International Journal of Remote Sensing and Earth Sciences Vol. 6: 46-56 (2009) © IReSES

DERIVATION OF POPULATION DISTRIBUTION BY COMBINING CENSUS AND LANDUSE DATA: AS AN INPUT FOR TSUNAMI RISK AND VULNERABILITY ASSESSMENT

M. Rokhis Khomarudin[1], Günter Strunz[2], Joachim Post[2], Kai Zoßeder[2], And Ralf Ludwig[3]

Abstract. Information on population distribution is crucial in disaster risk management. Every disaster such as flood, drought, volcanic eruption, storm, earthquake and tsunamis implies threats to people with respect to loss of live, injury, and misery. Therefore, the information on detailed population distribution in the disaster or hazard zone is important in order to mitigate the impact of natural disasters. Moreover, accurate information on people exposure will help the government to improve the evacuation planning and to decrease the amount of people at risk. The available information on population distribution is mostly based on statistical data related to administrative boundaries, e.g. village, municipal, district, province, or national level. Within the border of administrative boundaries, the population is assumed to be distributed homogenously within each unit area, even in the part of uninhabited areas e.g. lakes, forest, swamps, and areas with high slopes. Hence, this research focuses on the improvement of the available data on population distribution for the area along the west coast of Sumatera, south coast of Java and Bali. The results were used as an input for the tsunami risk assessment in the framework of the German-Indonesian Tsunami Early Warning System (GITEWS) project.

Keywords: *People distribution, spatial improvement, tsunami, remote sensing and GIS*

1. Introduction

In the historical record of UN-ESCAP (2007) more than 3.5 billion people around the world were affected by natural disasters during 1991-2006, approximately 981,000 people were. Therefore, information on population distribution in hazard or disaster zones can mitigate the impact of a disaster. Especially in the case of tsunamis in Indonesia, this population information is important because more than 80% of the main cities in Indonesia are located at coastal areas (Subandono, 2005). Detailed information on the population distribution will help the evacuation planning and risk assessment.

The available information on population distribution is mostly based on statistical data and is related to administrative boundaries, like village, municipal, district, province, or national borders. In most countries of the world and also in Indonesia population distribution data are available for villages as smallest administrative level. The number of people is assumed to be distributed homogenously within each unit area, even in the part of uninhabited areas e.g. lakes, forest, swamps, and areas with high slopes (Bielecka, 2005).

[1] Indonesian National Institute of Aeronautics and Space (LAPAN), Jakarta, Indonesia.
Email: Rokhis.Khomarudin@dlr.de
[2] German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, Germany.
[3] Department of Geography, University of Munich (LMU), Munich, Germany.

Alternative information on population distribution is provided by LANDSCAN and CIESIN data sets. These data sets use a different statistical basis, but they use also methods to derive population distribution depending on their activities at land use classes like settlement, agriculture, plantation, and so on. These data sets provide worldwide population distribution maps but the spatial resolution is not more than 30x30 arc seconds or less than a map scale of 1:1,000,000. For a reliable tsunami risk assessment, data with such spatial resolution is not sufficient. Hence, the current available population data for Indonesia have some limitations especially for tsunami risk assessment. An improvement of the spatial resolution of population distribution data is needed. The distribution of the population which is depending on their activities is the key of this approach.

To point out the requirement of a better spatial resolution for the population distribution a short example is given: For instance, there are two communities with similar size and similar amount of population but the location of settlement areas differs. The location of the settlement area in the first community "A" is 50 m away from the coastline and in the second community "B" about 1,000 m away from the coastline. With only the statistical data information, it is assumed that the people are distributed homogenously throughout the community area and hence, the vulnerability level of the two villages is more or less the same. But with a downscaling approach it becomes apparent that the people distribution in both communities is quite different and thus, the vulnerability of village "A" is much higher than in village "B".

In the framework of the GITEWS project this approach will be applied for the whole west coast of Sumatera, south coast of Java and Bali to provide population distribution information which is spatially explicit and provide an improved spatial resolution.

2. Previous Studies

Several methodologies have been developed for improving the information of population distribution. The method was first developed and named by the Russian Cartographer, Tyan-Shansky (1827-1914), who in the 1920s developed and published a multi-sheet population density map of European Russia at a scale of 1:420,000 and popularized by Wright (1936). As well as the other forms of thematic mapping, the dasymetric method was created and used because of the need for accurate visualization methods of population data. Now dasymetric mapping is widely used for producing finer population distribution maps. The process of this technique became easy and faster with automatic tools such as geographic information systems (GIS) and remote sensing.

There are two main focuses in this context of the previous investigations. The first is to estimate the population distribution based on inhabited and uninhabited area derived from Census data (as main data) and earth observation as ancillary data, such as land use/cover, Digital Elevation Model (DEM), street network, and night time lights (Landscan: Dobson et al., 2000; CIESIN: Balk et al., 2005; Mennis, 2003; Mennis and Hultgren, 2006; Tatem et al., 2007; Schneiderbauer 2007; Gallego, 2007 and Bielecka, 2005). The second is to estimate population distribution based only on the characteristics of remote sensing data. The investigations show correlations between settlements, impervious zones, and built up area with population distribution. Some correlations have been developed by Sutton et al. (2001), Martinuzzi et al. (2006), Liu et al. (2006), and Lu et al. (2006).

This study uses all available data such as the number of people per village from census data, land use data from topographic maps and also people growth rate to estimate the current population distribution.

3. Data Availability

Land use classification, census data, and Earth Observation (EO) data are available for the whole west coast of Sumatera, south coast of Java and Bali. One problem in using such data integration approach is the different scales of the data and the occurrence of some missing data (blank areas) of the statistical data set in the west coast of Sumatera. The land use information derived from topographic maps on scale 1:25,000 (1999) which is is available for Java and Bali. For Sumatera there is only available topographic information on scale of 1:100,000 (2002), except for Padang and Banda Aceh (scale 1:50,000). The CENSUS data set is available for the year 2000 and the potential of village data (PODES) is available for the year 2005. To estimate population in the year 2007 or in the future, the population growth rate data set (from 2000-2007) is used. The statistical databases are provided by the Indonesian Statistical Bureau (BPS) and the topographic information is provided by the Indonesian National Coordinating Agency for Surveys and Mapping (Bakosurtanal).

4. Concept and Methodology

4.1. General Concept

The applied concept can be classified as a so called top-down method of people distribution (for example Taubenböck, 2008). The disaggregation of the population from CENSUS data to land use classes is the core of this analysis. Statistical analysis of the activities of the people is used to allocate weights for the disaggregation. Figure 1 points out the influence of the weighting factors by the disaggregation method.

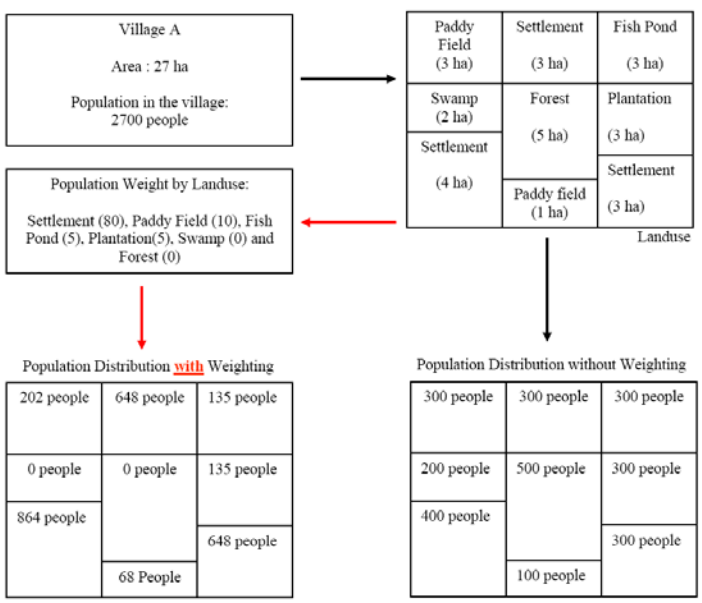


Figure 1. Synthetic example of the used disaggregation method to improve the spatial population data.

The formula for population disaggregation is given below:

$$X_d = \sum_{i=1}^{n} P_i \quad (1)$$

$$P_i = \sum_{J=1}^{k} P_{ij} \quad (2)$$

$$P_{ij} = \frac{S_{ij}}{\sum_{j=1}^{k} S_{ij}} . W_i . X_d \quad (3)$$

where:

X_d = Number of people in administrative unit

P_i = Number of people in land use i

P_{ij} = Number of people in polygon j in land use i

S_{ij} = Size of polygon j in land use i

W_i = Weighting of land use i $\sum_{i=1}^{n} W_i = 100\%$

K = Number of polygons in land use i

n = Number of land use in a village d

Eq. (1) explains the distribution of the population from one administrative unit to several land use polygons inside that administrative boundary. Eq. (2) explains that the population of one certain land use is the accumulation from several polygons of the same land use in the administrative boundary. For instance, one land use (for example settlement) could have more than one polygon with different sizes inside an administrative boundary. Eq. (3) explains that the population of one polygon is proportion of the area of the polygon to the total area of the polygons of the same class. Besides that every polygon is weighted depending on the people activity which characterises a land use class (W_i) which will be discussed in the following.

4.2. Weighting Factor Derivation

The weighting factor is the value of proportion to of the population in several land use polygons within an administrative boundary. The sum of all weighting factors in one administrative unit is 100%. If an administrative unit (e.g. village) contains only one type of land use the spatial improvement is, of course, not possible, in fact it is not necessary. The weighting factors depend on the functional land use for human activities. For example, for the land use class “settlement” it is expected to have a high weighting factor, because most of human activities take place in the settlement area. But for classes like “forest”, “swamp”, and “water body”, the weighting factor is nearly zero because of very rare human activities there. To determine the weighting factor for each land use class, statistical data was analysed.

There are two kinds of statistical data available, the potential village (PODES) and the CENSUS data. The PODES data set contains information on the main sources of income in a community. This represents the number and kinds of human activities in the certain communities. The information on the number of workers and non workers in the PODES data are used as indicator of the number of people staying at home (settlement area) and the number of people going out to work. The percentage number of people of certain occupation in each community derived from CENSUS data is used to calculate the potential number of people staying in a certain land use. This value is determined by the distribution (more specifically the median) of the number of people of certain occupation characterising a land use class (e.g. occupation of farmer will refer to the land use class of agriculture area) for all villages at the observed region. To improve the determination of the weighting factors and to investigate probable regional variations, different characteristics of Sumatera, Java and Bali, of urban and rural areas and also of coastal and non coastal areas are analysed. Moreover, the weighting factors are determined differently for day and night time. The significance of the aforementioned difference is caused by the change of the number of workers or non-workers for certain land use class, which for example the change of the weight factors of settlement will be less than the change of the weight factors of other land use classes.

4.3. Multi Scale Disaggregation

The available population data per village of the CENSUS data is not complete, especially at the west coast of Sumatera. The multi scale disaggregation is an approach to combine the disaggregation at village level and at district level to fill in the data of areas with data gaps. In these areas the uncertainty of the population distribution modelling will apparently be higher.

4.4. Error Analysis

As in every approach with many spatial databases as input, there are some sources of errors in the present study, which should be studied. These errors are related to the databases themselves, the weighting factor determination, and the multi scale disaggregation. Here, the focus of the error analysis is on the uncertainty of the population modelling such as the error caused by weighting factors and multi scale disaggregation. The error of the input data set here is not taken into consideration. The following error calculations are used:

1. We assumed that the proportion of people who are located in a certain land use category from occupation data (CENSUS) calculation is the real number of people in that certain class (example class i) and indicated as Pi_real which will be compared to the proportion of people who are located in the same land use category (class i) calculated from models which is indicated as Pi_mdl. The error is calculated using Eq. (4) below.

$$Error(\%) = \frac{|Pi\_mdl - Pi\_real|}{Pi\_real} X100\% \qquad (4)$$

   These errors were calculated for 9,100 villages.
2. A comparison between the people number for village level from CENSUS data and the population number estimation based on multiscale disaggregation was done. A RMSE (Root mean square error) analysis is performed to show the error of multiscale analysis for village level disaggregation and districts disaggregation. This analysis is done only for the west coast of Sumatera where the multiscale disaggregation method was applied.

5. Results and Discussion

5.1. Weighting Factors

For the disaggregation of population per village to land use classes weighting factors are calculated following the method explained before based on the database of PODES 2005 and CENSUS 2000. The weighting factors are divided into day time and night time. One of the results of the statistical analysis is that the characteristics of occupation in Sumatera, Java and Bali are not significantly different. Therefore the weighting factors for different set of occupations have not been considered. The significant differences of occupation occur at urban and rural areas, at settlement and non settlement areas and at coastal and non coastal area. The weighting factors generally divide the land use classes into two categories: inhabited and uninhabited areas. Inhabited areas mean that some people do their activities there and so the weighting factor is not zero. Uninhabited areas means that no one is in that land use class. The weighting factor is then more or less equal to zero. For example during day time inhabited land uses are settlement, agriculture, plantation, beach, fishpond, and open field/grass and uninhabited land uses are forest, swamp, water, shrub, and river.

The results of the analyses of the weighting factors, depending on different statistical databases, are shown in Table 1. There are two models of weighting factor which had been determined. In the first model, weighting factor model 1, we differentiate weighting factors into urban coastal, urban non coastal, rural coastal, and rural non coastal. In the second model, there is no differentiation in those village categories. To evaluate which model is better, the error and uncertainty analysis is conducted.

Table 1. Weighting factor during daytime based in two models [%].

Land use	Weighting Factor Model 1				Weighting Factor Model 2			
	Urban		Rural		Urban		Rural	
	Coastal	Non Coastal	Coastal	Non Coastal	Coastal	Non Coastal	Coastal	Non Coastal
Settlement	84	84	79	79	75	75	75	75
Agriculture	4	8	11	15	13	13	13	13
Plantation	2	4	2	3	3	3	3	3
Fishpond	1	1	1	2	1	1	1	1
Beach	8	0	6	0	6	6	6	6
Open field/grass	1	3	1	1	2	2	2	2
Uninhabited	0	0	0	0	0	0	0	0

5.2. Spatial Improvement

The results of the established population distribution model can be successfully used for the spatial improvement of people distribution of the west coast of Sumatera, south coast of Java and Bali. Figure 2 shows one result of the spatial improvement of the population distribution data in Cilacap (south coast of Java). It is demonstrated that for example population information at river structures, near the beach, at settlement or in the forest are more realistically represented compared to the CENSUS data set.

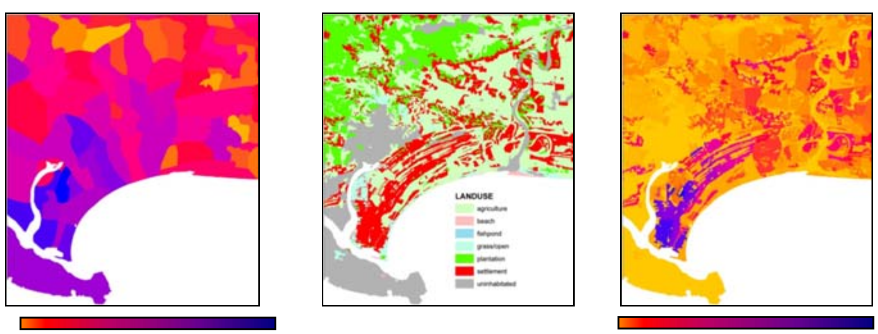
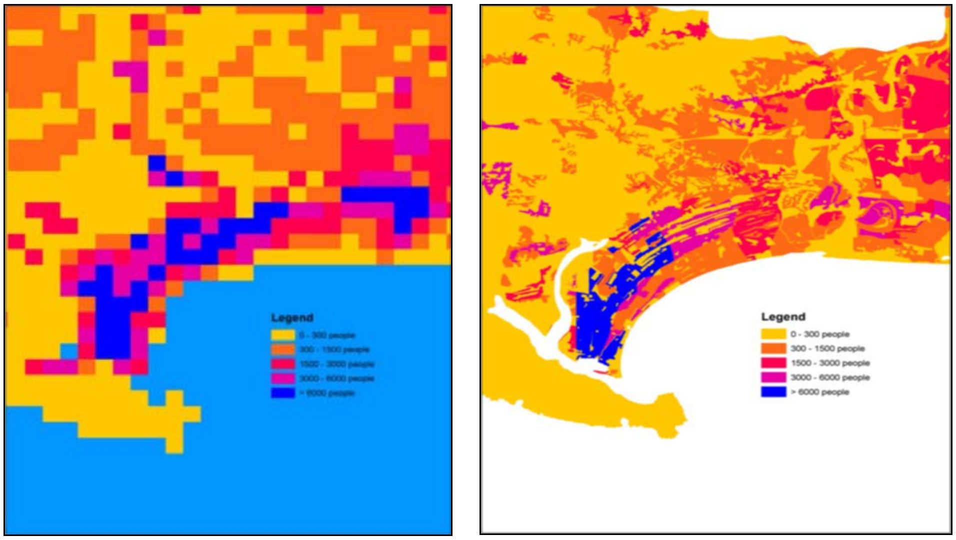


Figure 2. Spatial improvement of population distribution for Cilacap.



Landscan product. Research result.

Figure 3. Comparison of research result with Landscan result.

The comparison of resulting population distribution data with Landscan data is shown in Figure 3. Regarding the spatial resolution, this research approach can considerably improve the information from the Landscan product. According to the land use data with a scale of 1:25,000, this research result could have 30 m x 30 m spatial resolution.

5.3. Error Analysis

To evaluate which sets of weighting factors should be used an uncertainty analysis is performed. The results of error analysis in 9100 villages and land use categories are shown in Figure 4.

The error curves of the weighting model 1 and weighting model 2 (Figure 4) in settlement, agriculture, and open land differ significantly, but in plantation, fishpond, and beach do not different significantly. It means that the differentiation of villages categories as urban coastal, urban non coastal, rural coastal, and rural non coastal in settlement, agriculture, and open land is necessary, but for other land uses (plantation, fishpond, and beach) is not necessary. The error of weighting model 1 is better than the error of weighting model 2. It is clear because of the differentiation of village categories in weighting model 1 is done. The error of people distribution in settlement area is lowest than other land uses. Almost 100% from 9,100 villages have error less than 30%, and 90% have error less than 10%. In the other land uses, 80% from 9,100 village have error less than 40% and 10% have error more than 60%.

This result means that fluctuation of people distribution in settlement area is not high compare to other land uses. It shows that the differences of people which have occupation in settlement (e.g. service, trading, etc.) and proportion of non workers in Sumatera, Java and Bali is not much. People in Sumatera, Java and Bali have similar characteristics or proportion of non worker and people which have occupation in settlement.

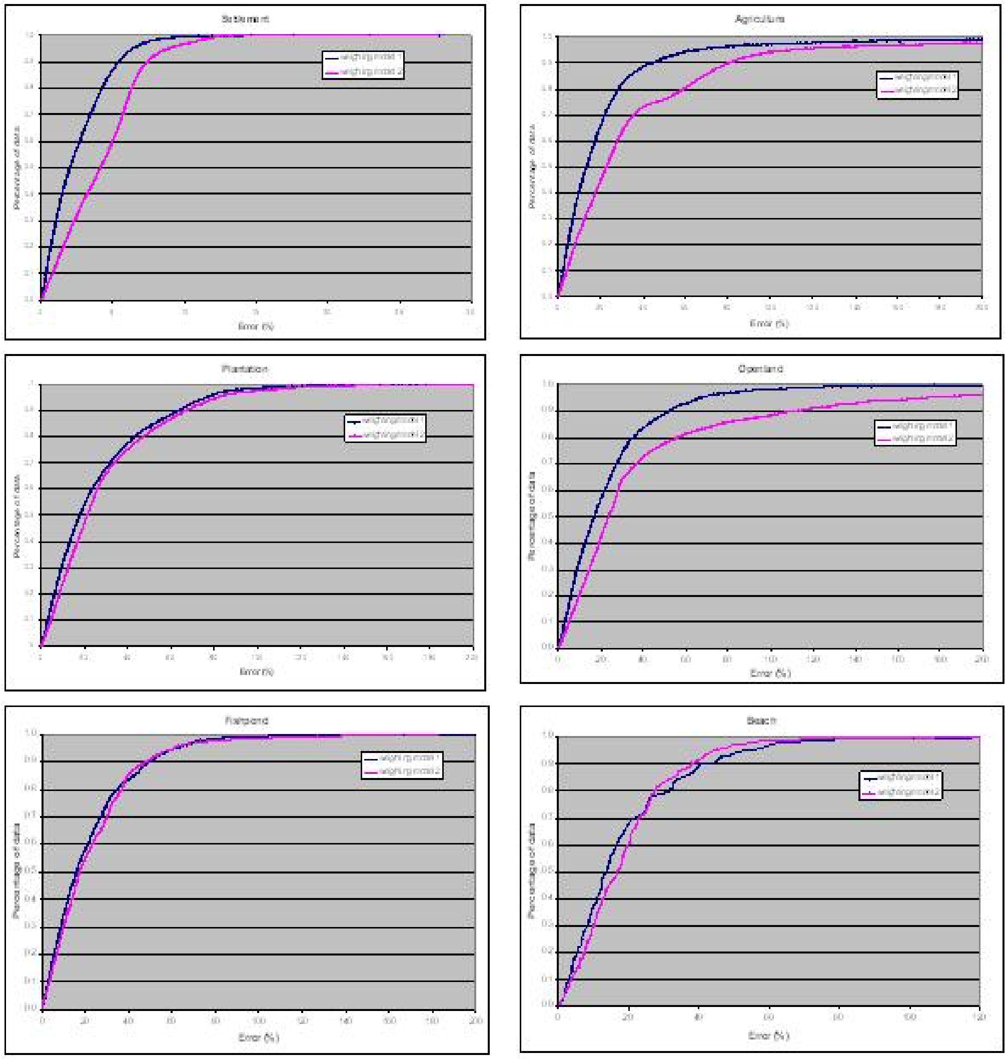
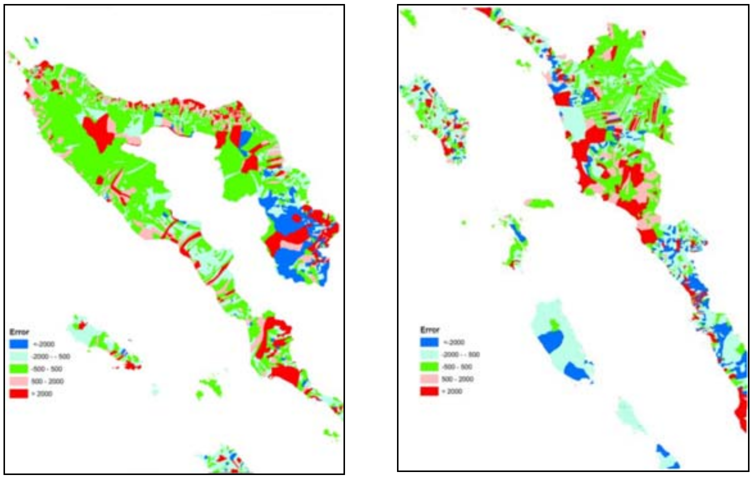


Figure 4. Error of people distribution in several land uses using two models of weighting factors.

5.4. Error Analysis of Multi Scale Disaggregation

In the case of Sumatra, a multiscale disaggregation is needed to fill the missing population data on village level. As a result of the multiscale disaggregation, the minimum error in the Jambi province is about 660 people and in the Aceh province about 710 people. The maximum error is in the Lampung province about 6,690 people. These errors of multiscale disaggregation are too high for population distribution suitable for risk analysis. The population distribution modelling at this region should be improved in the future by using more detailed land use data and CENSUS information. Figure 5 shows the spatial error of districts at the west coast of Sumatera.



Aceh. West Sumatera.

Figure 5. Spatial error of district disaggregation.

6. Conclusions

The results showed that improvement of spatial information on population distribution using land use classes and statistical data analysis (PODES, CENSUS) has been successful. The method was performed using weighting factors which describe the people's activities in certain land use classes. The output of the established population model can provide more detailed information than the single information of the CENSUS data. Also the comparison between the LANDSCAN data set and the present model showed similar population distribution but an obvious improvement of the spatial resolution in the presented model results.

The results of errors using two models showed relatively low errors especially in settlement area which are considered the most important land uses in this analysis. The model with differentiation of village categories; urban coastal, urban non coastal, rural coastal, and rural non coastal, gaves better result which means produced less error than the model without differentiation of village categories. The error by using district disaggregation methods leads to an error in the population distribution model which is too high for a reasonable usage.

In the remote sensing perspectives, the development of methodology to classify land use categories which related to population location such settlement is necessary to produce a more accurate detection of land use classes. The better accuracy of land use classification will yield the population distribution analysis more accurate. The freely available of Landsat TM data in internet, provided by USGS, is one will give opportunity to fully fill in the blank of land use data in Indonesia of certain scale, especially in Sumatera. Additionally, this research has also the significance of the finding which basically to transform from the tabular population data into population data with regard to landuse.

Along the process of transformation, the finding suggest the relation between occupation (e.g. farmer) to appropriate land use/cover which is agriculture land. The finding also show that there is improvement of scale of result. The suggestion of using multiscale disaggregation, i.e using district level, because of missing data is a frequent occurence in developing countries such as Indonesia. Although the findings show that the multiscale disaggregation does not work properly which need more work to apply this method in the future, however the basic idea is benefecial worth further study.

7. Acknowledgments

This research work has been performed in the framework of the GITEWS project. The GITEWS project is carried out through a large group of scientists and engineers from the German Research Center for Geosciences (GFZ, consortium leader) and its partners from the Alfred Wegener Institute for Polar and Marine Research (AWI), the German Aerospace Center (DLR), the GKSS Research Center, the German Marine Research Consortium (KDM), the Leibniz Institute for Marine Sciences (IFM-GEOMAR), the United Nations University (UNU), the Federal Institute for Geosciences and Natural Resources (BGR), the German Agency for Technical Cooperation (GTZ), as well as from Indonesian and other international partners. Funding is provided by the German Federal Ministry for Education and Research (BMBF), Grant 03TSU01.

Main parts of this research have been carried out under a PhD scholarship provided by United Nations University, Institute for Environmental and Human Security (UNU-EHS), and with support from DLR and the Indonesian National Institute of Aeronautics and Space (LAPAN).

8. References

Bielecka, E., 2005, A Dasymetric Population Density Map of Poland, Institute of Geodesy and Cartography ul. Modzelewskiego 27, 02-679 Warsaw, Poland.

Balk, B. L., U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay and A. Nelson, 2005, Determining Global Population Distribution: Methods, Applications and Data, CIESIN (Center for International Earth Science Information Network), Columbia University, PO Box 1000, Palisades, NY 10964, USA, dbalk@ciesin.columbia.edu.

Dobson, J. E., E. A. Bright, P. R. Coleman, R. C. Durfee, and B. A. Worley, B. A., 2000, LandScan: A global population database for estimating populations at risk, *Photogrammetric Engineering and Remote Sensing,* 66:849-857.

Diposaptono, S., 2005, Tsunami Disaster Mitigation in Indonesia, APEC-EqTAP Seminar on Earthquake and Tsunami Disaster Reduction. September 27-28, 2005, Jakarta, Indonesia.

Gallego, J., 2007, Downscaling Population Density with Corine Land Cover, Integrating Socio-Economic and Remote Sensing Information for Food Security and Vulnerability Analyses, Technical Seminar, 12-13 Oktober 2007. (Power Point presentation)

Liu, X., K. Clarke and M. Herold, 2006, Population Density and Image Texture: A Comparison Study, *Photogrammetric Engineering & Remote Sensing,* 72(2): 187-196.

Lu, D., Q. Wenig, and G. Li, 2006, Residential population estimation using a remote sensing derived impervious surface approach, *International Journal of Remote Sensing.*

Martinuzzi, S., W. A. Gould, and O. R. M. Gonzalez, 2006, Land development, land use, and urban sprawl in Puerto Rico integrating remote sensing and population CENSUS data, *Landscape and Urban Planning*, 79:288-197.

Mennis, J., 2003, Generating Surface Model of Population Using Dasymetric Mapping, *The Professional Geographer*, 55(1):31-42, Association of American Geographers. Blackwell Publishing.

Mennis, J. and T. Hultgren, 2006, Intelligent Dasymetric Mapping and Its Application to Areal Interpolation, *Cartography and Geoinformation Science*, 33(3):179-194.

Sutton, P., D. Roberts, C. Elvidge, and K. Baugh, 2001, Census from Heaven: an estimate of the global human population using night-time satellite imagery, International Journal of Remote Sensing.

Schneiderbauer, S., 2007, Population Densities in Zimbabwe. Integrating Socio-Economic and Remote Sensing Information for Food Security and Vulnerability Analyses, Technical Seminar, 12-13 October 2007. *(Paper and power point presentation).*

Tatem A. J., A. M. Noor, C. von Hagen, A. Di Gregorio, and S. I. Hay, 2007, High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa, *PLoS ONE*, 2(12): e1298. doi:10.1371/ journal.pone.0001298.

Taubenböck, H., A. Roth, and S. Dech, 2008, Linking Structural Urban Characteristics Derived from High Resolution Sattelite Data to Population Distribution, Urban and regional Data Management – Coors, Rumor, Fendel & Zlatanova *(eds)* @2008 Taylor and Francis Group, London, ISBN 978-0-415-44059-2.

UN ESCAP. United Nation Economic and Social Commision for Asia and Pacific. www.www.unescap.org .

Wright, J. K., 1936. A method of mapping densities of population with Cape Cod as an example, *Geographical Review*, 26:103–10.