Articulating Uneven Regional Development: artificial intelligence as a tool in development planning

SASHI SIVRAMKRISHNA and RAMAKRUSHNA PANIGRAHI

Sashi Sivramkrishna is Chairman of the Foundation to Aid Industrial Recovery, New Delhi/Bangalore, and Ramakrushna Panigrahi is Faculty Member, ICFAI Business School, Bangalore

Abstract There is an urgent need for more efficient and effective design, targeting and implementation of interventions to reduce regional imbalances in development. To do so, development agencies and practitioners need to articulate uneven regional development as regional inequalities in, and patterns of, development. The widespread popularity of composite indices like the Human Development Index has led to the acceptance of regional inequalities as a basis for intervention. However, in computing composite indices of development like the Human Development Index, information that could be of great utility to planners is lost. This is especially important when planners work on smaller spaces and several indicators of development. There is then a need to also articulate patterns of development for optimal intervention. Unfortunately, the conventional statistical methods to discern patterns in development are complex and have not found widespread acceptance like composite indices. Artificial intelligence, in particular the Kohonen Self-Organizing Map, is a user-friendly tool for development planners and practitioners to explore patterns in development. An application with several indicators over 399 Indian districts illustrates the need to study development patterns. This paper also makes clear the versatility of the Kohonen Self-Organizing Map technique in exploring these regional patterns of development.

Key words: Development, Regional Development, Human Development Index, Artificial Intelligence

Introduction

Development is rarely, if at all, evenly distributed across regions, whatever may be the level of aggregation or disaggregation: country, state, district, subdistrict (*taluka* in India) or village. Development agencies and practitioners, both governmental and non-governmental, are increasingly concerned by this uneven development given its repercussions on political stability, social relations, as well as the full and efficient utilization of economic resources. At the same time, it is now recognized that development is multi-dimensional and cannot be reduced to a single indicator like Gross Domestic Product per capita. Targeting of interventions by state and non-governmental organizations (NGOs) to reduce regional imbalances then becomes more complex since it is necessary to compare regions across several indicators of development, not just a single indicator.

The scarcity of resources at the disposal of many governments and civil society organizations in the less developed world is pressurizing them to target their interventions more efficiently and effectively. To facilitate the optimal targeting of resources, we argue that policy-makers and development practitioners need to articulate uneven regional development as:

- regional *inequalities*; and
- regional *patterns* of development.

With the widespread acceptance of the United Nations Development Programme (UNDP) Human Development Index (HDI) as a measure of development, it has become standard practice to discern uneven regional development, not only at the international level, but also at subnational levels (state, district or subdistrict), through a study of regional *inequalities*. We explore the limitations of such a unilateral approach, and argue that it is useful and necessary to also look at regional *patterns* in development. This becomes all the more vital when regional imbalances are studied at the level of smaller spaces like districts and subdistricts, often the level of greater relevance to development planners and practitioners.

An important reason for the 'popularity' of the HDI and HDI-type composite indices of development as a basis for policy is the simplicity of the concept, computation and interpretation of index values. On the contrary, the methods used to explore regional *patterns* of development, like factor analysis, require more specialized skills. This fact has limited its appeal among a larger audience. Artificial intelligence, in particular the Kohonen Self-Organizing Map (K-SOM), as we will see, is not only a proficient tool to decipher patterns in development, but its user-friendliness could promote its acceptance among policy-makers and development practitioners in targeting their activities.

To illustrate the difference between and the importance of both spatial inequalities and patterns of development, we apply our concepts and methods to a study of 399 districts in India. On the basis of our results, we propose further applications in the study of uneven regional development.

Studying regional or spatial *inequalities* in development

The concept of 'inequality' is a complex one, being implicitly quantitative (numerical) and, at the same time, conveying a notion of injustice or unfairness. In development theory and practice, identifying inequalities usually means first measuring development quantitatively and then ranking regions based on the measured values. The measure of development also provides a measure of the degree of inequality between rich and poor, advanced and backward, developed and underdeveloped regions. Furthermore, regions can be clustered together as high, medium or low development on the basis of arbitrary cut-off index values. For instance, the UNDP clusters countries as follows (UNDP, 1999): high human development, where HDI $\geqslant 0.8$; medium human development, where $0.8 < \text{HDI} \leqslant 0.5$; and low human development, where HDI < 0.5.

The most easily quantifiable measure of development is income per capita. However, when one recognizes development as a multi-dimensional process, measuring it is no longer straightforward or uncontroversial. Here it becomes necessary to compute a composite or aggregate index like the HDI or a basic needs index. The process of aggregation involves computing a weighted average of several indicator index values. Two criticisms are commonly leveled on such composite indices.

1. The assignment of weights is arbitrary (Ravallion, 1996; Deichmann, 1999; Lok-Dessallien, no date[1]). It is true that any index constructed as an average of more than one indicator cannot overcome this problem in assigning weights. Even in a Borda count-based ranking (Dasgupta, 1993; Atkinson *et al.*, 1999), an implicit equal weight is assigned to each indicator. However, the essence of an index like the HDI is to break away from the notion that income and development are one and the same. It recognizes the multi-dimensionality of development and has been successful in realizing this objective.
2. In the process of averaging indicator index values to yield a composite index, information is lost or wasted (Ravallion, 1996). We will examine this aspect in the next section. For now it suffices to state that it is this drawback in composite indices that makes it necessary to also discern uneven regional patterns in development for optimal resource allocation.

These criticisms have led some to argue that composite indices should be ignored, and instead relevant indicators must be looked at *separately* to measure inequalities and rank regions. Deichmann, for instance, considers:

> The most promising route of inequality is therefore to recognize the multiple dimensions of deprivation but to describe these dimensions separately. (1999, p. 12)

This position is also taken by Lok-Dessallien:

> Composite indicators ... hide important policy and programme messages inherent in their constituent variables. For poverty monitoring within countries, it is therefore not advisable to combine different indicators into composites for policy purposes, but to let each set of indicators speak for themselves. (no date, p. 7)

However, mapping each indicator separately need not be the second best solution since, as we will see, both approaches waste useful information.

In spite of these drawbacks, measuring inequalities could be important for some purposes. For instance, in the disbursement of non-specific equalization grants and budgetary allocations, or for advocacy purposes[2] (Lok-Dessallien, no date), it is useful for government organizations, NGO funding organizations and NGOs to have a ranking of regions based on a composite index, at least as a first step. Single indicator-based development indices and maps also provide important information for planners.

Information loss in composite indices and single indicators of development

The construction and application of composite indices of development are extensively discussed in economics literature (Henninger, 1998; UNDP, 1999). Given the simplicity in construction and interpretation, such indices are also finding extensive use in varied subjects (Granados and Peterson, 1999; *Ecologist*, 2001). However, in the construction of a composite index, the process of averaging indicator values leads to wastage of information; in particular, information that may be of specific use to development organizations. As pointed out by Ravallion,

> aggregation wastes information; it can be important to know that region A is doing well in the income space, but not in basic health and schooling, while in region B it is the reverse. (1996, p. 9)

Diechmann also points this out in text for the World Bank's website for 'Inequality, Poverty and Socio-economic Performance':

> Aggregating individual measures also hides important sectoral information that can be used to select specific policy intervention". (1999, p. 14)

Let us illustrate what this argument means for a development agency formulating a specific program or project. We use a contrived dataset (Table 1) for nine regions (R_1, $R_2, \ldots, R_9$) and two indicators, X_1 and X_2, given equal weights. The corresponding composite index, I, is given by $(X_1 + X_2)/2$ —

TABLE 1. A contrived dataset, composite index value (I), Borda score (B) and rank

Regions	X_1	X_2	Composite index value	Borda score	Rank
R_1	0.05	0.95	0.5	10	1
R_2	0.1	0.9	0.5	10	1
R_3	0.15	0.85	0.5	10	1
R_4	0.55	0.45	0.5	10	1
R_5	0.5	0.5	0.5	10	1
R_6	0.45	0.55	0.5	10	1
R_7	0.85	0.15	0.5	10	1
R_8	0.9	0.1	0.5	10	1
R_9	0.95	0.05	0.5	10	1

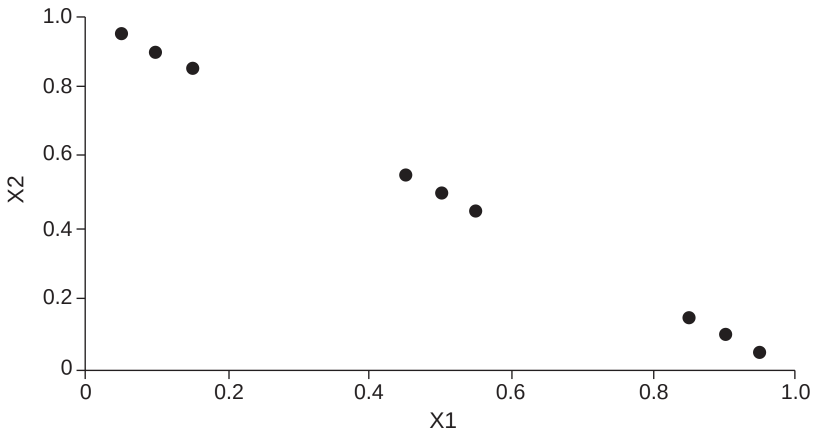

FIGURE 1. Scatter plot of X_1 and X_2 from Table 1.

similar to the method adopted by the HDI. In the next column, the Borda score, B,[3] is calculated and presented. Using either of the two methods, we find that, on average, all nine regions are equally developed. This may be important information to development organizations. However, there is another piece of information so easily apparent in Figure 1 (a plot of the data in Table 1) but not extracted by the composite indices or Borda count; namely, the existence of three clusters of homogeneous regions. In other words, using composite indices we lose information on *similar* regions within a cluster and *differences* between clusters.

A case in point here is that of Kerala and Punjab States in India, where the composite indices of development are almost equal: 0.775 for Kerala and 0.744 for Punjab, based on 1991 data (Krishnan, 2000). However, what remains concealed in this index is the fact that the per-capita state domestic product of Kerala is less than one-half that of Punjab (Indian rupees 4618 and 9643, respectively in 1991–1992). On the contrary, female literacy and infant mortality rates in Kerala are 86.9% and 17%, respectively, whereas in Punjab they are 49.7% and 61%, respectively (Krishnan, 2000). These wide differences in development variables are not captured by the composite index; instead, they get averaged out. Capturing the regional differences between Kerala and Punjab States could be useful and important information to development organizations. For instance, a health project could have a different impact in Kerala and Punjab due to the differences in education levels in the two states.

It is important to point out that the process of averaging does not distort or conceal information where data is distributed as in Figure 2;[4] where regions are usually more developed than others for all indicators. Our observation is that development indicators, in general, are more likely to be so distributed for large spaces: for example, at a country level or state level (although the earlier example of Kerala and Punjab highlights problems that could arise even for large spaces). It is for this reason that country-level

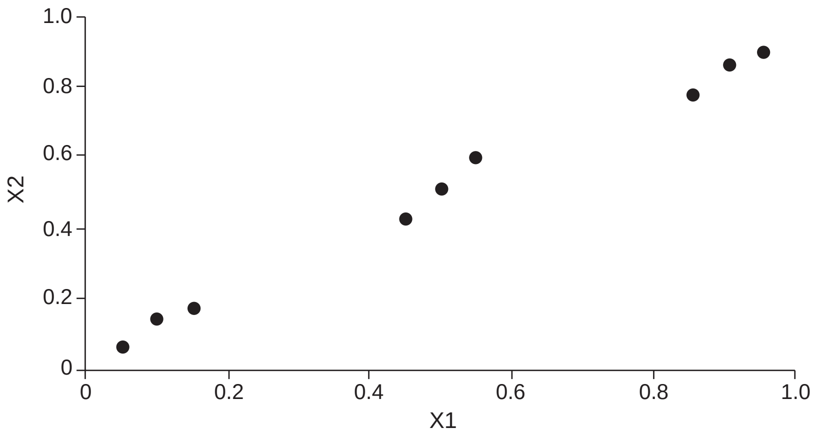

FIGURE 2. Scatter plot of X_1 and X_2 for a positively correlated distribution.

studies like the UNDP HDI (UNDP, 1999) do not face problems of information loss in averaging indicator index values, and they provide satisfactory results in so far as ranking and clustering countries by their levels of development. Constructing composite indices from several indicators of development, especially for small spaces, like districts or *talukas*, are more likely to face problems from averaging, and render rankings quite irrelevant to micro-level interventions (Lok-Dessallien, no date).

As a simple exercise, we calculated the Pearson's correlation coefficient between income, health and education indicator values at the international, India state and district level (Karnataka and Madhya Pradesh). The results are presented in Table 2. As we go down to smaller spaces, the data distribution becomes more irregular and will be less likely to resemble Figure 1. Here significant information loss can occur from averaging in the construction of a composite index.

Concentrating only on single indicators of development is important but does not 'reduce' data — the very purpose of constructing a composite index. With reference to our example in Table 1, disaggregating information would mean mapping X_1 and X_2 separately using arbitrary cut-off index values to cluster regions into different levels of development.

TABLE 2. Pearson's correlation coefficient between indicators at country, state and district levels

	Country	State (India)	District (Karnataka)	District (Madhya Pradesh)
Income–health	0.810	0.607	0.461	0.177
Income–education	0.730	0.219	0.683	0.247
Education–health	0.779	0.671	0.436	0.302

Sources: Country, UNDP (1999); state, Shivkumar (1994); district (Karnataka), Government of Karnataka (1999); district (Madhya Pradesh), Government of Madhya Pradesh (1997).

If required, Deichmann then proposes that:

> Spatial maps can be used to overlap income-poverty, malnutrition, and access maps to assess joint correlations, or disparities. (1999, p. 3)

Even in our simple two-variable contrived case, the method and result from 'overlapping' maps are unclear. For large multivariate data sets, such a method cannot reduce data meaningfully. Information in the dataset pertaining to the *inter-relationship* between variables is lost.

To illustrate our argument, we refer readers to a study by Rajakutty *et al.* (1999) for the National Institute for Rural Development, where several independent indicator maps have been constructed across 55 Indian regions. The study also overlaps two indicators at a time on a single map. However, mapping the inter-relationship between more than two variables or all the indicators taken together becomes impractical. The study then resorts to computing composite indices; namely, a Social Development Index and an Infrastructure Development Index. These composite indices face the problem of wasted information from averaging. One alternates in a vicious cycle between independent indicator maps and composite indices.

Returning to our example in Table 1 and put simply, we need is a technique to reduce the dataset into three clusters; each cluster containing regions with similar combinations of X_1 and X_2, and at the same time segregating different clusters. However, before discussing techniques and methods, we must understand the nature of information that we are trying to extract from the dataset. The notion of development *patterns* is introduced for this reason.

Regional *patterns* in development

Patterns in development neither rank regions nor measure their levels of development, only that regions with similar combinations of development indicators are extracted from the data. In other words, we need to construct a summary map that relates data to locations, provides a truly geographical representation of information, and identifies or illustrates spatial patterns and relationships (Cowlard, 1998). As Deboeck points out, focusing entirely on individual indicators (or composite indices) could be:

> inadequate for effective and efficient detection of patterns in data . . . the complexity and interwoven nature of poverty remains largely unexplored. (2000, pp. 3)

As apparent, deciphering development patterns is essentially cluster analysis. However, it must be emphasized that mere clustering of data does not always capture patterns in development. For instance, a development map based on clustering regions according to high, medium and low development, based on composite index values, does not qualify as a study of development patterns since inter-relationships among variables in the dataset remain unexplored. In Rajakutty *et al.* (1999), the Social Development Index-based region-wise cluster map cannot be taken as a pattern map because information on the inter-relationship between indicators considered in the Social Development Index are not explored.

Development planners and practitioners are often concerned not only with development or poverty indicators, but also their inter-relationship with a region's social, demographic, cultural and physical attributes; attributes that cannot always be categorically classified as good or bad, better or worse, more or less developed. In other words, they cannot be ranked. Unlike in the study of inequalities, these are easily brought into a study of regional patterns of development since we are not intent on measuring development or poverty or ranking regions in terms of their level of development; we are only interested in identifying relationships between the variables across regions.

The HDI focuses on overall development of a region by looking at *output* indicators of development. However, for micro-level and macro-level planning, development agencies often have to consider a large set of *input* indicators of development, even where output indicators may be available — although measurement of income becomes difficult and controversial for smaller spaces. Once again, composite indices are prone to problems from averaging of indicators indices. The study of development patterns could retain valuable information in the dataset for the planner.

Our experience in the NGO sector has been that exploring multivariate data could reveal certain interesting and useful underlying patterns in the spatial distribution of development. Consider, for instance, a children's health project. Its effectiveness will benefit from knowledge of regional patterns in demography, education, health, income, gender, urbanization, women's occupational structure, child labor and social (caste/tribe) parameters. Areas with high incomes, but low education and women's status, may require a different program design and implementation strategy as compared with a region where education levels and status of women are better, but incomes are low. Policy design requires not only identification of the poorest or least developed regions, but also those that are most likely to benefit from intervention, thereby making it efficient and effective. The study of development patterns is essential to such efforts.

What we then often look for is a reduction in data, *keeping intact* information on regional *differences* — without these differences getting averaged out. As we have shown earlier, composite indices and single indicator mapping fail in identifying patterns in development. In fact, patterns in data are concealed, wasted or ignored by the methods used to identify regional inequalities. In the context of Table 1, the three distinct clusters in the data set must be identified; that is, neither reduced to a single index value in the process of averaging (composite index) nor ignored (single indicator mapping).

Having articulated the need to discern patterns in development, one could resort to standard statistical clustering techniques to do so — like the K-means clustering algorithm or factor analysis. We must make it clear that we are not making a mathematical evaluation or comparative study of these techniques, but only making some comments on their application from a practical point of view.

The K-means clustering algorithm

The three clusters in the dataset (Table 1) can be effectively identified using the K-means clustering algorithm. The major limitation in using the K-means algorithm is that the number of clusters must be specified, *a priori*, by the researcher. When an 'incorrect' number of clusters is specified, the clustering becomes vague and so also does the resulting development map. Just as in the case of the HDI or the Borda count, the use of K-means in recognizing regional disparities in development could actually prove counter-productive when an arbitrary number of clusters is specified. The problem of specifying the 'correct' number of clusters *a priori* for large multivariate data sets, as in the case of our applications in the following (399 districts $\times$ 12 indicator variables), limits the usefulness of this technique in constructing useful pattern development maps.

Factor analysis as a clustering technique

Factor analysis can be used as a clustering technique (see Appendix 1). However, there are some major hurdles that must be faced by development practitioners in using this technique.

- The mathematical inputs required for a rigorous understanding of the technique are a definite constraint to its acceptance by development practitioners. Moreover, there is no clear and unique methodology to cluster cases in the dataset.
- The difficulty of using a 'black box' approach with factor analysis; that is, to go from the raw dataset to obtaining clusters of homogeneous regions without a rigorous understanding of the technique. A *practical* clustering technique is what practitioners look for: after all, these are taken as inputs in planning and are not to accept or reject a theoretical hypothesis. A 'black box' approach could sometimes suffice given the nature of application in context.
- Planners often need an exploratory data analysis technique, adding and deleting variables according to their specific requirements. General academic studies may not be relevant to their purpose.

Consider, for example, a study by Rao and Babu (1996) of regional disparities in the Hyderabad-Karnataka region of India. The delineation of homogeneous regions (*talukas*) is carried out separately for structural and sectoral factors using factor analysis. The structural factors pertain to physical attributes of regions like forest area and land under cultivation, whereas the sectoral attributes take into account agricultural, industrial, financial, transport, communication, education and health development. However, to capture the overall region-wise relationship between factors, the authors construct a composite index of development as an average of sectoral indices, the latter in turn based on 'factor loadings'. Sectorally, *taluka*-wise delineation of similar regions is carried out using factor analysis, but in terms of 'overall' development, a composite index is resorted to.

What we wish to argue pertains not to the methodological validity of such studies, but to see their usefulness to development practitioners. Often, development practitioners need to add and delete variables to suit their own needs. Not only are such studies difficult to manipulate, but also the expertise required to replicate the analysis makes them unappealing to a wider audience.

These complexities in studying development patterns (using factor analysis) as opposed to the simplicity of studying development inequalities (using composite indices) have relegated application of the former to the sidelines, and mainly to academic discourse. In the process, the development practitioner loses invaluable information for more effective development action.

The Kohonen Self-Organizing Map

Although artificial intelligence, in particular neural network techniques, has found widespread application in the sciences and engineering, its use has remained rather limited in economics and confined to specific areas like finance (Skapura, 1995; Deboeck, 1998; Deboeck and Kohonen, 1998; Shumsky and Yarovoy, 1998). An in-depth introduction to artificial intelligence and neural networks is beyond the scope of this paper and can be found elsewhere[5] (Ginsberg, 1993; Aleksander and Morton, 1995; Skapura, 1995; Nilsson, 1998). The artificial intelligence technique chosen for our study here is the K-SOM, an unsupervised learning technique that clusters data based on a distance function without any *a priori* information on the number of clusters. The (artificial) intelligence of the algorithm is that it discerns something similar to what the human brain sees in the dataset. In the present context, the algorithm is able to group or cluster regions with similar combinations of indicators based on information within the data set itself. Once again, a technical understanding of the K-SOM algorithm is beyond the scope of this paper. Interested readers may refer to Beale and Jackson (1990), Kohonen (1990), Aleksander and Morton (1995), Kaski and Kohonen (1996), Beveridge (1996), Frohlich (1999), Germano (1999), and Deboeck (2000). A brief description the K-SOM technique is presented in Appendix 2.

Applying the K-SOM technique to the dataset presented in Table 1[6] clusters the data into three distinct sets — namely, (R_1, R_2, R_3), (R_4, R_5, R_6), and (R_7, R_8, R_9) — which can be readily transformed into a development map.

It is important to reiterate here that the number of clusters was not specified *a priori* as in the K-means algorithm. Moreover, the difficulty encountered by non-specialists in using and interpreting the results of factor analysis is absent. The development practitioner can take a 'black-box' approach to obtain the clusters of homogeneous regions, adding and deleting variables according to their specific needs.

In our contrived example, a composite index (I or B) then enables ranking of regions (in this case, equal rank of 1). The K-SOM, on the contrary, neither measures development nor ranks regions; it only identifies the spatial pattern of development. Moreover, average indicator values for each cluster could provide information on the general level of development of regions in the cluster. The K-SOM algorithm, by extracting information on regional differences in development from the dataset, could be a useful tool[7] in development program formulation.

A district-level analysis for India

The K-SOM technique has been used in the study of country-level development by Kohonen and Kaski (1996) and Deboeck (2000). However, as we have stated earlier, the data distribution of country-level indicators is likely to follow a pattern as in Figure 2. This would mean that results obtained using a composite index and the K-SOM are quite similar. Moreover, these country-level studies do not articulate the essential difference between inequalities and patterns of development, the latter forming the *raison d'etre* of using the K-SOM technique.

We use the K-SOM technique for a study of regional disparities at a level of relatively smaller spaces using Census of India (1994) data, and data compiled by the Center for Monitoring the Indian Economy (2000). India with a population of over 1 billion has approximately 500 districts. Each district, with an average population of 2 million, is an administrative unit of state governments. In this study, we have taken into consideration 399 districts from 16 Indian states. The remaining districts have been excluded due to large number of missing data.

At a country level, the HDI is one of the most widely accepted composite indices. This index is an average of three indicators of development — income, health and education — each given equal weights. More recently, the HDI has been calculated at the state level for India (Shivkumar, 1994) and substate (district) level (Government of Madhya Pradesh, 1997; Government of Karnataka, 1999) to study regional inequalities in India.

To our knowledge, official all-India district-level HDI studies have not been undertaken given the non-availability of Gross Domestic Product per-capita and Life Expectancy at Birth (LEB) data for all states. For this reason, we use multiple input surrogates for each indicator of development. To maintain the *spirit of HDI*, we make adjustments in weights assigned to each surrogate for income, LEB and education so that, in the aggregate income, health and education have a one-third share. We have used five surrogates each for income and health to avoid problems caused by outliers in any single indicator value. Each surrogate for income and health has been assigned a weight of $(0.33)/(5) = 0.066$. In the case of education, we have assigned equal weights ($= 0.33/2$) to the ALR and the female literacy rate.[8] All variables chosen have a definite bearing on development, either positive or negative.

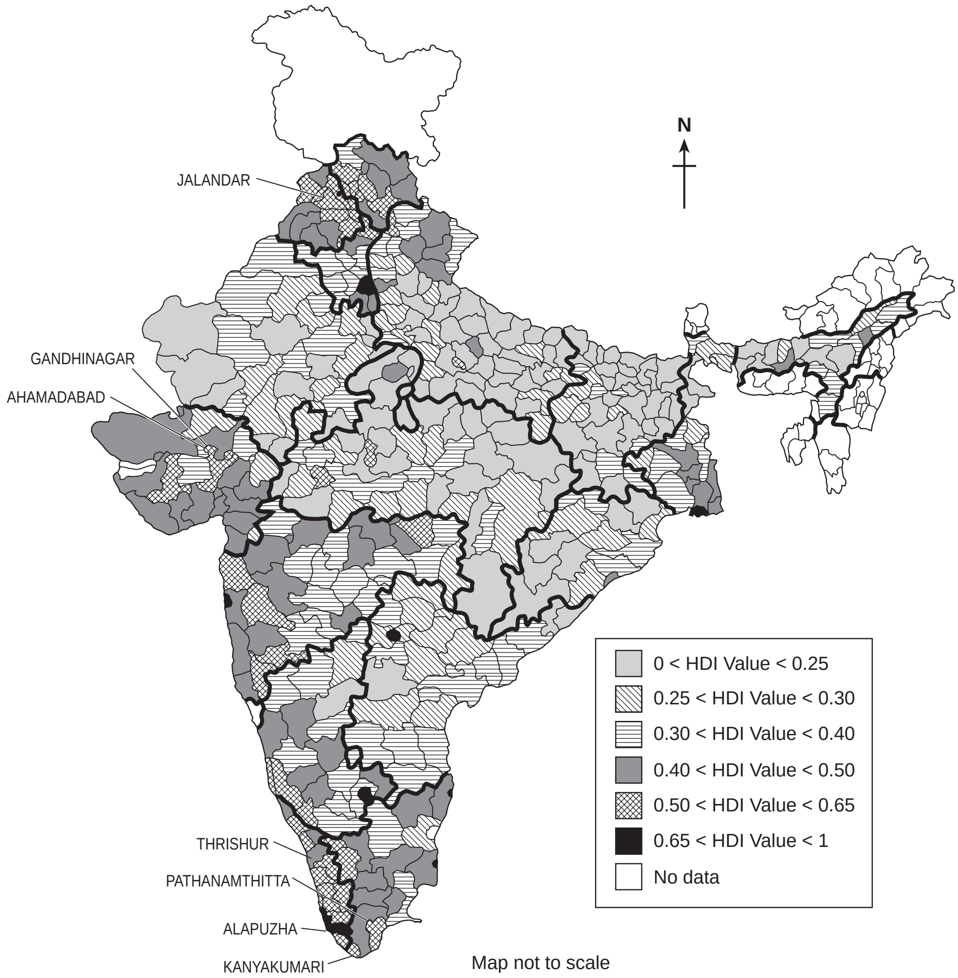

FIGURE 3. Poverty (inequality) map of 399 Indian districts.

We first construct the development (inequality) map presented in Figure 3, where the districts have been assigned six levels of development: very high, high, high-middle, middle, low-middle and low. The cut-off points to define these levels of development are arbitrary and are stated in the key to Figure 3.[9]

The inequality map provides information of average or overall development of each district, but does not delineate homogeneous regions. Take, for example, three districts, Thrissur (Kerala State), Ahmedabad (Gujarat State) and Jalandhar (Punjab State), ranked 13, 14 and 15, respectively. Table 3 presents the indicator values and their composite index values for the three districts. Although their composite index values are almost equal (i.e. there is no significant *inequalities* between these districts), the differences in independent indicator values are significant or, in other words, they do

TABLE 3. Indicator values and composite index for three districts

Indicator	Thrissur	Ahmedabad	Jalandhar
Deposits	7357	11153	20202
Electricity	57.63	74.89	90.22
Banks	11.48	10.82	16.41
Credit	2269	7214	43.57
Housing	57.89	74.41	90.72
Drinking water	18.32	83.62	94.45
Toilet	63.63	55.91	36.18
Hospital beds	147.16	328.19	143.3
Total fertility rate	2.10	3.55	3.64
Infant mortality rate (per 1000 population)	29	64	53
Adult literacy rate	0.791	0.617	0.5815
Female literacy rate	0.77	0.53	0.52
Composite index	0.6273	0.619	0.6152
Rank	13	14	15

Key: Deposits, deposits in Rupees per capita; electricity, proportion of households having electricity facility; banks, bank branches per *lakh* (100,000) population; credit, credit disbursed in Rupees per capita; housing, percentage of households occupying *pucca* houses; drinking water, proportion of households having safe drinking water facility; toilet, proportion of households having toilet facilities; hospital beds, hospital and dispensary beds per *lakh* population.

not exhibit a similar *pattern* of development. Deposits in Jalandhar are almost three times those of Thrissur, whereas mortality rates in Ahmedabad and Jalandhar are significantly 'worse than' in Thrissur.

The question we then ask is whether consideration of mere rank or composite index value is sufficient for optimal targeting of development resources? As development practitioners, do we not need information on how these three districts are *different* from each other and which districts are similar to each other?

Figure 4 is a poverty (pattern) map constructed using the K-SOM technique, with the same indicators and weights as taken earlier in the construction of composite indices so as to compare the results from the two methods. The K-SOM algorithm, without any *a priori* information on the number of clusters, identified six distinct groups of regions. Table 4 presents the average values for the variables in each cluster. It is clear that Cluster 1 has a higher development level than most others clusters, but when we look at Clusters 2 and 3 no definitive ranking is possible. Cluster 2 is better off for some indicators (like drinking water), whereas Cluster 3 is better off for others (like the female literacy rate). A ranking of clusters with a Borda count of average values of indicators could be performed to indicate regional levels of development.

We can make good judgment of the K-SOM algorithm from the results of our present study of 399 Indian districts. Consider the following.

- The K-SOM-based map has been able to identify all districts in the southwest State of Kerala[10] in a single cluster, whereas with the composite index approach averaging eliminates the uniqueness of Kerala, with its districts

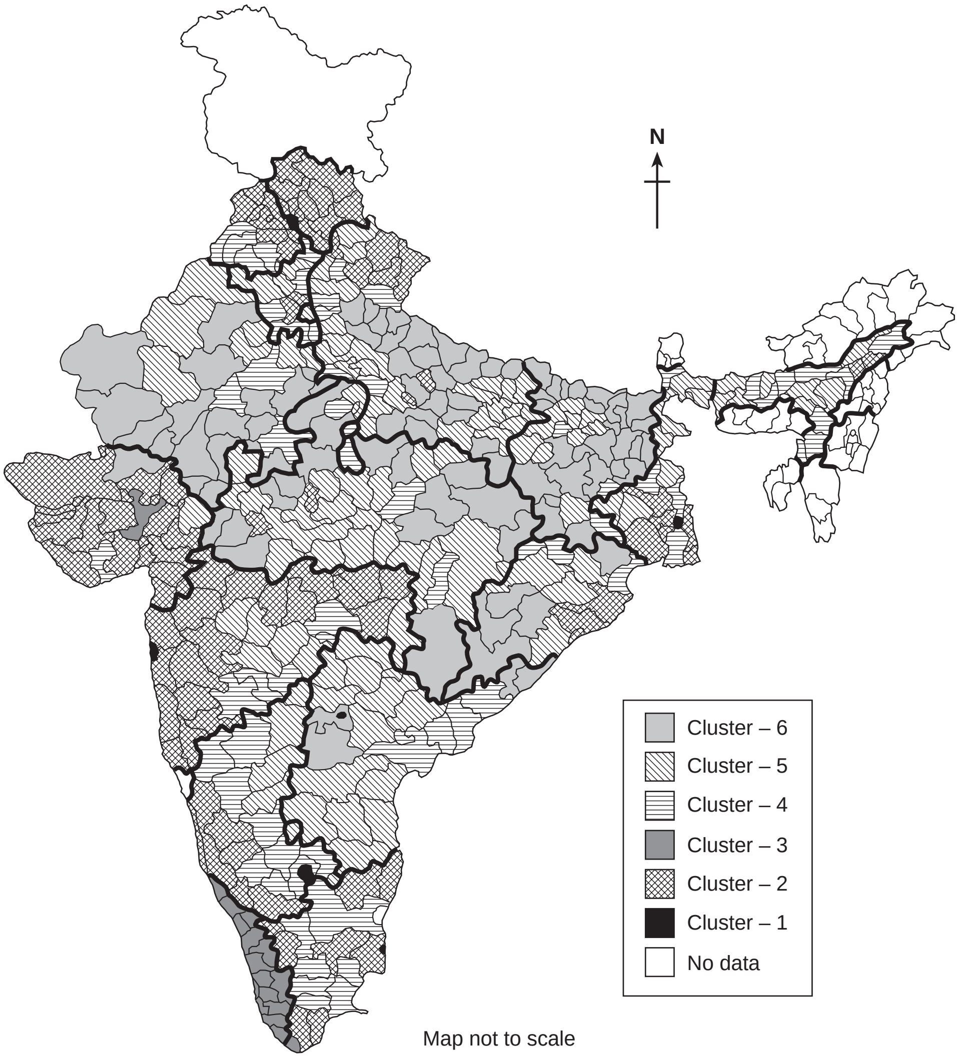

FIGURE 4. Poverty (pattern) map of 399 Indian districts.

falling into three distinct levels of development. The only other districts in India with a similar pattern of development to that of Kerala are the bordering district of Kanyakumari (Tamil Nadu State) and Gandhinagar (Gujarat State). Cluster 3 in Figure 4, which includes the districts of Kerala, shows low levels of development for income variables but they are significantly better when we consider health and education levels (see Table 3).

- Another interesting result in Figure 4 is that the entire northern State of Himachal Pradesh forms part of a single cluster; although in terms of inequalities this is not so (Fig. 3), where its districts again fall into three levels of development.

TABLE 4. Cluster-wise average value for each indicator

Indicator	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Deposits	28336	5068	5847	2528	1594	1202
Electricity	84.32	65.64	49.6	46.87	32.23	23.29
Banks	14.9	9.68	10.16	6.73	5.84	5.62
Credit	20740	2312	2681	1435	777	515
Housing	82.85	53.24	52.05	40.82	35.93	30.4
Drinking water	86.98	69.09	28.47	67.34	58.57	51
Toilet	78.32	26.87	46.57	23.44	15.53	1141
Hospital beds	227	149.1	123.5	65.9	49.5	33.7
Total fertility rate	3.441	3.7	2.707	4.055	4.827	5.276
Infant mortality rate (per 1000 population)	38.9	66.8	39.6	69.8	88.3	101.8
Adult literacy rate	0.65	0.548	0.762	0.446	0.354	0.266
Female literacy rate	0.59	0.453	0.73	0.34	0.227	0.143

Key: Deposits, deposits in Rupees per capita; electricity, proportion of households having electricity facility; banks, bank branches per *lakh* (100,000) population; credit, credit disbursed in Rupees per capita; housing, percentage of households occupying *pucca* houses; drinking water, proportion of households having safe drinking water facility; toilet, proportion of households having toilet facilities; hospital beds, hospital and dispensary beds per *lakh* population.

- An important difference between inequalities and patterns of development is also noticeable in the southwest State of Karnataka. Along its eastern border with the State of Andhra Pradesh,[11] no two adjoining districts have the same level of development in terms of inequalities (Fig. 3). However, when we consider patterns of development (Fig. 4), a strong trend emerges with adjoining districts showing greater homogeneity. These districts in fact form part of the Hyderabad-Karnataka region.
- For planners, the pattern map (Fig. 4) is useful in pointing out that (excluding Bangalore urban district), the state of Karnataka falls into three zones: coastal, Hyderabad-Karnataka and a central region (Cluster 2, Cluster 5 and Cluster 4, respectively).
- The K-SOM algorithm also groups seven major metros in a single cluster, without having specified the percentage of urban population in the district. The inequality map based on the HDI method, however, fails to capture the distinctiveness of these major cities. These seven urbanized districts are spread over ranks one to 11, with non-urban districts like Alappuzha and Pathanamthitta (Kerala State) appearing in-between. Obviously, in interventions, the non-urban districts may need distinct strategies from urban regions. The pattern map captures this information whereas the inequality map does not.
- A large part of central and north India in both Figures 3 and 4 are, by and large, similar since they follow a data distribution more like that in Figure 2; that is, they have *worse*[12] values for most of the indicators. This is evident from the average values of the indicators in Clusters 5 and 6 (see Table 4).
- The recently demarcated states of Uttaranchal, Chattisgarh and Jharkhand are not shown on Figures 3 and 4 since the dataset pertains to 1991.

However, it is interesting to observe that, in the case of Uttaranchal, the districts are fairly homogeneous and distinct from the development pattern of Uttar Pradesh.

This application illustrates the difference in results obtained when we consider development inequalities and development patterns. In practice, development planners and practitioners often have to work with smaller spaces with several development variables as well as socio-cultural, environmental, physical and other indicators relevant to their needs. As Rao and Babu argue:

> One type consists of those which are resource poor and do not possess adequate development potential. The other type consists of those which have rich natural resources ... but owing to historical and political factors could not exploit the resources for development purposes and, therefore, remained backward. These differences in the nature of the sub-regions are important while formulating a regional plan ... (1996, p. 47)

The flexibility offered by considering patterns of development allows practitioners to take into account variables that could be of relevance to them.

In 2000, the Government of Karnataka announced a study of regional disparities at the *taluka* level. Data on some 42 indicators of development are expected to be available at the time of the final report. Focusing only on regional inequalities through a composite index of a large number of indicators at the *taluka* level could mean a significant loss of information. Moreover, in constructing a composite index there may exist a problem of dealing with indicators that are not directly related to the *level* of development — including demographic variables, area under certain crops, urbanization, caste structure, and so on — although these indicators are often of importance to development planners and practitioners in their interventions. For instance, a project may need to consider resource endowments like forests, water, mines as well as topography, weather and other natural factors, and their interaction with socio-cultural and development indicators across regions. These variables can be considered simultaneously with development indicators and the pattern mapped using the K-SOM algorithm.

Conclusion

Development plans, policies and projects to reduce regional imbalances need to study *both* inequalities in and patterns of development. The composite index has become an attractive tool to development practitioners to study inequalities. On the contrary, the complexity in the techniques to study patterns of development has limited its application in development planning. The K-SOM artificial intelligence algorithm is a user-friendly tool that could provide insights into development patterns, an invaluable input for optimal targeting of interventions.

Notes

1 Date of publication not found on the web page.
2 An example of such advocacy could be state re-organization within a country.
3 The Borda score or Borda count of a region is the sum of its ranks for each indicator; the higher the score, the lower the rank of a region in terms of overall development.
4 Here $R_i + 1$ is 'better than' R_i (for all i) with respect to all indicators X_j (in this case, X_1 and X_2).
5 Several interesting and informative articles are also available on the Internet.
6 The VISCOVERY® SOMine Standard Edition package was used for the K-SOM analysis. The authors are grateful to Chemols Infotech Private Limited for the data analysis.
7 The specialized VISCOVERY® SOMine package gives users scope for exploratory data analysis such as, for example, 'nearest' regions in development levels, component maps, and so on. These could be of practical use to development agencies.
8 The female literacy rate, although included in the estimation of the adult literacy rate, is considered a separate variable in our analysis given the gap in gender inequality in third-world economies, particularly India. There are districts with almost the same adult literacy rate, but with very high differences in the female literacy rate. For example, Bhojpur (Bihar) and Marigaon (Assam) have 37.47% in the adult literacy rate, while the female literacy rate in Marigaon is 30% and in Bhojpur is 21%. As argued earlier, if the objective of the interventions is to reduce the gap between the educational attainment levels of males and females, the female literacy rate must be taken into consideration. For this reason, we have included both the female literacy rate and the adult literacy rate, and it may be noted that there is no double counting if both the variables are included in the analysis.
9 Districts referred in the text have been marked on Figure 3 only.
10 Kerala is known or its distinct pattern of development; namely, its less than average per-capita income but high levels of health and education development.
11 Many of the districts along this border, called the Hyderabad-Karnataka region, have been demanding special recognition and government support.
12 We use *worse* instead of lower since higher infant mortality and fertility rates mean less development.

References

Aleksander, I. and Morton, H. (1995) *An Introduction to Neural Computing*, 2nd edition, International Thomson Computer Press, London.

Atkinson, G. R., Duborg, R., Hamilton, K., Munasinghe M., Pearce, D. and Young C. (1999) *Measuring Sustainable Development: Macroeconomics and the Environment*, Edward Elgar, Cheltenham.

Beale, R. and Jackson T. (1990) *Neural Computing: An Introduction*, Adam Hilger, Bristol.

Beveridge, M. (1996) *Biological Neural Networks* [http://www.dcs.napier.ac.uk/hci/martin/msc/node6.html].

Census of India (1994) 'Housing and amenities: a database on housing and amenities for districts, cities and towns', *Occasional Paper No. 5*, Government of India, New Delhi.

Centre for Monitoring Indian Economy (2000) *Profiles of Districts*, CMIE Private Limited, Economic Intelligence Service, Mumbai.

Cowlard, K. A. (1998) *Decision-Making in Geography: A Handbook of Method and Practice*, 2nd edition, Hodder & Stoughton, London.

Dasgupta, P. (1993) *An Inquiry Into Well-Being and Destitution*, Clarendon Press, Oxford.

Deboeck, G. J. (1998) *Financial Applications of Self-Organizing Maps* [www.dokus.com].

Deboeck, G. (2000) *Self-Organizing Patterns in World Poverty Using Multiple Indicators of Poverty, Repression and Corruption*, 1818 Society, World Bank, [http://www.dokus.com/Papers_on_the_Web/Poverty.htm/].

Deboeck, G. and Kohonen T. (1998) *Visual Explorations in Finance Using Self-Organizing Maps*. Springer, Berlin.

Deichmann, U. (1999) *Geographical Aspects of Inequality and Poverty, Text for the World Bank's Site on Inequality, Poverty and Socio-Economic Performance*, [http://www.worldbank.org/poverty/inequal/povmap/].

Ecologist (2001) 'Keeping Score', *Ecologist*, 31(3), pp. 44–47.

Frohlich, J. (1999). *Neural Networks with Java* [http://rfhs8012.fh-regensburg.de/~saj39122/jfroehl/diplom/e-index.html].

Germano, T. (1999) *Self-Organizing Maps* [http://davis.wpi.edu/~matt/courses/soms/].

Ginsberg, M. (1993) *Essentials of Artificial Intelligence*, Morgan Kaufman Publishers, San Mateo.

Government of Karnataka (1999) *Human Development in Karnataka 1999*, Planning Department, Bangalore.

Government of Madhya Pradesh (1997) *Madhya Pradesh Human Development Report*, Government of Madhya Pradesh, Bhopal.

Granados, A. J. and Peterson, P. J. (1999) 'Hazardous waste indicators for national decision makers', *Journal of Environmental Management*, 55(4), pp. 249–263.

Henninger, N. (1998) *Mapping and Geographical Analysis of Poverty and Human Welfare — review and assessment*, report prepared for the UNEP/CGIAR Initiative on GIS, World Resources Institute, Washington, DC [http://www.povertymap.net/pub.htm].

Kaski, S. and Kohonen T. (1996) 'Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world', in A.-P. Refenes, A.Y. Mostafa, J. Moody and A. Weigend (Eds), *Neural Networks in Financial Engineering*, World Scientific, Singapore.

Kohonen, T. (1990) *Self-Organization and Associative Memory*, 3rd edition, Springer-Verlag, Berlin.

Krishnan, T. N. (2000) 'The route to social development in Kerala: social intermediation and public action', in S. Mehrotra and R. Jolly (Eds), *Development with a Human Face: experiences in social achievement and economic growth*, Oxford University Press, Oxford.

Lok-Dessallien, R. (no date) Review of Poverty Concepts and Indicators, *SEPED Series in Poverty* [http://www.undp.org/poverty/publications/pov_red/].

Nilsson, N. J. (1998) *Artificial Intelligence: a new synthesis*, Morgan Kaufman Publishers, San Mateo.

Ravallion, M. (1996) 'Issues in measuring and modeling poverty', *The Economic Journal*, 106, pp. 1328–1343.

Rajakutty, S., Acharya, S. and Haque, T. (1999). *India's Rural Development Report 1999: regional disparities in development and poverty*, National Institute of Rural Development, Hyderabad.

Rao, H. and Babu, M. Devendra (1996) 'Intra-regional disparities in Hyderabad-Karnataka region: an analysis', in A. Aziz and S. Krishna (Eds) *Regional Development: problems and policy measures*, Concept Publishing Company, New Delhi.

Shivkumar, A. K. (1994) 'Human Development Index for Indian states', *Economic and Political Weekly*, XXVI, pp. 2343–2345.

Shumsky, S. A. and Yarovoy, A. V. (1998) 'Kohonen atlas of Russian banks', in G. Deboeck and T. Kohonen (Eds) *Visual Explorations in Finance Using Self-Organizing Maps*, Springer.

Skapura, D. M. (1995) *Building Neural Networks*, ACM Press, New York.

UNDP (1999) *Human Development Report 1999*, Oxford University Press, New York.

Appendix 1 A Note on Factor Analysis

In our analysis of mapping 399 districts on the basis of development indicators, cluster analysis using artificial intelligence techniques is preferred to factor analysis due to the inadequacy of factor analysis to take into account *all* the variables. The main application of factor analysis techniques can be classified as:

- to reduce the number of variables; and

- to detect structure in the relationships between variables (i.e. to *classify* variables).

Factor analysis is applied as a data reduction or structure detection method, and is useful in the context of combining variables based on factor scores and to determine the spatial hierarchy in the case of multiple variables. For example, if there are 20 variables, factor analysis classifies these variables into four or five groups, each containing four or five variables on the basis of their similarity. The major criticism of factor analysis methods of clustering include the implausible use of a linear model across cases, the problem of multiple factor loadings (what is to be done in case of high loadings on more than one factor) and the double-centering of data. When it comes to mapping development using many variables, factor analysis is inadequate as it fails to classify regions based on all the variables simultaneously without any loss of information. Instead, it classifies variables based on their similarity and, therefore, becomes useful when one studies the regional pattern of development for groups of variables separately instead of all variables. Cluster analysis using artificial intelligence proves to be more relevant in the context of development mapping as articulated in the present analysis.

Appendix 2 A Note on Kohonen Self-Organizing Map

The goal of clustering is to reduce the data by grouping data items together. The K-SOM does exactly this, by replicating, in a simple way, what the human brain does. The human brain is understood to contain more than 1 billion neural cells or neurons. Information is transported between neurons in the form of electrical stimulations along dendrites, when the stimulations exceed a certain threshold. When the stimulations do not exceed this threshold, information will not be transported. Information is transported in this way to a destination where some reactions will occur. Moreover, connections between neurons are adaptive, the connection structure changing dynamically, and our learning ability is based on this adaptation, enabling faster recall of information and more efficient reactions.

Artificial intelligence techniques model the functioning of the brain using such neural networks, where the connections between neurons are 'weights'. Initially, these weights are assigned specific values. Depending on the actual output and targets, an error value is computed, which is then used for adjusting the weights between neurons. When new input data is presented to the network, the output is determined from the learning that has already occurred from the original data. In the case where no target data is available to the network, it is necessary to use unsupervised learning techniques. K-SOM is one such technique, where neural cells organize themselves in groups, according to incoming information.

Each output neuron is connected to the input vectors by a weight vector. The activation of a neuron is a dot product of the weight vector with the input vector:

$$\text{Output} = \Sigma_{i=0,\ldots,n}(w_i x_i)$$

The neuron having maximum activation is declared the winner, and the weights are updated in such a way that it will react to this particular input even more strongly next time, thus strengthening its winning position. The weight vectors will eventually converge to a point of stability where the training is set to be complete. After training, similar input vectors excite similar regions of the self-organizing map.

It should be noted that each neuron in the input layer is connected to each neuron on the map (output layer). The resulting weight matrix is used to propagate the net's input values to the map neurons. Additionally, all neurons on the map are connected among themselves. These connections are used to influence neurons in a certain area of activation around the neuron with the greatest activation received from the input layer's output.

In the beginning, the activation area is large and so is the feedback between the map neurons. This results in an activation of neurons in a wide area around the most activated neuron. As the learning process progresses, the activation area is constantly decreased and only the neurons closer to the activation center are influenced by the next activated neuron. In K-SOM, the map neurons do not change their positions on the map. This 'arranging' is simulated by changing the values in the weight matrix.

The K-SOM maps the input data from an n-dimensional space to a lower dimensional plot (usually one-dimensional or two-dimensional) while maintaining the original topological relations. The physical locations of points on the map show the relative similarity between the points in the multi-dimensional space. The idea is to repeatedly present a set of input data and update the network weights via the K-SOM training algorithm, until the network reaches some stable final configuration, usually characterized by a two-dimensional topographic representation of the n-dimensional input data.