(Abstract)

Agricultural Land Information System (ALIS) to support area sample survey

Shoji KIMURA
ASEAN Food Security Information System, email: wood_v@yahoo.co.jp

ABSTRACT: There was an issue on fundamental differences of reliability on statistical data among ASEAN member countries. It obviously appears that the implementation of statistical sample survey to used proper framework to produce accuracy data. However, chronic lack of labor and budget on statistic task with shortage of staff experience would be considered as the main factors of blocking the conduction of sample survey. Therefore, AFSIS Expert has considered a tool by using personal computer (PC) to make an area framework, estimate agricultural land area, and developed the statistical system called ALIS. As case studies in this connection, Lao PDR and Kingdom of Cambodia were selected countries in which agricultural statistical survey technique are developing which in turn enable to get area data with a small number of labor, low budget based, and sustainable.

1. ASEAN Food Security Information System (AFSIS)

At the 1st Meeting of ASEAN Ministers on Agriculture and Forestry Plus Three (AMAF + 3) held on 5 October 2001 in Medan, Indonesia, the Ministers approved the ASEAN Food Security Information System Project, which was commenced in 2003. Since its commencement, the ASEAN Food Security Information System Project (hereafter referred to AFSIS) has been implementing by ASEAN and Japan with participation open to China and Republic Korea to compile and synergize agricultural statistics and information in the region.

The main activities of AFSIS Project are concerned with the construction of information network system and human resource development. AFSIS Project is currently providing food security related information gathered from member countries to public through its own created database and website.

2. Improvement of accuracy on statistics in country level

There was an issue on fundamental differences of reliability on statistical data among member countries. It was necessary to meet the requirement of data users who want to grasp regional situation.
In this regard, AFSIS Project has tackled the above issues with member countries from 2011 through an improvement of the capacity of the countries in which encountered the issues on accuracy of statistical data for further development of the utilization of database in collaboration with the Expert of ASEAN Food Security Information System (hereafter referred to "AFSIS Expert") whom the project task is assigned.

3. Consideration for improvement of data accuracy

Although it obviously appears that there are several methods for improvement of data accuracy, for instances the holding of seminar and training or improvement of survey manual, survey definition, and etc. However AFSIS Expert has targeted the improvement of core statistics in particular area of target countries. It was expected to have an impact for improvement of data accuracy on whole productive statistics by improvement of area data accuracy and enable the capacity building of sample survey and statistical theory for staffs of target countries by conducting of area sample survey.
In addition, it was necessary to consider the main factors of blocking the conduction of sample survey are the chronic lack of labor and budget on statistics work with shortage of staff experience.

Hence, AFSIS Expert has considered a tool by using personal computer (PC) to estimate agricultural land area and developed the statistical system that enable to conduct area sampling with a small number of labor and low budget based on aforementioned problem.

4. Logic for area estimation using PC

In principle, we can estimate the statistical data by using the following estimation formula.

$$Y = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} y$$

The sign notation used to describe this formula appears below (in case of area estimation);

Y = Estimated area of this year
y = Total area of last year
x_i = Area in "i" number sample of this year
y_i = Area in "i" number sample of last year
n = Number of survey sample

If we can obtain the data that can match above all signs concept and area framework of "n", we could possibly estimate "Y" easily by using this formula. However, in reality, some countries have undeveloped area survey cannot obtain any data out of doubtful "y".

On the other hand, we can seek the possibility of replacing the sign concept by using PC for "Y". If we can make the area framework (N), select the indicated sample (n) randomly and calculate the area of sample (y_i) on the PC, we can estimate a tentative agricultural land area (y) as a replacement for area of last year. In fact, we can approach to "Y" by using the below formulas.

$$Y = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} y \quad \text{at that time} \quad y = \frac{\sum_{i=1}^{n} y_i}{n} N$$

The sign notation used to describe this formula appears below (in case of above approach);

Y = Estimated area of this year
x_i = Area in "i" number sample of this year
y = Tentative Agricultural land area made on the PC
N = Number of framework population made on the PC
y_i = Area in "i" number sample made on the PC
n = Number of sample selected by PC

Remarks: only remains x_i as unknown data in this formula.

5. Development of Agricultural Land Information System (ALIS)

ALIS stands for Agricultural Land Information System. ALIS has been developed to accomplish the above replacement, in particular to create the area framework, to select the indicated sample randomly, and to calculate the area of sample on the PC.
In addition, ALIS is designed as a total supporting system for area sample survey by attaching the supporting function of field survey, database function and etc.

The outline of ALIS functions are as follows, however, the detail of functions and operations are shown by ***ALIS Program Operation Manual***.

5.1 The function for the development of agricultural land mesh framework

ALIS can make area mesh framework out of picture-map, for this purpose, ALIS has the following functions;

5.1.1 To scan picture-map;
ALIS scans a target province picture-map out of Google Map.
The province border line on ALIS is drawn based on Google information which may have an accuracy problem. Therefore ALIS has the map border registration function to modify it correctly in case of false border line existing.

5.1.2 To make area mesh;

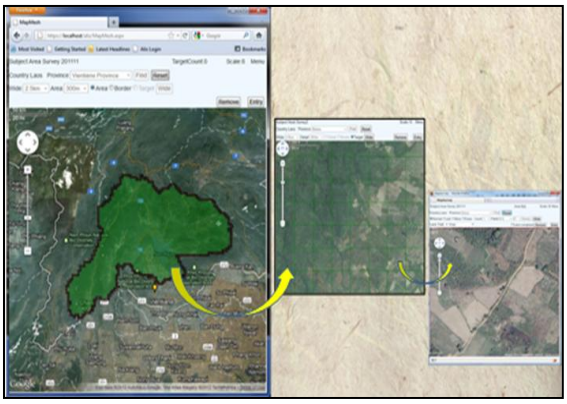

Figure 1: ALIS function for make area mesh

Figure 1 shows the actual operation screen of ALIS of Vientiane province in Lao PDR. ALIS makes an area mesh framework by indicated provinces, it means that ALIS estimates the area by province. ALIS has a function to make 4 kinds of mesh, 5km mesh and 2.5km mesh as wide mesh and 600m mesh and 300m mesh as area mesh. Operator can make area mesh by selecting target province and mesh size and hold down "Entry" button on the screen. A total of 121,152 area meshes of 300m are registered in ALIS by this operation.

5.1.3 To add a mesh location information (latitude, longitude) and a mesh ID to each mesh;

To add the ID to each mesh becomes an important operation in order to change from just a mesh aggregate to a statistical framework. ALIS adds mesh ID to each wide mesh and area mesh from top left to bottom right in series for example like 415-27 by province. These meshes ID are displayed on the

monitor during operation and capable to search a mesh by input mesh ID.

5.1.4 To create the sample framework

ALIS program does not assume all area mesh which made on Google Map as the sample framework but assumes area meshes including cultivated land of those as the sample framework. Because of the sample framework should be treated as a class of "measurements" not a factor to measure. For this purpose, the operator implements the visible judgment on the Google Map screen, divides into "area meshes including cultivated land" and "area meshes not including cultivated land" and registers "area meshes including cultivated land" to system as the sample framework.

On the other hand, the unified criteria for judgment area with visible judgment are an essential rule for making of sample framework. The rules of making of sample framework in ALIS are shown by a study report the ***Consideration for making of area framework in ALIS*** written by AFSIS Expert.

5.1.5 To build the mesh database

ALIS saves the position information (latitude and longitude) of corners of area mesh, agricultural land area data and crop planted area data by sampling based on mesh ID to the database. And ALIS makes mesh list sheet, agricultural land area data sheet and crop planted area sheet by using CSV file.
In addition ALIS saves the borderline information of agricultural land area on each area mesh. This means ALIS capable to reflect the borderline information to updated Google Map pictures.

5.2 The function for area calculation

ALIS can calculate the agricultural land area by tracing its border line on the PC monitor. For this purpose, ALIS has some following functions;

5.2.1 To classify agricultural land and planted crops land (rice, soybean, maize, sugarcane, cassava) by land indications

5.2.2 To calculate area by the land indication;

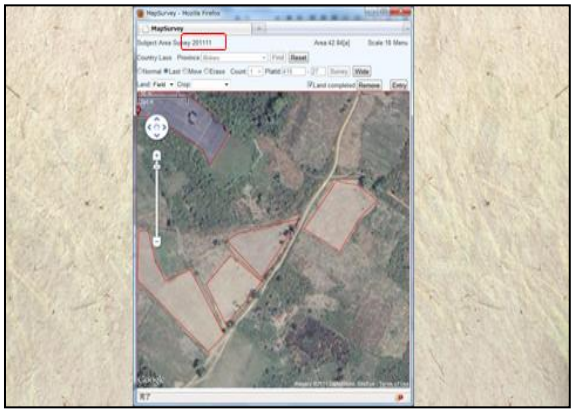

Figure 2: ALIS function for area calculation

We need to attach the mother population information to the mesh for making a statistical framework. In this case, we need to attach the area information to mesh. Figure 2 shows ALIS calculates the agricultural land area by tracing its borderline on the PC monitor. The calculated area is displayed on the monitor with registration to database.

5.3 The function for survey support

ALIS can support the field area survey based on sample method. For this purpose, ALIS has some following functions;

5.3.1 To conduct mesh sampling randomly;

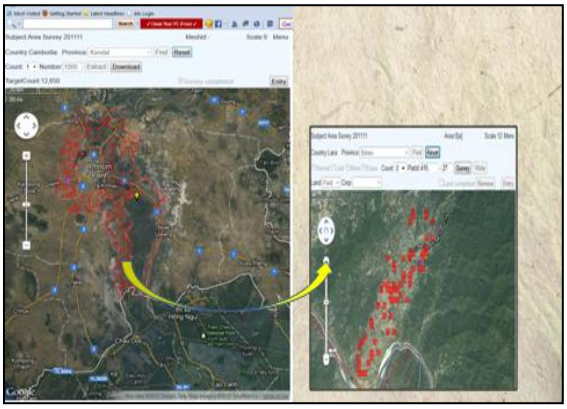

Figure 3: ALIS function for mesh sampling

Figure 3 shows the actual operation screen of ALIS in Kandal province in Cambodia. ALIS can extract the indicated number of sample mesh randomly for the purpose of sample survey.

5.3.2 To make wide map for guidance and survey sheet for researcher;

ALIS makes wide map and survey sheet of indicated sample mesh, or in other word, for making "xi" remained as unknown data on estimation formula

5.3.3 To estimate area using estimation formula and to compile result sheet.

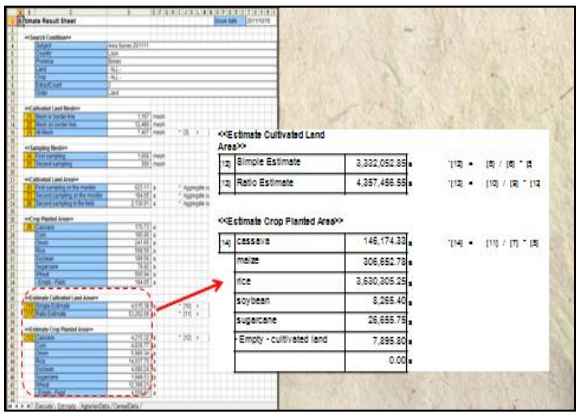

Figure 4: ALIS result sheet

Finally, ALIS makes a result sheet by using CSV file. In addition, ALIS saves all data areas by each mesh ID to database.

6. Master Sample Method

ALIS adopts the "master sample method" to reduce the operation work. The master sample method is one of the statistical sample design method which is considered to reduce the labor of frame maintenance. Under this method, the master sample meshes are extracted from original mesh file (first sampling), gives framework information (calculated area of agricultural land by tracing its border line) to these extracted master sample meshes, and then extracts survey sample meshes (second sampling) from this master sample mesh file.

The detail of Master Sample method in ALIS is shown by a study report the ***Consideration of statistical accuracy on Master Sample method*** written by AFSIS Expert

7. ALIS operation flow

ALIS operation flow is determined based on Master Sample Method. Concretely, it becomes following operation flow, Confirmation of target province border line – Making wide mesh and area mesh (N) – Registration of area mesh which include agricultural land (N') – First sampling (n) – Addition of area information to "n" – Second sampling (n') – Print out wide map and survey sheet = Field survey = Input the result of field survey to "n" – Estimation of agricultural land area and crops planted area.

ALIS makes four files of mesh i.e. original mesh file (N), framework mesh file (N') master file of mesh (n) and sample file of mesh (n') by taking master sample method. The figure 4 shows the outline of ALIS operation flow. We can see that every task in estimation of area data has completed by ALIS excepting field survey.

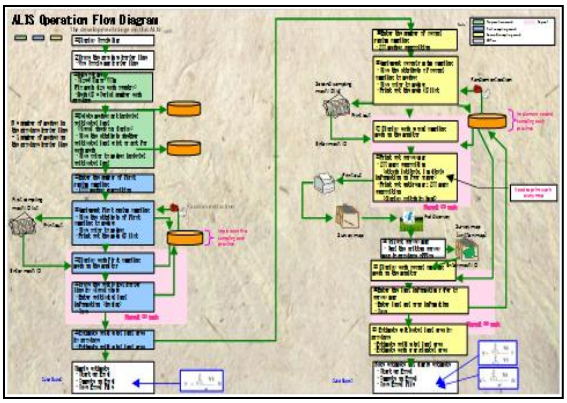

Figure 5: ALIS operation flow

8. Estimation of area data

ALIS can estimate the agricultural land area by each province and the planted area of rice, maize, soybean, sugarcane, cassava and other set two crops by countries. ALIS itself does not have the area estimation function. Instead, the officials do the estimation by using CSV files that have been downloading from ALIS. Agricultural land area is calculated based on comparison ratio estimation and the planted area by each crops are estimated by a simple estimation. Estimation result is printed out as an estimation sheet.

Incidentally, Area Meshes in Wide Meshes including province border line become out of sampling target on initial setting of ALIS. This is a result of consideration about the survey condition of surrounding national border line. However, these area meshes become the target of area estimation by rational estimation from the cultivated conditions of Area Meshes which is not including province border line. Moreover, ALIS has a specific function which it changes area meshes in wide meshes including province border line to the sampling target mesh for small area province.

The agricultural land area data are calculated by ALIS with following formula.

$$Y=\frac{\sum_{n'=1}^{n'} x_i}{\sum_{n'=1}^{n'} y_i} y \qquad \text{at this time} \qquad y=\frac{\sum_{n=1}^{n} y_i}{n} N$$

The sign notation used to describe this formula based on ALIS operation appears below;

Y = Estimated agricultural land area of the province
y = Estimated agricultural land area by measurement of Google map
N = Total number of meshes including agricultural land
n = Number of first sampling meshed
n' = Number of second sampling meshes
yi = Agricultural land area on i-th number mesh of first sample by measurement of Google map
xi = Agricultural land area on i-th number mesh of second sample by field survey

The crop planted area data are calculated by ALIS with following formula

$$Y'=\frac{\sum_{i=1}^{n'} x_i}{n'} N$$

The sign notation used to describe this formula based on ALIS operation appears below;

Y'= Estimated crops planted area of this province
N = Number of meshes including agricultural land
n' = Number of second sampling meshes
xi = Crop planted area on i-th number mesh of second sample by field survey

9. Conclusion of ALIS function

The feature of ALIS is to provide the tool of area sample survey method as a package. The person in charge for area survey can make the framework, design the survey scale according to target accuracy, extract sample, make the survey sheet and estimate the statistical area data by using ALIS. On the other hand, ALIS is a centralized management system. The central office can conduct overall management of sample survey from survey design to estimate of statistical area data while local office has remained responsibility to only conduct the field survey.

ALIS is particularly useful for the countries where neither area survey nor area framework was developed.

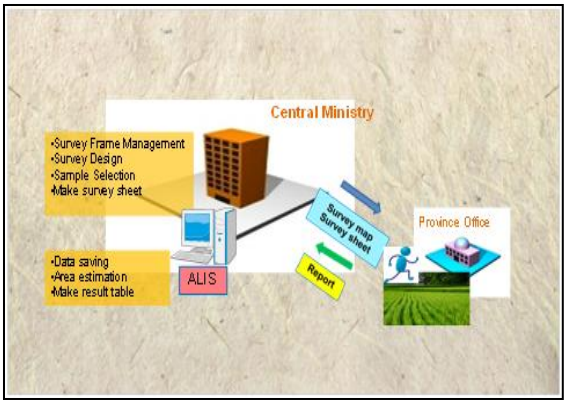

Figure 6: ALIS Centralized Management System

10. Implementation of feasibility study

The feasibility study of ALIS was implemented in 2011 in order to confirm ALIS operation and verify the estimated data by ALIS. Lao PRD and Kingdom of Cambodia, in which target countries are requested to designate a consultant for implementation of the feasibility study. The consultant was tasked to support the activity of AFSIS expert and it was fully responsible for field survey, researcher meeting and ALIS operation in their countries. Figure 7 shows the operation schedule of feasibility study in Lao PDR. It has been 4 months for operation in order to get the estimated area data. Surprisingly, these operations have completed by almost one staff excluding field survey.

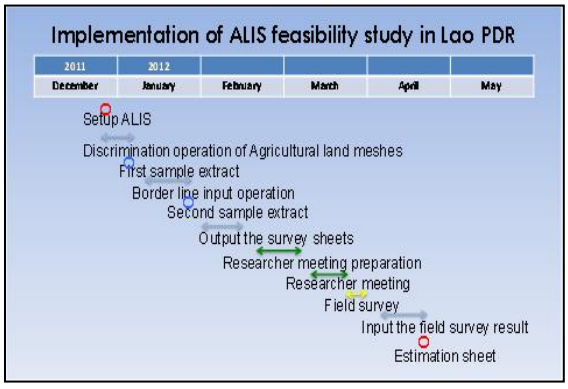

Figure 7: ALIS operation schedule in Lao PDR

11. Field survey

The field survey is the most important process in this feasibility study and it is required the correct survey to ensure the data accuracy. The ***Survey operating manual for national consultant*** gives the specific consideration for field survey.

The field survey of feasibility study has been implementing in Vientiane province, Lao PDR and

Kandal province in Kingdom of Cambodia. Fortunately, we could implement the feasibility study on two provinces that have different land location. In fact, we can see that Vientiane province is located in mountain area. On the other hand Kandal province is located in plain field area if we compare with each "agricultural land mesh rate"

Fundamental indicator	**Vientiane province**	unit	**Kandal province**	unit
Number of total mesh (300m mesh)	**121 152**	mesh	**35 456**	mesh
Number of mesh including agricultural land (Area framework)	**5 500**	mesh	**13 509**	mesh
Agricultural land mesh ratio	**4.5**	%	**38.1**	%
Number of first sample mesh *	**957**	mesh	**837**	mesh
First sample extraction ratio	**17.4**	%	**6.2**	%
Average area of first sampling	**6.33**	ha	**7.99**	ha
Number of second sample mesh *	**194**	mesh	**173**	mesh
Second sample extraction ratio	**3.5**	%	**1.3**	%
Average area of second sampling	**7.46**	ha	**8.45**	ha
Average area of second sampling by field survey	**7.92**	ha	**8.86**	ha

Table 1: Fundamental indicator of target provinces on feasibility study

** It shows the number of sample meshes which actually has been conducted survey.*

AFSIS Expert has prepared the ***Area Survey Manual for Researches*** for field survey.

12. Verification of estimated area data on feasibility study

12.1 Formula of accuracy rate

The accuracy rate of agricultural land area data are calculated by using the following formulas through standard error. These standard errors are led by the rate of sample size for framework size and standard error of data average.

a) The formula of accuracy rate on estimation agricultural land area (Y) on first sample

$$\sigma^2 = \frac{\sum(y_i - \bar{y})^2}{n-1} \qquad SE\bar{y} = \sqrt{\frac{N-n}{N}}\frac{\sigma}{\sqrt{n}}$$

b) The formula of accuracy rate of area change rate on second sample

$$\sigma^2(xi - Ry) = \frac{\sum(x_i - \hat{R}yi)^2}{n-1} \qquad SE\bar{x} = \sqrt{\frac{N-n}{N}}\frac{\sigma}{\sqrt{n'}}$$

c) The formula of whole accuracy rate calculation on agricultural land area

$$SEY = \sqrt{SE^2\bar{y} + SE^2\hat{x}}$$

The accuracy rate of crop planted area data are calculated by standard error using the following formula.

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n'-1} \qquad SE\bar{x} = \sqrt{\frac{N-n'}{N}}\frac{\sigma}{\sqrt{n'}}$$

12.2 Standard error of estimated area data

	Vientiane province			**Kandal province**		
	$\bar{y}$	SEY	SERY	$\bar{y}$	SEY	SERY
Agricultural land area	7.92ha	0.250ha	3.4%	8.38ha	0.212ha	2.4%
Crop planted area	$\bar{x}$	$SE\bar{x}$	$SER\bar{x}$	$\bar{x}$	$SE\bar{x}$	$SER\bar{x}$
Rice	6.60ha	0.248ha	3.8%	8.11ha	0.226ha	2.8%
Maize	0.56ha	0.081ha	14.4%	0.38ha	0.115ha	30.2%
Cassava	0.27ha	0.082ha	30.8%	…	…	…
Sugarcane	0.05ha	0.028ha	58.5%	0.10ha	0.053ha	51.6%
Soybean	0.02ha	0.015ha	98.7%	…	…	…

Table 2: Result of target provinces on feasibility study

Table 2 shows the standard error and standard error rate of estimated agricultural land area and each targeted crop planted area by provinces. The results of each crop planted area excluding rice are led by the small appearance rate for samples. It needs to consider the solution and statistical approach for the minor crops based on this result.

In addition, we have to consider that these accuracy rates are not for only area data but also ALIS method is the beginning of Master Sample Method. In other word, the accuracy rate for the true value is assumed the totally completed ALIS operation and the error-free implementation of field survey, however, the standard error rate itself appears good result.

The detail of accuracy calculation including the consideration for target accuracy and number of sample in ALIS are shown by a study report the ***Accuracy Calculation in ALIS*** written by AFSIS Expert

13. Adjustment function of ALIS

Generally speaking, it is a quite rare case that the result of sample survey becomes possible to use itself as public announcement data. Because the statistical data is on data trend of annual, on the other hand it would be requested statistical basement. The final goal of ALIS program targets the public announcement of area data. So data adjustment work is required in the middle of statistical operation in some cases.

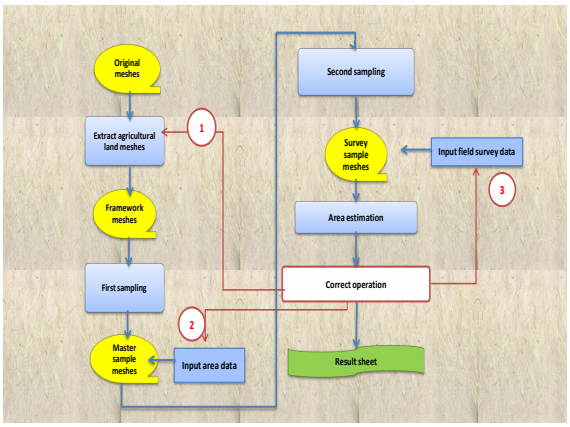

Figure 8: Adjustment functions of ALIS

Figure 8 shows the simple outline of ALIS operation. ALIS operation was one way operation in original design. After this feasibility study, ALIS has three routes for data correct operation. One is for adjustment of the number of extracted agricultural land meshes. Second is for adjustment of input area data for first samples. Third is for adjustment of input field survey data. ALIS operation style changes from one way operation to workflow operation by this renewal.