of internal cohesion and making the descriptions more specific. The original Health Care Products and Services cluster, for example, was a prime candidate for segmenting because of its broad definition and large total employment. The original cluster included industries from a variety of fields such as biotechnology, pharmaceuticals, healthcare service providers, and industrial chemicals. The IO model analysis determined which industries within this broad cluster exhibit the strongest relationships. Biotechnology industries show a stronger relationship with other biotech industries than to biomedical product manufacturers. A clear pattern arose from this analysis, and the Healthcare Products and Services cluster was divided into three narrower cluster groupings. Even if a cluster does not require a split, it is still useful to identify the inter-related cluster sub-groups. The narrower the definition of the cluster and its sub-groups, the more specific the policy focus can be. However, it is important to strike a balance between a narrow cluster definition and excluding an important industry sector from the group.

Cluster Dependency Factor

The transactions table of the input-output model is used to produce a set of ratios called Cluster Dependency Factors (CDF). The regional I-O model's transactions table shows both the flow and direction of the relationships and the relative strengths of the relationships between industries; strong relationships representing a possible cluster group. Cluster Dependency Factors use this information to determine the direction and relative strength of the buyer-seller relationships within clusters. Also, by analyzing the relationships from the input-output model, SANDAG can identify any industries that may

be missing from the cluster definitions as well as eliminate those that do not belong with the others in the group.

The Cluster Dependency Factors are developed through a number of steps. The basic concept is to look at one particular export-oriented industry and determine with whom it has the strongest relationships. Once a sufficient amount of strongly related industries are grouped together, they can be addressed as a preliminary cluster and CDFs can be calculated. There are two Cluster Dependency Factors: CDF (Cluster) and CDF (Industry). CDF (Cluster) focuses on the cluster's overall demand on a single industry, whether it is part of the cluster or not. For example, the Electronics Manufacturing cluster's demand on the printed circuit board industry. CDF (Industry) represents a cluster's supply to a particular industry. A CDF (Industry) figure could represent the amount that the Electronics Manufacturing cluster supplies to the printed circuit board manufacturing industry. Cluster Dependency Factors can be calculated for both industry to industry and cluster to cluster relationships.

It is almost certain that there will be a strong relationship between a cluster industry and an industry not included in the current cluster definitions. Therefore, it is useful to construct the Cluster Dependency Factors for any industry not presently included in the cluster but that shows a significantly above-average relationship with a cluster industry. This analysis acts as a safety mechanism to ensure that no relevant industries have been missed. It also enables the analyst to identify non-cluster support relationships. The choice of which additional industries to focus on should be based on how the results are to be used. One goal of the analysis may be to target specific industries or to attract new industries to the region. In this case, industries that show a strong relationship to the cluster but have low concentration factors will be included and a will be a primary focus of the analysis. Another policy may be to identify only the businesses in a cluster that meet the export-orientation requirement.

Economic Prosperity Factor

The final step in defining the San Diego region's industrial clusters is to determine which of the export-oriented industries contain high paying jobs. The high paying industries are identified through the calculation of the third technical factor, the Economic Prosperity Factor (EPF). The EPF is a measure of the economic significance of an industry and it equals an industry's annual payroll per employee over the San Diego average payroll per employee. If the ratio is greater than one, that industry's economic significance in the region is above the regional average. If the cluster analyst is only trying to understand regional industry dynamics and is not focusing on high wage industries this step is not mandatory. The Economic Prosperity Factor describes the wage characteristics of an industry. If policy makers wish to focus on certain types of industries, this factor allows for a narrower, wage-based focus.

Relevant findings from the Economic Prosperity Factor analysis are any industries that may have a relationship to the cluster (determined by input-output analysis or industry knowledge), a high and significant economic prosperity rating but a low employment concentration factor. These industries, containing the high paying jobs beneficial to the regional economy, are relatively weaker in terms of employment than their cluster colleagues, despite the local demand for their product. For some reason, the other cluster industries are highly concentrated while they are not. These industries, which SANDAG calls "weak industries," may have strong growth potential and are ideal areas for attraction and development policy focus.

Emerging Cluster Analysis

The emerging cluster analysis begins by identifying either a driving, export-oriented industry or a fast growing "emerging" industry. The export-oriented industries are determined using the Employment Concentration Factor while the emerging industries are determined by calculating an industry's employment growth rate over a certain time

period.[1] These are the primary emerging cluster industries and constitute the core of the new group. The core industries and the most strongly related additional industries constitute the preliminary cluster grouping. Additional industries will then be added on the basis of their relationship to the core. Once assembled, the preliminary grouping is similar in appearance to the original cluster definitions SANDAG used from Collaborative Economics.

SANDAG analyzed the following potential clusters: optics and lasers; advanced materials; environmental technology; power supplies and systems; higher education; several agricultural sectors; and other niche industries. Some of these industries were added to the regional cluster list, others were not. Many of the potential clusters consisted of a number of industries that were already included in an existing cluster. If the IO analysis failed to show that they belong elsewhere or as a separate group, they remained in their current cluster. Also, a few of the industries turned out to be general support institutions. These industries were related to many clusters but did not meet the three factor criteria for becoming their own cluster group.

Conclusion

The three technical factors were used as a basis to identify, define and describe the region's industrial clusters. The factors act as a guide and provide the information necessary to determine which industries should be grouped together to form industry clusters. It is important that additions and subtractions to the clusters be made with an idea of what the end result is going to be. Occasionally, it may be necessary to make certain additions to the preliminary definitions that have negative impacts on the characteristics of the final cluster group. For example, a low paying industry that does not meet the Economic Prosperity Factor criteria may be added to the cluster because of its strong relationship with the group. This can be a positive addition because sometimes it is not the aesthetics but the relationships and industry dynamics of a cluster that are

[1] SANDAG used Employment Development Department data for the years 1990 and 1995 to determine marginal growth rates. Many of the potential clusters analyzed by SANDAG were by recommendation from individuals with knowledge about the particular industries.

important to policy makers. Ultimately, the cluster analyst needs to determine the goal of the project and organize the analysis of the cluster industries accordingly.

As mentioned previously, the current lack of standardized rules for grouping industry sectors results in unique and individualized clusters that vary from region to region. Some clusters are even defined differently within the same region by different people. Though general industry patterns and relationships exist, each area's unique output will create different results. If clusters are to be used as a basis for regional, state and national planning, there must be an objective procedure developed to identify and define them. SANDAG hopes that within the near future some form of state-level standardization of cluster definitions will occur. This would be useful since in the early stages of cluster identification it is useful to refer to a standardized definition or other reports on clustering of industries. Broad definitions of clusters, such as those done at a state level, may be helpful in identifying and clarifying some of the more obvious industry relationships. It is much easier to refine and customize existing definitions than create new ones altogether. Regions would be able to calculate employment concentration factors and identify concentrated export-oriented industries. Then the identified industries would be matched up with their standardized cluster group. These relationships can later be refined using region specific input-output models and the criteria described in this report.