ESTIMATION OF POPULATION IN INFORMAL SETTLEMENT COMMUNITIES USING HIGH RESOLUTION SATELLITE IMAGE

Florence A. Galeon

University of the Philippines, Diliman, Quezon City 1102 Philippines - florence.galeon@up.edu.ph

KEY WORDS: High Resolution; Quickbird; Informal Settlement; Estimation; Mapping; Urban

ABSTRACT:

This paper envisioned the creation of a highly accurate base map from high resolution QuickBird satellite imagery for the University of the Philippines. A compilation of land titles was used to define the extent of university property. The lots are plotted using CAD software then overlayed on the base map image. There were two identified qualified categories of informal settlements for this study such as the "slum" and the "semi-formal" type. Fieldworks ensued to validate and update the mapping result with the aid of GPS. Stratified random sampling was employed to verify samples and a high 95% identification accuracy was achieved. Base from the QuickBird satellite image, around 16% of the total 493 hectares UP Campus or roughly 79 hectares may be labeled as informal settlements. This 2004 UP Campus informal settlements map will serve as the baseline data for monitoring further encroachment. Individual houses were digitized and a fieldwork was conducted to gather sample data for the regression analysis. A total of 160 samples were collected and processed using the Grid and Theme Regression in the AVENUE programming language. The result show that estimation of the number of residents in informal settlements is possible for slum areas but not in semi-formal areas. Likewise, population of informal settlement communities can be estimated from a high resolution satellite image using first order equation.

1. INTRODUCTION

Urban areas in the world are the most dynamic region on earth. They serve as the social and economic centers of our modern life. They promise a secured and much better life conditions as compared to the opportunities found in rural areas. As a consequence, the population of cities and urban centers has tremendously increased in the past and have become the densest populated regions in the world over the past few decades.

Meinel et al (2001) has observed this strong trend of urban migration which is much experienced in less developed countries. The Philippines, with its peace and order problems in the countryside, has witnessed the exodus of people to the urban centers particularly the National Capital Region. The Metro Manila Urban Services for the Poor Project (MMUSP) reported in 2002 that Metro Manila has joined the small group of mega cities worldwide with populations of more than 10 million. This concentration of people in very dense populated urban areas presents tremendous challenge to managers and planners. It requires the introduction of sophisticated monitoring systems for planning purposes such as a satellite acquired data providing a regularly observed synoptic overview of large areas.

High resolution satellite images offer abundance of information of the earth's surface for remote sensing applications including mapping and monitoring of urban areas. Hoffman (2001), Herold (2002), and Shackelford and Davis (2003) are just some of the works done in this particular field. Thomson and Hardin (2000) integrated remote sensing and GIS to identify site specific information on land cover for mapping urban residential land use. Although a number of experiments have been performed for urban areas, the number of applications dealing with informal settlements is very limited. One possible reason for this seemingly lack of attention and interest for informal settlement studies is the fact that most researchers come from developed countries that does not experience housing shortage. To illustrate the housing problem in Metro Manila, approximately 3.5 million residents live in depressed neighborhoods across the nation's capital (MMUSP Report, 2002). Most of them remain undocumented by the government making it difficult to gather demographic data. Information particularly on the number of people living in informal settlements is often limited, since inhabitants are often inadequately covered by formal censuses (Uzun and Cete 2004).

This paper envisioned the creation of a highly accurate large-scale natural color image map from high resolution satellite imagery to be used as a base map suitable for the University of the Philippines as a reference for physical planning, management and other mapping purposes. In addition, this paper aim to map from the established base map the informal settlements inside the campus and determine if population of these communities can be effectively estimated using the high resolution satellite image.

2. LITERATURE REVIEW

Studies in the past have demonstrated the ability of remotely sensed data to extract socio-economic attributes directly or indirectly by means of surrogate informations. According to Cowen and Jensen (1998), population estimates are being made every 5–7 years in western countries requiring the need for high-spatial resolution remotely sensed data. They described various uses of remote observation including measuring the area, height and volume of buildings, measuring traffic patterns and road conditions, estimation of residential energy demand and building predictive models of residential expansion utilizing high-spatial resolution (0.3–0.5 m) panchromatic and/or color infrared aerial photography.

The operational potential of high resolution satellite data depends on its ability to respond to the requirements of the planners and how it is integrated with other secondary data related with everyday activities. Jain (2007) presented a study for linking remotely sensed data with property tax related issues and evaluated the capabilities of remote sensing technology to measure these attributes accurately at parcel level. Quincey et al. (2007) in his study has demonstrated that fine spatial resolution imagery facilitates land cover mapping at an unprecedented level of detail.

Other works focused directly on the possible extraction of population using remotely sensed data. Rindfuss et al. (1996) used remotely sensed data for population migration and social change, mainly with micro level data sets in which individual/households are the unit of analysis. Lo (2003) evaluated the use of satellite data for zone-based estimation of population and housing units from land use/land cover maps. Dasymetric mapping using the land use/land cover information extracted from remotely sensed images to obtain an improved estimation of where people actually live was done by Liu (2004). Mennis (2003) applied an areal interpolation technique to disaggregate census population data into spatial units with homogenous land use.

3. STUDY AREA

The study area is the University of the Philippines, Diliman which is the flagship campus and the largest Constituent University of the University of the Philippines System. It is located in Quezon City, the most populated city in the nation’s capital, making its 493 hectares prime property a hot commodity.

According to an official paper of the university, the squatting problem in the campus started in 1970. It was initially tolerated and continued to grow until it became a complex problem to the administration. One inescapable fact which partly explains this occurring problem is that land will always have a market value simply because it is an increasingly finite resource in urban areas. The shanties left unattended in the 1970s was commodified by market forces so that with each transfer of ownership, it has increased its market value until eventually, it approximated prevailing real estate market values, notwithstanding that the land on which it stands has been indefinitely reserved by law exclusively for education and education-related uses. Several years ago official reports estimate over 20,000 squatter families occupy 66 hectares, approximately 13 percent of the 493 hectare property claiming 15 major areas of the campus.

4. DATA AND SOFTWARES USED

The satellite data available and suitable for this research was QuickBird Satellite Image acquired on April 25, 2004. Imagery resolution is sub-meter with the panchromatic (Pan) band having an effective ground resolution of 0.61 meters and 2.44 meters for the four multispectral (XS) bands.

The image that was purchased from DigitalGlobe for basemap production is a *‘Standard Product’ type Level 2A* which means that standard radiometric and sensor corrections have been applied to the raw imagery by the image supplier. This product has also been geometrically corrected and referenced to a standard local map projection- in this case a Universal Transverse Mercator Zone 51 projection compatible with accepted map standards in the Philippines. A coarse DEM was used together with RPC (Rational Polynomial Coefficients) values in a special polynomial rectification process to correct for geometric distortions.

In this study, ENVI 4.1 software has been used to carry out sophisticated image processing operations such as image fusion, enhancement, and georeferencing. The initial digitization of informal settlements has been done using an open source GIS software called fGIS. Arc view 3.2, a GIS software with extensive analytical tools and customized regression analysis capability have been used for the later works.

5. METHODOLOGY

Figure 1 illustrates the detailed methodology adapted in this study.

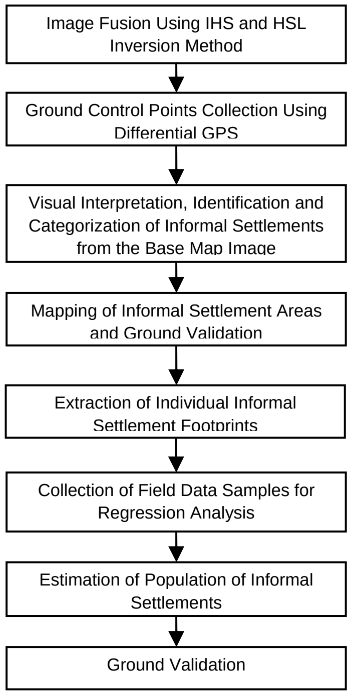

Figure 1. Methodology flow diagram

5.1 Image Fusion

Image fusion is the process of combining multiple images into a composite product of which some desired characteristic or property of the original images is preserved in the resulting image. For this research project, it is desired to combine a higher spatial (ground) resolution panchromatic (B/W) image with a lower resolution but multi-spectral (natural color and NI) image to create a composite image which has the higher resolution of the panchromatic image while retaining the multi-spectral properties required to produce a natural color image map.

There are only several methods that have been developed over the past few decades to accomplish data fusion effectively. Each of these methods has their own advantages and disadvantages. These different methods have been tried and visually evaluated. Most remote sensing image processing systems common in the market lack the proper tools for image editing and retouching and those that have are of high cost. Using improved methods developed by GeoAnalytika, an affiliate local company of the author, the desired accuracy and image quality has been achieved. The final method used was actually a modified multistep process using IHS and HSL Inversion Method.

5.2 Ortho-Image Production

After creating the pan fused natural colour satellite image, the next step is image rectification and registration. This stage involves warping the image to correct for geometric errors and conform it to a selected map shape in the correct coordinate system. These five major steps summarizes the whole ortho-rectification process namely selection of map projection and coordinate system, identification of ground control points (GCPs) from the image, collection of ground control points with a GPS, rectification using a suitable mathematical model, reprojection to a standard map projection and quality control and assessment.

A total of 33 identified GCPs were successfully identified, occupied, and established in a local map projection and datum as required for the final base map. The distribution of points is shown in Figure 2. In order to meet the accuracy requirements for the basemap, a high accuracy differential GPS survey was conducted. Occupation time per control point was set from 15-20 minutes, which ensures that enough satellite observation data is acquired to produce a high accuracy position per point. Results of the Differential GPS survey results were summarized as follows: Maximum RMS error = 0.017 meters, Average RMS = 0.009 meters, Maximum Standard Deviation = 0.143 meters, and Average Standard Deviation = 0.05 meters. The high accuracy survey results ensures that positional errors will be kept to a minimum and shall not affect the accuracy of the final basemap.

After all GCPs have been collected and processed, a rectification model is then created. The average GCP RMS error is 0.89 meters. In addition, the RMS error of the GCPs exhibited an unusually high deviation from the mean error with differences from a high of 4.78 meters (8 pixels) to a low of 1.13 pixels (0.68 meters) even with repeated checks of their plotted position in the image. It was deduced that the errors were inherent in the image because of pre-rectification done by the image supplier. An alternative method is using a delauney triangulation. This approach fixes and preserves the image point locations and uses a linear stretch to warp the pixels in between the control points. Thus, there is virtually zero (0) error at the control points and the errors are distributed linearly between control points. This method assumes that the GCP image points have been chosen as carefully as possible to match the actual ground feature with minimum doubt.

The result is a geometrically sound UP Campus base map image with sub-meter accuracy and visual quality that can rival conventional color aerial photography suitable for large scale mapping purposes up to 1:2,000 (Figure 2).

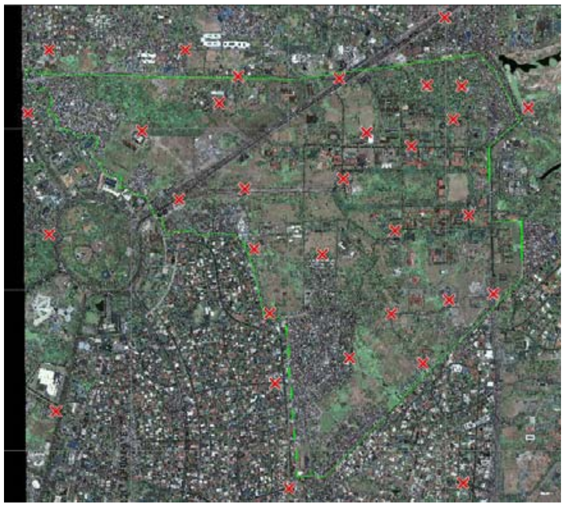

Figure 2. UP Campus Base Map Showing Boundary and Distribution of GCPs in the Project Area

5.3 Visual Interpretation of High Resolution QuickBird Imagery

A compilation of land titles was used to define the extent of university property (Figure 2). The lots are plotted, checked for traverse errors, and then consolidated to form the whole coverage of the campus using CAD software. The boundaries were overlayed on the base map image to determine the extent of the study area. Remote sensing softwares were used to view and visually interpret the image to determine locations of areas occupied or directly utilized by informal settlers.

Ward and Peters (2007) used visual interpretation of high spatial resolution multispectral IKONOS satellite image to identify low-income informal homestead subdivisions (IFHS, also known as colonias) in peri-urban areas of US metropolitan areas. This process of visual interpretation of high spatial resolution satellite data is not automated, but requires the systematic search, identification, and delineation of the target features by the analyst.

The object of study which is informal settlements is readily identifiable on the image because of the high resolution and excellent visual quality of the image. There were two identified qualified categories of informal settlements for this study. The first is the “slum” type characterized by MMUSP (2002) to have lack of spatial pattern, smaller structures, irregular boundary demarcation, clustering and uneven spread, different reflectance and locational attributes. These settlements are considered to be the miserable or depressed areas, normally lacking in basic services such as electricity, water, and communication lines. The second type is a “semi-formal” type characterized by a more decent neighborhood similar to a low-cost housing with provisions of the basic services and amenities. These settlements, although not physically informal, lack necessary tenure or permit to legally occupy the property.

5.4 GIS Mapping

Abbott (2003) has pointed out the value of GIS for the evaluation of informal settlements. In his study, vegetation cover and land use have been quantified by visual interpretation of 1:8000 scaled air photos using the Autocad 14 (AutoDesk) and ArcView 3.2 (ESRI) softwares. The digitization of the subjects is straightforward. Thanks to the excellent visual quality of the image and the effectivity of displaying the satellite image even at 1:2000 scale. An open source GIS software called fGIS was first used to delineate the areas occupied and controlled by the informal settlements. The choice of the software was mainly due to its availability at the time of mapping.

One particular challenge in mapping informal settlements is the existence of houses partially or totally under large trees. Significant portion of the campus is vegetated particularly the northwestern part called the "arboretum" which is considered to be the last remaining forest in the city. Mapping informal settlements in this situation required some innovations involving actual field boundary demarcation and the aid of positioning devices to plot the peripheries of the community. A handheld Garmin Global Positioning System (GPS) attached to the pocket computer provided instantaneous positioning information and markings on the satellite image which was useful not only in documentation of the paths that were taken by the field crew but also as a navigational aid.

Fieldworks ensued to validate and update the mapping result with the aid of GPS. Stratified random sampling was employed to verify samples from all informal settlement clusters. A high 95% identification accuracy was achieved. The misidentification is attributed to the one year temporal difference of the data and the actual fieldwork, during which eviction occured as a result of the continuous effort of the campus administration to recover the control of the land. Base from the QuickBird satellite image, around 16% of the total 493 hectares UP Campus or roughly 79 hectares may be labeled as informal settlements. This 2004 UP Campus informal settlements map will now serve as the baseline data for monitoring further encroachment.

After the establishment of the UP Campus informal settlements map, the individual houses were digitized to prepare for the other phase of this research which is to estimate the population of informal settlement communities. In the high resolution satellite image, the semi-formal houses with an average surface area of 30 square meters are large enough to be easily delineated. Problem is apparent only in mapping the roofs of each slum type houses. The very cramped area even as small as an unimaginable 10 square meters is very difficult to visually delineate especially when they are placed so closed with each other forming a seemingly one large continuous roof. Jain (2008) experienced the same in extracting information in old developments as well as in informal settlements where dwelling size is considerably small and building are placed adjacent to each other.

6. RESULT AND ANALYSIS

6.1 Regression Analysis

A fieldwork was conducted to gather sample data for the regression analysis. A stratified random sampling method was employed to assure complete representation of the total population throughout the image. A total of 160 samples were collected bearing the identity number, number of residents in each house, informal settlement category whether slum or semi-formal, roof-derived surface area of the house, and the type indicating whether single or multilevel. The field data gathering is a challenging task. There is security to consider and there is the hesitation of residents to share informations for fear that anything they say might be used against them.

The sample data was processed using the Grid and Theme Regression, a program made by Jeff Jenness in the AVENUE programming language, an extension of the ArcView 3.2 software. The initial plot of the raw data in Figure 3 showed low correlation of the two variables in all polynomial order with the third order producing the highest. The summary of correlation is shown in Table 1. This may be attributed to the undivided type of data (slum and semi-formal) and the existence of multilevel houses showing large number of residents with very small area, producing an inconsistent trend or random occurrence.

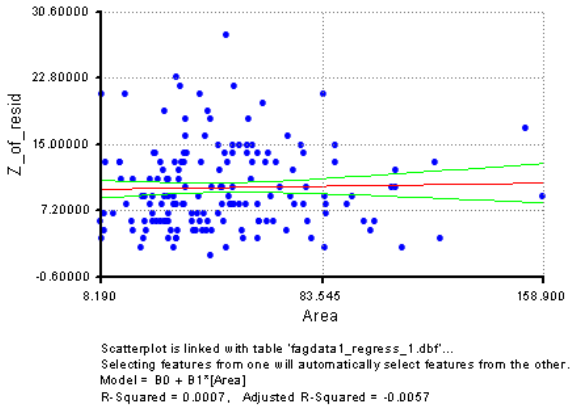

Figure 3. First Order Regression Plot

	1st Order	2nd Order	3rd Order
R^2	0.000689	0.002756	0.013938

Table 1. Summary of Correlation

Having multilevel floors among the houses appears to be the biggest problem in this endeavor to establish an effective way of estimating population. This exposes the limitations of the satellite image being not able to recognize this particular natural phenomena. However, this should not hinder us from formulating solutions to discover and make something useful. According to Rindfuss and Stern (1998), each data source has its imperfections, but combining sources with different limitations might provide a better picture of the entire phenomenon. In this way, remote sensing even with its imperfections, can make a contribution to social scientific measurements by improving on some measures and cross-checking others.

Modifying the data entries by removing the multilevel houses and separating the sample data of the two informal settlement types, the number of samples was reduced from 160 to over a hundred. For the two informal settlement types, slums and semi-formal, the same processing using the Grid and Theme Regression in AVENUE has been repeated. The correlation of the variables for both type has dramatically increased in all polynomial order with the third order producing the

SLUMS	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**Community**
Roof Area	15.22	29.79	75.83	26.79	9.96	19.24	10.96	302.989
Expected Population (1st Order)	4	7	14	6	3	5	4	71
Expected Population (3rd Order)	3	6	-664	7	7	4	5	35
Actual Population	4	5	9	6	4	4	6	70

Table 4. Validation Data Showing Number of Residents in Slum Area

highest as usual. The sample plot in the third order, correlation factor, and the regression equations for slums are shown in Figure 4 and Table 2 below. The area is measured in square meters. The sample plot for semi-formal is shown in Figure 5 and Regression Equation in Table 3.

Figure 4. Third Order Regression Plot of Slum Area

	1st Order	2nd Order	3rd Order
R^2	0.090750	0.154538	0.160570
Regression Equation	Y-hat = 2.3306145 + 0.1662026*[Area]	Y-hat = 26.2106264 + 2.4328515*[Area] - 0.0442838*[Area^2]	Y-hat = 39.5889753 - 5.6473600*[Area] + 0.2826836*[Area^2] - 0.0043599*[Area^3]

Table 2. Slums Regression Equation

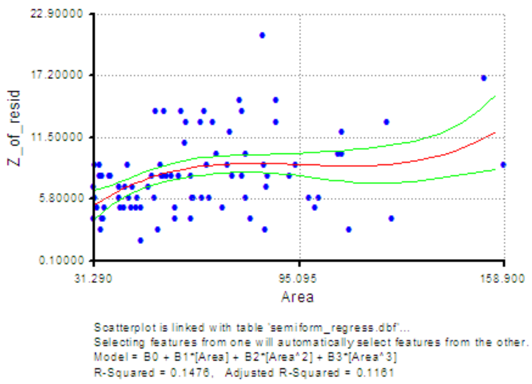

Figure 5. Third Order Regression Plot of Semi-formal Area

6.2 Validation

The formulated equations for the two informal settlement types have been validated by randomly selecting a number of individual houses and several clusters of houses in the field. The actual number of validation data, however, has been determined by the feasibility and security involved in field data gathering. Due to the limited time and resources for fieldwork, two sets of data each for slums and semi-formal were collected. These data were used to check the accuracy of predicting first, the number of residents in one house and second, a selected community or cluster of houses.

	1st Order	2nd Order	3rd Order
R^2	0.107333	0.118824	0.147635
Regression Equation	Y-hat = 5.2650832 + 0.0432388*[Area]	Y-hat = 3.3147666 + 0.1010005*[Area] - 0.0003618*[Area^2]	Y-hat = -4.5233895 + 0.4459212*[Area] - 0.0047731*[Area^2] + 0.0000166*[Area^3]

Table 3. Semi-formal Regression Equation

6.2.1 Slum Area

The result of the validation shows that the third order function did not consistently predicted with accuracy the number of individual houses particularly those with large areas. This is also true for the prediction of the cluster or community, with an expected large number of people, because of the exponential nature of higher order polynomials. The low or first order equation provided a more realistic estimates both for the individual houses and the community. Results is summarized in Table 4.

6.2.2 Semi-formal Area

The result is the same as in the slum area showing the third order function not consistently predicting with accuracy the number of individual houses. The prediction of the cluster or community with large number of people is also erroneous (Table 5). The low or first order equation provided the same realistic estimates for the individual houses and the community. The use of this equation, however, is restricted by the fact that this may be effective only for single level houses which actually defined the premise for its formulation. All the field samples for the regression analysis and the validation data indicated a great percentage of multilevel houses. This will indeed render the formulated equation for estimating population in semi-formal houses useless.

SEMI-FORMAL	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**Community**
ROOF AREA	40.271	52.246	58.869	61.982	94.443	211.551	250.684	770.046
Expected Population (1st Order)	7	7	7	7	9	14	16	67
Expected Population (3rd Order)	6	8	8	8	9	33	68	140
Actual Population	7	4	9	17	10	7	9	63

Table 5. Validation Data Showing Number of Residents in Semi-formal Area.

7. CONCLUSIONS

This study concludes that the QuickBird imagery can be used to produce an accurate base map of excellent visual quality capable of being displayed even at 1:2000 scale. The use of a high resolution satellite image when integrated with GIS mapping capabilities can produce an accurate map of informal settlements. The estimation of the number of residents in informal settlements is possible for slum areas but not in semi-formal areas. Likewise, population of informal settlement communities can be estimated from a high resolution satellite image using first order equation from a regression analysis of sufficient number of field samples. This may help alleviate the difficulty of gathering field demographic data in the future.

REFERENCES

Abbott, J. (2003). The use of GIS in informal settlement upgrading: Its role and impact on the community and on local government. Habitat International, 27, 575–593.

Cowen, D.J., Jensen, J.R., 1998. Extraction and modeling of urban attributes using remote sensing technology. In: Diana, L., Emilio, F.M., Ronald, R.R., Paul, C.S. (Eds.), People and Pixels: Linking Remote Sensing and Social Science. National Academic Press, pp.164–188, http://www.nap.edu/catalog/5963.html.

Herold, M., Mueller, A., Guenter, S. and Scepan, J., 2002. “Object-oriented mapping and analysis of urban land use/cover using IKONOS data,” Proceedings of the 22nd EARSEL symposium, Prague, June 2002.

Herold, M., Scepan, J., Müller, A. and Günther, S, "Object-oriented mapping and analysis of urban land use/cover using IKONOS data". In: Proceedings of 22nd EARSEL Symposium “Geoinformation for European-wide integration. Prague, 2002.

Hofmann, P., 2001. Detecting urban features from IKONOS data using an object-oriented approach. *The Remote Sensing & Photogrammetry Society Proceedings*.

Jain, S., 2007. Remote sensing application for property tax evaluation. International Journal of Applied Earth Observation and Geoinformation 10 (2008) 109–121.

Jain, S., 2007. Use of IKONOS satellite data to identify informal settlements in Dehradun, India. Int. J. Remote Sens. 28 (15), 3227–3233.

Liu, X., 2004. “Dasymetric mapping with image texture,” ASPRS annual conference proceedings, Denver, Colorado, May 2004.

Lo, C.P., 2003. Zone-based estimation of population and housing units from satellite-generated Land use/land cover maps. In: Victor, M. (Ed.), Remotely Sensed Cities. Taylor & Francis Publications, London and New York, pp. 157–180.

Meinel, G., Neubert, M. and Reder, J., "The potential use of very high resolution satellite data for urban areas – first experiences with IKONOS data, their classification and application in urban planning and environmental monitoring," In: Jürgens C. (Eds.): Remote Sensing of Urban Areas/Fernerkundung in urbanen Räumen (=Regensburger Geographische Schriften, Heft 35). Regensburg, 2001, pp. 196-205.

Mennis, J., 2003. “Generating Surface Models of Population Using Dasymetric Mapping,” *The Professional Geographer*, 55(1), pp. 31-42.

Quincey D.J. et al. Fine-resolution remote-sensing and modeling of Himalayan catchment sustainability. Remote Sensing of Environment 107 (2007) 430–439

Rindfuss, R.R., Stern, P.C., 1998. Linking remote sensing and social science: the need and the challenges. In: Diana, L., Emilio, F.M., Ronald, R.R., Paul, C.S. (Eds.), People and Pixels: Linking Remote Sensing and Social Science. National Academic Press, pp. 1–27, http://www.nap.edu/catalog/5963.html.

Shackelford, A.K., Davis, C.H., 2003. A combined fuzzy pixel based and object based approach for classification of high-resolution multi-spectral data over urban areas. IEEE Trans. Geosci. Remote Sens. 41 (10), 2354–2363

The Metro Manila Urban Services for the Poor Project (MMUSP) Report, 2002.

Thomson, C.N., Hardin, P., 2000. Remote Sensing /GIS Integration to Identify Potential Low income Housing Sites. Cities 17 (2), 97–109.

Uzun, B., and Cete, M., A Model for Solving Informal Settlement Issues in Developing Countries, FIG Working Week 2004, Athens, Greece.

UNCHS (Habitat) 1995 Monitoring human settlements, abridged survey. Indicators Programme. Nairobi: United Nations Centre for Human Settlements.

Ward, P.M., Peters, P.A. Self-help housing and informal homesteading in peri-urban America: Settlement identification using digital imagery and GIS. Habitat International 31 (2007) 205–218