ESTIMATION OF ECONOMIC DATA FOR TRAVEL DEMAND FORECASTING BASED ON PHYSICAL INFORMATION

Kazuaki MIYAMOTO
Professor
Center for Northeast Asian Studies
Tohoku University
Kawauchi, Aoba-ku, Sendai,
980-8576 Japan
FAX: +81-22-217-7477
E-mail: miyamoto@plan.civil.tohoku.ac.jp

Takashi UCHIDA
Associate Professor
Department of Civil Engineering
Graduate School of Eng., Tohoku University
Aoba 06, Aramaki, Aoba-ku, Sendai,
980-8579 Japan
Fax: +81-22-217-7479
E-mail: uchida@civil.tohoku.ac.jp

Yunika NISHIMURA
Civil Engineer / Planner
Mitsubishi Heavy Industries, LTD.

2-5-1, Marunouchi, Chiyoda-ku, Tokyo,
100-0005 Japan
FAX: +65-386-9772
E-mail: 18723@notes.mmw.mhi.co.jp

Keiichi KITAZUME
Research Associate
Center for Northeast Asian Studies
Tohoku University
Kawauchi, Aoba-ku, Sendai,
980-8576 Japan
FAX: +81-22-217-7477
E-mail: kitazume@cneas.tohoku.ac.jp

Abstract: In the study of transportation in developing countries/areas, economic data such as population and employment is not always available. However, the conventional survey which requires many surveyors and many hours of fieldwork, does not necessarily guaranty accuracy of the data. On the other hand, physical data such as remote-sensing data are in general available even in little developed area and also represent economic activities. The purpose of the present study is to develop a system to estimate economic data based on physical data for transportation study in developing countries/areas. The system requires a limited number of economic data samples to calibrate estimation equations of economic data. We apply this system to the Sendai Metropolitan Area to test its validity in a situation where a complete set of data is available. The result shows that the system can substitute to some extent the conventional survey of economic data.

Key words: Remote-sensing, Aerial photography, Physical data

1. INTRODUCTION

Regional and transportation planning always requires exact economic data such as population and employment of a target region. In developing countries/areas, however, there are many cases that economic data are not available, since no system that surveys these economic data is developed. Because of this, it sometimes becomes needed to survey them newly.

The survey to obtain economic data needs not to be carried out with conventional methods in advanced countries, but to be done in accordance with the actual circumstances of developing countries/areas. In these countries/areas, undeveloped transportation networks cause some difficulties of the survey in all of regions. High costs with many surveyors are often required especially in rural areas. Rapid economic development causes survey results separated largely from the present condition, since it will be summarized very at the later time. The surveyors who are not skilled well bring about bad influences to the reliability of results. The simplicity of methods, prompt reports and reliability are required in surveying economic data in the developing countries/areas.

On the other hand, remote-sensing data obtained from an artificial satellite is an excellent data resource in terms of simplicity of methods, prompt reports and reliability. Remote-sensing data is available in large regions in most of countries. This is also able to obtain the latest data of area instantaneously. Furthermore, this is able to disregard errors in the field data collection stage besides it is easy to process the data because of digitalized data. Remote-sensing data seems to include socio-economic information of the areas to some degree, although that itself is not the possibility economic data. Therefore, this provides useful information needed for regional and transportation planning by the method as described later.

The simplicity of methods, prompt reports and reliability of survey are also requested even in advanced countries. In addition to these points, remote-sensing data also takes a low cost. Especially this may be useful very much in the trend prediction in a short term.

The purpose of this paper is to develop a method to estimates social economic data from remote sensing data. In this paper, the data represented to remote-sensing data, which are seen to eyes and also are able to measure physically, can be defined as physical data. Population, output of industries, commercial sales needed for regional and transportation planning are social economic data stemmed from each activity of cities. These data are not seen to eyes and are not able to be measured physically. We can also say that the purpose of this paper is to develop a method by which we can estimate social economy data from physical data.

Population is estimated from aerial photographs by using the simple regression analysis. Many attentions to relation between distance from center of a city and population density are directed, and for example, distance from DID, population density and possession of cars are connected by some functions. These researches seem to be insufficient to estimate economic data represent of complicated urban activities. Miyamoto (1988) expanded the estimation object even to the quantity of economic activities and theoretically developed a method which is able to estimate various city activities by using a principal component analysis.

High reliable regression analysis was made, especially, in the estimation of population, and indicated the possibility of basic survey with simplicity and promptitude of report by adjusting the scale of sampling in accordance with accuracy required. The fuzzy theory was adopted to consider the ambiguousness of the data and used remote-sensing data by an artificial satellite from the viewpoint of prompt report and easiness of processing data. The accuracy of the estimation is obtained to some degree. However, the relation between the accuracy of data and acquisition cost of physical data is not clear. In other words, the cost performance such as level of difficulty of acquisition and a sample ratio for the improvement of accuracy of an estimation system are not clear, when an appropriate physical data is selected from various ones. Because of this, the present paper makes clear the relation between the accuracy of physical data and its acquisition cost and the effect of remote-sensing data as a physical data. Furthermore, the present paper studies the influence of classification for decreasing the number of sampling data in a block random sampling.

2. DATA AND METHOD

2.1 Physical Data and Economic Data

The difficulty of acquisition of data for surveying condition of a city is various. The data that we can read from a map or an aerial photograph is easy to acquire. The data which we can obtain only by visiting houses is to be difficult to acquire. In this paper, the data obtained from a city is divided to two and we consider the problem that economic data that is difficult to acquire is estimated by physical data that is easy to acquire.

Physical data and economic data are in relative relation from a viewpoint of the possibility of acquisition. Physical data in this paper is the data that is seen to the eye and possible to read from an aerial photograph or a map. In the sense, physical data can say in other words as visible data. On the other hand, economic data indicates the physics quantity from the various activities of the human beings such as population, outputs of industries, commerce sales amount, transportation density and so forth. Economic data is located in the opposite pole to physical data, since this is generally not seen to the eye and difficult to read from an aerial photograph or a map. Actual data are located between these both data and are classified according to the difficulty of acquisition, cost restrictions, purpose and so on. We can define level of physical data as the position on the axis between physical data and economic data. Level 1 indicates the easiest data to seen to the eye and to obtain.

The performance of an estimation system in this paper should be evaluated comprehensively by the accuracy, cost, period of the estimation. The survey of economic data such as population and employment requires long time and many costs, since this is the hearing surveys by many surveyors. In spite of long time and many costs, high accuracy of survey can not be expected, especially, in developing countries. On the other hand, the physical data is easily obtained, since this is from remote-sensing data, aerial photographs or simple field surveys. The field survey here is thinking the survey that is not influenced to the ability of surveyors. This kind of survey dose not require high skilled surveyors and can be easily controlled, since main works are done by a few experts. The period of survey, therefore, becomes short and the cost becomes inexpensive.

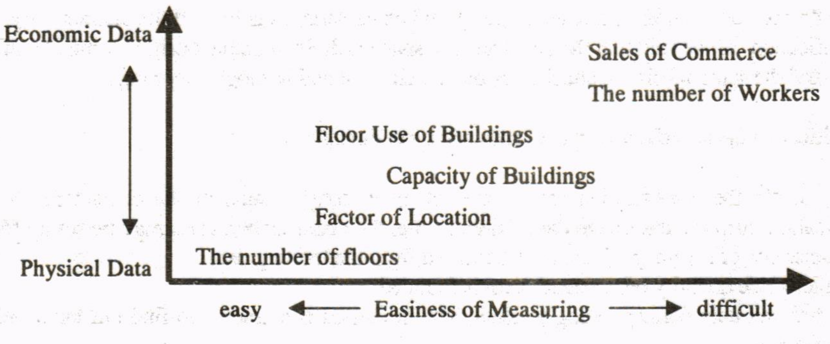

Figure 1. Physical Data and Economic Data in terms of Easiness of Measuring

The physical data is ranked from a viewpoint of the easiness of acquisition as follows. First of all, level 1 is Thematic Mapper data itself that is obtained from the artificial satellite, LANDSAT. Thematic Mapper data is able to be processed by a few experts, since it is digital data. Level 2 is the group of data which can be obtained from aerial photographs. Although these are more expensive data in comparison with level 1 and are analogue data, furthermore the smaller data are obtained than the TM data in terms of the definition in the present. Needless to say, it becomes a premise that there are aerial photographs in the area. Level 3 makes the group of data that is obtained from simple field survey and aerial photographs. Actually, each data is located consecutively on an axis. The economic data are also located from a viewpoint of acquisition. Therefore, each data is located on these two axes. For example, data about building are located in Figure 1.

2.2 Remote-Sensing Data

The remote-sensing data in this paper is obtained from TM (Thematic Mapper) data of artificial satellite LANDSAT. LANDSAT, an earth observation satellite, saves observation data with a regular cycle about almost all area on the earth. This is radiating radio waves of the different frequency toward the ground surface and receives the reflection from ground surface. The difference of reflection rate indicates some information regarding ground surface. The data are used by more researchers than the one of other satellite because of the upgrading disclosure. The data are now used not only as land cover data, but also as the survey data of natural resources, environment surveillance and so on.

There are two reasons why TM data of this LANDSAT are used in this paper. The first, this data is saved as digital data. This data is Computer Compatible Tape and the difference of reflection rate is given by values with 256 gradients every pixel of the size of about 30m × 30m. This is indicating that it is possible to treat this data easily in a computer like other physical data. The second, this data is convenient to be observed the change of information periodically, since the satellite comes back on the same point every 16 days. There is an important meaning that we can observe the rapid change of the city area in developing countries.

2.3 Classification

Many researchers, for example, Hutchinson (1982) and Mason *et al.* (1988), examined classification of remote-sensing data or TM data. However, in this paper, the purpose of classification of remote-sensing data is to identify each land use which is included to the data. The characteristics that we perceive in the classification are density, texture and so on of each picture element. The space defined with the constitution vector with such plural characteristics is termed the characteristic space. The classification can be said that the characteristic space is divided on the basis of some classification standards and the same labels are attached as picture element and homogeneity range.

The classification is carried out with the following procedure almost.

(1) First of all, the classification classes are set up in consideration of the characteristics of the application purpose and image data. Furthermore, the classification class may be set up from the characteristic of the image data that is extracted from the training data.
(2) Characteristic quantity represented them is founded.
(3) The training data corresponding to classification classes is extracted to find out the standard of classification.
(4) It is adopted as a method that the parts represented given classification class are sampled from the image and the characteristic of the population is estimated.
(5) Every picture element is classified to the class by using the classification standard set up above. There are two methods that the elements are classified every 1 picture element and are done every the homogeneous range where it is divided previously. There are multi level slice

classifier, decision tree classifier, minimum distance classifier, maximum likelihood classifier and so on, and the present paper adopts the minimum distance classifier as described later.

The accuracy and reliability of classification are confirmed by comparing the estimated data and the training data and the data whose classification classes are known.

2.4 Minimum Distance Classifier

Minimum distance classifier is the method that resemblance degree between the picture element data and the characteristic of classification class is represented with the distance in characteristic space and these data are classified by the shortest distance. As the distance in the characteristic space, the Euclid distance, standard Euclid distance and Mahalanobis' generalized distance are used.

In this paper, the neural network is adopted as the method and compared with the conventional regression analysis to estimate the accuracy. The urban economic data such as the population, output of industries are the objective variables and physical data are the explanation variables.

2.5 Block Random Sampling

This estimation system tries to develop the model that estimates the urban economic data of all area on the basis of the ones in the points that were sampled. The urban economic data of high accuracy should be done with small number of sample preferably in terms of cut of the labor and cost.

For regarding this point, the block random sampling is adopted on the basis of the trend of land uses with regard to the zone of the target area that is the result of classification analysis above mentioned. Because the unit of results of the classification analysis is a picture element, the data are aggregated in the unit of the zone. The land use that occupies the majority of a zone is regarded as the representative land use of the zone.

The flow that estimates economic data is shown in Figure 2. We sample several points from the area where we want to obtain economic data and combined economic data and physical data in the points, followed by the regression analysis. Hereafter, inputting all the physical data of the area to the regression equation, we estimate the economic data of the whole area.

There are three land uses of residence, commerce and industry. As the selection standard of physical data, the characteristic of the land use of each zone is adopted other than the level of physical data. In other words, we select it as the difference between the characteristics of the residence, commerce and industry of each zone, become evident.

For example, in zones where the elements of residence are major, such as population and automobile possession rate are estimated significantly. In zones where the elements of commerce are major, sales amount and the employment of service industry are estimated significantly. In zones where the elements of industries are major, outputs of industries are estimated significantly. Table 1. shows an example. The number in the table is respectively value of physical data. As for zone 1, the value of the economic data with regard to the proper land use is large, because the value of physical data 3 is larger in comparison with other 2 zones. Thus, the difference of the characteristic of residence, commerce and industry area is shown clear between zones.

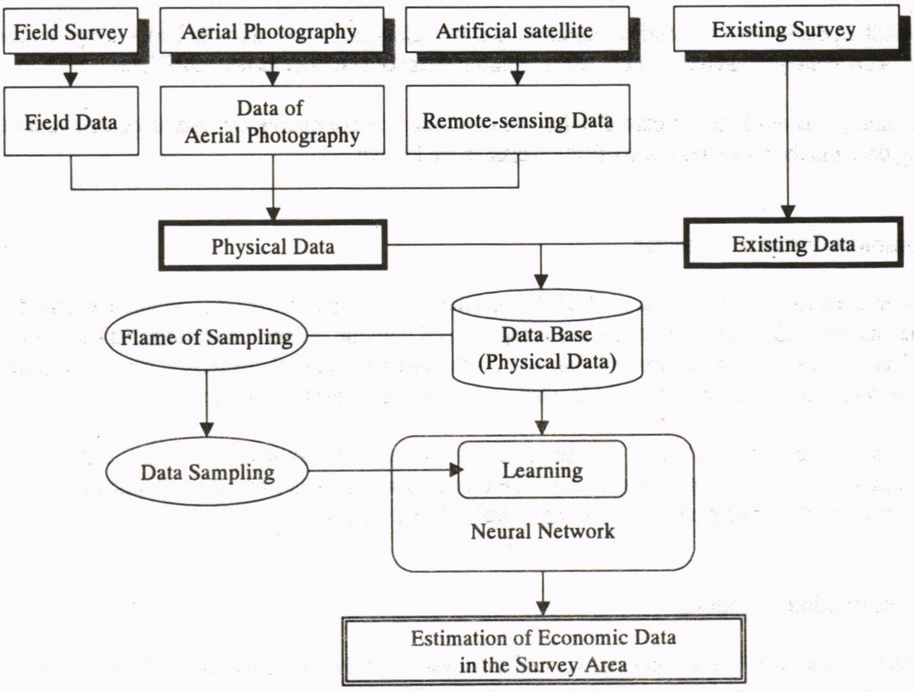

Figure 2. Flow of This Research

Table 1. Physical Data and Zone Data

	Physical 1	Physical 2	Physical 3
Zone 1	1	3	12
Zone 2	2	6	5
Zone 3	1	5	3

In the case that the survey result of different economic data is obtained to the area, we can improve accuracy of estimation by inputting the data to the database as explanation variable.

Several zones are sampled from the database composed of such physical data. The block random sampling is adopted as sampling method, since this is able to express the constitution of database well and makes the number of sample data small. We do a simple field survey economic data in the sampled zones. The learning is carried out to the neural network regarding physical data and economic data that are obtained about the zones as teaching data. Substituting the physical data for each zone to the neural network that finished the learning, we estimate economic data of all the zones.

3. Estimation

We adopt population and employment as economic data, in this research. In this chapter, at first, we make clear what physical data are the most effective for estimating these economic data. Then, we estimate the most optimal combination of physical data, in other words, we examine an excellent method to the cost performance.

The data of Sendai city in Japan is used as the inspection of the system in this paper. Although the main purpose of this system is the estimation of economic data in developing countries, this is applied to the area where is arranged all the economic data and physical data to conduct the inspection of this system. City planning foundation surveys are carried out in Sendai city and divide in 756 zones the target area with the data that is able to become such physical data candidate as population, employment and do on in each zone. We would use the data in 636 zones which are thoroughly saved as inspection data of this estimation system. The observation scene of the TM data is 107-033 on April 11th in 1990 and the band is used from 1 to 7.

Actually available physical data should be chosen when economic data are actually estimated. Because the data which price is expensive or is difficult to be obtain, is not suitable as the physical data in this paper. Considering the situation that we actually estimate economic data in developing countries where we estimate them with only the data that we can obtain from simple surveys, remote-sensing data and aerial photographs, we use only these survey data in this paper. A database of physical data for each zone is required to estimate economic data from them. The unit of physical data shall show the characteristic of the same zone. In the case that there are some stocks of such data that is useful as the physical data in developing countries fortunately, we need to pay attention to this point.

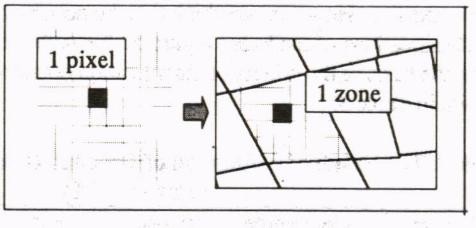

Figure 3. Pixel and Zone

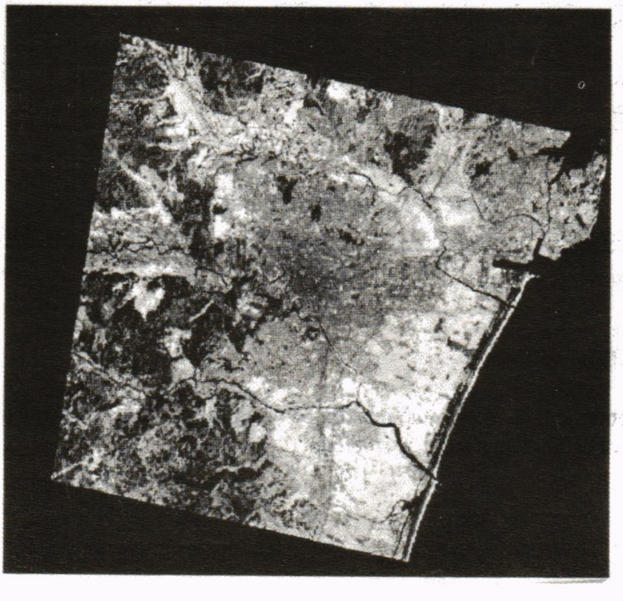

Figure 4. TM Data in Area and the Result of Discrimination

The unit of analysis is the zone of the city planning foundation survey as shown in Figure 3. The data is based on the Sendai city foundation survey, 1/25,000 map. Physical data are collected and aggregated in every zone and then economic data are estimated with the value of density to every on the basis of them as shown in Figure 4.

3.1 Classification of Land Use

Classifications of land use foe sampling are residence, commerce, industry, farm, water and green, and then the unit is a pixel of remote-sensing data. The characteristics are used remote-sensing data and are given mean value and standard value of the TM data in 5×5 pixels, 180 ㎡ next to the target pixel. Because there is a research result that the accuracy of estimation rises by considering the periphery land use. Twenty data are sampled every land use as the sampling data and then the classification method is the minimum distance classifier.

Table 2 shows the accuracy of classification of training data themselves by TM data. Although the classification between urbanized area and non-urbanized area is almost satisfactory, the one of land use in urbanized area has 5～15% of. The reason is recognized that there are many places where commerce and residence are coexisting. However, we think that the accuracy is not so low comparing to the one of conventional classifications of land use. Figure 5 shows the result of the estimation of picture elements of all area on the basis of the criteria of classification obtained. The constitution ratios of land uses of zones are shown in Table 3.

Table 2. The Result of Estimation of Training Data

		Estimation (%)					
		Residence	Commerce	Industry	Farm	Water	Green
Real Data	Residence	95	5	0	0	0	0
	Commerce	10	85	5	0	0	0
	Industry	0	5	95	0	0	0
	Farm	3	0	0	97	0	0
	Water	0	0	0	0	100	0
	Green	0	0	0	0	0	100

Hit Ratio 0.954

Table 3. The Share of Land Use by Zone

	Residence	Commerce	Industry	Farm	Water	Green
Share(%)	63.5	13.5	5.0	0.3	0.2	17.5

3.2 Selection of Physical Data

In estimating economic data of city, the relation between the difficulty of acquisition of physical data and the accuracy of estimation should be studied.

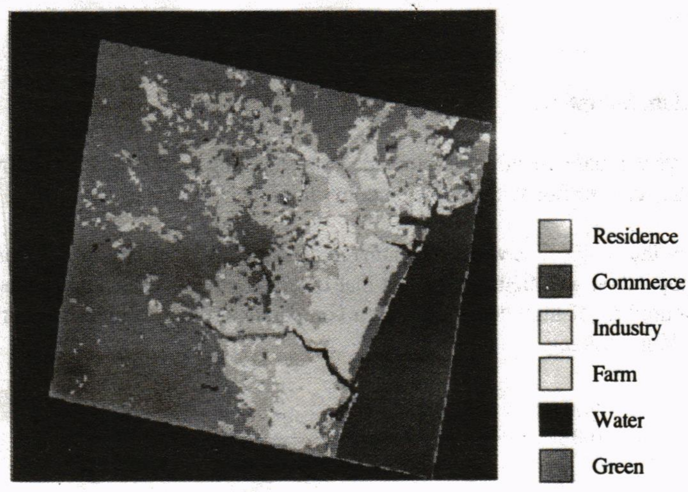

Figure 5. The Result of Classification

The physical data are ranked in terms of the difficulty of acquisition. Then we estimate economic data by using the physical data of each rank to check the improvement of the accuracy according to the level of difficulty. The data ranked L1 is the average of TM data inside the same zone. The data ranked L1' are mean and standard value of the TM data. Then the data ranked L1" is the number of pixel inside the zone treated the result of classification of land use on the basis of those data. The data ranked L2 are the data such as the distance from the center of Sendai city, the distance from the next rail station and the area of roads inside the. The data ranked L3 is the data such as the total number of buildings, the number of buildings higher than 3 floors, the number of industrial facilities, the number apartment houses as the data of building. The result is shown in Figure 6. The data of L3 over is necessary to obtain the result whose coefficient of correlation is 0.9 or more.

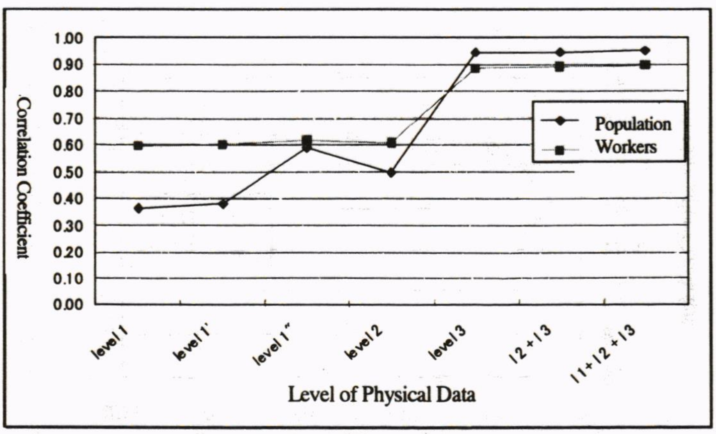

Figure 6. Correlation Coefficients according to Level of Physical Data

The result that the correlation coefficient of L1" is higher the one of L2 in estimation of population suggests the necessity of the classification of land. In the case that we use the only remote-sensing data, we can not estimate well. It is because remote-sensing data dose not have significant information about upper layers structure mainly. The definition of TM data that we used this time is 30 meters. Then there is the possibility that the high estimation value is obtained, if remote-sensing data of higher definition is used.

3.3 Block Random Sampling

In this section, we compare the effect of block random sampling on the basis of the trend of land use obtained from the classification with simple random sampling. The physical data of L3 is used.

Figure 7. shows the relation between the number of sampling data and correlation coefficient in estimation of population and figure 8. shows that of employment. Both figure show that we can estimate economic data of high accuracy with a comparatively less sample number in comparison with simple random sampling.

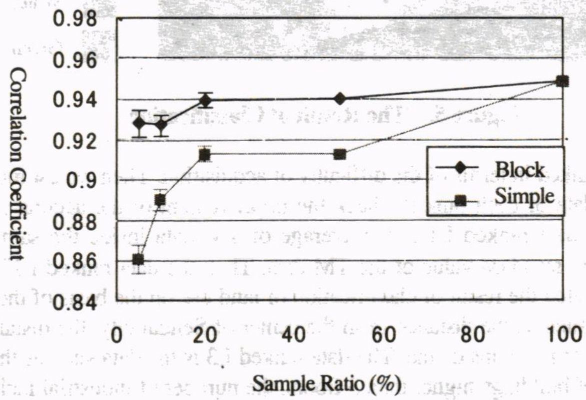

Figure 7. Sample Ration and Correlation Coefficient in Estimation of Population

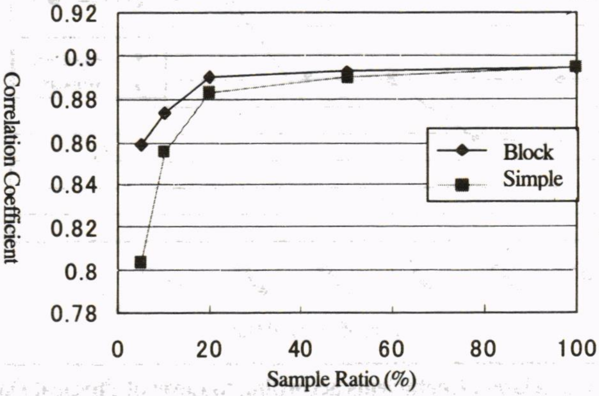

Figure 8. Sample Ration and Correlation Coefficient in Estimation of Employment

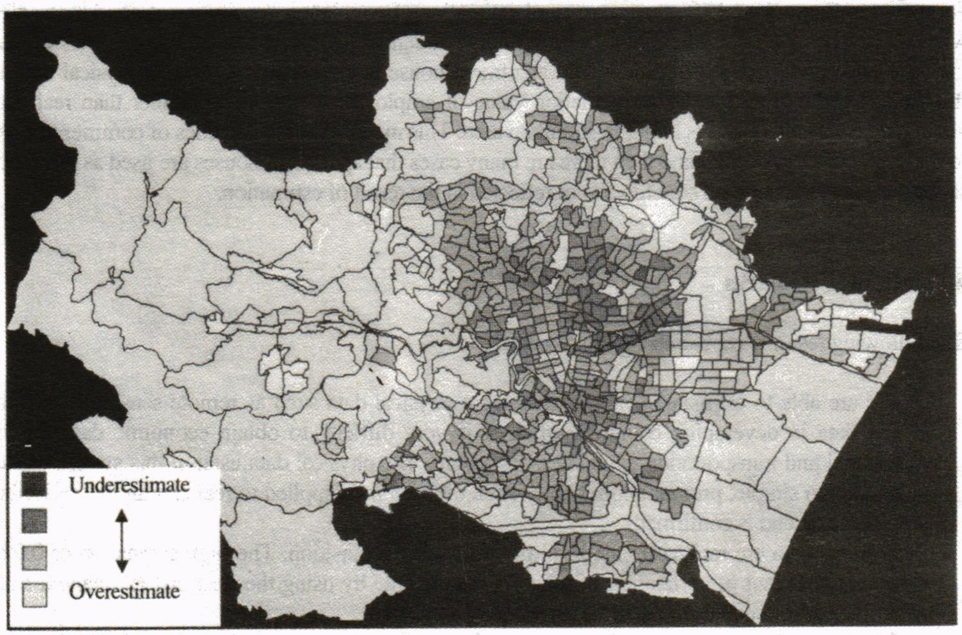

Figure 9. Estimation of Population Density

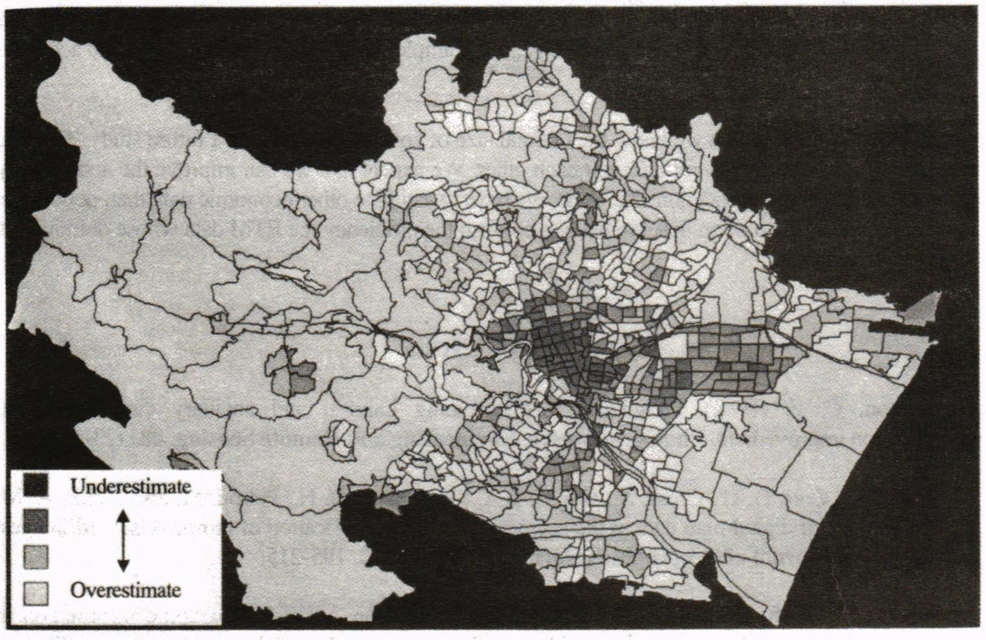

Figure 10. Estimation of Employments (Service)

As shown in figure 9. and 10., the accuracy of estimation of population density is not good in the area where there are many apartment houses with many residential people in some suburbs. It is because the difference of capacity of an individual apartment house is not considered in the physical data in this time. On the other hand, the estimation value of employment density is smaller than real data about the center of city and is larger in suburban area. This means that the densities of commerce close to the center of city are high and then there are many cases that apartment houses are used as residence in the periphery. This urban structure is not reflected in the result of estimation.

4. Concluding Remarks

The following result was obtained by this research.

(1) We are able to estimate economic data from physical data such as remote-sensing data from TM data in developing countries where it is very difficult to obtain economic data for the regional and transportation planning by a system. The physical data used in this system makes estimation simple, prompt and reliable. Then the system is applied to real city and its reliability to some extend is confirmed.

(2) Physical data are ranked by the level of easiness of acquisition. The high estimation of 0.9 or more coefficient of correlation of reliability is possible by using the data that is obtained from simple survey such as the general view of city and aerial.

(3) Block random sampling enables the accuracy of estimation higher. Then about 20% of sample rate is enough to obtain high reliable result.

We apply this system to Sendai city, since we are able to obtain both economic data and physical data for the confirmation of reliability. In case of the application this to developing countries/areas, we should consider difference of social conditions , for example, family size, types of economic activities and so forth.

To develop the method that we know the optimal size of zone is necessary as a future study. The size of zone at present study is very different each other. We expect that we can improve the accuracy by using the zones of the unification size. Also, we have to estimate other economic data than population and employment. In addition that, we have to confirm the efficiency the ETM data whose definition is higher than TM data.

REFERENCES

Hutchinson, C. F. (1982) Techniques for combining Landsat and ancillary data for digital classification improvement, **Photogrammetric Engineering and Remote Sensing, 48**, 123-130.

Mason, D. C., Corr, D. G., Cross, A., Hoggs, D. C., Lawrence, D. H., Petrou, M. and Tailor, A. M. (1988) The use of digital map data in the segmentation and classification of remotely-sensed images, **International Journal of Geographic Information Systems, 2**, 195-215.

Miyamoto, K. (1988) Estimation of Economic Data Based on Aerial-Photographic Information in Developing Countries, **International Archives of Photogrammetry and Remote Sensing, 27-B10,** 452-461.