INTEGRATION OF LANDSAT THEMATIC MAPPER (TM) DATA AND U.S. CENSUS DATA FOR QUALITY OF LIFE ASSESSMENT

C. P. Lo
Professor of Geography
University of Georgia
U.S.A.

Commission VII, Working Group 7

KEY WORDS: Quality of Life, Thematic Mapper Data, Census Data, NDVI, Principal Components Analysis

ABSTRACT:

The quality of life (QOL) of the Athens-Clarke county of Georgia was assessed by integrating biophysical variables extracted from Landsat Thematic Mapper (TM) data with socio-economic variables obtained from the U.S. Census data. The biophysical variables extracted include per cent of urban use, normalized difference vegetation index (NDVI), and apparent surface temperatures. The socio-economic variables extracted are population density, per capita income, median home value, and per cent of college graduates. These data are integrated by means of principal components analysis (PCA) and GIS overlay functionality. It was found that PCA has been particularly successful in linking the two sets of variables through the formation of two Tasseled Cap data planes of "Greenness" and "Economic Well Being" so that transition between the two planes can be interpreted. The results indicate that high-resolution satellite image data can complement census data in providing more realistic QOL assessment with an environmental perspective.

1. INTRODUCTION

Quality of life (QOL) is a subjective measure (Andrew, 1986). Most QOL studies make use of socio-economic data from the census, and environmental data are seldom employed. The objective of this research is to explore the integration of the biophysical data extractable from high-resolution satellite image data with the socio-economic data from the census for QOL assessment, using a computer-assisted remote sensing/GIS approach. It is noted that according to the French sociologist, Chombart de Lauwe (1952), who introduced the concept of "espace social" (social space), a remotely sensed image can resolve a total environment into its components ("sub-spaces") by interpreting different facets of the same area. Social space, the total environment in which people live, comprises of two main dimensions: (1) the morphological environment and (2) the socio-cultural environment (De Haas, 1966). The morphological environment is made up of biophysical and demographic sub-spaces, while the socio-cultural environment refers to the material infrastructure of social life created by the human mind. Clearly, an integration of the various sub-spaces in different environments is possible by means of the overlay functionality of GIS.

2. STUDY AREA AND DATA

The study area is Athens-Clarke County in Georgia which is located about 100 km northeast of Atlanta and 160 km northwest of Augusta in the United States. It has a total area of 31,379.19 ha and a population of 87,594 in 1990. The major city in the county, Athens, had a population of 45,734 and is home to the approximately 28,000 student campus of the University of Georgia. There is a high degree of contrast between high and low income groups in this county. The 1990 per capita income was $15,715, which is 82 per cent of the national average and 90 per cent of the state average. The poverty rate as defined by the U.S. Bureau of Census is 24.5 per cent. The majority of the county population is urban (82.4 per cent in 1990). Only 0.19 per cent of the population live on rural farms. The remaining population is rural residential (Hodler et al., 1994).

There are three primary data sources used in this research: (1) Landsat TM data acquired on July 16, 1990, (2) black-and-white panchromatic aerial photographs acquired on January 10, 1990 at 1:18,000 scale, and (3) 1990 Census and Housing Data from the U.S. Bureau of Census. The areal unit used to aggregate the population data is the block group. There are 52 block groups covering the Athens/Clarke county. In order to register the Landsat TM images with the population census block group data, two 1:24,000 USGS topographic map sheets, Athens East and Athens West Quadrangles, were used to provide ground control points. These were used to rectify the Landsat TM data of the study area with nearest neighbor resampling at 25-metre resolution, giving a RMSE of planimetry of ± 0.33 pixel. The census block group boundaries were also converted into UTM coordinates and raster format.

The biophysical data extracted from Landsat TM data were: (1) land use/cover, particularly the percentage of urban use, (2) normalized difference vegetation index (NDVI), and (3) apparent surface temperatures.

The land use/cover was extracted using a supervised approach with the maximum likelihood classifier. Eight land use/cover types were extracted: (1) water, (2) forest, (3) commercial /industrial, (4) transportation, (5) agriculture in pasture, (6) agriculture in crop, (7) low-density residential, and (8) high-density residential. The misclassifications were later manually corrected area by area with the aid of a parallelpiped classifier and ground truth from aerial photography, thus achieving an overall accuracy of 99.1 per cent. The "commercial /industrial" cover class constituted "urban use".

NDVI, which measures vegetation amount, was extracted using TM band 4 (0.76-0.90 μm) and TM band 3 (0.63-0.69 μm) as follows:

$$NDVI = (TM4-TM3)/(TM4+TM3)$$

The value varies from -1 to +1 with an increase in vegetation amount. The raw NDVI exhibited a range of values from -0.4778761 to 0.9148936. The NDVI image very clearly delineated the built-up areas from the vegetated areas.

The apparent surface temperatures were extracted from the thermal infrared TM band 6 (10.3-12.5 μm) using a quadratic regression model in the form:

$$T(K) = 209.831 + 0.834\ DC - 0.00133\ DC^2$$

where DC is the digital counts between 0 and 255, and T(K) is absolute temperature in Kelvin. The resulting image of surface temperatures revealed a range from 21°C to 38 C, with high surface temperatures clearly delineating the built-up areas.

From the population census data, the following socio-economic variables were extracted or derived: (1) population density, (2) per capita income, (3) median home value, and (4) per cent of college graduates, by block groups. Each one of these variables produced a map which was then converted into raster format for analysis by the IDRISI software.

3. INTEGRATION APPROACHES

3.1 Principal Components Analysis (PCA)

PCA was applied to the seven layers of bio-physical and socio-economic image data described above all aggregated at the census block group level. The correlation matrix of these variables indicated particularly strong negative correlation between (1) NDVI and surface temperatures ($r=-0.88$ at 0.0001 rejection level), and (2) NDVI and per cent urban ($r=-0.85$ at 0.0001 rejection level) (Table 1). The implication is that NDVI by itself is a versatile environmental quality variable. NDVI also showed moderately positive correlation with per capita income, median home value, and per cent college graduates, as well as negative correlation with population density. The first of the three principal components extracted explained about 54 per cent of the variance (Table 2). This component showed strong positive loadings on four variables, namely, NDVI, per capita income, median home value, and per capita income, and per cent of college graduates, but also strong negative loadings on three variables, namely, per cent of urban use, surface temperatures, and population density. The second principal component which explained 21 per cent of the variance, showed only one weak negative loading on NDVI, but positive loadings on the remaining six variables (Table 2). A plot of the component pattern indicates two clusters of variables: (1) population density, per cent of urban use, and surface temperatures verus (2) per capita income, median home value, per cent of college graduates, and NDVI (Figure 1). These two clusters of variables suggest that the two principal components represent two planes of data similar in nature to those defined by the Tasseled Cap features of Greenness and Brightness, which are related to the plane of vegetation and the plane of soils respectively (Crist and Cicone, 1984). The first cluster of variables is "environmental" in nature while the second cluster is "socio-economic" in nature. It is interesting that NDVI falls in the "socio-economic" cluster. These first plane can be labelled as "Greenness" because of its strong positive correlation with NDVI and the second plane can be called "Economic Well Being" because of its strong correlation with housing and education variables. Clearly, PCA has succeeded to integrate the biophysical environment with the socio-economic characteristics of the population. NDVI as an effective measure of "Greenness" can be regarded as an excellent quality of life indicator.

A map of the first principal component scores by block group for the study area is therefore an excellent map of the quality of life of the population living in each block group (Figure 2). The map revealed high life quality areas in the southeast and southwest edges of the county and the low life quality areas in the middle of the county near the downtown area.

3.2 GIS OVERLAY

The GIS overlay approach involves addition of NDVI and the socio-economic layers together in the development of the quality of life (QOL) indicator. The values in each data layer were equally divided into 10 classes and a rank of "life quality" was computed for each block group. There were 10 ranks (10 being the highest rank score). In the case of "surface temperatures", "per cent urban", and "population density", the higher the value, the less desirable for the QOL, and hence high values were assigned low ranks. On the other hand, for "NDVI", "per capita income", "median home value", and "per cent college graduates", the higher the value, the more desirable for QOL, and hence high values were assigned high ranks. These seven layers of ranked data of "life quality" were added one on top of another by the GIS overlay method. The composite score ranged from 18 to 61. The roughly upper twenty per cent of QOL scores were found in those block groups along Clarke County's southern, western, and eastern borders, while low QOL scores were found in block groups located in the center of the county, very close to the downtown area of Athens. The spatial pattern mapped by the GIS overlay approach is very similar to that produced by the first principal component scores. Therefore, the two approaches to the integration of Landsat TM data with the census data for QOL assessment are equally valid.

4. CONCLUSION

This research has demonstrated that a strong relationship exists between data relating to the biophysical environment and those relating to the socio-economic environment. The vegetation index in the form of NDVI, apparent surface temperatures, and urban land use can be extracted from the Landsat TM data. NDVI, which shows strong negative correlation with apparent surface temperatures and per cent of urban use, is a particularly useful measure of greenness in the biophysical environment. As such, it is a good integrator of the two and is a good measure of the quality of life by itself. Principal components analysis of the biophysical and socio-economic data revealed two Tasseled Cap data planes of "Greenness" and "Economic Well Being", depicting possible transition between the two planes. The integration approaches reported in this paper provides an environmental perspective in the assessment of quality of life and can be easily achieved with the GIS technology. In the absence of up-to-date census data, high-resolution satellite images data can be employed to assess the condition of socio-economic development in a region.

5. ACKNOWLEDGEMENTS

The author is very grateful to EOSAT for the research grant awarded to him, making the research reported in this paper possible.

6. REFERENCES

Andrew, J.R., 1986. Research on the Quality of Life. Survey Research Center, Ann Arbor, MI.

Chombart de Lauwe, P.H., 1952. Paris et l'agglomeration parisienne; L'espace social dans une grande cite.

Crist, E.P. and Cicone, R.C., 1984. Application of the Tasseled Cap concept to simulated Thematic Mapper data. Photogrammetric Engineering and Remote Sensing, 50, pp.343-352. De Haas, W.G.L., 1966. Integrated Surveys and the Social Sciences. Publications of the ITC-UNESCO Centre for Integrated Surveys, Delft, The Netherlands.

Hodler, T., Lawson, N., Schretter, H. and Torguson, J., 1994. The Interactive Atlas of Georgia. Institute of Community and Area Development, Athens, Georgia, U.S.A.

Table 1: Correlation Matrix of Variables

	popden	percap	medhome	percoll	NDVIblk	perurb	surtemp
popden	1.0000	-0.3008	0.0398	-0.0885	-0.4002	0.1964	0.4885
percap	-0.3008	1.0000	0.5907	0.7176	0.5637	-0.5302	-0.4537
medhome	0.0398	0.5907	1.0000	0.5501	0.4284	-0.4605	-0.2899
percoll	-0.0885	0.7176	0.5501	1.0000	0.3065	-0.2996	-0.1227
NDVIblk	-0.4002	0.5637	0.4284	0.3065	1.0000	-0.8498	-0.8838
perurb	0.1964	-0.5302	-0.4605	-0.2996	-0.8498	1.0000	0.7747
surtemp	0.4885	-0.4537	-0.2899	-0.1227	-0.8838	0.7747	1.0000

Key:
popden = population density, percap = per capita income, medhome = median home value, percoll = per cent of college graduates, BDVIblk = NDVI by block groups, perurb = per cent of urban use, surtemp = surface temperatures

Table 2: Principal Component Analysis Loadings

	COMPONENT	LOADINGS	
VARIABLES	C1	C2	Communality
popden	-0.42881	0.50695	0.44
percap	0.80670	0.36255	0.78
medhome	0.64172	0.54280	0.71
percoll	0.57380	0.66469	0.77
NDVIblk	0.90438	-0.28840	0.90
perurb	-0.85299	0.16968	0.76
surtemp	-0.82029	0.48673	0.91
Eigenvalue	3.7941	1.4738	
% variance	54.20	21.05	

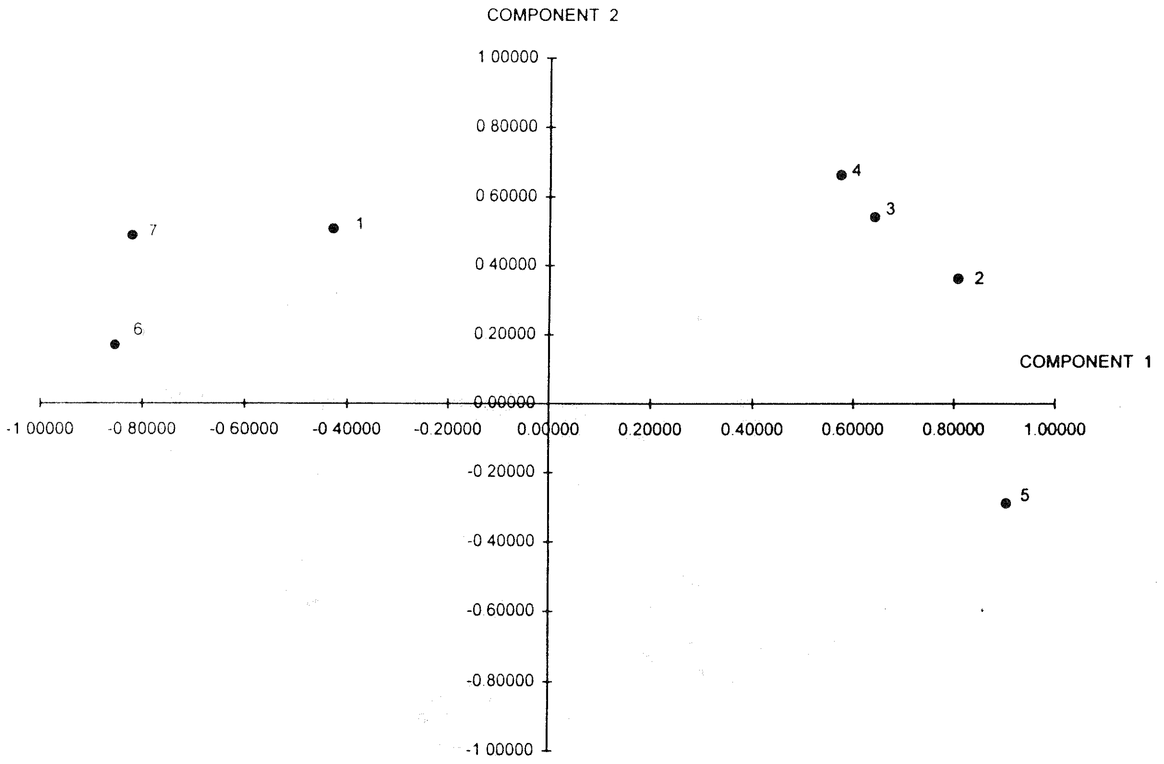

Figure 1. Plot of component pattern for principal components 1 and 2.
Key: #1: population density, #2: per capita income, #3: median home value, #4: per cent of college graduates, #5: NDVI, #6: per cent of urban land use, #7: surface temperature.

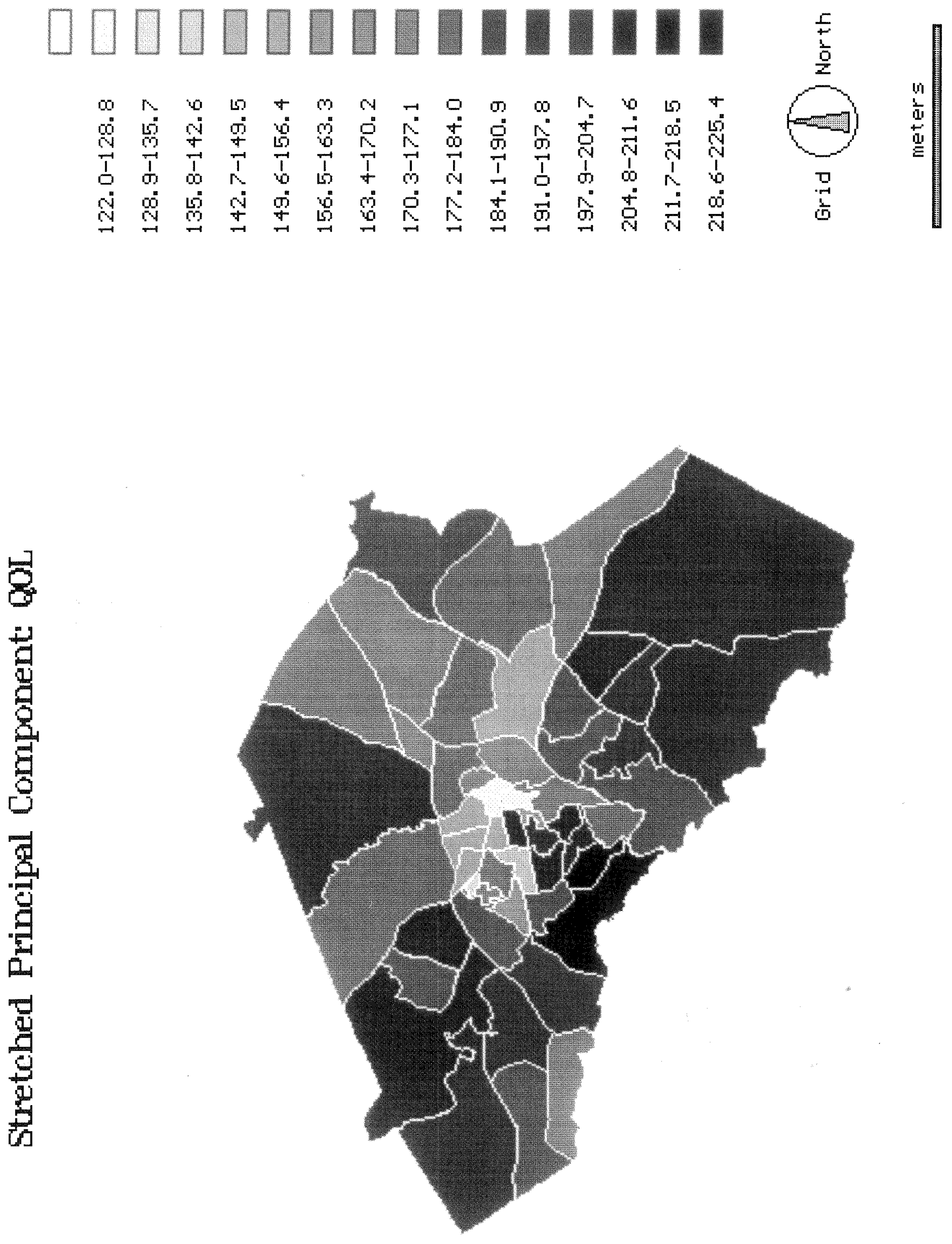

Figure 2. The first principal component scores of Athens-Clarke County by block group.