INTEGRATING SPATIAL STATISTICS AND GIS FOR REGIONAL STUDIES IN THAILAND

TRAN Hung, Dr.; Yoshifumi YASUOKA, Prof.
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
Tel: +81-3-5452-6415 Fax: +81-3-5452-6410 E-mail: tranhung@iis.u-tokyo.ac.jp

Abstract: This paper utilizes spatial statistical techniques integrated with GIS to analyze regional development problems associated with the rapid expansion of the regional city of Chiang Mai – Lamphun in the Northern Thailand during the 1986–1994 period. A set of combined GIS procedures is used to integrate up-to-date information from remotely sensed, physio-graphical and socio-economic data in order to create a comprehensive spatial database for the study region. Then, with a combination of logical and statistical operations, reliable cross-sectional spatialized study variables describing physical, socio-demographic and economic aspects of the regional development at the sub-district level are derived for subsequent spatial statistical analysis. The spatial patterns of regional development in the Chiang Mai – Lamphun area over time are traced out in various development aspects using spatial autocorrelation and spatial association statistics. The unbalanced regional development pattern and significant urban-rural disparity associated with rapid urban growth in the region are explored. Then, the spatial statistical modeling is adopted to simultaneously model spatialized physical and socio-economic factors affecting the urban-rural disparity and urban-rural interaction in order to provide insights into the region-wide spatial impacts on rural surroundings of the rapid economic growth in Chiang Mai – Lamphun urban centers. This paper demonstrates that GIS can help to solve the "data barriers" problem in developing countries, while its integration with spatial statistics is useful in gaining further insights into regional development problems.

Keywords: Spatial Statistics, GIS, Regional Development, Developing Countries, Thailand.

1. INTRODUCTION

Up-to-date and reliable information is vital for the management of a region's human and natural resources and for dealing with regional development decisions that have a spatial context (Klosterman, 1995). A comprehensive information base could reduce uncertainty and enhance decision-making. Managers and policy makers may wish to integrate social, economic and environmental data in order to formulate strategic development plans (Kliskey, 1995). In developing countries, however, the *data barriers* are still obvious due to both institutional and technical reasons. As institutional issues are being recognized and governments start to invest millions of dollars in collecting data, the data management and usage are still far from satisfactory level. Information on various aspects of regional development – social, economic and environmental data – is originally collected for different purposes, at different scales, at different time frames and with different underlying assumptions about the nature of the phenomena. This creates technical difficulties to the integration of social and environmental data, and explains the scarcity of successful empirical researches on regional development analysis in developing countries.

In recent years, Geographic Information Systems (GIS) have become an important tool for regional and urban research. As about 80-90% of data collected and used for regional and environmental information systems are related to geography (Huxhold, 1991), GIS provides such an integrated computing environment for social and environmental data integration. It is widely recognized that GIS provides a large range of analytical capabilities to operate on topological relationships or spatial aspects of the geographical data, on the non-spatial attributes of such data, or on non-spatial and spatial attributes combined. GIS facilitates the integration of disparate data sets, creation of new and derivative data sets, and development and analysis of spatially explicit variables. Furthermore, the integration of GIS with spatial statistical analysis has the potential to become a powerful analytical toolbox, enabling regional and social scientists to gain fundamental insight into the nature of spatial structures of regional development (Brown, 1996). Many efforts have been made to apply GIS, spatial statistics and modeling to regional studies. Key contributions to this emerging literature include those by Getis and Ord (1992), Anselin (1994, 1995), Chou (1995), and Bao *et al*. (1995), who contributed to the building of theoretical concepts.

The objective of this empirical study is twofold: (1) to demonstrate GIS capability in integrating disparate datasets and creating a comprehensive regional spatial database, and (2) to analyze regional development problems associated with the rapid expansion of the regional city of Chiang Mai – Lamphun in the Northern Thailand. Specifically, this study attempts to provide answers to the following questions:

1. What is the spatial pattern of urban growth centers in the Chiang Mai - Lamphun area?
2. To which spatial extent could the urban growth centers have impacts on rural surroundings?
3. What are the socio-economic factors explaining the pattern and intensity of urban core - rural periphery interactions?

2. STUDY AREA

The Chiang Mai - Lamphun area is defined as the first regional growth center for the North, according to Thailand's Fourth National Development Plan (1977-1981) in the line with decentralized/regionally balanced urban development. It has been continuously included as one of the Industrial Promotion Zones of subsequent Plans. The expected development impacts of those growth centers are to provide services to and induce growth in the hinterland through diffusion of innovations and strengthening forward and backward linkages (Lo, 1981). However, as indicated by Sharma (1984) and Potter & Unwin (1989), the tendency for the polarization forces is stronger than trickle-down forces, which may cause spatial structure of a dominant core with a dependent periphery, and widen income inequalities.

The study area is located approximately between latitude $18^{0}08'$ N and $19^{0}06'$N and longitude $98^{0}30'$ E and $99^{0}25'$ E, with a total area of 5806 km^2. Administratively, the study area is composed of ten districts of Chiang Mai Province and six districts of Lamphun Province, resulting in a total of 146 subdistricts or *tambols*[1] (Figure 1). Topographically, the area covers most of the Chiang Mai basin associated with the Ping River, surrounded by hilly to mountainous terrain. This natural condition allows one to consider the area as an independent *functional economic area*[2] in spatial regional analysis. The transportation network is concentrated around Chiang Mai City and Lamphun Municipality, indicating a pattern of region with two major urban growth centers (Figure 2). In fact, Chiang Mai City has become the monocentric economic, financial and cultural center for the whole region. Moreover, during the last two decades, the area has experienced rapid urban expansion, with ever more rapid industrial establishments (Suwan *et al*., 1992). From 1986 onward, the Northern Industrial Estate with 87 projects implemented (as of December 1994) was built in Muang Lamphun District. Lamphun Municipality, thus, has emerged as a new industrial center in the area. The average income was rising as labor shifted from the agricultural to the manufacturing sector. After large investments had been made in the Chiang Mai - Lamphun urban centers during recent decades, it is of interest to investigate what impact they have had on rural areas, and whether any 'trickling-down' effect has occurred. Moreover, an understanding of development patterns, phases and constraints and an appraisal of how far urbanization and industrialization could contribute to the development of its rural hinterland is necessary to arrive at recommendations for development planning in the region, which might be applicable to other regional cities as well.

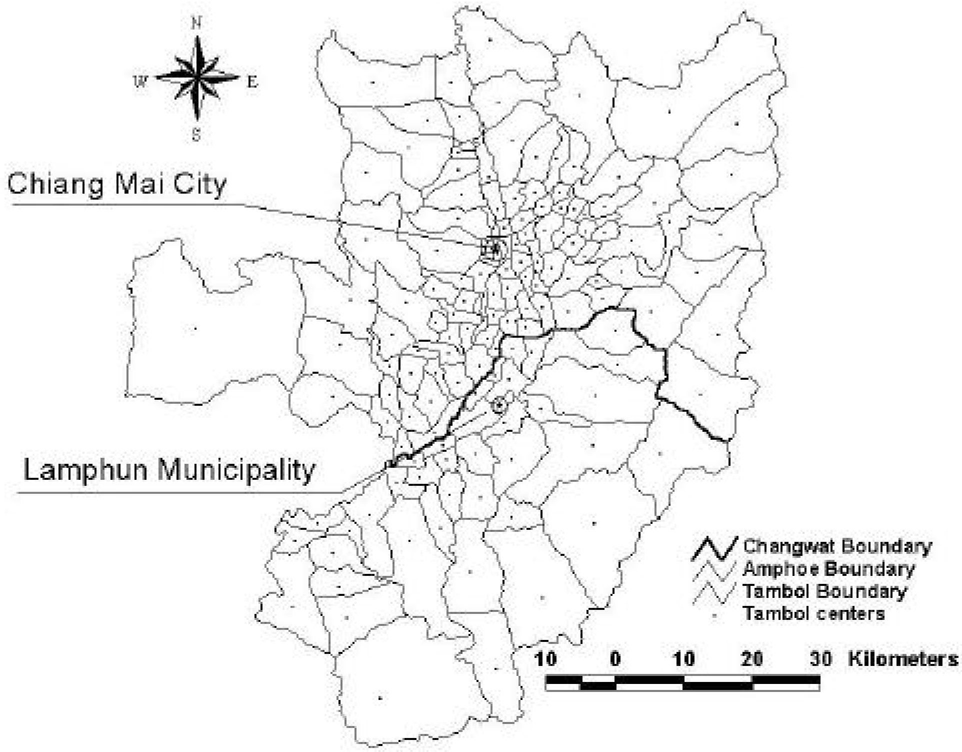

Figure 1. Administrative boundaries in the Chiang Mai – Lamphun area

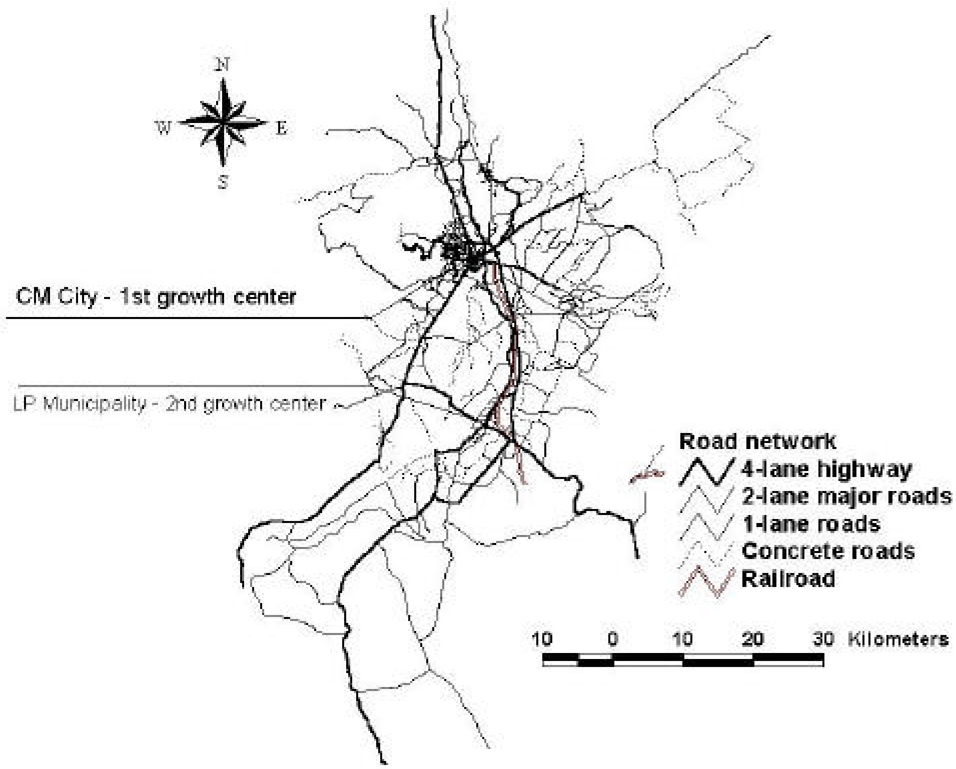

Figure 2. Transportation network in the Chiang Mai – Lamphun area

3. GIS SPATIAL DATA INTEGRATION

3.1 Data Collection and Database Building

With regional development issues in the focus, data in Chiang Mai – Lamphun area are collected from various government offices in the form of physio-graphic data (e.g., topographic, administrative, land-use, industrial locations as well as transportation network maps) and socio-economic indicators. Satellite TM images of different dates (1986 and 1994) are used to provide up-to-date land cover and land use information. The image processing procedures are used to classify georeferenced remotely sensed images and to produce updated classified land-use maps and transportations networks maps in raster format (ERDAS Imagine). The spatial physio-graphic data sets from paper maps are classified, digitized and fed into vector GIS (Arc/Info). The classified images are, then, integrated with other environmental and societal data sets through raster-to-vector data conversion to update and build time series data. As the loose data integration approach was used in this research, data are managed in both raster and vector format and convert from one to another when data analysis required.

The major source for socio-economic data is the National Rural Development Database (NRD-2C), which provides surveyed data at village level after each two years from 1986 composing of more than 100 economic and social indicators. The data are also collected from other government documents, statistical records at provincial and municipal offices. They are selected, reclassified and combined based on the basic administrative unit IDs – village code number in dBase IV format. A program in Visual FoxPro is written to automate the process of extracting, normalizing and combining socio-economic indicators including population, income, education, health, natural environmental conditions, services, agriculture and industrial activities, work force, capital investment, employment, etc. The detailed database building procedures are discussed in Tran (1998).

3.2 Integrated GIS Database Management

As data management in GIS facilitates the integration of diverse data sets and determines the analyses possible with those data, some data transformation routines are built in this research to facilitate the conversion of physio-graphic and socio-economic data to a common spatial structure (e.g., set of areal units). The spatial physio-graphic data in GIS database are to summarize/regionalize by administrative units in order to be compatible with socio-economic data. Based on internal homogeneity criterion, *tambol* is chosen as basic spatial unit for data integration for this study.

Compilation of socio-economic data: The socio-economic data are aggregated from village to *tambol* level, and are normalized as relative shares of the total population of each respective *tambol*, in order to further reduce the effect of unequal sizes of *tambol*.

Creating Spatial Indicators within GIS: The spatial physio-graphic data such as land use types, road networks, irrigation networks, industrial factories are aggregated to tambol level using spatial overlay and logical-statistical analysis functions in GIS (Arc/Info). Some accessibility measures such as distance from residential areas to nearest roads and nearest factories are derived through GIS spatial joins functions utilizing the locational information of data. The common summarizing / regionalizing procedures are presented in Figure 3.

Comprehensive spatial GIS database: With all aggregated socio-economic data and regionalized spatial physio-graphic data to common tambol level, the GIS join function through a key item – *tambol* ID – is used to complete the comprehensive spatial GIS database for Chiang Mai – Lamphun area. The GIS database, thus, contains comprehensive spatial information characterizing development states of the Chiang Mai – Lamphun area (in 1986 and 1994) for each tambol in terms of:

- *spatial physio-graphic data*: % of urban land-use, industrial land-use, agricultural land-use, road length density, irrigation length density, median distance from industrial land to closest residential areas, median distance from residential area to the nearest road;
- *demographic aspect*: population density; *economic aspect*: average household property taxes, travel time to nearest town and commercial center, % of farmer, per capita number of vehicles, number of factories, per capita industrial capital investment, % factory employees, average household income, % people working far from home; and *social aspect*: level of primary education, secondary education, illiterate rate, etc.

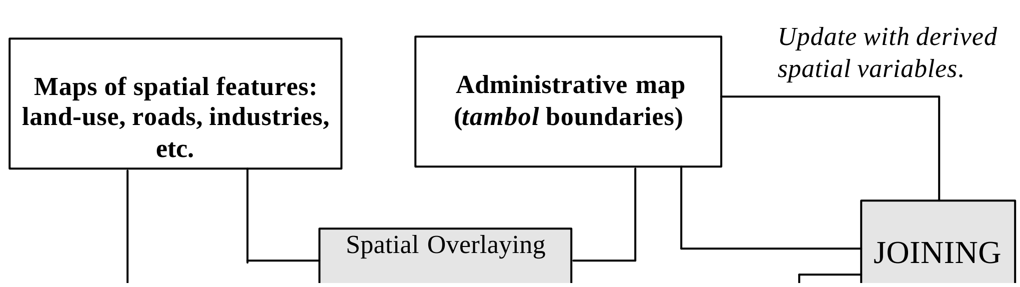

Figure 3 GIS procedure to regionalize the spatial physio-graphic indicators into *tambol* level from Chiang Mai - Lamphun spatial database.

4. SPATIAL STATISTICAL ANALYSIS

The resulting spatial data sets in unique format are useful for further empirical analysis of regional spatial development patterns using various multivariate statistical and spatial statistical analysis techniques. According to Potter and Unwin (1989), we explore development impacts of an urban growth center on rural surroundings focusing on three main headings: demographic, economic or social aspects. We chose population density to represent demographic aspect, as population pressure could be one of the important factors pushing rural people from their village to look for employment in other places. Based on available data, the social aspect is represented by different levels of education attainment (*illiteracy*, *primary education*, *secondary education*), as education is an essential qualification for rural people to find employment in urban areas (Sriboonruang, 1992). The set of available numerous economic variables representing primary, secondary and tertiary sectors is submitted to factor analysis, in order to identify the underlying dimensions, or factors of the existing economic structure. As a result, the economic structure of the study area is represented by three major *composite economic factors* having respective groups of high-factor-loading original variables, as summarized in Table 1. The detailed procedure to derive three major economic factors and their interpretation are beyond the scope of this paper; they are discussed in Tran (1998).

As an important building block in spatial information theory, the concept of spatial autocorrelation provides insight into spatial patterns and association of spatial data. As in its most general sense, spatial autocorrelation[A-1] (Moran *I*) is concerned with the degree of clustering of similar objects, or indicates the extent to which the occurrence of one feature is influenced by the distribution of similar features. In a regional setting, it can be seen as an indicator of a causal process, which suggests the degree of influence exerted by an urban center upon its rural periphery (e.g., spatial interaction processes, externalities, spatial diffusion, copy-catting, spill-overs, etc.). Furthermore, LISA statistics [A-2], suggested by Anselin (1994), can provide further insights into the nature of core – periphery structure as the local Moran statistic allows for the identification of spatial agglomerative patterns, while the local Geary allows for the identification of spatial patterns of similarity/dissimilarity (interactions). As exploratory spatial analyses reveal strong spatial dependence in core – periphery structure, spatial modeling[A-4] (which accounts for spatial effects) is chosen to have insights into mechanism of influence of significant socio-economic factors upon core – periphery interactions. Supplementary technical notes on measurement of spatial statistics and spatial modeling adopted in this study are described in the Appendix.

Using the derived set of spatial variables, the integrated spatial statistical analysis and GIS are applied involving three main steps: (1) exploring the overall spatial structure of regional development in terms of various socio-economic indicators using spatial association statistics; (2) exploring the spatial extent of impact zones of the two major growth centers; and (3) spatial modeling of socio-economic factors to understand the roles of growth centers in spreading development to their rural hinterland. The GIS selection and manipulation functions of Arc/Info 7.0 utilize spatial information such as location, topology and distance to create spatial weight matrices for exploratory SDA statistical modules. The SpaceStat 1.80 developed by Luc Anselin (Anselin, 1995) has been used for analyses of global / local spatial patterns, spatial cross-correlograms and spatial modeling, while the Regional Analysis System (RAS) module developed by Shuming Bao (Bao *et al*., 1995) have been applied for *ring analysis*. Then, location-specific results of the spatial statistical analyses are transferred back to the GIS (ArcView 3.2) for visualization and mapping.

Table 1 Factor characteristics and respective groups of high-factor-loading economic variables.

• ***Factor 1*** (*Index of Urban-biased Economy)* highly positively correlates with *percentage of urbanized and residential areas, road density, property taxes* and *proportion of trading population*. • ***Factor 2*** (*Index of Industrial-based Economy)* highly positively correlates with normalized *total number of industrial employees, number of employees in large-scale factories, number of factories, total capital investments*, and *percentage of industrial land-use*. • ***Factor 3*** (*Index of Lacking Opportunity)* highly positively correlates with *travel time to nearest town* and *market centers, median distance to industrial centers* and *nearest roads*, and *farmer population*, and negatively correlates with *percentage of agricultural land*.

5. SPATIAL IMPACTS OF URBAN CENTERS

5.1 Spatial Patterns of Urban Growth Centers

To identify the spatial agglomeration of urban and industrial development for whole region, the global Moran indexes[A-1] for levels of *Urban-biased* and *Industrial-based Economies* are calculated. Visually, the study area is characterized by significant clustering patterns of transportation network, population density and average household income around Chiang Mai City and Lamphun Municipality, as shown in Figures 2, 4 and 5 respectively. The clustering patterns of *Urban-biased* and *Industrial-based Economies* are confirmed by significant positive Moran indexes of 0.66 and 0.26, with significance levels lower than 0.01%. The Moran scatterplots[A-1] (Figure 4) show good of fit for level of *Urban-biased Economy* (i.e., highly concentrated urban development around Chiang Mai City) and relatively lack of fit (albeit significant) of regression line for level of *Industrial-based Economy* (i.e., relatively more scattered industrial development in the study area).

To have further insights into localized urban core – rural periphery inter-dependencies in terms of urban and industrial development, LISA statistics[A-2] are calculated. The calculated LISA statistics indicate a local, positive spatial association (++) based on the level of *Urban-biased Economy* (significantly positive I_i and G_i) around Chiang Mai City. This confirms that the urbanization process has spread through the growth[A-2.1] of the Chiang Mai City core onto the rural periphery. On the other hand, the LISA statistics indicate a local, positive spatial association (++) of *Industrial-based Economy*, but negative association (-+) of *Urban-biased Economy* for Lamphun Municipality. It confirms that Lamphun Municipality, with its moderate level of urbanization (*spread through decentralization*[A-2.2]), serves only as a growth center to spread industrial development. Its urban infrastructure seems not to catch up with the rapid industrial development at the Lamphun Industrial Estate.

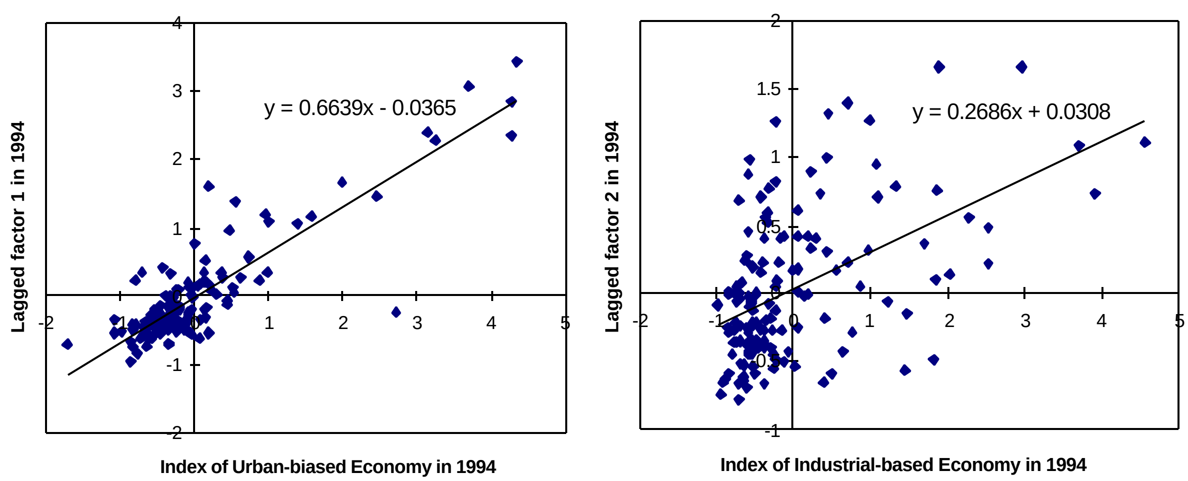

Figure 4. Moran’s scatterplots for *Urban-biased Economy* and *Industrial-based Economy* indexes by *tambol* in 1994

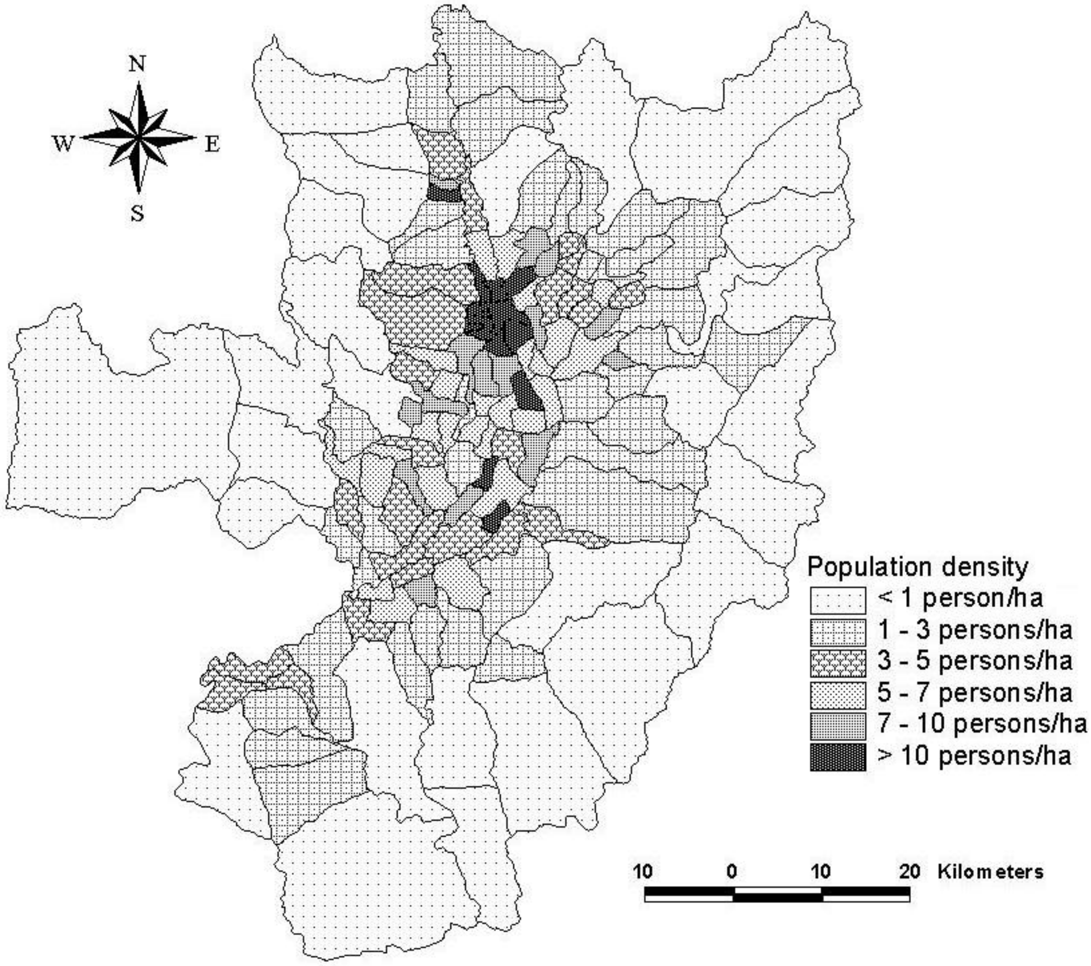

Figure 5. Population density by *tambol* in the Chiang Mai – Lamphun area in 1994

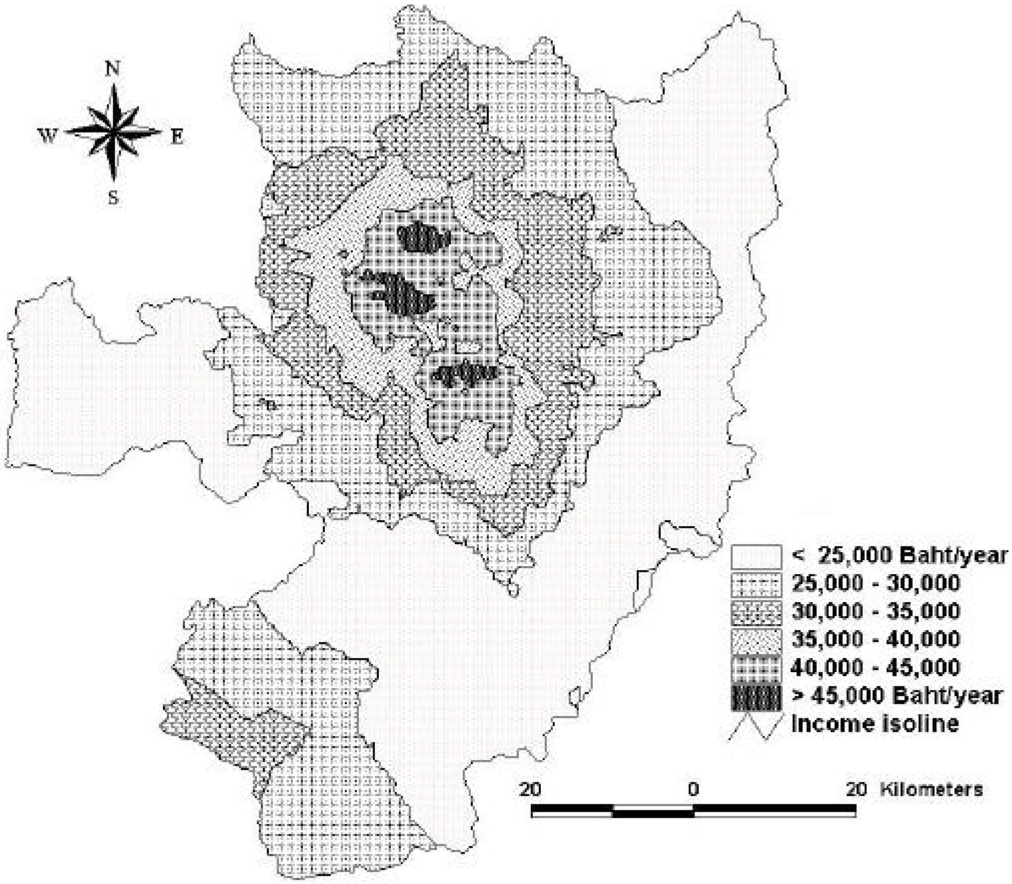

Figure 6. Interpolated surface of average household income by *tambol* in the Chiang Mai – Lamphun area in 1994 (using *spherical kriging* method)

5.2 Spatial Extent of Urban/Industrial Centers by Ring Analysis

The spatial extent of urban centers (by exploring the urban core – rural periphery relationship spatially) can be addressed using LISA statistics in the modified ring analysis (Bao *et al*., 1995). Starting with the assumption of an isotropic structure of space (classical *central place theory*), the space could be divided into rings at different distances from urban cores. By comparing the local Moran indexes for the urban core within different rings, we are testing the extent of spatial association between development in the city core and its effects on the rural periphery. Based on the nature of local spatial associations described above, the indexes of *Urban-biased* and *Industrial-based Economies* are used in studying the spatial spreading effects of Chiang Mai City and Lamphun Municipality, respectively; results are shown in Table 2.

For Chiang Mai City, the center of the rings is chosen at Chang Klan *tambol* centroid (in the center of the inner city). All the *tambol* of the study area are then divided into five rings, centered at the urban core according to the adjacency criterion. To identify the scope of spatial association of the urban core with the rural areas, the local Moran and local Geary indexes ($p < 0.05$) within different rings are calculated. A significant local Moran index value indicates that relatively high values are associated with the core area. As evident in Table 2, there is a significant positive spatial relationship (++) between core *tambol* and adjacent *tambols*. Chiang Mai City can be confirmed again as the *spread through growth*[A-2.1] type of growth center, i.e., the Chiang Mai suburban *tambols* get spreading effects of the growth of the city core in terms of *Urban-biased Economy* (e.g., spatial expansion of urban facilities). In addition, the local Moran index value for the urban core is significant ($p < 0.05$) within the 3rd ring. This suggests that the economic development of the Chiang Mai City core is strongly associated with the growth of rural areas, within a radius of about eleven kilometers. For Lamphun Municipality, the centroid of Nai Muang *tambol* is chosen as the center of the rings. A significantly negative spatial relationship (-+) is observed between the municipality core and adjacent *tambols* known as *spread through decentralization*[A-2.2], meaning the growth of adjacent *tambols* is associated with the slow growth in the Municipality core. The local Moran index value for the Municipality core is significant ($p < 0.05$) within the

1st ring of seven kilometers in radius. This has delineated the impact zone of any significant association between urban core and rural periphery of this industrial cluster, which exerted spatial influences on rural areas within seven kilometers. The combined impact zones of Chiang Mai City and Lamphun Municipality are then mapped as shown in Figure 7, which resembles the interpolated surface of the average household income distribution (Figure 6). In addition, using ANOVA, it is found that there are significant differences between inside and outside delineated impact zones in terms of average household income ($F = 6.809$ and $p < 0.001\%$), proportions of population working outside their home *tambols* ($F = 2.942$ and $p < 0.5\%$) and proportions of populations having secondary and higher education ($F = 9.203$ and $p < 0.001\%$). This suggests that, people living inside the delineated impact zones, close to major urban centers, have benefited from recent urban and industrial development significantly more than people outside the zones.

5.3 Spatial Extent of Industrial Center by Spatial Cross-Correlograms

From another perspective, effects of spatial configuration on the measure of spatial autocorrelation could be examined using spatial correlograms [A-3], which show the variations of the coefficient over a higher-order spatial relationship (Chou, 1995). While spatial autocorrelation has been used for characterizing the spatial pattern of a phenomenon, the concepts of spatial cross-association (multiple spatial correlation[A-3]) could be useful to characterize the relationship of two or more phenomena in the spatial domain. In the regional development context, spatial relationships (interaction) tend to go beyond immediate neighbors (or very close distance). Certain variations in spatial patterns, thus, may not be detected using statistics derived from the direct spatial relationship alone. To combine multivariate spatial correlation into spatial correlograms, we propose the so-called spatial cross-correlograms, where instead of using Moran indexes in Chou’s spatial correlograms, the multivariate spatial correlation coefficients [A-3] are used (Tran, 1998).

While strong effects of urban development on socio-economic life in the study area were detected by significant Pearson’s correlation coefficients between index of *Urban-biased Economy* and socio-demographic indicators, no significant direct effect was found for industrial development. However, significant positive *first-order spatial correlation coefficient*[3] ($r = 0.133$) between indexes of *Urban-biased* and *Industrial-based Economies* suggests possible indirect impact of industrialization on other aspects of development. To shed more lights on spatial effects of an industrial center exerting upon its rural surroundings in different social and demographic aspects, the spatial cross-correlograms are exploited.

Table 2 Identifying the urban core - rural area linkages for the Chiang Mai - Lamphun area using local Moran and local Geary indexes of concentrated rings

1. Impact Rings of Chiang Mai City

The urban core at centroid of *tambol* Chang Klan - Var. ***Factor 1*** CORE_Z: 3.70841
Local Moran and Geary indexes within the distance of the ith Ring - aggregate rings

Ring	Dist(km)	ZMEANi	Ii	E(Ii)	V(Ii)	Z(Ii)	Ci	E(Ci)	V(Ci)	Z(Ci)
2	3	0.1567	0.5811	-0.0046	0.0040	9.2609**	0.0679	0.7171	0.0076	-7.4597**
3	6	0.2069	0.7675	-0.0183	0.0136	6.7477**	1.7728	2.8684	0.0257	-6.8360**
4	11	0.1191	0.4159	-0.0582	0.0206	3.9973*	8.2118	9.1173	0.0391	-4.5810*
5	15	0.0061	0.0227	-0.0877	0.0061	1.4127	13.5566	13.7272	0.0116	-1.5865

2. Impact Rings of Lamphun Municipality

The urban core at centroid of *tambol* Bang Klang - Var. ***Factor 2*** CORE_Z: 3.72787
Local Moran and Geary indexes within the distance of the ith Ring - aggregated rings

Ring	Dist(km)	ZMEANi	Ii	E(Ii)	V(Ii)	Z(Ii)	Ci	E(Ci)	V(Ci)	Z(Ci)
2	7	0.0324	0.1207	-0.0053	0.0046	1.8617*	0.5697	0.8276	0.0105	-1.93904*
3	11	0.0317	0.1181	-0.0463	0.0219	1.1098	7.0319	7.2413	0.0504	-0.9327
4	17	0.0116	0.0432	-0.0760	0.0144	0.9927	11.7461	11.8964	0.0331	-0.8256
5	21	-0.0306	-0.1142	-0.0945	0.0012	-0.5690	14.8379	14.7929	0.0027	0.8603

*Note: * indicates pseudo-significance at p<0.05, ** at p<0.01.*
Zi are standardized values converted from the original inputs Factor1 or Factor2 – Factor1 ~ N (0,1), Factor2 ~ N (0,1)).

ZMEANi are the average values of the units surrounding unit i within designed distance. The value of unit i is not included

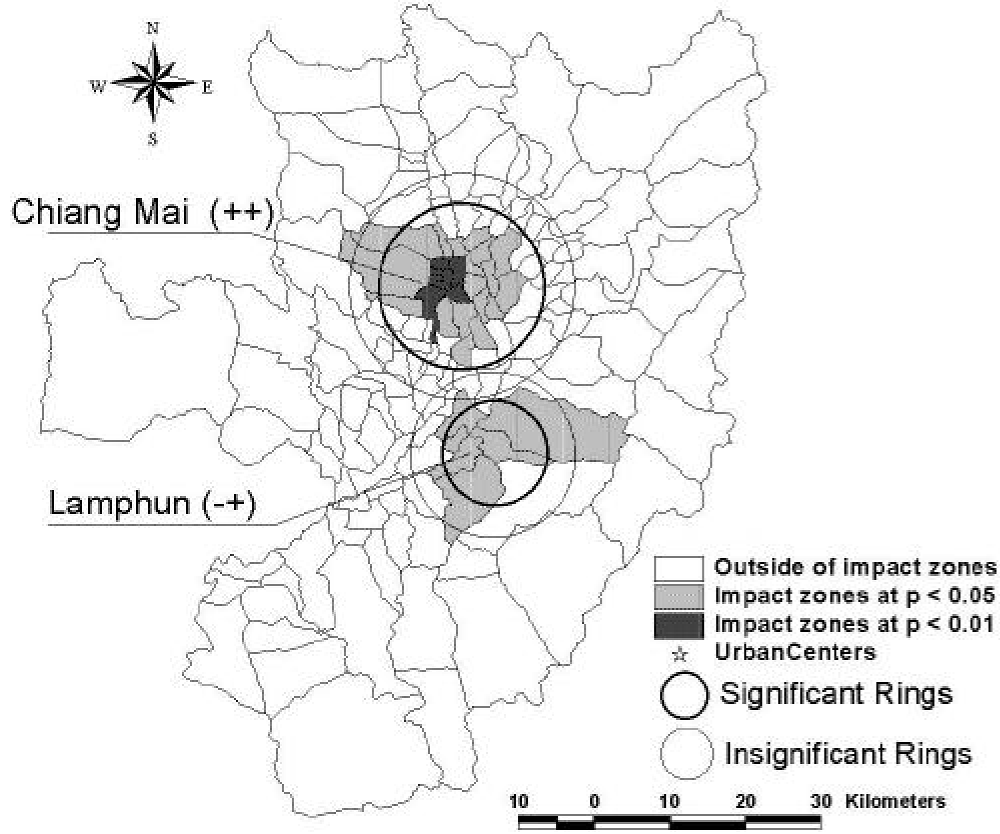

Figure 7. Combined significant impact zones of urban growth centers in the Chiang Mai – Lamphun area.

As distance is essential in spatial interaction models, spatial weight matrices based on distance criteria would well represent the possibilities of interaction between pairs of points in space. Hence, the spatial correlation coefficients for each distance would indicate the intensity of its spatial interaction. In this study, the centers of industrial establishment groups within each *tambol* are assigned as polygon label points for distance calculation. Based on distances between neighboring *tambol* centers ranging from four to ten kilometers, the spatial weight matrices for distance bands of 2, 4, 5, 10, 15, 20, 25, 30, 35 and 40 km are calculated. (Spatial weight $w_{ij}(d)$ here is set to one if centroid of *tambol j* falls within a given distance *d* from centroid of *tambol i*, and zero otherwise). Then, multiple spatial correlation coefficients[A-3] between index of *Industrial-based Economy (Factor 2)* and index of *Urban-biased Economy (Factor 1), Population Density, Proportion of Primary-educated Population,* and *Proportion of Working-out Population* for 1994 are calculated using SpaceStat 1.80. The spatial cross-correlograms are constructed as graphs of multiple spatial correlation coefficients by distance bands (Figure 8).

Based on both amplitude and wavelength, the behavior of the cross-correlograms provides much more reliable information than any single Moran index / spatial correlation coefficient, revealing less-evident spatial impacts of industrial development. For all development aspects, the constructed spatial cross-correlograms show no significant spatial impacts of industrial development on rural areas at any distance farther than 20 kilometers, even for highly mobile labor-flow. Moreover, the maxima on the cross-correlograms show the distance where the most intensive spatial interaction is possibly taking place. The spatial cross-correlograms for index of *Urban-biased Economy* and *Population Density* have maxima at around five to ten kilometers (which apparently equals worker's commuting distance at current times), showing spatial relationships between industrial establishments and inner urban centers. As for urban-rural linkages, the spatial cross-correlogram for *Proportions of Working-out Population* shows weak albeit significant relationships (correlation) with the level of industrial development, within 15 to 20 kilometers, and a maximum at four to 15 kilometers for 1994. Revealing this hidden spatial relationship provides valuable input into spatial modeling of core-periphery interactions below.

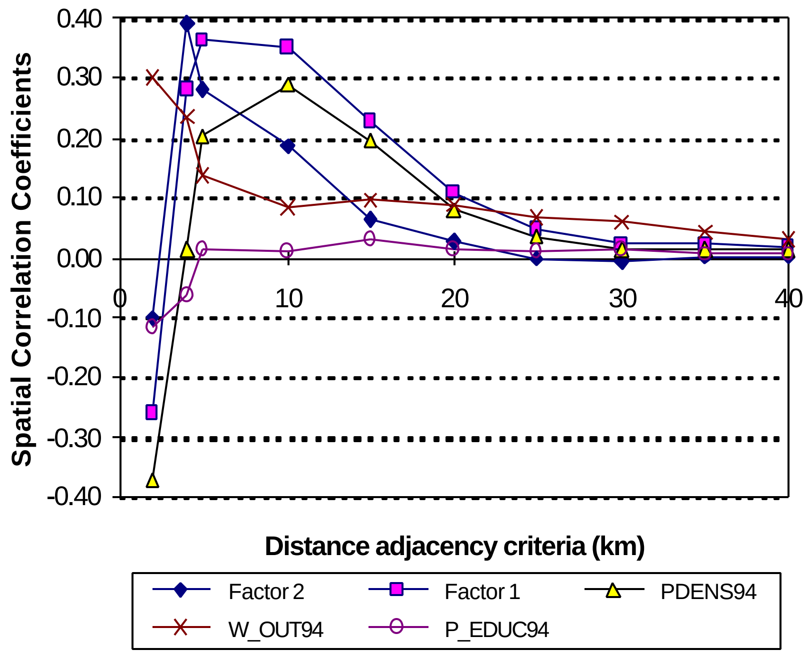

Figure 8. Spatial cross-correlograms based on multiple spatial correlation coefficients varying by different distance-based adjacency weight matrices for level of *industrial-based economy* with other aspects of development.

5.4 Spatial Modeling of Core-Periphery Interactions

Given the "push" and "pull" factors in urban-rural linkages, rural population tends to look for employment outside the place of their residence (Kaur, 1995). Thus, the rural labor outflow indicating the job attraction of urban centers as well as the excess of free labor released from the agricultural sector could represent the intensity of urban-rural linkages, as a result of regional development. The *Proportion of Working-out Population* of each *tambol* is used as response variable to model socio-economic factors significantly influencing the intensity and extent of core-periphery interaction. The *Proportion of Working-out Population* may initially be formulated as a linear regression model of demographic (*population density*), social (*various levels of education attainment*) indicators and three major economic factors (Table 1).

To avoid possible non-normal errors, the dependent variable is transformed using natural logarithm function and, then, submitted to classical OLS regression analysis using SpaceStat 1.80. The insignificant explanatory variables are excluded from the model based on *t-value* ($p = 0.1$), and the eventual OLS linear regression model is derived as shown in Table 3. The regression diagnostics show a significant spatial autocorrelation error (at significance level of 0.001 %), indicating a significant deviation from the basic assumption for linear regression analysis on spatial independence of sample observations and, thus, reducing the validity of significance tests. Moreover, as the flow of labor is a spatially dependent process (indicated by a positive, strong Moran *I* of 0.4096), the explanation is not complete without some characterization of spatial interaction. Therefore, in order to improve model estimates and account for spatial effects, the spatial-lag regression model[A-4] is adopted using SpaceStat 1.80, with its output shown in Table 4. The spatial lag term is highly significant and, more importantly, its addition reduced the spatial autocorrelation in the model residuals to an insignificant level (p = 0.125). The adjusted R^2 of 0.4921 (vs. 0.4567 for OLS model), the log-likelihood of –88.428 (vs. –101.514) and Akaike Information Criterion of 122.855 (vs. 171.027) show significant improvement in overall fit and more reliable parameter estimates of the spatial lag model.

According to the sign of estimated parameters for explanatory determinants and the meaning of the model (Table 4), significant 'pull' factors are levels of *Urban-biased Economy* (*Factor 3)*, levels of *Industrial-based Economy* (*Factor 2)*, and *Population Density*, while significant 'push' factors are *Rural-Urban Indicator* and the *Spatial Lag-of-response-variable* itself. A closer look at the original economic variables constituting *Factor 3* (Table 1) reveals that accessibility is, indeed, a crucial factor in providing rural people the opportunity to move out and seek employment, i.e., in intensifying the urban-rural interaction. Moreover, the model is also in support of the conventional wisdom that land pressure is the real force affecting the outflows of free agricultural labor (Table 1). However, for the study area, it is found that population pressure is not a factor affecting the outflow of rural labor. The rural rather than the urban population tend to rush out to seek employment, as the urban centers are providing sufficient employment opportunity. Concerning the economic factors, the level of *Urban-biased Economy* (*Factor 1*) appears not significantly affecting the outflow of labor from rural areas, while the level of *Industrial-based Economy* (*Factor 2*) is significantly attracting rural labor. These findings have spatial relevance since urbanization, in fact, is concentrated mostly around Chiang Mai City, and the rapid industrialization process in the study area since 1986 appears to have a favorable impact on employment generation for the rural population.

Table 3 Results of traditional regression analysis (OLS) with test diagnostics

Dependent Variable: *ln(Working-out Population + 1)*

R^2 = 0.4567 R^2-adj = 0.4187 Log-likelihood = -101.514 AIC = 171.027

Variable	Coefficients	Std. Err.	t-value	Prob
Constant	1.94702	0.197472	9.859745	0.000000
Factor 1	0.1777	0.100939	1.760476	0.080511
Factor 3	-0.452775	0.0591491	-7.654810	0.000000
Illiteracy rate	-0.00745726	0.00679591	-1.097315	0.144387
Pop. Density	-0.000138259	5.54217E-05	-2.494679	0.013770
Rural-Urban Indicator	0.630054	0.205611	3.064298	0.002618

Regression Diagnostics

Multicollinearity condition number = 7.165699
Kiefer-Salmon (error normality) = 11.435 (p = 0.003)
Koenker-Bassett test (heteroskedasticity) = 33.138 (p = 0.000)
Moran's I (error) = 0.276 (p = 0.000)
Lagrange Multiplier (error) = 29.165 (p = 0.000)
Lagrange Multiplier (lag) = 37.499 (p = 0.000)

Table 4 Results of spatial-lag regression analysis solved through maximum likelihood, with diagnostics of residuals.

Response Variable: *ln(Working-out Population + 1)*
pseudo R^2 = 0.4921 Log-likelihood = -88.428 AIC = 122.855

Variable	Coefficients	Std. Err.	z-value	Prob
Lagged variable of *Ln(Working-out Population+1)*	0.0497432	0.0126568	3.930157	0.000085
Constant	1.39591	0.232631	6.000540	0.000000
Rural-Urban Indicator	0.464627	0.187835	2.473583	0.013377
Factor 3	-0.322413	0.0597546	-5.395623	0.000000
Pop. Density	-9.88531E-05	5.02207E-05	-1.968374	0.049025
Lagged Factor 2	0.186305	0.105104	1.772585	0.076297

Regression Diagnostics
Breusch-Pagan (heteroskedasticity) = 36.347 (p = 0.000)
Lagrange Multiplier test (spatial error dependence) = 1.966 (p = 0.125)

6. CONCLUDING REMARKS

The paper has indicated that GIS is a useful technical platform in integrating social and environmental data for regional development studies in Thailand. It facilitated the manipulation of large amounts of geographic data, generating spatial variables from GIS database to supplement available spatialized socio-economic indicators, and constructing the topological structure, which altogether facilitated the spatial analysis of complicated spatial phenomena. Furthermore, with the developed spatial databases, GIS can serve as an efficient technical vehicle for spatial analysis and spatial modeling functions to gain insights into regional development problems, e.g., the patterns and roles of two major growth centers in the regional development of the Chiang Mai – Lamphun area have been empirically explored. The combination of exploratory and explanatory spatial data analyses has revealed the impact of demographic, economic and social factors upon the spatial relationship between urban core and rural periphery. Specifically, the exploratory analyses based on LISA statistics, ring analysis and spatial cross-correlograms have revealed indirect spatial neighborhood relationships between various indicators, which had hardly been researched so far. Accounting for the spatial association inherent in the data resulted in a spatial model that better extracts information from the variables and has more precise estimates of model coefficients than does the OLS model. With core-periphery interaction explored, development impacts of urban and industrial expansion are evaluated in order to enhance growth center development strategies through facilitating various scenarios. Finally, findings from this meso-scale study provide an overall picture of regional development, which has additional importance opening up a vast scope for further detailed research work in the region as decentralization planning is on increase in developing Thailand.

ENDNOTES

1 *Tambol*, equivalent to sub-district in other countries, is the smallest Thailand's administrative unit with clearly defined spatial border. Since 1994, *tambol* has its own elected local government, the so-called *Tambol Administrative Organization*, with certain administrative decision-making autonomy.

2 Functional Economic Area is defined as a relatively self-contained labor market, which contains a metropolitan central city and hinterlands within commuting distance (Bao *et al.*, 1995).

3 First-order spatial correlation coefficient is calculated following equations (A5) – (A7) with spatial weight matrix defined by direct adjacency criterion, i.e., w_{ij} is set to one if *tambol j* is adjacent to *tambol i*, and zero otherwise.

REFERENCES

Anselin L., (1994). "Local Indicators of Spatial Association – LISA". *Research Paper 9331*, Morgantown, WV: Regional Research Institute, West Virginia University.

Anselin L. (1995). *SpaceStat, A Software Program for the Analysis of Spatial Data, Version 1.80*. Morgantown, WV: Regional Research Institute, West Virginia University.

Bao S., Henry M. S. & Barkley D., 1995. "RAS: A Regional Analysis System integrated with Arc/Info". *Computers, Environment and Urban Systems*, Vol. 19, No.1, pp. 37-56.

Brown, D. G., 1996. "Spatial statistics and GIS applied to internal migration in Rwanda, Central Africa". *Practical Handbook of Spatial Statistics, Arlinghaus S. L. (Eds.)*. New York: CRC.

Chou Y. H., 1995. "Spatial patterns and spatial autocorrelation". *Spatial Information Theory. A Theoretical Basis for GIS*, Frank A.V. & Kuhn W. (eds.), Springer, New York.

Getis, A. and Ord, J.K., (1992). "The analysis of spatial association by the use of distance statistics". *Geographical Analysis*, Vol. 24, pp. 189-206.

Huxhold W.E., 1991. *Introduction to Urban Geographic Information Systems*. New York: Oxford University Press Inc.

Kaur R., 1995. *Urban-Rural Relations*. A geographic analysis. Anmol, New Dehli.

Kliskey A.D., 1995. "The role and functionality of GIS as a planning tool in natural-resource management". *Computer, Environment and Urban Systems*, vol. 19, No. 1, pp. 15-22.

Klosterman R.E., 1995. "The appropriateness of geographic information systems for regional planning in the developing world". *Computer, Environment and Urban Systems*, vol. 19, No. 1, pp. 1-13.

Lo F-C. (ed.), 1981. *Rural-urban relations and regional development*. United Nations Centre for Regional Development, Nagoya, Japan.

Potter R.B. & Unwin T. (eds.), 1989. *The geography of urban-rural interaction in developing countries*. Routledge, London.

Setty D. E., 1991. "Rural Industrialization, Small-scale and Cottage Industries in Asia". *UNCRD Working Paper* No.91-2. United Nations Centre for Regional Development, Nagoya, Japan.

Sharma P. R., 1984. "Growth centres and regional development: Aspects of theory and policy". *HABITAT International*, Vol. 8, No. 2, pp. 133-150.

Sriboonruang, S., 1992. *Chiang Mai province and its emerging development problems*. Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand.

Suwan, M. *et al*., 1992. *Impacts of industrialisation upon the villager's life in Northern Thailand*. Faculty of Social Science, Chiang Mai University, Chiang Mai, Thailand.

Tran H., 1998. *Integrating GIS with spatial data analysis to study the development impacts of urbanisation and industrialisation: Case study of Chiang Mai - Lamphun area, Thailand*. Ph.D. Dissertation No. SR-98-3, Asian Institute of Technology, Bangkok, Thailand.

Wartenberg D. (1985). "Multivariate spatial correlation: a method for exploratory geographical analysis". *Geographical Analysis,* Vol. 17, pp. 263-283.

APPENDIX: MEASUREMENT OF SPATIAL STATISTICS AND MODELING

[A-1] Moran Index and Moran Scatterplot

To measure the global spatial autocorrelation, one of the most popular indicators is the Moran I that is defined by:

$$I = \frac{N}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_j (x_i - \bar{x})^2} \qquad \text{(A1)}$$

where N is the number of observed geographic units; w_{ij} denotes the spatial relationship between the i^{th} and j^{th} geographic units, which equals 1 for adjacent units and 0 otherwise; $S_0 = \Sigma_i \Sigma_j w_{ij}$ is the total number of adjacent pairs. The value of the Moran I is generally between -1 and 1, indicating the spatial clustering patterns of a phenomenon. The Moran I is positive when nearby objects tend to be similar in attributes, and negative when they tend to be more dissimilar than what is normally expected. It is approximately zero when attribute values are arranged randomly and independently in space. The equation (A1) shows that the Moran I is calculated based on specification of spatial weight matrix $\{w_{ij}\}$, which can be defined by either continuity and/or distance criteria. In this study, spatial weights are defined by the spatial adjacency (i.e., w_{ij} is set to one if *tambol j* is adjacent to *tambol i*, and zero otherwise) for all spatial analyses, unless otherwise indicated.

On the other hand, Moran I can be expressed in matrix notation as (Anselin, 1995):

$$I = \frac{N}{S_0} \frac{y' W y}{y' y} \qquad \text{(A2)}$$

where y is a vector of observations in deviation from the means, W is spatial weight matrix (when W is row-standardized, $N = S_0$), and Wy is the associated spatial lag, which is a weighted average of the neighboring values. Thus, the Moran I gives a formal indication of the degree of linear association between a vector of observed values y and its spatial lag Wy. To visualize and summarize the overall pattern of linear association, Anselin (1994) suggested bivariate spatial lag scatterplot of spatial lag Wy against y, which is referred to as a *Moran scatterplot*. The Moran I, here, can be interpreted as the slope of a regression line of spatial lag Wy on y and a lack of fit would indicate important local pockets of nonstationarity. In addition, the Moran scatterplot can be used as a means to identify "outliers" – locations with extreme values with respect to the central tendency reflected by the regression slope.

[A-2] Local Spatial Statistics and Core-periphery Inter-dependencies

Decomposing global indicator into the contribution of each observation in order to assess the influence of individual locations, Anselin (1994) proposed LISA as measurements of local spatial associations, which include the local Moran and local Geary. The local Moran and local Geary statistics for each observation i is defined as follows (Anselin, 1994):

$$I_i(d) = Z_i \sum_{j \neq i}^{n} w_{ij} Z_j, \qquad \text{(A3)}$$

$$C_i(d) = \sum_{j \neq i}^{n} w_{ij} (Z_i - Zj)^2 \qquad \text{(A4)}$$

where the observations Z_i and Z_j are in standardized form (with mean of zero and variance of one). The spatial weights W_{ij} are in row-standardized form. So, I_i is a product of Z_i and the average of the observations in the surrounding locations. Significant local Moran with consistent signs between Z_i and its standardized value suggests that location i is associated with relatively high values in surrounding locations and otherwise. On the other hand, C_i is a measure of the weighted sum of squared differences between Z_i and those of its surrounding locations. A small and significant C_i suggests a positive spatial association (similarity) of observation i with its surrounding observations, while a large and significant C_i suggests a negative spatial association (dissimilarity).

In a regional analysis context, Bao *et al*. (1995) extended Anselin's work by analyzing urban core – rural periphery interdependencies based on combinations of local Moran with local Geary statistics. The spatial association between urban cores and their surrounding rural areas may suggest one of the following four types:

[A-2.1] *Spread through growth* (++): Rural growth is associated with rapid growth in the urban core, i.e., a significant local Moran index with consistent signs between Z_i and its standardized value, and a small and significant local Geary index;

[A-2.2] *Spread through decentralization* (-+): Rural growth is associated with slow growth in the urban core, i.e., a significant local Moran index with consistent signs between Z_i and its standardized value, and a large and significant local Geary index;

[A-2.3] *Backwash* (+-): Urban core growth is associated with slow growth or decline in the rural areas, a significant local Moran index with inconsistent signs between Z_i and its standardized value, and a large and significant local Geary;

[A-2.4] *Independence* (?): Growth in rural areas is not closely associated with changes in economic activity in the urban core, i.e., the local Moran and local Geary indexes are not significant.

[A-3] Spatial Correlation and Spatial Correlograms

To extend the concept of Pearson's correlation between two variables, the spatial correlation is taking into account the spatial effects of adjacent areas. For irregularly spaced data (areal features), the multivariate measure of spatial correlation computed in SpaceStat 1.80 follows the approach suggested by Wartenberg (1985). First, all variables are standardized:

$$z_k = (x_k - m_k) / s_k \qquad \text{(A5)}$$

where the subscript k refers to the vector of observations on the k^{th} variables, m_k is the mean for variables k, and s_k is its standard deviation. Also, the spatial weight matrix is converted to a stochastic matrix, i.e., a matrix for which all elements sum to one. The resulting matrix (W^s) is always symmetric, with elements

$$w^s_{ij} = w_{ij} / S_i S_j w_{ij} \qquad \text{(A6)}$$

where w_{ij} are the elements in the unstandardized weights matrix. A matrix of coefficients of spatial association is constructed as:

$$M = Z' W^s Z \qquad \text{(A7)}$$

where Z is a matrix with the values for the standardized variables as columns. The association represented in this matrix is similar in form to a bivariate Moran index between two variables (for more technical details see Anselin, 1995).

Similar to Moran *I*, the spatial correlation coefficient is calculated based on the specification of spatial weight matrix. Chou (1995) proposed spatial correlograms as a graph of Moran *I* against different-order spatial relationships (specified by either continuity or distance criteria), based on which one could define the wavelength and amplitude of spatial patterns of phenomenon under study, e.g., clustering or agglomeration effects of regional industrial development.

[A-4] Spatial Lag Regression Model

The linear models are widely used to demonstrate the co-variation of a response variable with its major socio-economic independent determinants. The presence of spatial dependence in cross-sectional geo-referenced data could be utilized to interpret the form of spatial interaction, the precise nature of spatial spill-over and the economic and social processes that lie behind this. The transactions occurring near each other may exhibit an adjacency effect, which could be incorporated into the model as an additional explanatory variable in form of spatial lag. Formally, a mixed regressive-spatial-autoregressive model includes a spatially lagged variable, *Wy*, as one of the explanatory variables (Anselin, 1995):

$$y = \rho W y + X\beta + \varepsilon \qquad \text{(A8)}$$

where y is a vector of observations on the response variable, Wy - spatial lag for y, r - spatial autoregressive coefficient, X is a matrix of observations on the (exogenous) explanatory variables with associated vector of regression coefficients β. The estimate for ρ can clearly be considered as an indication of spatial autocorrelation, for example, as an alternative to the use of Moran *(I)*, Geary *(c)*, or spatial association *(G)* statistics. As the correlation of the lag W*y* (as one of the explanatory variables) with the error term invalidates the optimality of OLS as an estimator for this model, ML approach needs to be used instead. The estimates of the coefficients in a mixed regressive-spatial-autoregressive model can be interpreted in several ways. The inclusion of *Wy* in addition to other explanatory variables allows one to assess the degree of spatial dependence in the model, while controlling the effect of other explanatory variables. Hence, the main interest is in the spatial effect. Alternatively, the inclusion of *Wy* allows one to assess the significance of the other (non-spatial) explanatory variables, after the spatial dependence is controlled for.