# Estimation of Area of Paddy Field Using Vegetation and Soil Indices Derived from MODIS Data
# - A Case of Heilongjiang Province, China -

S. Uchida
Japan International Research Center for Agricultural Sciences (JIRCAS)
1-1 Ohwashi, Tsukuba, Ibaraki, Japan
uchidas@jircas.affrc.go.jp

**Abstract:** In this study, the author attempted to develop a method to estimate the area of paddy field for Heilongjiang Province of China using MODIS data. Firstly, he selected the scene of which the cloud coverage was small enough for the period of transplanting rice. Discrimination of paddy field using Ch.1 and Ch.2 only or NDVI was often unsuccessful due to the confusion with bare soil. Then, he introduced NDSI (Normalized Difference Soil Index) using Ch.1 and Ch.7 of MODIS, similarly as for NDVI using Ch.1 and Ch.2. As a result, a diagram drawn in 2-dimensional space of NDVI and NDSI showed a unique feature that almost pure pixel with bare soil and vegetation were allocated at the specific points and pure pixel of paddy was located around another corner of a triangle. In order to obtain the formula of estimating fractional coverage of paddy field in a pixel, he examined a number of distances defined in the triangle and found a capable form, which could be correlated with the constitutional ratio of paddy field in a pixel. By using this relationship, he estimated the distribution of paddy field for some selected counties in Heilongjiang Province. The results showed a pattern of distribution which was compatible with the data classified from LANDSAT-ETM+ data. This also indicated that a considerable improvement of the accuracy of estimation was attained compared with the pixel based classification of MODIS data.
**Keywords:** MODIS, Paddy Field, Fractional Estimation, NDVI, NDSI.

## 1. Introduction

Satellite remote sensing has been expected as an effective tool to monitor the changes of land cover or land use for wide area. And often it is disputed about the trade-off relationship of spatial and temporal resolution of satellite remote sensing data. For the purpose of producing land cover/use map, which could characterize the shape of lot of agricultural land for example, a high spatial resolution data, say about 1 meter, would be required. However, the swath width of high spatial resolution data is generally too narrow to cover the wide area within a limited period of time and it would not be adaptable to produce the land cover/use map in regional scale. On the other hand, a series of high temporal resolution satellite data would have a specification of wide swath width, but its spatial resolution would not be suitable to characterize detail structure of land cover/use.

Although there are some constraints of producing land use information by using high temporal resolution satellite data, acreage estimation of the specific land use would be the target of its feasible application. Pixel based classification often resulted in mapping of significantly low accuracy due to the mixture of various types of land cover in a pixel. This issue about mixel was one of the major obstacles for the development of methods to discriminate land use using coarse spatial resolution satellite remote sensing data. Then, researchers have examined the techniques of sub-pixel classification or fractional estimation for the specific land use. The author also proposed a method of sub-pixel classification mainly for the purpose of monitoring winter wheat sown area in the major grain crop production area of China using multi-temporal NDVI obtained from NOAA-AVHRR data [1]. This method and additionally conventional method using 2-temporal NDVI data could characterize the temporal changes of winter wheat sown area for the past several years by county or arbitrary defined area [2]. For winter wheat, its temporal profile of NDVI in the spring season would show considerably distinctive pattern from other land use types and this point leaded to bring a successful result for the case.

The method mentioned above might not be optimally adaptive to the case of summer crops including rice because of the existence of multiple land use types which showed similar temporal profile of NDVI during the crop growing period. An attempt applying linear un-mixing technique using channels 1, 2 and NDVI of MODIS showed successful estimation of fractional coverage of paddy fields over East Asia [3]. This study also suggested that the upland fields tended to be falsely discriminated as paddy fields when the source of data was limited to channels 1, 2 and NDVI. In some regions such as Heilongjiang treated in this paper, the allocation pattern of paddy fields and upland fields is complexly mixed in a small area or the land has not been consolidated in the large scale. Therefore, in order to estimate the area of paddy field, utilization of mid-infrared channel data as well as visible and near-infrared channel data of MODIS was examined in this paper. The study site was selected from the representative rice production area under cool climate condition in the northeastern part of China, where cropping season is limited to one time in the summer for both rice and upland crops.

## 2. Materials and Methods

The study site, Heilongjiang Province, is located in the northeastern part of China and is one of the major grain crop production areas of China. Climate condition is too severely cold to cultivate crops in winter season but is normally warm enough to cultivate rice as well as soybean and maize in summer season, although cold damage on rice production would be happened in some years. The reason of cultivation of rice under rather marginal environmental condition is that the quality of rice is high and valued as the commercial product in this region. Then, the land has been exploited for extending paddy fields even in recent years.

MODIS data is provided by the SIDaB (Satellite Image Data Base) operated by Agriculture, Forestry and Fisheries Research Information Center [4]. Another satellite data used in this study is LANDSAT-ETM+, of which the Path is 117 and Row is 27 to 29 acquired on 31 May 2003. Almost all paddy fields were inundated by water during the period from late May to early June when the cultivation stage was transplanting. LANDSAT data taken in this period would be suitable to discriminate paddy field from other land cover/use types, so that the classification results from LANDSAT data were assumed as the reference information of actual distribution of paddy fields.

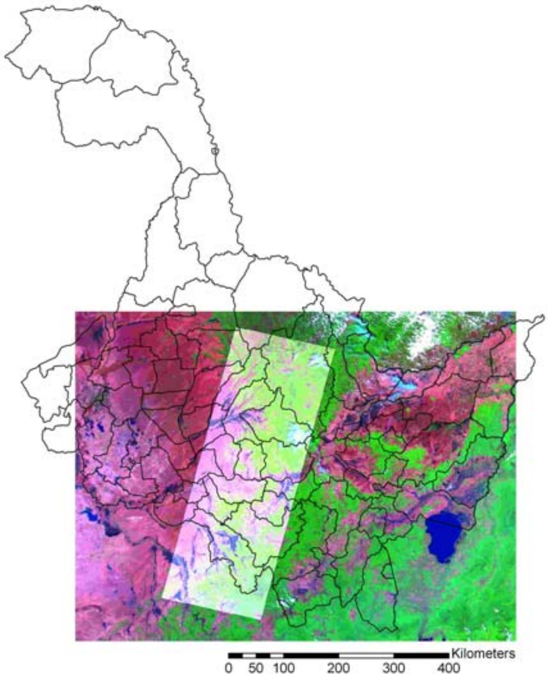



**Fig.1 Extracted coverage of MODIS data and LANDSAT data (brighter part) shown by the scenes on 31 May 2003 overlaid by the boundary of County in Heilongjiang Province.**

Firstly, the author examined the condition of cloud cover over the MODIS data during the period of transplanting rice and selected the cloudless scenes. Then he extracted the area of 800 km west to east and 600 km north to south, which covered the most of major rice production area of Heilongjiang Province, with the pixel size of 250 meters. Fig.1 shows the coverage of extracted MODIS data overlaid by the range of LANDSAT scenes with brighter tone. Both of MODIS and LANDSAT data represented in the figure were acquired on the same date, 31 May 2003. Line feature in the figure indicates the boundary of County in the Province.

The next step was to produce reference information of distribution of paddy field by the process of supervised classification to LANDSAT data. The classified paddy field data was modified to the value of probability density in the window with the size of 9 pixels by 9 pixels. This value would be used for the examination of relationship between fractional percentages of paddy field in a pixel of MODIS with indices derived from MODIS data. In this study, indices of NDVI (Normalized Difference Vegetation Index) and NDSI (Normalized Difference Soil Index) were considered as probable factor to estimate the area of paddy field, where NDVI was obtained by the formula as (Ch.2-Ch1)/(Ch.2+Ch.1) and NDSI as (Ch.7-Ch.1)/(Ch.7+Ch.1). Here, Ch.1 to Ch.7 denotes the values converted to the reflectance at the upper atmospheric layer. Thereafter, the author tried to find out a systematic pattern of allocation of pixel with certain fractional percentages of paddy field in the 2-dimensional space of NDVI and NDSI, and then to formulate the relationship.

## 3. Results

### 1) Condition of Cloud Coverage

The condition of cloud coverage could be strongly influenced by the climate of the site. In this case study site, the atmosphere would be stable around transplanting and harvesting period of rice. Then, it is expected to obtain almost cloudless scene, say less than 5 %, covering the area as shown in Fig.1 during these periods. The author identified optimal scenes in 2003 as follows: 17 May, 27 May, 31 May, 2 June, 20 September, 22 September and 29 September. Similarly, a number of cloud free data could be available during the transplanting period in other years, for example, 26 May, 2 June, 5 June and 7 June in 2004. According to the agricultural survey carried out at the representative rice production areas in the Province, the date of transplanting would be concentrated in about 2 weeks. This suggests that there is rather high probability of acquiring cloudless conditioned data with covering the whole parts of the major rice production area in Heilongjiang Province during the period of transplanting rice. This also suggests that if it is successful to develop the estimation method of area of paddy field, which is inundated with water, by using MODIS data, the information of provincial scale would be produced by combination of the estimated results using small number of MODIS scenes.

## 2) Relation between Paddy Area in a Pixel and Indices Derived from MODIS Data

Fig.2 shows the land cover classification map obtained from LANDSAT-ETM+ data. This figure also indicates the locations named Qingan, Fangzheng and Wuchang, which are representative rice production areas and employed as the site to compare the estimated results by using MODIS data in the following parts. In the figure, some isolated clouds were presented in the northeastern side and the parts of southeastern fringe were falsely classified due to the influence of haze. But the most of the covering area of image could be optimally classified and the results would be applicable to setting as reference information of actual distribution of paddy fields.

In this case, the author analyzed the data of Qingan to develop a model to estimate the area of paddy fields. Fig.3 shows the scattergram of NDVI in the horizontal axis, and NDSI in the vertical. The pure pixels of densely vegetated area were concentrated around at 0.73 of NDVI and 0.05 of NDSI. Another concentrated cluster of distribution could be found around at 0.25 of NDVI and 0.42 of NDSI, which was corresponded to the pixels of bare soil areas. The author examined the location of point in the scattergram in consideration with the probability density of paddy in a pixel, then he could recognize that a pixel with the higher probability density of paddy tended to be allocated at the more approaching place to the lower left corner of the triangle formed by the points.

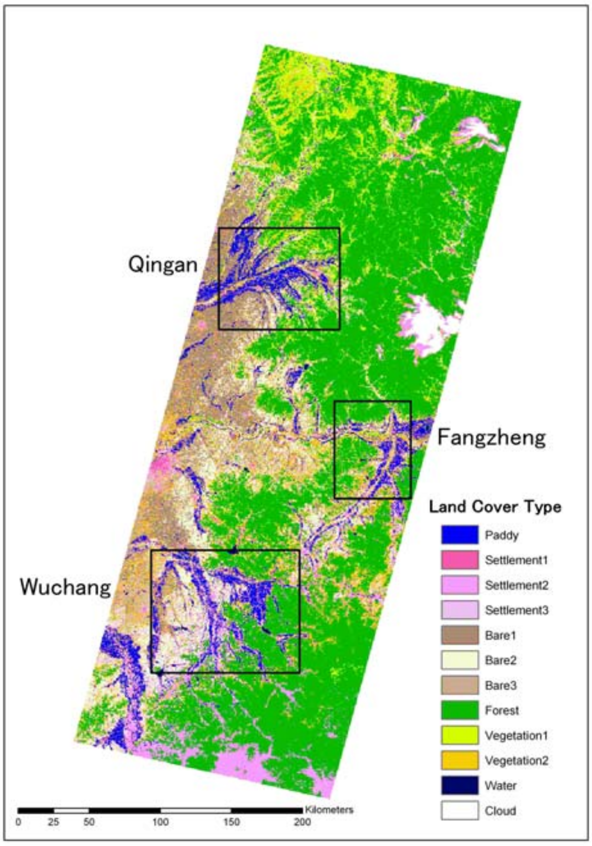



**Fig.2 Land cover classification map obtained from LANDSAT-ETM+ data.**

In order to construct the formula of estimating area of paddy field, several types of parameters defined in the scattegram were investigated. Fig.4 schematically represents the expression of 2 types of parameters. One was called as Normalized Distance (ND) in this paper, which was defined by the following procedures. First, he drew 2 set of enveloping lines outside the point distributed parts along non-vegetation end and non-bare soil end, and cross point of both lines was set as P. Next, he identified the center of pure pixels of density vegetated area as V and the center for bare soils as S. The conversion ratio to Y axis could be obtained so as to have the same distance (D') from P to S with P to V. By using the same conversion ratio, a distance from P to arbitrary point in the figure (A) would be calculated as d'. Thus, the form of ND=d'/D' could be defined. Another parameter was called as Paddy Index (PI) in this paper. This index was obtained by using distances from the enveloping line at the non-vegetation end to V (V') and from the same line to the point A (v'). The form was defined as PI=(1-v'/V')x(1-ND).

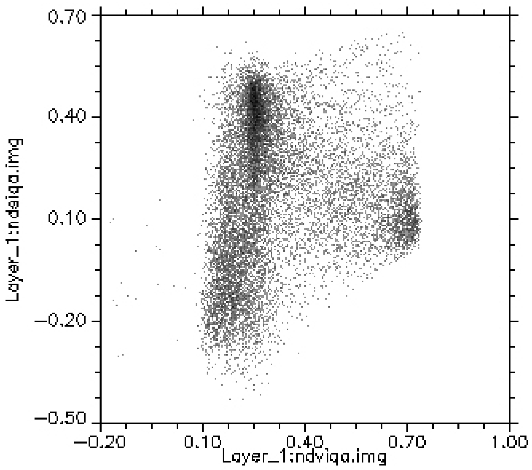



**Fig.3 Scattergram of NDVI and NDSI for Qingan area calculated from MODIS data.**

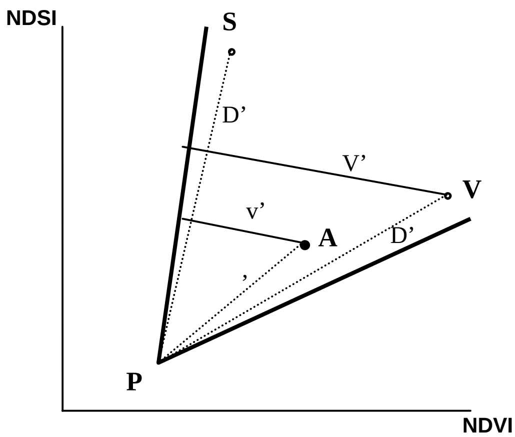



**Fig.4 Schematic representation of obtaining parameters used in the estimation models.**

The author modified the referenced data of distribution of paddy fields derived from LANDSAT-ETM+ to level sliced data with the interval of 10 % of paddy field area in a window with the size of 9 pixels by 9 pixels. Then he examined the relationship between the mean value of sliced data and ND or PI defined above. Fig.5 shows the relationship for the case of ND. From this figure we could recognize an evidentially correlated feature between ND and the fractional percentage of paddy field in a pixel. Accordingly, he employed the formula shown in the figure to estimate the area of paddy field for wide area as the first attempt. The results would be generally acceptable, however, the portion of mixing with paddy and vegetation tended to be overestimated as explained in the next section. Therefore, the second version of formula was constructed by using PI and its relationship is shown in Fig.6. This figure also exhibits the correlation between the parameters and the coefficient of determination of fitting the formula would show sufficiently high value.

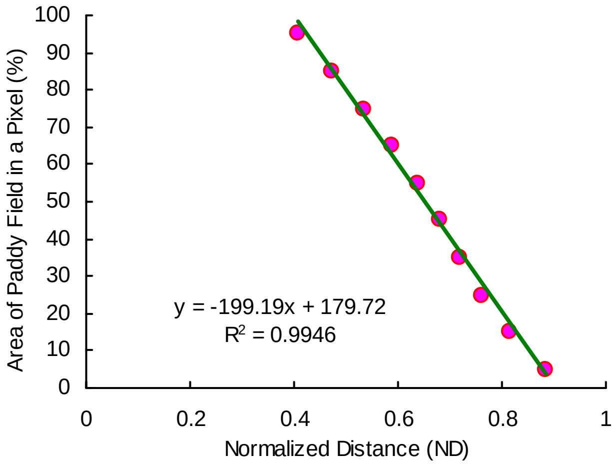



**Fig.5 Relation between Normalized Distance (ND) and area of paddy field in a pixel.**

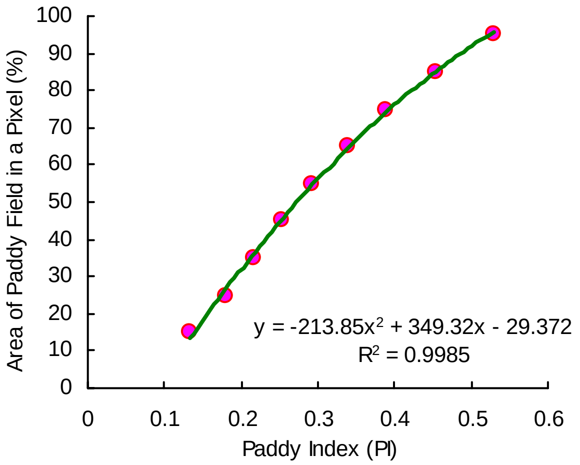
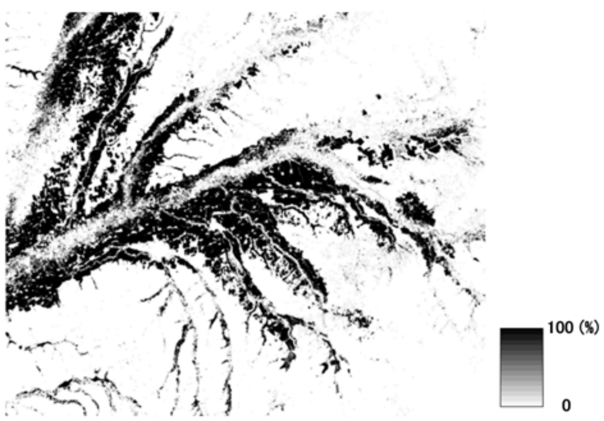



**Fig.6 Relation between Paddy Index (PI) and area of paddy field in a pixel.**

### 3) Estimation of Spatial Distribution of Paddy Field

The author estimated the distribution of paddy field for Qingan County and surroundings by applying the formula shown in Fig.5 and Fig.6. Fig.7 shows the estimated results of area of paddy field. This figure also compares their features of appearance by different methods; (a) classification of LANDSAT-ETM+ data, (b) pixel based classification of MODIS data, (c) fractional estimation using ND, and (d) fractional estimation using PI. The total area of paddy field in this area was 116,960 ha based on the results of classification of LANDSAT data, whereas the area by pixel based classification was about 40 % underestimated. On the other hand, the estimated value using ND resulted in 7.75 % more than the reference amount and the value using PI resulted in 1.98 % less than the reference value. The pattern of distribution of estimation represented in (c) and (d) generally matched to that of (a). In the case of (c), apparently estimated values could be presented in the eastern side of area, especially along the valleys and this error could be more or less mitigated in the case of (d).

(a)




(b)


Fig. 7 continued.

(c) (d)

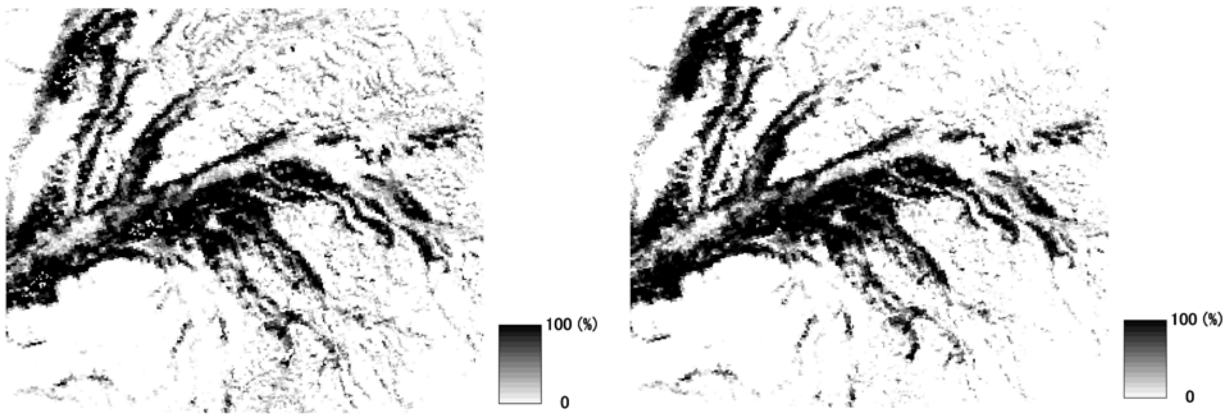



**Fig.7 Comparison of distribution of paddy area; (a) classification of LANDSAT-ETM+, (b) pixel based classification of MODIS, (c) fractional estimation of MODIS using ND, (d) fractional estimation of MODIS using PI.**

Although the general pattern of distribution of paddy field was sufficiently reproduced compared with the reference data, the estimated value might not be accurate in some portions. Then, the author examined the characteristics of locations, where the values were considerably overestimated or underestimated. A typical overestimated area was the middle of extensively developed paddy fields. Actually this kind of area consisted of not only paddy fields but also facilities or infrastructures such as roads, canals, constructions. However, the spectral reflectance of these materials contributed to the pixel value of MODIS could be rather complex and might not be optimally distinguished from paddy fields. On the other hand, a typical underestimated part was placed in the area of existing small scaled paddy fields surrounded by bare land. This was supposed to be caused by the specification of MODIS that the spatial resolution of Ch.7 was 500 meters and NDSI with the pixel size of 250 meters could not be appropriately represented the condition of mixture of bare soil and paddy field.

The parameters of formula shown in Fig.6 were obtained in the case of Qingan area and it should be testified the same parameters would be applicable to the case of other areas. Fig.8 shows the case applied the method of fractional estimation of MODIS using PI to Fangzheng area and Fig.9 to Wuchang area, both of which are placed the classification data by LANDSAT-ETM+ in the left hand side. As discussed above, the parts of extensively developed paddy fields tended to be overestimated as to fill the small blanks. Additionally, we could be notified that there was a tendency of overestimation at the mountainous areas especially in the case of Wuchang. This was supposed to be caused by the effects of haze over the area. If the haze is presented, the value of Ch.1 would be more strongly affected and increased than the value of Ch.7, therefore, NDSI would apparently decrease. This change would bring to increase the estimated value of area of paddy field in a pixel followed by the formula.

The method introduced here is applicable to the wide spatial range observed by MODIS. Fig.10 shows the example of application to the area of 800 km west to east and 600 km north to south. Cloud covered area typically located in the northeast part revealed to be estimated falsely. However, there is rather high possibility to obtain cloudless image during the period of transplanting rice and this error would be compensated by combining the multiple MODIS data taken in this period.

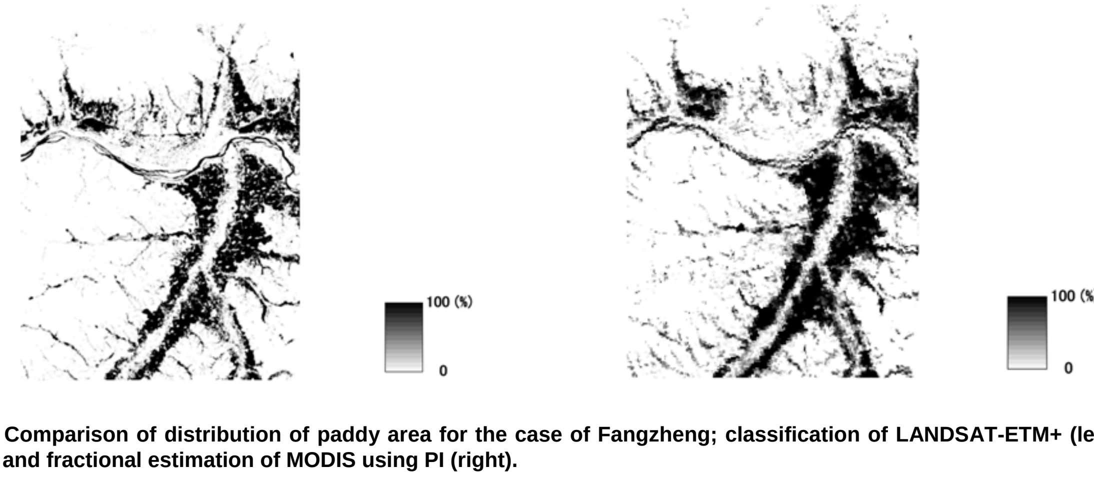



**Fig.8 Comparison of distribution of paddy area for the case of Fangzheng; classification of LANDSAT-ETM+ (left) and fractional estimation of MODIS using PI (right).**

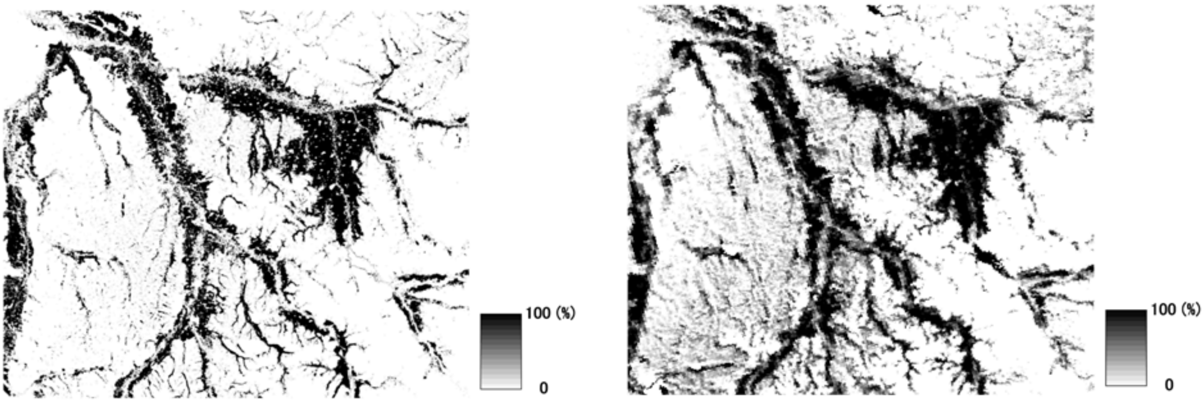



**Fig.9 Comparison of distribution of paddy area for the case of Wuchang; classification of LANDSAT-ETM+ (left) and fractional estimation of MODIS using PI (right).**

## 4. Conclusions

MODIS data which equipped moderate spatial resolution sensor data was expected to be applied to estimate the area of paddy field in Heilongjiang Province located in the temperate-cold climate zone with competent accuracy, if an appropriate method was developed using the combination of visible, near-infrared and mid-infrared spectral data. In this article, the author demonstrated an attempt to introduce the formula of estimation of fractional percentage of paddy field in a pixel by using NDVI and NDSI derived from MODIS data. Although the results were not completely satisfied for various cases of mixing features of land cover types, the spatial pattern of distribution in wide area could be reproduced in general and the accuracy of estimation of area in regional scale was improved considerably to the case of pixel based classification. The method developed in this study would be applicable to the other sites, where the period of transplanting rice was limited in the short period. Similarly, MODIS data would probably be applied to estimate other major agricultural land use by constructing an appropriate method. The final target would be to provide the quantitatively reliable information of land use in regional scale for the purpose of developing the scheme of proper management of land resources.

**Fig.10 Estimation of distribution of paddy field for the major part of Heilongjiang Province in 2003; darker part shows higher fractional percentage of paddy field in a pixel.**

## Acknowledgement

This study has been implemented as a part of China-Japan collaborative research project entitled “Stable food supply systems for mitigating the fluctuations of production and markets in China”. The author thanks Dr.Tang Huajun, Director of Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, and Dr.Jiao Jiang, Director of Institute of Cultivation, Heilongjinag Academy of Agricultural Sciences, for their contribution to the programs.

## References

[1] Uchida, S., 2001. Sub-pixel Classification of Land Use Using Temporal Profile of NDVI, *J. the Japanese Society of Photogrammetry and Remote Sensing*, 40(1): 43-54.

[2] Uchida, S, 2005. Development of A Method to Analyze Agro-environmental Changes in the Typical Food Crop Production Area Using Geographic Information System (GIS), *JIRCAS Working Report*, 42:19-31.

[3] Takeuchi, W. and Y. Yasuoka, 2004. Mapping of Fractional Coverage of Paddy Fields over East Asia Using MODIS Data, *J. the Japanese Society of Photogrammetry and Remote Sensing*, 43(6): 20-33. (in Japanese)

[4] **URL:** SIDaB (Satellite Image Data Base). Agriculture, Forestry and Fisheries Research Information Center. Available at: http://rms1.agsearch.agropedia.affrc.go.jp/sidab/index-en.html.