Integrating Remote Sensing and Social Science - The correlation of urban morphology with socioeconomic parameters

H. Taubenböck[1,2], M. Wurm[1,2], N. Setiadi[3], N. Gebert[3], A. Roth[1], G. Strunz[1], J. Birkmann[3] & S. Dech[1,2]

[1] German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Wessling-Oberpfaffenhofen, Germany
hannes.taubenboeck@dlr.de
[2] Julius-Maximilans University Würzburg, Geographical Institute, Earth Observation, 97074 Würzburg, Germany
[3] United Nations University, Institute for Environment and Human Security (UNU-EHS), Bonn, Germany

Abstract—**The alignment, small-scale transitions and characteristics of buildings, streets and open spaces constitute a heterogeneous urban morphology. The urban morphology is the physical reflection of a society that created it, influenced by historical, social, cultural, economic, political, demographic and natural conditions as well as their developments. Within the complex urban environment homogeneous physical patterns and sectors of similar building types, structural alignments or similar built-up densities can be localized and classified. Accordingly, it is assumed that urban societies also feature a distinctive socio-economic urban morphology that is strongly correlated with the characteristics of a city's physical morphology: Social groups settle spatially with one's peer more or less segregated from other social groups according to, amongst other things, their economic status. This study focuses on the analysis, whether the static physical urban morphology correlates with socioeconomic parameters of its inhabitants – here with the example indicators income and value of property. Therefore, the study explores on the capabilities of high resolution optical satellite data (Ikonos) to classify patterns of urban morphology based on physical parameters. In addition a household questionnaire was developed to investigate on the cities socioeconomic morphology.**

I. INTRODUCTION

More and more people are living in cities. According to the latest United Nations' projections, virtually all of the world's population growth over the next 30 years will be absorbed by urban areas. During 2007, for the first time in history of the world, the proportion of the population living in urban areas exceeded 50 per cent [1]. Especially the observed dynamics – reflected in rapid spacious sprawl, changing urban morphology and socioeconomic patterns – need new ideas for sustainable planning and management issues. Thus, a paradigm shift from singular research perspectives to value-adding inter- and transdisciplinary perspectives has to take place for a more holistic understanding of urban systems.

Remote sensing generally enables direct measurements of the earth's surface and the spatial distribution of its physical objects. Social science is generally more concerned with why things happen than where they happen [2]. "Socializing the pixel" is to take remote sensing imagery beyond this use in the applied sciences and toward its application in addressing the concerns of the social sciences [3]. According to [4], "Integrating social science and remote sensing will require the fusion not only of data, but also of quite different scientific traditions." The integration of both disciplines in our study aims at extending research for multidisciplinary perspectives to analyze and understand complex systems like urban environments. Using the synoptic view from space that only remote sensing can provide we try to cover phenomena over large areas, thus broadening the scope of social science inquiry.

The physical urban appearance is a reflection of the society that created it. Thus, an isolated analysis of social questions detached from geospatial questions does not do justice to the capabilities of the two research disciplines – remote sensing and social science. In this study we derive spatial contextual information on the urban anthroposhere – e. g. number of buildings, their shape, size and height or built-up density and vegetation fraction – from remotely sensed data. From the social sciences a field survey on 1000 georeferenced households has been conducted. The hypothesis is based on the assumption that populations living in urban areas showing almost similar physical housing conditions will have homogeneous social and demographic characteristics [5] [6] [7] [8] [9]. The integration of different disciplines aims at the above mentioned value-adding by synergistically using the capabilities of both perspectives. Thus, this study focuses on the question how remote sensing might inform social surveys and how social surveys might inform remote sensing.

II. STUDY AREA AND DATA SETS

Padang is the capital city of the Sumatera Barat province (West Sumatra), and with almost one million inhabitants the third largest city on the island Sumatra, Indonesia. Padang features supra-regional relevance with an international airport, a port as well as binding to the rail network. Furthermore, the city possesses an important economical role for the coastal region and the mountainous back-country. The multi-faceted functions of the large city are also reflected in a very heterogeneous physical urban appearance.

The analysis of small-structured, heterogeneous and dynamic urban environments requires high resolution satellite data. Ikonos imagery features four spectral bands (blue, green, red, nir) and a geometric quality of 1 m for the panchromatic

978-1-4244-3461-9/09/$25.00 ©2009 IEEE

band, 4 m multispectral, and 1 m pan-sharpened. Fig. 1 shows the diverse heterogeneous structure of the coastal zone in Padang and the local differences in urban morphology.

A field survey was conducted in the city of Padang to obtain information on socioeconomic parameters of the population. This study focuses on two parameters enquired from the questionnaire: Income and value of the property. The income was queried based on four predetermined categories in accordance with the minimum standard salary 2008 of the province West Sumatra 800.000 Rupiah per month [10] starting with income per month lower than 800.000 Rupiah (circa 60 Euros), between 800.000 and 1.600.000 Rupiah, between 1.600.000 and 4.000.000 Rupiah and more than 4.000.000 Rupiah. The surveyed buildings were selected by reason of five different attributes to cover the spatial and thematic spectrum of urban morphology:

(i) *the geographical location*; close to the beach, city centre, interior areas and suburbs;

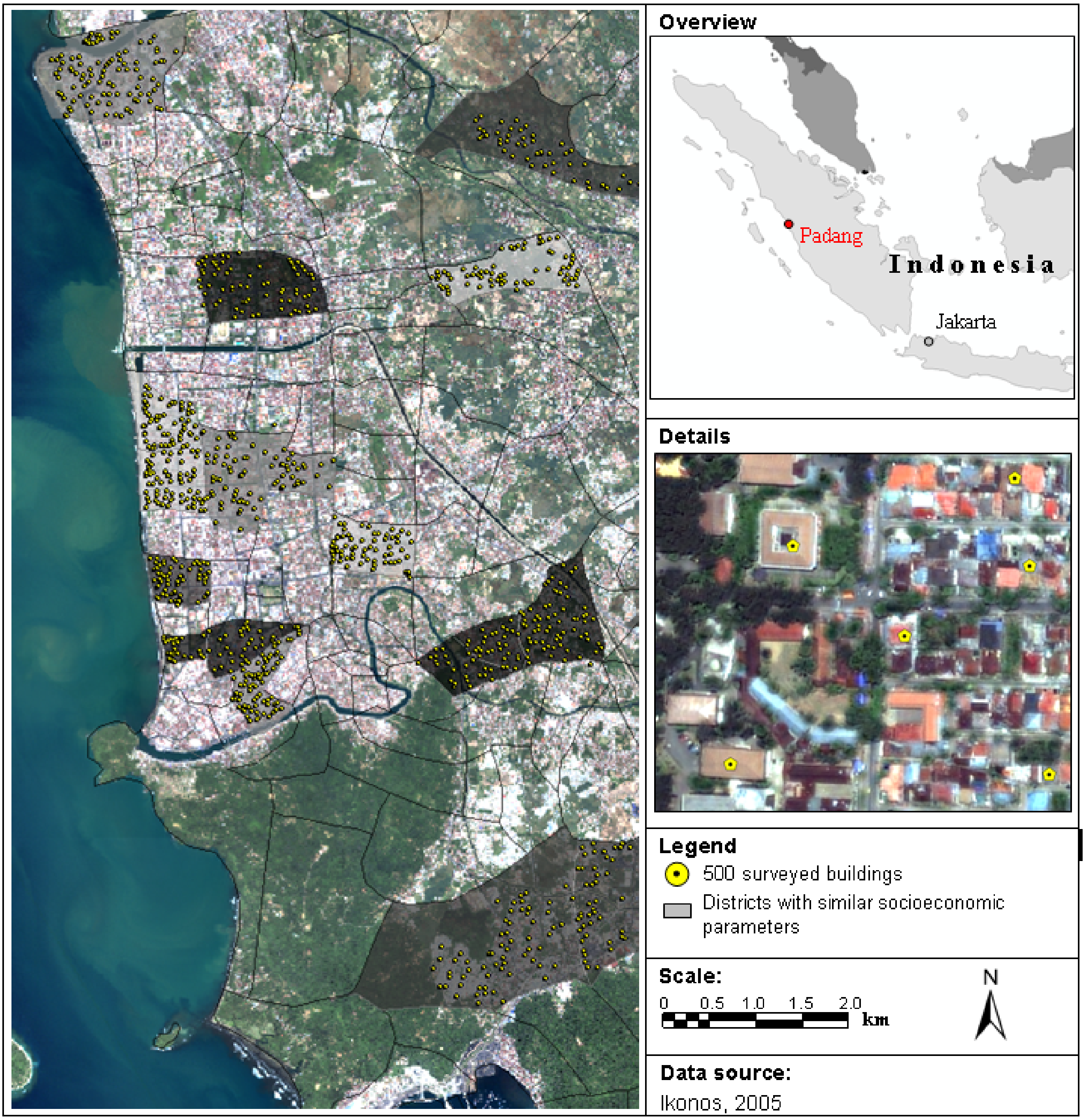

Figure 1. Ikonos imagery of Padang, Indonesia, pre-classification of socioeconomic differences and distribution of 1000 surveyed buildings

(ii) *socioeconomic disparities*; data at Desa- (village) level were available to analyse socioeconomic large-area clusters in the city based on dependency ratio, average household size, economic diversity, primary education, poverty rate [11] [12];

(iii) *physical building types;* disparities in building types – defined by size and height using the classification results from the Ikonos data;

(iv) function of the building; residentially, mixed or commercially used;

(v) per selected Desa area 60-120 sample buildings were consulted; to have a large population per district available. However, the sample size was reduced during implementation due to various reasons – e. g. due to inaccessibility or people refusing to answer – to 49 – 119 per Desa that make a total number of surveyed buildings of 933.

Thus, the distribution of surveyed buildings was systematic to cover the complete spectrum different physical and socioeconomic components within the urban landscape of the city. The selection and distribution of survey samples clearly benefits from remote sensing data and results. Fig. 1 displays the distribution of the selected buildings.

III. METHODOLOGY

A land cover classification derived from the high resolution Ikonos image provides area-wide and up-to-date knowledge on the urban environment. With a combination of an automatic object-oriented methodology [13] [14] and subsequent manual enhancement the accuracy of the results is 97 %. This product displays the basic information to know 'what' is 'where' within the complex urban landscape. The results are eight classes —'individual houses', 'streets', 'sealed areas', 'grassland', 'trees', 'wetland', 'bare soil', and 'water'— mapping highly detailed urban morphology. Thus, the capabilities of the available remote sensing data sets enable an analysis of the urban morphology on house / block level.

We use the land cover classification to calculate physical parameters classifying the urban structure. The street network structures the urban landscape into blocks. The ratio of area of buildings to area of the particular block results in built-up density. The building height is derived through an intersection of the building mask with the calculated elevation difference between the Digital Surface Model (DSM) and the Digital Terrain Model (DTM) [15]. We assessed the number of stories per building calculating the average of available height points per building resulting in an accuracy of 86.7 %. Furthermore individual building sizes are calculated and in addition, average building sizes for particular blocks are derived [16].

Using theses physical urban morphology parameters a semantic classification has been performed. The idea of semantic classification aims at a first assumed interrelation between physically homogeneous sectors within the complex urban morphology and the socioeconomic characteristics of people residing there. The combination of the area-wide available statistical physical parameters describing the building stock of Padang per sector – built-up density, average house size, average building height, location – enables to identify physically homogeneous areas. This approach for semantic classification is generic, aiming at transferability on any urban area throughout the world with similar physical parameters available. Therefore, we used descriptive statistic values as Quartile (Q1, Q3), Median (Med) or Mean (M) to subdivide the different classes. The six resulting semantic classes are defined by the following statistical borders. An example, we classify 'slums' using built-up density values higher than the third quartile (Q3) of the complete spectrum of the built-up density values classified in Padang. Analogous, the buildings of slums are assumed to be smaller than Q3 and lower than Q1 of the particular spectrum of values classified for the city of Padang. Utilizing these statistical parameters the classification is not affected by cultural or regional characters of urban morphology occurring worldwide. With respect to this methodology, only the semantic nomenclature has to be adjusted on the particular structures and locations.

Terminology of the semantic classes is based on housing quality and location. The housing quality is assumed to be higher with rising building size or height and lower built-up density. We classify six different semantic classes – 'slums', 'suburbs', 'low class areas' (LC), 'middle class areas' (MC) and 'high class areas' (HC). For every semantic class we assume typical physical conditions. As an example, slum areas are defined by the highest built-up density measured within the urban environment with mostly one storey buildings, with the smallest buildings sizes. With the nomenclature 'slum' the classification by solely physical parameters sets a first hint in the direction of assumed socioeconomic relevance.

A city is the physical and architectonic reflection of the society that created it. Thus, a change in urban morphology is assumed to involve altering socioeconomic or demographic characteristics. The main idea is therefore based on the assumption that semantic classes defining physical urban structure zones correlate with field work information on socioeconomic parameters. If a correlation between semantic classes and socioeconomic parameters becomes apparent, extrapolation of the punctual survey results would become possible for an area-wide assessment of the whole city of Padang.

The analysis if the socioeconomic parameter derived from the survey correlate with the semantic classification result from remote sensing data is two fold: First of all we test if the complete 1000 surveyed buildings show a basic interrelation with the classification based on the physical parameters and its subsequent semantic assumptions (1). Therefore we use the average values for the particular socioeconomic parameter for every semantic class as well as the standard deviation as a measure of homogeneity within one semantic class. In a second step we analyse if certain semantic classes, like slums or low class areas, although located at different areas of the urban environment (e. g. at the coast close to the city centre or at a suburban area), show homogeneous socioeconomic parameters (2). Therefore we use location-based analysis of the same semantic classes for testing homogeneity.

IV. RESULTS

The result of the semantic classification shows an allocation of 7 % slums, 19 % suburbs, 27 % low class, 46 % middle class and only 1 % high class areas for the complete building stock of Padang. The complete building stock contains 88.004 individual buildings. Fig. 2 displays the semantic classification.

The survey resulted in 445 households providing information on the assessed value of their property if they would sell it now. The missing households refused to answer or did not know the value of the property. 8.7 % of the households of the respondents are located in slums, 12.8 % in suburbs, and 22.0 % in low class, 52.8 % in middle class, and 2.7 % in high class areas. In contrast 826 households from the interviewed households provided information on their income. 7.6 % of the interviewed households are located in slums, 11.5 % in suburbs, and 22.2 % in low class, 56.6 % in middle class

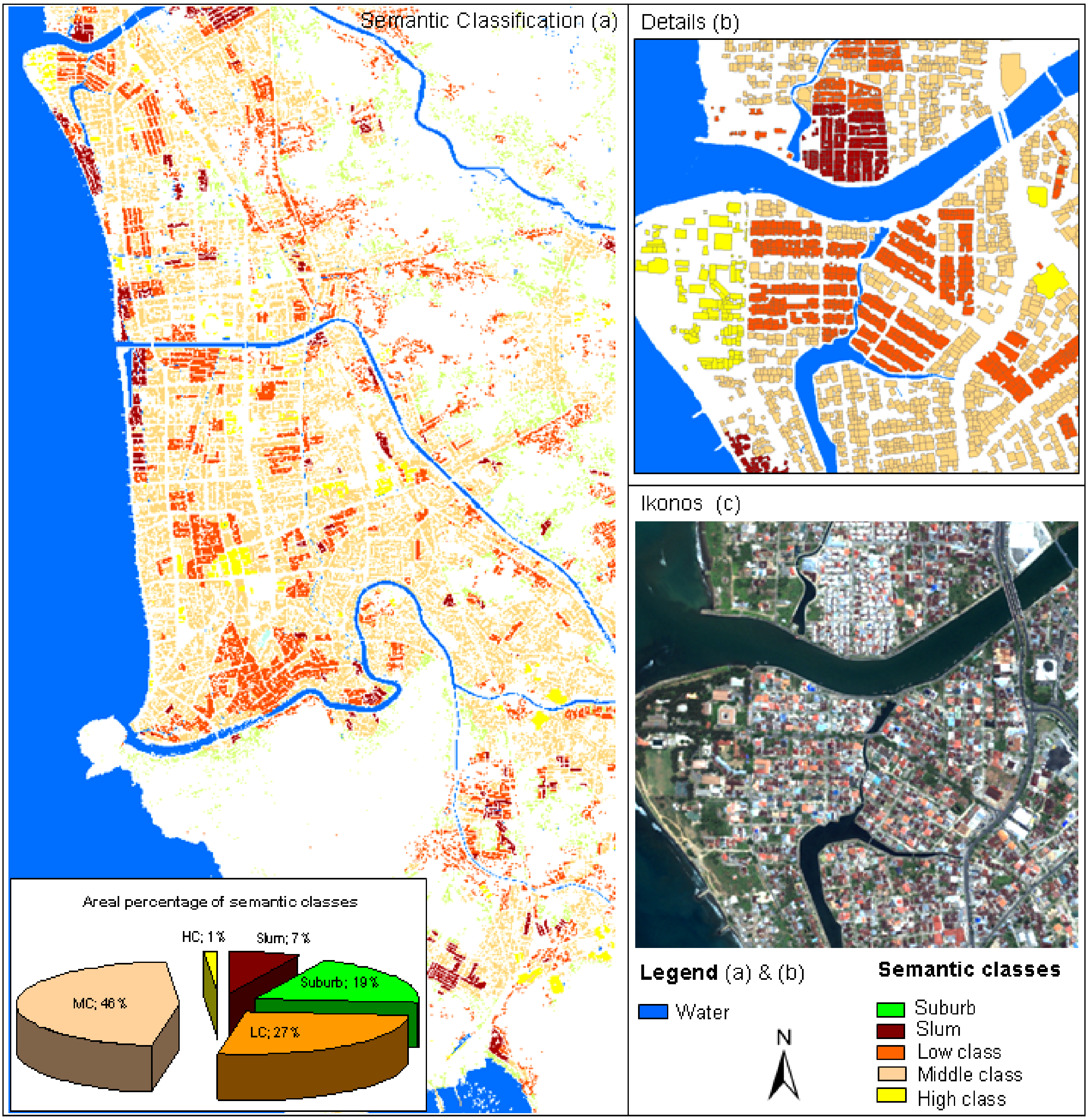

Figure 2. Semantic classification and areal percentage for the building stock of Padang

and 2.1 % in high class areas. The distributions correspond to the frequency of occurrence of the semantic classes in Padang.

As starting point we use Pearson's correlation coefficient, which is a common measure to indicate the strength and direction of a linear relationship between the two variables – income and value of the property. Thus, we see that both socioeconomic parameters, applied to all households providing information on both variables, have a low correlation of 0.21.

At first we analyse if the income and value of property meets our assumptions made by the semantic classification derived from the physical parameters defining urban morphology. Fig. 3 displays the mean of the two socioeconomic parameters –income and value of the property– with respect to their membership to the semantic classes. It becomes apparent that the hypothesis that physical characteristics of buildings basically correlate with both income and property parameters. We observe that in the suburbs income and value of the property is lower than in slums, which are basically all near the city centre. Thus, we reason that the location also has a high impact on the socioeconomic parameters analysed in this study. We also notice that an improvement of the physical structures – the semantic classes 'low class' to 'middle class' – show higher income and value of property than the two previous classes and thus, confirm the hypothesis. The best physical structures – the semantic class 'high class' – contradict to the hypothesis with lower income and values of property than the middle class area.

Figure 3. Mean value correlation of semantic classes and socioeconomic parameters

In a second step we evaluate if individual semantic classes show homogeneity within themselves, or if the location of the classes plays a greater role on the socioeconomic parameter. Fig. 4 displays the statistical analysis of the different semantic classes in three different areas –a northern, central and southern area– within the urban landscape of Padang.

We observe that the trend from the overall analysis is basically confirmed. In detail, the analysis shows that the classified slum areas as well as the classified suburb areas reveal lowest income values independent from their location within the urban landscape. We also found consistent rising income levels to the semantic classes 'low class' and subsequently to 'middle class' areas, and lower income for the 'high class' is confirmed in the northern area, while for the central area we have no high class buildings. Interestingly, in the southern area we observe what was stated in the initial hypothesis – a rising income. This contradicts the overall analysis which showed decreasing incomes for 'high class' areas. The analysis of the 'high' semantic class generally has a high uncertainty, due to very small number of sample buildings with complete information and also the fact that some of these buildings were found to have other functions as merely dwelled houses (e.g. storage, school, etc.). Thus we state that this consistent homogeneity of socioeconomic parameters within semantic classes is not applicable for the high class areas and that the location has higher impact on the income level at this semantic class.

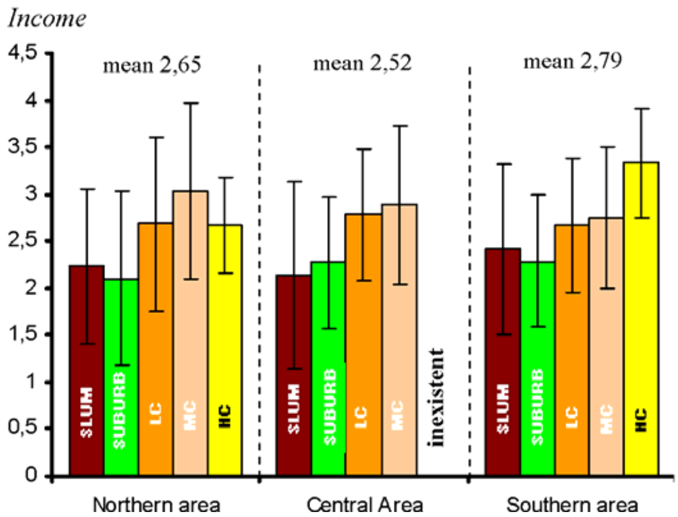

Figure 4. Location-based mean value correlation of semantic classes and the socioeconomic parameter income and their standard deviation

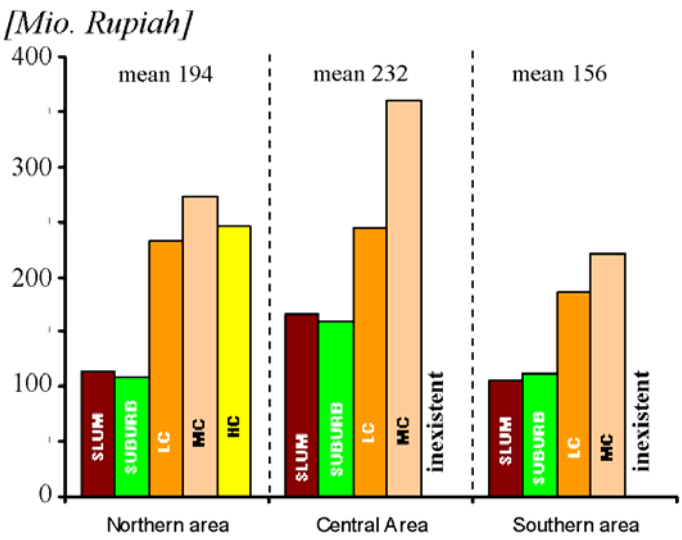

Figure 5. Location-based mean value correlation of semantic classes and socioeconomic parameter value of property

978-1-4244-3461-9/09/$25.00 ©2009 IEEE

With respect to the parameter ‘value of property’ the results basically reveal higher values of the properties for the central area in Padang. The highest mean values of the properties in central Padang demonstrate this observation. As example, the properties in the slum areas are highly conform in the northern or southern parts of Padang. In comparison real estate values of similar structures in the center of Padang are 1/3 higher. Thus, we conclude that the location of the real estates has a much higher influence on the parameter ‘value of the property’ than on the income level of the people at the particular areas.

V. DISCUSSION

Extensive field surveys are highly time- and cost consuming. The combination of social science and remote sensing aims at identification of correlations between physical urban morphology and socioeconomic parameters in order to seek the potential of extrapolation and even spatial transferability of socioeconomic information on the whole area of a city or even across cities.

The results in this study show, that the correlation between physical urban morphology and socioeconomic parameters of the people is basically given. The interdisciplinary approach contains many sources of error: The derivation of physical parameters of urban morphology from high resolution satellite data shows very high accuracies (cp. section. III.). However, misclassifications as well as natural limitations of remote sensing, like assessing wrong ground floors due to roof overhangs produce consequential errors. The idea of semantic classification based on statistical thresholds is a generic and transferable approach, but at the same time the determination of the thresholds still contains subjectivity. Also, the acquisition of field data contains various uncertainties. The large amount of surveyed buildings requires different people conducting the survey interpreting differently or misinterpreting the targeted location of the sample. This already influences ground truth data results. In addition, the interviewees had to assess the amount of people for a household or multiple households for larger buildings, which contains subjectivity of the individual asked person. Another aspect to consider is land or house ownership status that cannot be extracted through remote sensing analysis, e.g. we found out that some of the buildings classified as ‘high class’ were inherited from family or without certificate or traditional land. Especially with the parameter ‘value of property’ the subjective assessment of the interviewees showed a high variance.

Independent from uncertainties funded from various sources of error, the study shows encouraging results. From it, extrapolation can now be calculated on the complete building stock of the urban landscape. This opens up a wide field of applications starting from the support in census questions to urban and regional planning or risk and vulnerability analysis. Within the current project [17], the area-wide physical urban morphology classification in combination with the socioeconomic parameters will be used for evacuation modelling [18].

VI. CONCLUSION AND OUTLOOK

This study shows that physical parameters of the urban landscape like building size, building height, built-up density, vegetation fraction or location clearly correlate with socioeconomic parameters, which were in our case ‘income per month’ and ‘value of the property’. It is interesting to link spatial data with qualitative information regarding the urban development, in order to interprete the socioeconomic characteristics as well as semantic classification in a more comprehensive manner. Thus, remote sensing may provide a cost-effective method to reduce, but not replace, expensive ground data collection. Future analysis will aim at a clear identification if further socioeconomic parameters, like e. g. education level or age pattern, are also correlating with the physical urban landscape. Although to date, social scientists are sceptical that remote sensing can measure anything considered improtant in their fields of study, our first approach opens up a wide field of applications. Our study proves that socioeconomic parameters of the population have a strong spatial component and thus, the integration of remote sensing data and social science aims at interdisciplinary value-adding for applications like urban and regional planning or risk- and vulnerability assessment.

The study was conducted for the Indonesian city of Padang. The question arises, if the identifed correlation between semantic classes and socioeconomic parameters will also be existent in other cities of Indonesia, or even throughout the world? Thus, future analysis also has to take transferability into account.

VII. ACKNOWLEDGEMENT

The authors would like to thank the DFG/BMBF special Programme “Geotechnologies” - Early Warning Systems in Earth Management. Sponsorship Code: 03G0643A-E. We would also like to thank our partners in Padang Indonesia from Andalas University and the city municipality of Padang.

REFERENCES

[1] United Nations, *World Urbanization Prospects. The 2007 Revision – Highlights*, New York, 2008, 244 pp. http://www.un.org/esa/population/publications/wup2007/2008WUP_Highlights_web.pdf

[2] Turner, B. L. II., *Frontiers of Exploration: Remote Sensing and Social Science Research.*, Pecora 13 Symposium, Humannteaction with the Environment: Perspectives from Space (August 20-22, 1996, Souix Falls, SD), pp. 15-19. 1998.

[3] Livermann, D., Moran, E., Rindfuss, R. & Stern, P., *People and Pixels – Linking remote sensing and social science.* ISBN-10: 0-309-06408-2, Natl. Academy Press, pp. 256, 1998.

[4] Rindfuss, R. & Stern, P., *Linking remote sensing and social science: The need and the challenges.* In: Liverman, D., Moran, E., Rindfuss, R. and Stern, P. (eds.). People and pixels. Washington, DC: National Academy Press. 1998.

[5] Anas A., Arnott R., Small K., *Urban Spatial Structure,* Journal of Economic Literature, American Economic Association, vol. 36(3), pages 1426-1464. 1998.

[6] Baudot, Y., *Geographical analysis of the population of fast-growing cities in the third world.* In: Donnay J., Barnsley M, Longley P. Remote

sensing and urban analysis, Taylor and Francis: London, pp.229-246. 2000.

[7] Hoffmeyer-Zlotnik J. H. P., *Das "Wohnquartier" als Hintergrundmerkmal zur Erklärung von Umfragdaten.* Festschrift für Max Kaase. Mannheim: ZUMA 2000, 232 Seiten, ISBN 3-924220-20-4. Mohler, P. Ph. & P. Luettinger (Hrsg.). 2000.

[8] Taubenböck, H., Roth, A., Dech, S., *Linking structural urban characteristics derived from high resolution satellite data to population distribution.* In: Urban and Regional Data Management. In: COORS, RUMOR, FENDEL & ZLATANOVA (Hrsg.). Taylor & Francis Group, London, ISBN 978-0-41544059-2. S. 35-45. 2007.

[9] Wlodarczyk D., *Structural Analysis of urban space in residential areas in "Methodologies in Housing Research"*, Newcastle upon Tyne, The Urban International Press. 2005.

[10] Depnaker (Ministry of Manpower and Transmigration of the Republic of Indonesia): Provincial Minimum Standard Salary 2008. http://www.nakertrans.go.id/pusdatin.html,14,163,pnaker

[11] BPS (Statistics Indonesia), *"Census 2000"*, http://www.bps.go.id/service/index.html; 2008.

[12] BPS (Statistics Indonesia): Poverty Survey for Direct-Cash-Grant Program 2005.

[13] Taubenböck, H., *Vulnerabilitätsabschätzung der Megacity Istanbul mit Methoden der Fernerkundung.* Dissertation. Universität Würzburg; p. 178. (Online-Publikation: http://www.opus-bayern.de/uni-wuerzburg/volltexte/2008/2804/); ISBN-10: 3639083180. 2008.

[14] Taubenböck, H., Roth, A., *A transferable and stable classification approach in various urban areas and various high resolution sensors".* In: Urban Remote Sensing Joint Event, Paris, France. S. 7. 2007.

[15] Intermap, *Product Handbook & Quick Start Guide*, Intermap Technologies, Standard Edition v4.0; 2007.

[16] Taubenböck, H., Post, J., Kiefl, R., Roth, A., Ismail, F., Strunz, G., Dech, S., *Risk and vulnerability assessment to tsunami hazard using very high resolution satellite data.* In: Proceedings of the EARSeL Joint Workshop: Remote Sensing: New Challenges of high resolution. (Eds., Carsten Jürgens). ISBN 978-3-925143-79-3. pp. 77-86. Bochum, Germany. 2008. http://www.earsel.org/workshops/HighRes2008/Artikel/09_Taubenboeck.pdf

[17] Birkmann, J., Dech, S., Goseberg, N., Lämmel, G., Moder, F., Nagel, K., Oczipka, M., Schlurmann, T., Setiadi, N., Siegert, F., Strunz, G. & Taubenböck, H., *Numerical Last-Mile Tsunami Early Warning and Evacuation Information System.* In: International Conference on Tsunami Warning (ICTW), Bali, Indonesia, November 12-14, 2008. pp. 12. 2008.

[18] Lämmel G., Rieser, M., Nagel, K., Taubenböck, H., Strunz, G., Klüpfel, H., Goseberg, N., Matheja, A., Schlurmann, T., Setiadi, N., Birkmann, J., *Emergency Preparedness in the case of a Tsunami - Evacuation Analysis and Traffic Optimization for the Indonesian city of Padang.* In: Proceedings of the Pedestrian and Evacuation Dynamics Conference, Wuppertal, pp.12, 2008.

978-1-4244-3461-9/09/$25.00 ©2009 IEEE