Social Area Analysis Using SOM and

GIS: A Preliminary Research

Yan Li

College of Asia Pacific Studies

Ritsumeikan Asia Pacific University

Subana Shanmuganathan

Visiting Research Fellow

Ritsumeikan Asia Pacific University

Social Area Analysis Using SOM and GIS: A Preliminary Research

Yan Li and Subana Shanmuganathan

Abstract: Social area analysis, pioneered in the 1950s concerns a city's discrete divisions with distinctive social traits. It is a useful tool in modeling a city's socio-spatial attributes. However, its applications in urban planning and policy-making are rare. The reasons for this being 1) the use of linear correlation methods tends to oversimplify the complexity involved and often produce patterns irrelevant to real world scenarios and 2) the vast amount of labor required for visualizing spatial patterns makes the task almost impossible in many cases. In this paper, a non-linear correlation method using Self Organizing Map (SOM) techniques and GIS, to investigate a city's social areas within Beppu City, Japan is outlined as there are not many studies on social areas in Japan.   This could be expanded to other larger cities as well.

1. Social Area Analysis

Even though the so-called "social area analysis" itself was not termed until the 1950s, the study of the spatial divisions in cities can be traced back to the early 1920s' efforts by the human ecologists of Chicago School. Burgess (1925) studied the residential distribution of cities and generated the well-known concentric zone model. In this model the author classified residential areas into several types, each forming a concentric band surrounding the city center and characterized by residents of increasingly high socio-economic status. In the investigation of Chicago city, Park and Burgess (1928) revealed that the significant processes underlying the spatial division of the then growing American industrial city as similar to those found in nature. Competition among land users resulted in the invasion of the most desired parts of a city and eventually the succession of existing land use by a more dominant activity. Under free-market conditions, certain parts of the city would be occupied by the function that could maximize the use of the site,

resulting in so called "natural areas" with homogeneous socio-economic or ethnic characters, such as slum or ghettos.

There are two main streams in the studies of urban socio-spatial patterns since that of the Chicago School. One can be seen in the reports on racial segregation research. Besides of those on the study of the phenomena and policy, a large quantity is on the modification of Duncan and Duncan's dissimilarity index (1955) of measuring the degrees of racial segregation. In one of recent examples, Wardorf (1993) proposed a simulation approach that demonstrated its responsiveness to changes in urban characteristics. (i.e, spatial variations in recent increase) and changes in preference structure (ie., declining importance of distance as an impediment to relocation). Using a numerical example, he as well proved the segregation experience of minorities as more severe than previously thought. Simpson (2004) as well explicitly condemned the earlier methods as "inadequate through lack of consideration of change over time and the confounding of population change with migration" (p1). He then mathematically defined a new index called Index of Segregation (IS) to overcome this inadequacy only to be disputed by Johnston (2004) Dawkins (2004) studied spatial patterns of residential segregation based on racial composition using what he referred to as 'spatial Gini index of segregation' developed based on techniques borrowed from literature on inequality and labor force segregation. Using this Gini index, Dawkins (2006) studied the spatial patterns of black-white segregation in a sample of 237 US metropolitan statistical areas. The latter study described the Gini index as a useful basis providing a multidimensional investigation of residential segregation.

Another approach on urban socio-spatial patterns is social area analysis, the efforts on understanding the variety of spatial differentiation of cities comprehensively. Shevky and Williams (1949), Bell (1953) and later Shevky and Bell (1955) generated a system of social area analysis by using multivariate classification procedures. They selected constructs based on a theory of social trends. They regarded societal trends as (1) increasing social stratification, (2) lifestyle changes and weaker family values, and (3) more mixed population, ethnic differences and continued migrations. They established three constructs, i.e. social rank or economic status, urbanization or family status and segregation or ethnic status, to express the changing character of the American society. They measured these three constructs respectively by a number of diagnostic indices based on variables taken from census data and then mapped them to visualize the social area

types, using the case of the San Francisco Bay region.

A number of papers published soon after the original Shevky and Bell system discussed its empirical validity and generality. For example, Duncan (1955) and Hawley and Duncan (1957) expressed doubts whether the rationale for social area analysis provided a satisfactory theoretical basis for describing social differentiation in geographically delimited areas. On the contrary, Van Arsdol *et al*., (1958) used factor analysis on five American cities and proved the empirical validity of Shevky and Bell constructs in those cities.

The methodology that Shevky and Bell used in social area analysis is deductive (based on theory). Contrary to this, in the 1960s, the application of multivariate statistical technique based factor analysis to social area studies facilitated in deriving the diagnostic factors inductively (i.e. by exploratory analysis of a data set) from a larger set of variables that represent the social economic and demographic characteristics of census tracts in a city. This was called factorial ecology and the methodology was first reported by Bell (1953) and then modified in a number of later studies (see reviews in Johnston, 1971). Instead of using several indices, they used dozens of variables from census, and through statistical methods of correlation and integration, they gained several components or factors to represent the whole dataset. Finally, they interpreted the factors or used further methods such as cluster analysis to establish more understandable scores, and projected the results onto a map.

Social area analysis has been applied or tested in countries other than the USA as well. For example, McElrath (1962) studied Rome in the early 1960s. Since the 1970s, a vast amount of applications and discussions have been reported on areas outside the USA, first in the western world (See the special edition of Economic Geography Vol.47, 1971), then in the third world countries as well (Abu-Lughad ,1969: Dangschat and Blasius, 1989: LO, 1972: Yeh *et al*. 1995).

Since the 1980s, there have been only a few published works in Western literature on social area analysis. People probably regarded the topic as outdated (Sit, 1999). Even though social area analysis is a very important tool that enables analysts to understand a city, there are not many applications on real world issues, such as urban planning or other urban policy-making analysis. There are several technical problems relating to this and they are; firstly, the social area analysis methodologies adopted in the past oversimplified

the complexities of urban residential structure, especially the use of linear correlation methods often result in a pattern irrelevant to the real world issues studied. Second issue is the labor-consuming work required in visualizing the spatial pattern. This paper aims to use the combination of SOM and GIS to solve these problems.

2. Methodology

2.1 Self Organizing Map (SOM)

The late twentieth century's explosive digital data growth led data warehouse specialists and IT professionals to embark on experimenting new techniques, such as machine learning, (i.e., neural networks) for pattern recognition and forecasting. Faster switching, data compaction and better technologies contributed significantly towards the need for better techniques to analyze the abundant digital data to its potential (Mitchell 1999). Statisticians blame themselves for their attitude towards data and technology and being slow in adapting to recent changes (Pregibon and DuMouchel 2001).

Artificial neural networks (ANN) are biologically inspired approaches to intelligent information processing methodologies. They provide a means to incorporating innovations and flexibility into conventional computing there by to solving real world problems that cannot be solved by conventional computing methodologies (Amari 1995). Increasingly, complex problems of modern day and human expectations from computers continue to demand innovations in this rapidly changing field of intelligent information processing and communications. (Aleksander 2000: Kasabov, 1999, 2000)

A Self Organizing Map (SOM) is a feed forward artificial neural network. It uses an unsupervised training algorithm to perform non-linear non-parametric regression as shown in figure 1. It is capable of projecting multidimensional datasets onto low, usually 1- or 2-D displays while preserving the useful information within the raw data, in doing so they enhance analyst ability to extract knowledge in the form of novel patterns, structures and correlations hidden in the data.

The kind of 'feature extraction' seen in SOM clustering is generally impossible by conventional methods that consist of limited abilities for revealing novel patterns in low dimensional datasets. Conventional statistical methods are useful in producing simple statistics, such as mean, standard deviation, maximum and minimum values, of low dimensional datasets. Even the complex conventional multidimensional data analysis

methods, such as clustering and projection techniques, are increasingly perceived to be inadequate in coping with the ever increasing digital data that doubles in volume every 20 months (Roland et al 1996)

The SOM algorithm, first introduced by Teuvo Kohonen was developed from the basic information processing modeling in the human brain's cortical cells, known from the neuro-physiological experiments of the late twentieth century (Kohonen, 1982). In the training process of the algorithm, initially the SOM output layer units or nodes (figure 2) are assigned with a random set of vectors, usually referred to as a code book. Each set of the input vectors is then presented to the SOM input nodes and matched against the output units to find the best matching unit (BMU) in the code book. Once a BMU is found, the particular set of input vectors is assigned to it, and then that output unit vector values are adjusted more close to the values of the input set. The neighboring units of this BMU are also adjusted close to the values of the latter. Similarly, the whole set of input data are assigned to their BMUs in the output layer, mapping the similar input data vectors together on the 2D display with most of the original attributes preserved. Hence, the trained SOM display enables analysts to view any implicit previously unknown useful knowledge within the raw data in the form of patters, structures and relationships.

Kohonen's SOM techniques provide an excellent tool for analyzing disparate multidimensional datasets from different sources. Using SOM techniques, multi sourced datasets, even with inconsistent labeling (i.e., different formats), could be collectively analyzed to discover implicit, previously unknown useful knowledge embedded in the complex data set. SOMs are increasingly seen as better data clustering, visualization, and reduction (with their topology preserved) tools in data mining and knowledge discovery applications. SOM applications to exploratory data analysis have produced considerable success in many fields, such as medical, financial market, customer segment, industrial engineering, manufacturing to name a few from the lists of thousands of SOM applications in Kaskiy et al (1998) and Oja et al (2002).

Bação et al (2005) investigated the most commonly used conventional k-means clustering techniques with different algorithms generally applied to solving GIS problems and proved Kohonen's SOM based clustering as the most convenient method under proper training parameters. They clearly explained in their publication that SOM could serve even under rigorous conditions, especially, in resolving complex k-means issues relating to

the correct use of initialization procedure that ultimately determines which part of the solution space should be searched.

In this paper, we illustrate how SOM techniques, together with GIS, could be applied to visualize social areas using population census data for the city of Beppu, Oita Prefecture, Japan.

2.2 Case study area and variables

Facing Beppu Gulf in the East and surrounded by mountains in all the other directions is the urbanized area of Beppu City which is about 25km^2 in area with a population of 123,000 (data in April, 2006). Population census is carried out every five years in Japan. Beppu 2000 census was conducted in October and consists of details on population aggregated in 23 tables for each census units. In this research, data on age, sex, household, housing, length of stay in the house (dwelling), mode of earning are used. The following 90 variables for 163 census units (excluding the 17 census units with 0 residents), are analyzed to study the social areas in the city (no details on the city's geo-referenced location was added in the analysis);

− Total population, female, male, number of households and density
− Female at 5 year age intervals. 22 groups in total. For the coding, we use F (0-4) for female aged 0-4, F (5-9) for female aged 5-9, etc.
− Male at 5 year age intervals, i.e., M (0-4): Male aged 0-4. 22 groups in total.
− Female foreigners F_F, and Male foreigners M_F.
− Housing data; own, rented and rented rooms; each consisting of House, Longhouse, 1&2 Floor apartment, 3-5 Floor apt, 6-10 Floor apt, >=11 Floor apt: 18 categories in total. For the coding, we precede "O_" for owned housings, "R_" for rented housings and "R_r_" for rented rooms for each type of housings.
− Dwelling details for female and male, each consists of (1) less than 1 year, (2) 1-4 years, (3) 5-9 years, (4) 10-19 years, (5) longer than 20 years, and (6) from birth, 12 categories in total. Coding "D_M_<1"means female population living in the census unit less than a year.
− Source of income, consists of salary only, pension only, salary+some, pension+some, allowance, agriculture, non-agriculture business (mainly family business), in-home work and others, 10 categories in total.

7

3. Social Area Analysis on Beppu

A 200 node SOM was created with the above-mentioned variables chosen from Beppu census 2000 using default software training parameters of Viscovery, commercial data depiction software (eudaptics software gmbh). The default training parameters were chosen as the software is automated to select the best parameters for the set of data being trained. The SOM was initially divided into 2 clusters and was further subdivided progressively into 8 clusters to see Beppu's social areas and their attributes. The two cluster SOM and its components for age, housing types, dwelling length and mode of earnings clearly distinguish the 163 census units into two broad clusters, C1 and C2 (Figure 3a).   The progressive clustering i.e., three, four and five cluster SOMs, further differentiate the C1 units into 3 sub clusters (Figure 3b). In the six cluster SOM, C2 is split into 2 clusters (Figure 3c). The seven and eight cluster SOMs again differentiate C1 units into further finer divisions (Figure 3d). The coding for each cluster is shown in figure 4, with C1 units divided into 6 subdivisions, C1a to C1f, and C2 units into 2 subdivisions, C2a and C2b.

3.1 The two cluster SOM

Figure 5 shows the spatial distribution of C1 and C2 units, the result of two cluster SOM, on GIS. The city is roughly divided into 3 parts, with C2 units located in the middle and C1 units located in the North and South. This echoes the developmental history of Beppu City, result of an annex of two old administrative areas, Kamegawa Town and Beppu City in 1933, and C2 units show a newly developed area emerged following the annexation. Since then C2 area gradually grew drawing more people, especially after the introduction of new land use planning projects led by the government since the 1970s.

As expected C2 units consist of dwellers with more under 25-44 age groups and high 0-14 children ratios (for both female and male, Figure 6a), and higher salary only households (see Figure 6b). On the other hand, the older settlements C1 units consist of higher ratios of 55-99 age groups and high pension only percentage. Housing is commonly owned 1-2 floor single houses and rented 3-5 floor houses. However, owned housing is higher in C1 units than that of C2 (see Figure 6c).   The difference of D_F_>20 and D_M_>20 (Females and males dwelling in the unit for more than 20 years) shown in Figure 6d also reveals that the two cluster SOM has divided the city in a meaningful

manner and classified the newly developed units from that of the older units.

3.2 Sub clusters of C1

As mentioned above, further clustering of C1 and C2 produced 8 clusters, in which C1 units are divided into 6 sub clusters. The spatial distribution of these subdivisions is shown in Figure 7a. The terrain in Beppu City is chiefly a slope from the mountainous of the west to the sea in the east. C1d, C1e, and C1f are distributed in the mountainous part on the west. In the main urban region, the coastal areas of the east, consisting of C1c and C1b are different from the surrounding C1a and C2 clusters.

C1f and C1c are distinctive with a high ratio of people living on allowance with regards to the income type (Figure 7b). Their age ratio shows high values within 15 to 29 year range (Figure 7c), and the housing types are mainly rental apartments (Figure 7d). These are the areas where Beppu's two main universities, Ritsumeikan Asia Pacific University (APU) and Beppu University (BU) are located. APU is in C1f and BU in C1c. APU, located in the isolated mountainous area, was launched only several months before the census. It consists of 50% foreign students and provides dormitories located in the immediate vicinity of the campus for the first year students. Meanwhile, BU is located in a residential area with students living amongst other citizens. This is the reason why C1f has extremely high proportion of young people and foreigners.

The most prominent characters of C1d are high pension income, single house and long dwelling periods (Figure 7b, d and e). This region is a farming area (also can be seen with a low density in Table 1) and there are a lot of senior citizens who were engaged in agriculture in their younger days but now living on pension.

C1e consists of a high ratio of senior citizens living there only for one to four years (Figure 7c and 10e). A lot of homes for the aged are distributed in this region.

In C1b, ratio of the non-agriculture business (mainly family business) income is seen to be relatively large (Figure 7b) and senior citizen's ratio is high only second to C1e (Figure 7c). Moreover, the dwelling length is the second longest next to C1d (Figure 7e). This region is the old center of Beppu, adjacent to the Japan Railway Beppu Station. It is a traditional hot spring tourism town where a number of aged individual proprietors, such as the hot spring, sightseeing, and shopping streets, reside.

9

C1a units are most widely distributed, and their social attributes are not so clear when compared with that of the other C1 subdivisions. It will be analyzed in the following section with subdivisions of C2.

3.3 C1a and subdivisions of C2

In the six cluster SOM, C2 units are separated into 2 groups (C2a and C2b), Figure 8a shows their distribution. C2b is located on the coastal side of the city. In this region, the land readjustment project of government was implemented in the 1970s, and a lot of rental apartments are located here. It is the youngest area, while aging is a feature in C1a., C2a is in between C2b and C1a in terms of their age groups (Figure 8c).

The salary only households decrease in the ascending order within C2b, C2a, and C1a, meanwhile, the pension only households show an increase (Figure 8b). Moreover, the ratio of owned single house is more and rented apartment is low in C1a (Figure 8d). C2b is the shortest, and C1a has the longest period for dwelling years.

The difference between C2a and C2b is that the latter consists of more nuclear families: young parents (aged 30-35) and children less than 10 years old. They depend more on salary type income and live in multi-floor apartments, with relatively higher proportion of dwelling length from 1 to 4 years. However, in C2a units, there is a high proportion (nearly 50%) of aged dwellers. They have less young children and comparatively with higher non-agricultural incomes, a high share (nearly 40%) of them live in owned single houses and with dwelling periods longer than 20 years. They had moved into the area when the government-led land use development program was introduced in the 1970s, probably with nuclear family status and settled there since then.

3.4 Social areas of Beppu

Beppu's social areas are summarized in Figure 9. This GIS figure also shows the outlines of buildings, which look like shadows underneath. From this figure, one can see the less populated C1d, homes of the aged and C1e, the formerly formers now aged areas. C1f, where APU is seen with its majority population, is also a very special social area, because it is new, located on a top of a hill and, isolated from the other residents. In the urbanized area, C1a and C2a share the greatest part. Through the above analysis, we could describe these areas as the host society of the city, distinctive of their ages due to the

government-led development project.

There are 3 social areas within the urbanized area that are also different from the host society. C1e, the BU area, has more young allowance dependent students. C2b, the multi floor apartment area that consists of more young nuclear families and this is the youngest area in the city. C1b is the old town center and has aged family tourism business owners living and working there.

Some other areas also have the similar social attributes to the above social areas. Circles A and B in Figure 9, identified as C1c, are the other main places for university students, because the former is near the bus center and the latter the railway station. Circle C is Kanawa hot spring area, another traditional tourism spot in the city. That is why it belongs to the same classification as C1b, the aged family tourism business owners' area in the old center of Beppu.

Conclusion

Even though social area analysis is an important tool that enables city planners to understand a city's socio-spatial attributes, there are not many applications in related disciplines, such as urban planning or other urban policy-making due to major technical issues. In Japan, social area analysis is seldom applied.

This case study explained an application of SOM and GIS methods to social area analysis on Beppu, a small city in Japan. In this research, all variables were obtained from the census data without checking the significance or making any other data refinement except for standardization. Not a single subjective parameter was used other than the standard training default settings of the software in using the unsupervised artificial neural network paradigm referred to as Kohonen's self-organizing map. All the variables were selected only because they were found to be relevant and best descriptive of the social attributes chosen for the analysis. Therefore, the analysis by SOMs and GIS could be described as more promising, and objective in compared with the traditional statistical techniques, as the later requires advanced knowledge in variable selection and data processing during the analysis.

One might have noticed that in the analysis, there were no geo-coding (location) variables included in the SOM clustering, and yet the discrete geographical distribution of Beppu's social areas was confirmative with that of the GIS screen. The proposed inductive

method explained in this paper provided a practical approach on how social areas observed in other countries could be established in Japanese cities as well. The research showed that in Japan, age or life stage is the most important factor that determines the socio-spatial divisions; the SOM clustering firstly separated the city into C1 and C2 units, due to the land development that was introduced in the 1970s, because development projects often bring certain types of families into the area. Occupation/income type is also important; C1a has more proportion of salary only, C1b more non-agricultural business/family business, C1c and C1f more allowance, C1d more pension only, and etc. Although foreign students are seen living in certain areas (C1f and C1c), no conclusion could be made on the influence of ethnicity from the analysis, due to the unavailability of data on it in the 2000 census and Japan has only about 1% of non-Japanese residents.

References

Abu-Lughod, J. L. (1969) 'Testing the theory of social area analysis: the ecology of Cairo, Egypt', *American Sociological Review* 134: 198-212.

Aleksander, I. (2000) 'I compute therefore I am', in *BBC Science and Technology*, http://news.bbc.co.uk/2/hi/science/nature/166370.stm, retrieved on December 7, 2006.

Amari, S. (1995) 'Foreword', in Kasabov, N. K. (ed.) *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*. The MIT Press: xi.

Bação, F., Lobo. V. and Painho, M. (2005) 'Self-organizing maps as substitutes for K-means clustering', in Bacao, F., Lobo, V., Painho, M., Sunderam, V.S., van Albada, G., Sloot, P. and Dongarra, J. J. (ed.), *International Conference on Computational Science 2005. Lecture Notes in Computer Science, Vol. 3516.* Springer-Verlag, Berlin Heidelberg, pp. 476-483.

Bell, W. (1953) 'The social areas of the San Francisco Bay Region', *American Sociological Review* 18(1): 39-47.

Burgess, E. W. (1925) 'The growth of the city', in Park, R. E. (ed.), *The city*, University of Chicago Press, Chicago.

Dangschat, J. and Blasius, J. (1989) 'Social and spatial disparities in Warsaw in 1978: an

application of correspondence analysis to a "socialist city"', *Urban Studies* 24: 173-91.

Dawkins, C. J. (2004) 'Measuring the spatial pattern of residential', *Urban Studies* 41: 833-851.

Dawkins, C. J. (2006) 'The spatial pattern of black-white segregation in US metropolitan areas: an exploratory analysis', *Urban Studies* 43:1941-1969.

Duncan, O. D. (1955) 'Review of Shevky and Bell, Social Area Analysis', *American Journal of Sociology*, 61:84-85.

Duncan, O. D. and Duncan, B. (1955) 'A methodological analysis of segregation indices', *American Sociological Review* 20:210-217.

Economic Geography, Vol. 47 (1971) Supplement: 'Comparative Factorial Ecology'.

Eudaptics software gmbh, Kupelwiesergasse 27, A-1130 Vienna, Austria http://www.eudaptics.com, retrieved on 13 March, 2007.

Kasabov, N. (1999) 'Evolving connectionist and fuzzy connectionist systems: theory and applications for adaptive, on-line intelligent systems', in Kasabov, N. and Kosma, R. (ed.) *Studies in Fuzziness and Soft Computing*, Physica-Verlag, pp.111-144.

Hawley, A. H. and Duncan, O. D. (1957) 'Social area analysis: a critical appraisal', *Land Economics* 33:337-345.

Johnston, R. J. (1971) 'In some limitations of factorial ecologies and social area analysis', *Economic Geography*, Vol. 47, Supplement: Comparative Factorial Ecology, pp. 314-323.

Johnston, R. J., Poulsen, M. and Forrest, J. (2005) 'On the measurement and meaning of residential segregation: a response to Simpson', *Urban Studies* 42(7):1221-1227.

Kasabov, N. K. (2000) 'On-line learning reasoning, rule extraction and aggregation in locally optimized evolving fuzzy neural networks', *Neurocomputing* 41(1-4):25-45.

Kaskiy S., Kangasz, I. and Kohonen, T. (1998) *Bibliography of Self-Organizing Map (SOM) Papers: 1981-1997*, http://www.cis.hut.fi/nnrc/refs/, retrieved on 13 March 2007.

Kohonen, T. (1982) 'Self-organised formation of topographically correct feature maps'. *Biological Cybernetics*, 43:59-69.

Koua1, E. L. and Kraak1, M. J. (2004) 'Geovisualization to support the exploration of large health and demographic survey data', *International Journal of Health Geographics* 3:12.

Lo, C. P. (1972) 'A typology of Hong Kong census districts: a study in urban structure', in SIT, V. F.S. (ed.) (1985): *Urban Hong Kong*. Summerson, Hong Kong, pp.26-43.

McElrath, D. C. and Dennis C. (1962) 'Social areas of Rome: a comparative analysis', *American Sociological Review*, 27(3): 376-391.

Mitchell, T. M. (1999) 'Machine learning and Data Mining', *Communications of the ACM*, 42(11):30-36.

Oja, M., Kaski, S. and Kohonen, T. (2002) *Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001*, http://www.soe.ucsc.edu/NCS/VOL3/vol3_1.pdf, retrieved on 13 March 2007.

Park, R. E and Burgess, E. W. (1928) *Introduction to the Science of Sociology*, Chicago, Univ. of Chicago Press.

Pregibon, D. and DuMouchel, W. (2001) *What Is Data Mining? (An Why Should Statistician Care?)*. 53rd Session of the International Statistical Institute, Seoul, Korea.

Roland, H., Keim, D. A., Panse, C., Schneidewind, J. and Sips, M (1996) *Finding Spatial Patterns in Network Data*, pp. (8-1)-(8-12).

Shevky, E. and Bell, W. (1955) *Social Area Analysis*. Stanford University Press, Stanford, CA.

Shevky, E. and Williams, M. (1949) *The Social Areas of Los Angeles*. University of California Press, Los Angeles.

Simpson, L. (2004) 'Statistics of racial segregation measures: evidence and policy', *Urban Studies* 41: 661-681.

Sit, V. S. F. (1999) 'Social Areas in Beijing'. *Geografiska Annaler. Series B, Human*

Geography 81(4):203-221.

Van Arsdol, M. D., Camilleri, S. F. and Schmid, C. F. (1958) 'The generality of urban social area indexes', *American Sociological Review* 23(3):277-284.

Waldorf, B. S. (1993) 'Segregation in urban space: a new measurement approach', *Urban Studies* 30(7):1151-1164.

Yeh, A. et al. (1995) 'The spatial structure of Guangzhou' *Urban Geography* 16(7): 595-621.
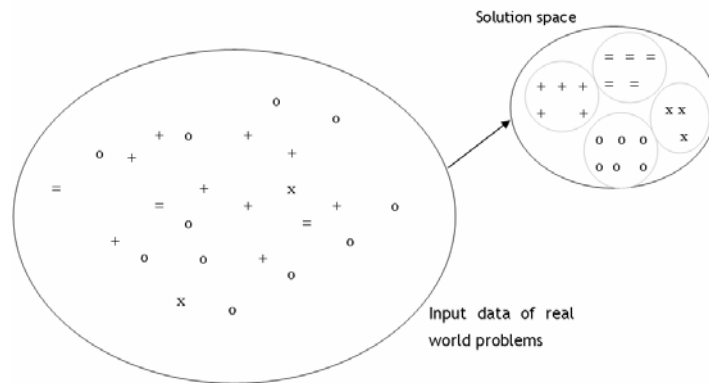
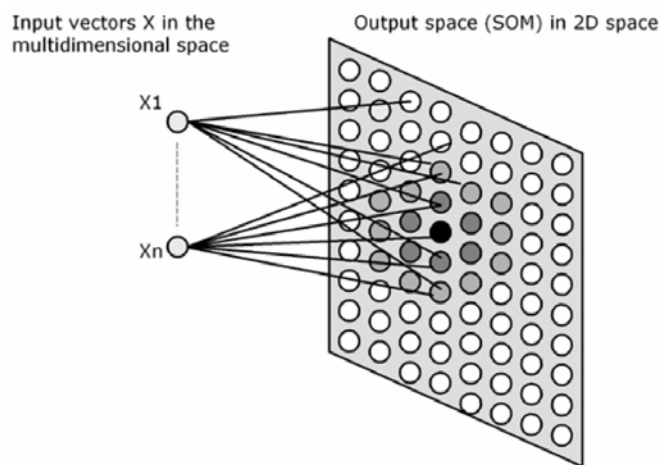*Figure 1: diagram showing a mapping process of input data onto a low dimensional solution space.*

Figure 2: Diagram of SOM layers with the training process. Source; Koual and Kraakl, (2001)
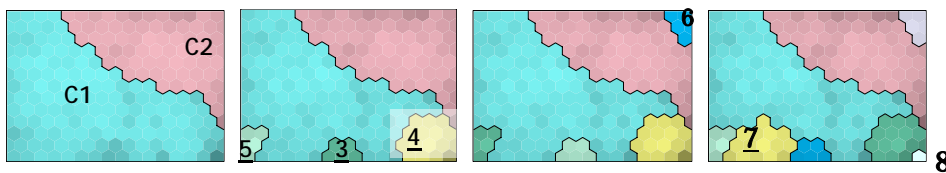
| *Figure 3a* | *Figure 3b* | *Figure 3c* | *Figure 3d* |

Figure 3a; The two cluster SOM split the units into C1 and C2.

Figure 3b; 3 new clusters of C1 identified in the three, four and five cluster SOM respectively

Figure 3c; The six cluster SOM, with C2 split into 2 clusters.

Figure 3d; The seven and eight cluster SOM, with 2 more new clusters are found in C1.

Figure 3; SOM clusters
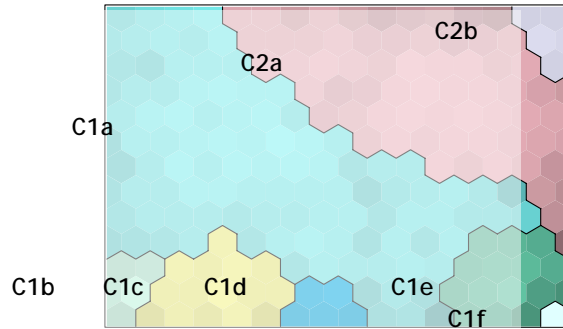
Figure 4; Coding for each clusters

The center of former Kamegawa town

The center of former Beppu City
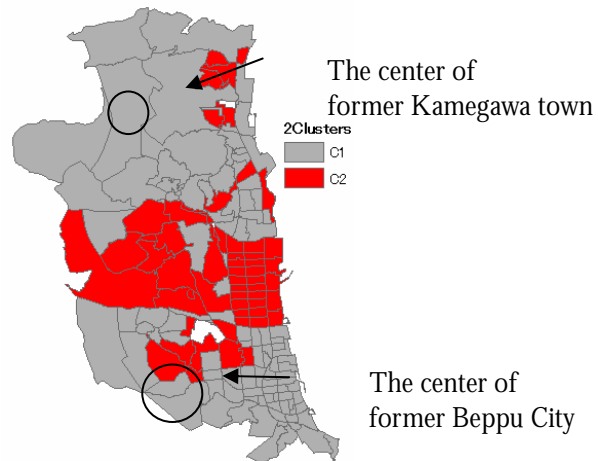
Figure 5; Distribution of C1 and C2 units on GIS
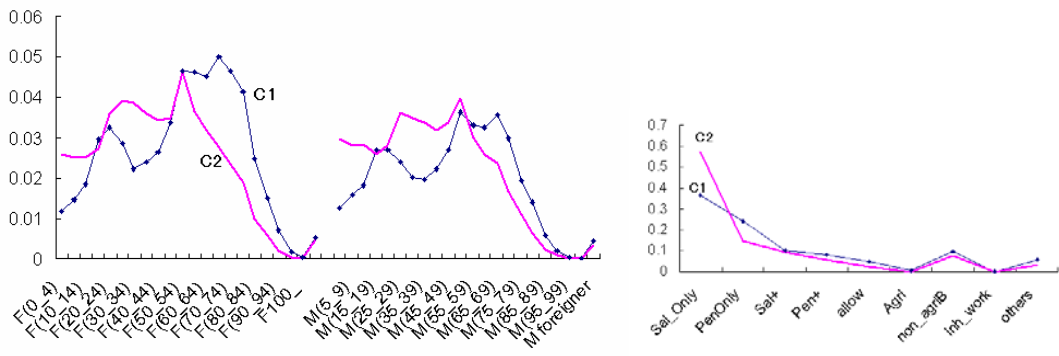
Figure 6a   5 year age group proportions (left: female and right: male) Figure 6b; Proportions of income types
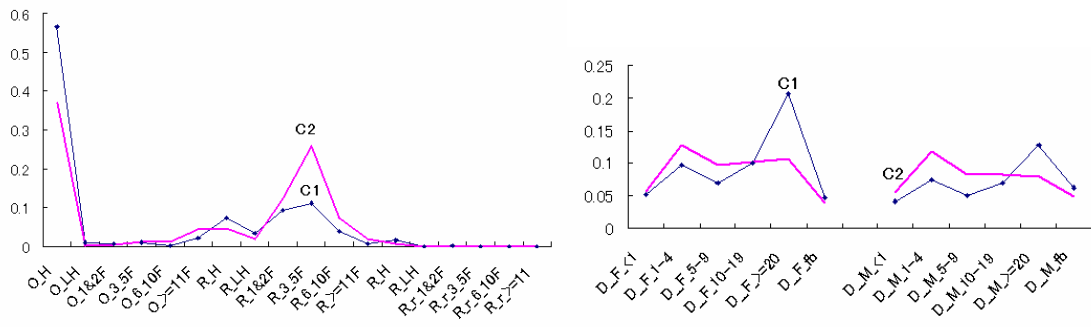
Figure 6c;   Housing types   Figure 6d; Dwelling length

Figure 6; Social distinctions of C1 and C2 units
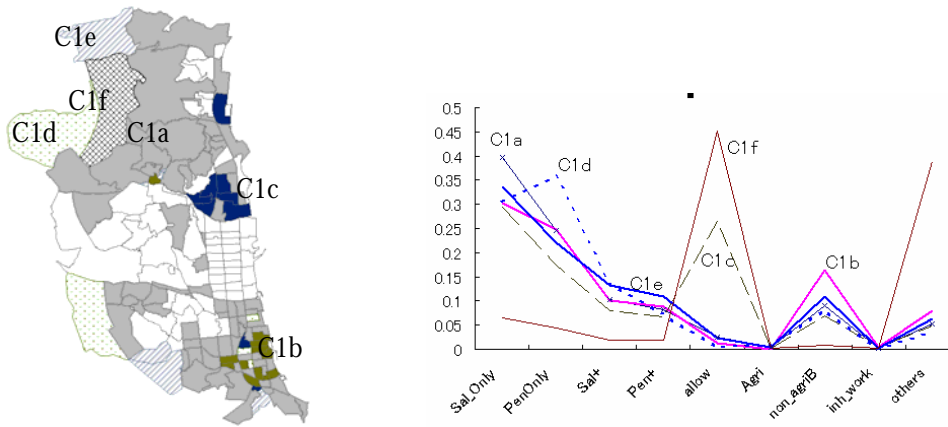
Figure 7a; Spatial distribution of C1 subdivisions. Figure 7b; Proportions of income types
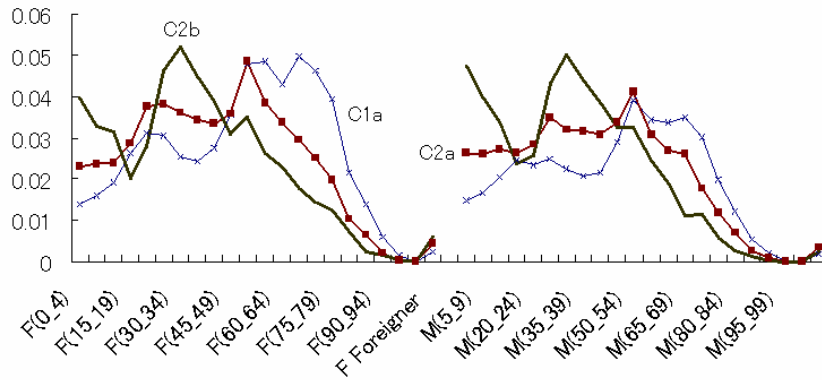
(White-color filled units are C2)
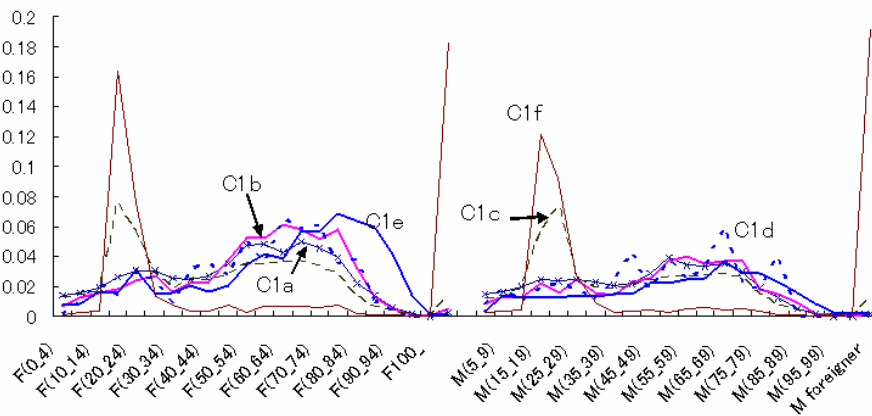
Figure 7c; 5 year age group proportions (left: female and right: male)
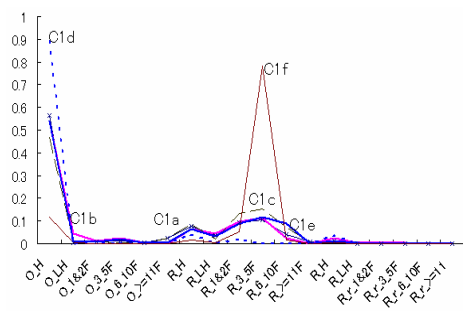
Figure 7d; Housing types *Figure 7e; Dwelling length*
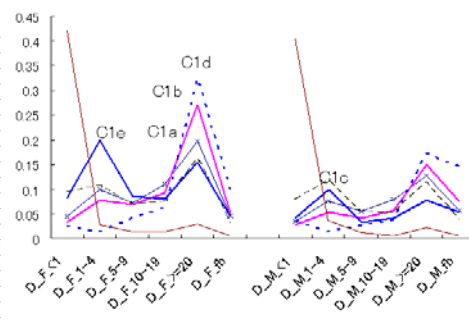
Figure 7; Social distinctions of C1 subdivisions

Table 1; Average population and density of each clusters

	C1a	C1b	C1c	C1d	C1e	C1f	C2a	C2b
Population (persons)	666	430	944	74	494	762	1,129	734
Households	283	200	453	29	161	91	429	288
Population Density (persons/ha)	76.4	105.3	87.8	15.5	99.5	30.7	81.5	87.7
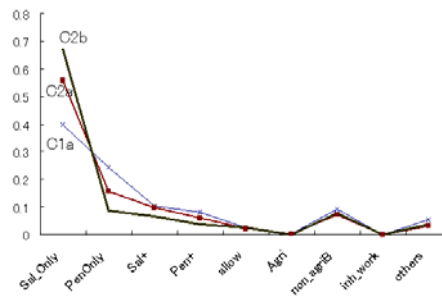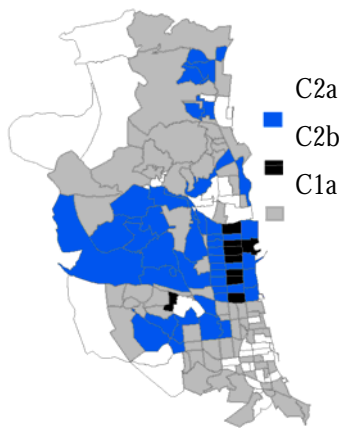
Figure 8a; Spatial distribution of C1 sub divisions   *Figure 8b; Proportions of income types*

Figure 8c; 5 year age group proportions (left: female and right: male)

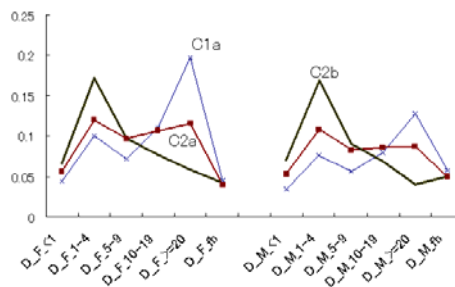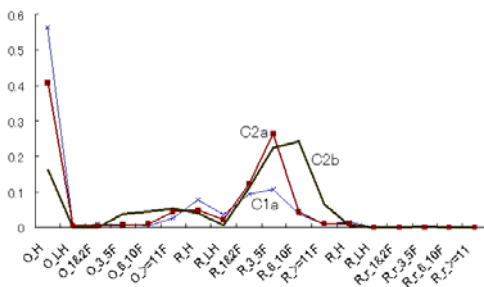*Figure 8d; Housing types*   *Figure 8e; Dwelling length*
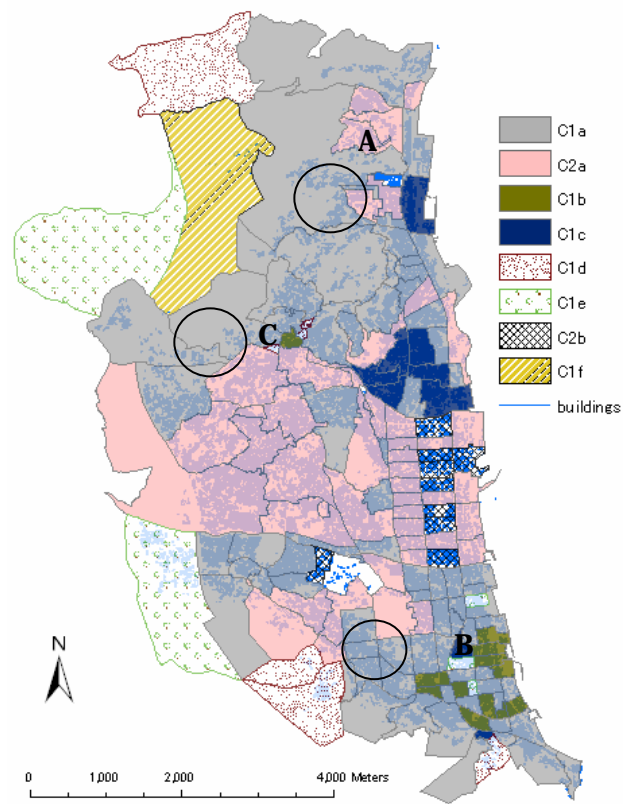
Figure 8; Social distinctions of C1 units C2 units

Figure 9; Social areas of Beppu City