Dot Sampling Method for Area Estimation

~ Basic concept, procedure, and results ~

Presented by Issei Jinguji.
Expert on Crop Production Survey.
Kamakura city, Japan
Expert meeting on 'Crop Monitoring for Improved Food Security' in Vientiane.
FAO Regional Office for Asia and the Pacific (FAO RAP)
Monday 17 February 2014, Vientiane, Lao PDR

1. Background

A Short Story (Episode) about the Dot Sampling Methods

(1) An old dot sampling method

Please estimate land use share

Around JICA Tsukuba in Japan

*Once upon a time , **an old statistician** in England suggested to apply a point sampling method to area survey on maps[1].*

*He says; You **put sample points** at random on accurate maps, and you **check the category of land use** at sample points, and **estimate the share of the category**. The points are selected clearly with PPS. The method does not require sampling list, it does not require measuring. It does not contact with farmers. It is very suitable for crop surveys[2].*

But, the method required accurate maps. So it was impossible to obtain such accurate maps which are new and reliable in a large area at that time. Few people were interested in the method.

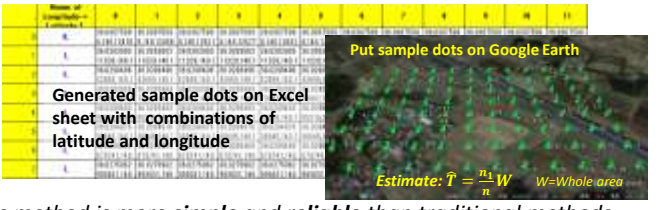
[1] ***Dr. Frank Yates** in England in 1949 . It is also called **"Monte Carlo Method"** (P.18) which was developed by Dr. John von Neumann , etc. in US 1940s*

[2] *Crop surveys are included not only area survey but also yield survey. The method can be applied to such as crop cutting survey .*

(2) A new dot sampling Method

*Recently, **young students group** in Asia learned the old method to improve current crop surveys[3], especially planted area survey, and wanted to realize the method on Google Earth[4]. The map is very accurate, and anyone can use it.*

*They studied how to generate **sample dots on Excel sheet**, and how **to put sample dots on Google Earth.** At last, connecting the old method with the latest information techniques, **they established a new dot sampling method.***



*The method is more **simple** and **reliable** than traditional methods.*

This is the short episode about the development of the dot sampling method, please enjoy next episodes.

The END.

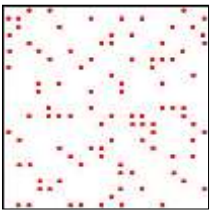
[3] *Current crop survey: area frame survey, or Remote sensing : a big system.* [4] *It is web-site maps, it was developed in US in 2005.*

2. Methodology

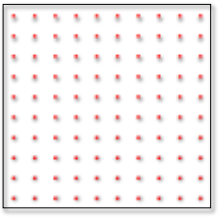
is a good tool

2.1 A "Dot sheet" to learn the dot sampling method

- The dot sheet has two types. One is a **simple-random-dot-sheet**, and another is a **systematic-random-dot-sheet**. The latter is more **useful and reliable**.



simple-random-dot-sheet



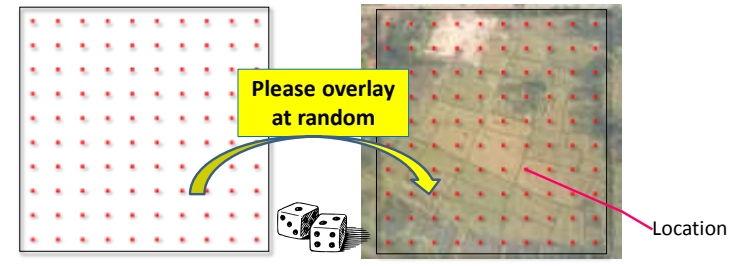
systematic-random-dot-sheet

- The dot sheet has three functions in crop surveys, first is "**random sampling**", second is "**share estimator**" and third is a special function to be "**area frame for variable survey**".

Note: The dot sampling method was suggested us by Mr. Kenji Kamikura who is a senior statistician in MAFF of Japan in May 2011. He always encourages us to develop the method and system.

2.2 First function

- The **dot sheet** can select sample dots **without list,** and **shows the locations** on the maps.

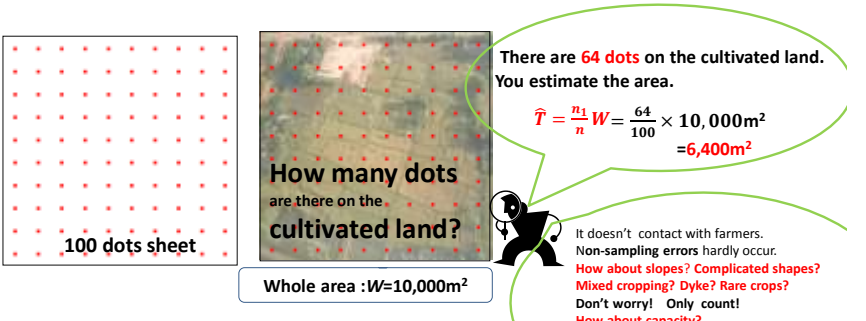


- A sample is selected **with PPS.**

Probabilities proportional to size of field. (land, crop area, every things).

2.3 Second function

- The dot sheet can **estimate** the **share by category.**



- **The method** is called **"attribute survey" .**
- The **formulas** are **simple.**

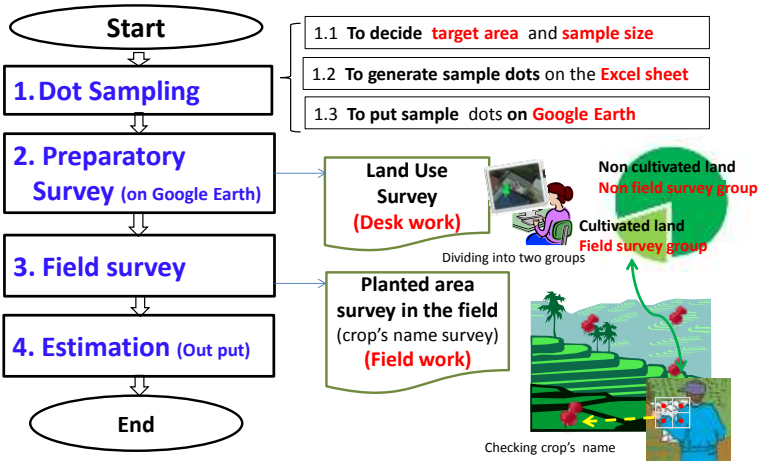
Estimate: $\hat{T} = \frac{n_1}{n} W$ (Whole area)

Standard error $S_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$ Where, $\hat{p} = \frac{n_1}{n}$ $\hat{q} = (1 - \hat{p})$ $\text{CV} = \frac{S_{\hat{p}}}{\hat{p}} \times 100$

Note: Mr. Akira Kato who is a mathematician and statistician discussed with me the theoretical back ground of the formula in Feb. 2013 .

3. Results obtained

3.1 Procedure of the dot sampling method to estimate planted area.



Next , we apply this method to a planted area survey **in a small area.**

3.2 Procedure in a small area is shown as follows:

3.2.1 You decide your target area and sample size

(The target area is **24 ha,** and the **sample size** is **96 .**

- Sample size is decided considering **aimed precision, budget, manpower,** etc. But when you decide a sample size, you should consider the **actual sample size for a field survey**.
- The model area was used for **JICA tanning courses** in 2011, 2012 and 2013,

Note: Mr. Sithixay Linglong who was a trainee from MOA of Lao PDR suggested me to use Google Earth in stead of Google Maps in Nov 2011. Google Earth is suitable maps for a land use survey.

3.2.2 Using Excel macro, you generate sample dots on the Excel sheet (Sampling without list).

- You **input information** which are required in the T-1.
- And sample dots are generated automatically on the T-2. The Table shows the location of the sample dots **with latitude and longitude.**

A part of Excel Sheet

T-1 Basic data to generate sample dots (Sampling Design)

Target area	Size of the Target area km²	Sample size	Starting point (latitude)	Starting point (longitude)	Finishing point (latitude)	Finishing point (longitude)	Interval in km (depend on (3))	Necessary Number of Lines	Necessary Number of Rows
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)=√((2)/(3))	(9)	(10)
JICAshimoyokob	0.24	98	36.030751	140.12411	36.027598	140.13022	0.05	8	11

T-2 Sample dots (Coodinate Values) **Don't change** the numbers on yellow cells because the numbers are used for the calculations.

Name of Longitude → Latitude ↓	0	1	2	3	4	5	6	7	8
0	0,	36.03075098,140.1241082	36.03075098,140.124663606033	36.03075098,140.125219012066	36.03075098,140.125774418099	36.03075098,140.126329824132	36.03075098,140.126885230165	36.03075098,140.127440636198	36.03075098,140.127996042231	36.03075098,140.128551448264
1	1,	36.03030031132 9,140.1241082	36.03030031132 9,140.124663602856	36.03030031132 9,140.125219005711	36.03030031132 9,140.125774408567	36.03030031132 9,140.126329811422	36.03030031132 9,140.126885214278	36.03030031132 9,140.127440617133	36.03030031132 9,140.127996019989	36.03030031132 9,140.128551422844
2	2,	36.02984964265 8,140.1241082	36.02984964265 8,140.124663599678	36.02984964265 8,140.125218999356	36.02984964265 8,140.125774399034	36.02984964265 8,140.126329798712	36.02984964265 8,140.126885198839	36.02984964265 8,140.127440598068	36.02984964265 8,140.127995997746	36.02984964265 8,140.128551397424
3	3,									29398973987, 128551372005
4	4,									28948305316, 128551346586
5	5,									28497636645, 128551321168
6	6,									28049679741 .128551295 75
7	7,	36.02759629930 31,140.1241082	36.02759629930 31,140.124663583 79	36.02759629930 31,140.125218967 58	36.02759629930 31,140.125774351 37	36.02759629930 31,140.126329735 16	36.02759629930 31,140.126885118 95	36.02759629930 31,140.127440502 75	36.02759629930 31,140.127995886 54	36.02759629930 31,140.128551270 33
8	8,	36.02714563063 21,140.1241082	36.02714563063 21,140.124663585 61	36.02714563063 21,140.125218961 22	36.02714563063 21,140.125774343 84	36.02714563063 21,140.126329726 45	36.02714563063 21,140.126885103 07	36.02714563063 21,140.127440485 68	36.02714563063 21,140.127995864 30	36.02714563063 21,140.128551244 91

Those sample dots are sent to Google Earth, using Excel macro.

- The distance between each sample dot is calculated from the size of target area and sample size. The interval in km is shown in (8) in T-1.
- The location of each sample dot which consist of the combination of latitude and longitude is calculated to be same distance in km considering the location (latitude and longitude) on the earth using trigonometric function on the Excel sheet.

3.2.3 Sample dots have just arrived at target area on Google Earth.

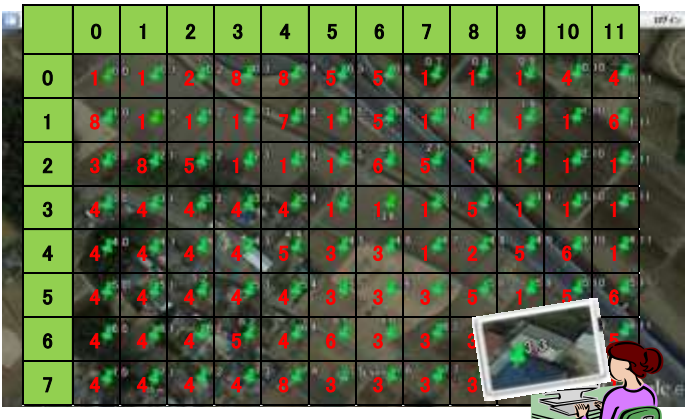
- Dots can show the **locations** and the **categories** at sample dots.
- Each dot is given **name systematically**.

*Note: The macro program **"Save Range as KML File"** was developed by Mr. Hakan Yuksel who was a JICA Expert in Tanzania in Oct 2012. KML file can be sent to others (e.g. local offices, enumerators) by e-mail, and displayed the sample dots.*

3.2.4 You conduct the preparatory survey

- The purpose of the preparatory survey is **to make the field survey efficient, dividing sample dots into two group**.
- While she is conducting the preparatory survey, we show you Google Earth.

Category		Code
Cultivated land	Paddy field	1
	Dyke	2
	Upland	3
Non Cultivated land	Residential Land	4
	Road (asphalt)	5
	Road(soil)	6
	Irrigation, river	7
	others	8
Tentative reserve		9



Note: **Tentative reserve** **makes your work efficiency**. For example, when you are conducting a preparatory survey , you may meet some sample dots which cannot be decided the category of land use quickly. In such cases it will be useful.

3.2.5 You summarize the result of the preparatory survey

- The number of sample dots which **you should go to field survey** is shown in the **"Cultivated land"** and "**Tentative reserve** ", total is 47. You don't have to go to remaining sample dots.
- **Those results can be land use statistics on Google Earth.**

Results of Preparatory survey

	0	1	2	3	4	5	6	7	8	9	10	11
0	1	1	2	3	3	5	5	1	1	1	4	4
1	3	1	1	1	7	1	5	1	1	1	1	6
2	3	3	5	1	1	1	6	5	1	1	1	1
3	4	4	4	4	4	1	1	1	5	1	1	1
4	4	4	4	4	5	3	3	1	2	5	6	1
5	4	4	4	4	4	3	3	3	5	1	5	6
6	4	4	4	5	4	6	3	3	3	5		
7	4	4	4	4	3	3	3	3	3			

Summarize

Frequency distribution of sample dots by category

	Category	Code	Frequency	Share (%)	Area (ha)	SE	CV
Cultivated land	Paddy field	1	30	31.3	7.5	4.7	15.1
	Dyke	2	3	3.1	0.8	1.8	56.8
	Upland	3	14	14.6	3.5	3.6	24.7
Non Cultivated land	Residential Land	4	24	25.0	6.0	4.4	17.7
	Road (asphalt)	5	12	12.5	3.0	3.4	27.0
	Road(soil)	6	5	5.2	1.3	2.3	43.5
	Irrigation, river	7	1	1.0	0.3	1.0	99.5
	others	8	7	7.3	1.8	2.7	36.4
Tentative reserve		9	0	0.0	0.0	···	···
Total			96	100.0	24.0	0.0	0.0

Note: "Dyke" has a special purpose to estimate real planted area separately. This function is important to **estimate accurate production**, concerning yield survey (crop cutting survey). The dot sampling method makes it possible. "Tentative reserve" has a special role to make preparatory survey efficient.

3.2.6 You conduct the field survey to estimate planted area

- The field survey should be conducted in the **crops' growing season.**
- **Google Earth guide you** at sample dots.
- You check the **category of the growing crop** at sample dots.

Category	code
Dyke	3
Residential land, (Include bulding, garden, parking,	4
Road (asphalt)	5
Road (soil)	6
Irigation, River	7
Others	8
Paddy	101
Sweet potato	102
Soybean	103
Vegitable	104
Fruite (Tree)	105
Turf (Lawn)	106
No plant (prepalation)	107

	0	1	2	3	4	5	6	7	8	9	10	11
0	101	3	7	8	6	5	101	101	101	101	8	104
1	106	101	101	101	8	101	5	101	101	101	101	3
2	107	8	4	101	101	101	6	5	106	101	101	3
3	4	6	105	4	5	101	101	101	5	101	101	101
4	4	4	4	4	107	101	107	101	101	5	101	101
5	4	4	4	106	5	107	106	106	3	101	5	101
6	8	8	4	5	107	5	106	5	106	5	101	5
7	8	4	4	4	107	107	104	6	106	106	5	101

ing season

- **When you go to field survey,** you **don't have to go to** sample dots which are at **non-cultivated lands.** This is a reason why you should conduct a preparatory survey before the field survey. It makes the field survey effective.
- But, we checked all sample dots in this training to learn the survey method exactly.

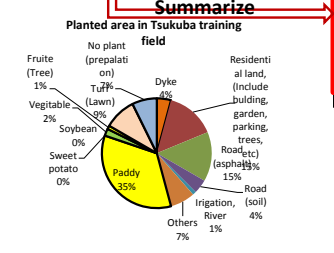
3.2.7 Finally, you summarize and estimate planted area by crop

- You will find **various crops** at sample dots. It means that you can estimate not only **core crops'** planted area but also **rare crops'** planted area, **setting category codes**

Result of the field survey

	0	1	2	3	4	5	6	7	8	9	10	11
0	101	3	7	8	6	5	101	101	101	101	8	104
1	106	101	101	101	8	101	5	101	101	101	101	3
2	107	8	4	101	101	101	6	5	106	101	101	3
3	4	6	105	4	5	101	101	101	5	101	101	101
4	4	4	4	4	107	101	107	101	101	5	101	101
5	4	4	4	106	5	107	106	106	3	101	5	101
6	8	8	4	5	107	5	106	5	106	5	101	5
7	8	4	4	4	107	107	104	6	106	106	5	101

Summarize



Estimate of planted area by crop (JICA training Sep 2012)

Category	code	Number of dots	Share (%)	Area (ha)	SE	CV
Dyke	3	4	4.2	1.0	2.0	48.9
Residential land, (Include bulding, garden, parking,	4	14	14.6	3.5	3.6	24.7
Road (asphalt)	5	14	14.6	3.5	3.6	24.7
Road (soil)	6	4	4.2	1.0	2.0	48.9
Irigation, River	7	1	1.0	0.3	1.0	99.5
Others	8	7	7.3	1.8	2.7	36.4
Paddy	**101**	**33**	**34.4**	**8.3**	**4.8**	**14.1**
Sweet potato	**102**	**0**	**0.0**	**0.0**	**0.0**	**…**
Soybean	**103**	**0**	**0.0**	**0.0**	**0.0**	**…**
Vegitable	**104**	**2**	**2.1**	**0.5**	**1.5**	**70.0**
Fruite (Tree)	**105**	**1**	**1.0**	**0.3**	**1.0**	**99.5**
Turf (Lawn)	**106**	**9**	**9.4**	**2.3**	**3.0**	**31.7**
No plant (prepalation)	**107**	**7**	**7.3**	**1.8**	**2.7**	**36.4**
Total		96	100	24.0	0.0	0.0

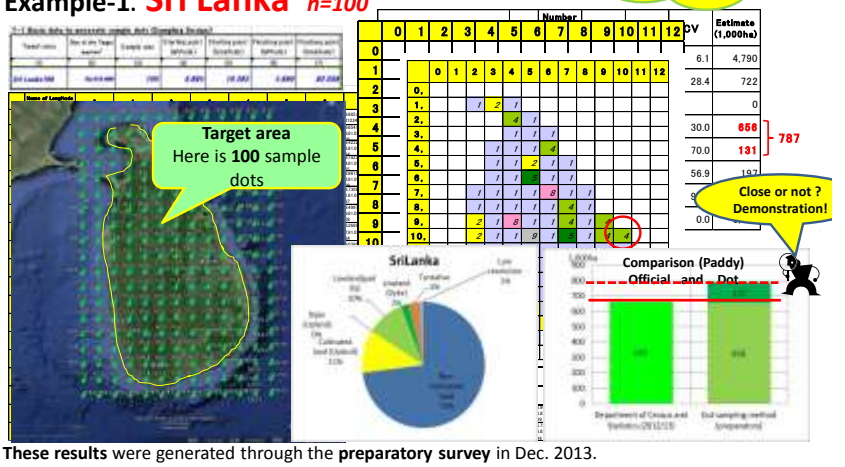
- We have just finished the planted area survey in **a small area**.
- Next, we try area surveys in **a country level**.

Procedure

3.3 Dot sampling in a country level

*After the training, **each trainee tries** a land use survey on Google Earth.*

Example-1. Sri Lanka *n=100*

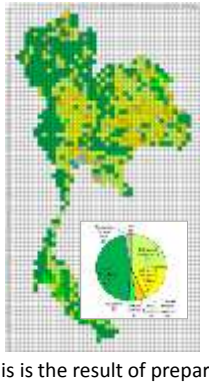


- **These results** were generated through the **preparatory survey** in Dec. 2013.
- The field survey has not conducted. If you want to estimate **planted area by crop,** you should **conduct a field survey at the sample dots.**
- It took **about 1 hour** to complete the preparatory survey.

Example-2 Thailand *n=999*

- Based on the same procedure, we estimated paddy field in Thailand.

Land use map



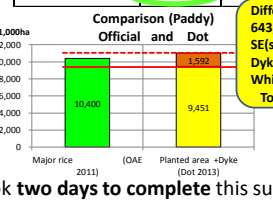
Paddy field in Thailand on Google Earth (Aug.2013)

Category		Number of dots	Rate(%)	Area(ha)	SE	CV
Paddy field	Pure planted area (Major rice)	184	18.4	9,450,859	1.2	6.7
	Dyke, Trees,	31	3.1	1,592,264	0.5	17.7
Total		215	21.5	11,043,123	1.3	6.0

Official planted area (Major rice), OAE, 2011

Category	Code
Cultivated land	...
Cultivated land(Paddy field) — Cultivated land(upland)	1
Cultivated land(Paddy field) — Cultivated land Paddy (Major rice)	11
Cultivated land(Paddy field) — Cultivated land in mountain	2
Cultivated land(Paddy field) — Dyke, Trees,etc	3
Cultivated land(Paddy field) — Maybe orchard	7
Cultivated land(Paddy field) — Aqua farm	8
Forest, River, Lake, etc	4
Residence, Factory, Road	5
Cloud	6
Total	

	Planted area(ha)	Rate(%)
Major rice	10,400,000	20.3



Difference: 643,000ha(1.2point) SE(standard error) 1.2 Dyke! Important factor. Which is better? Today, both are good.

- This is the result of preparatory survey. It took **two days to complete** this survey.
- **Even the area of dyke** (include trees, rocks, cottages in a field) **can be estimated**.
  Rate of dyke: 31/215=14.4% This rate suggests that dyke is an important category **to check planted area .**
- If you want to estimate reliable **planted area by crop ,** you should conduct **a field survey**. You can resolve the difference between the official and the dot estimate.
- Country levels have finished, we show you again Google Earth. Thailand.

3.4 Further discussion

You may have a lot of questions on our presentation.

① **Which is better "Random" or "Systematic"?**
② **Comparison of the dot estimate and complete survey.**
③ **Actual sample size for a field survey.**
④ **Category design.**
⑤ **Low resolution.**
⑥ **Update frequency.**
⑦ **How to use Google Earth.**
⑧ **Weak points of the method.**
⑨ **Etc.(e.g. GPS, Google permission……)**

①,② and③ *are important issues.*

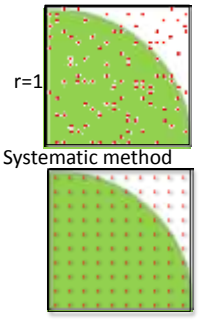
3.4.1 Which is better "random" or "systematic"?

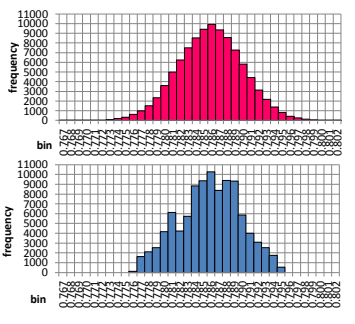
According to the result of our Monte Carlo simulation……

Random method

r=1

Systematic method





	Random	Systematic
min	0.7669	0.7756
max	0.8016	0.7949
average	0.7853906	0.7854058
true value	0.7853982	0.7853982
average/true value	0.9999904	1.0000097
SD	0.0041119	0.0039703

Theoretical (True)Share P = $\frac{A_{circle/4}}{A_{square}} = \frac{(\pi r^2/4)}{r^2} = \frac{\pi}{4} = \frac{3.14159265\ldots}{4} = 0.785398163$

Circumference rate can be estimated.

$\hat{\pi} = 4\hat{p}$=4×0.7854=3.1416

Monte Carlo simulation on the estimation of share of green part (a quarter circle).
Sample size 10,000. **Observation: 100,000 times**

The results show that the shape of the frequency distribution of random is more beautiful than that of systematic. But, **the min, max and average of systematic are better than those of random.** From the viewpoint of practical work, **systematic is better than random**, too.

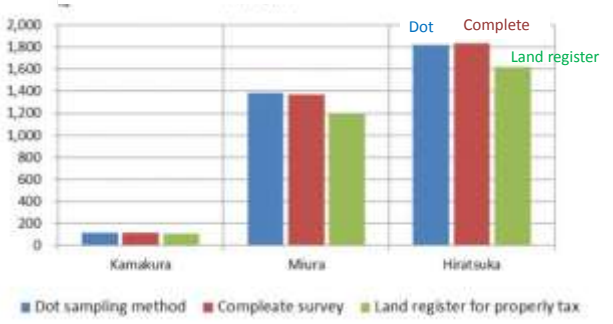
Note: This simulation was conducted by Mr. Nobunori Kuga who is a senior statistician in MAFF of Japan in Jun 2014.

3.4.2 Did you compare the results with complete surveys?

Yes, we did. The results are shown below.

Comparison of Dot sampling and Complete survey by GIS.

Three cities in Japan, December 2011, *n=2600*



- Results of a pretest on cultivated land estimation in Japan (Kamakura, Miura and Hiratsuka city in Kanagawa prefecture)
- This results show the **reliability of the survey.**

3.4.3 Please show us the relations of share, precision and sample size.

The relations are shown below.

Theoretical Sample size (Desk work)

Actual Sample size (Field work)

Share of Cultivated Land (%)	Share of Non cultivated land (%)	Needed sample size by CV(%) (Preparatory survey)			Actual sample size by CV(%) (Field survey)		
p	q=(100−p)	3	5	10	3	5	10
1	99	110,000	39,600	9,900	1,100	396	99
2	98	54,444	19,600	4,900	1,089	392	98
3	97	35,926	12,933	3,233	1,078	388	97
4	96	26,667	9,600	2,400	1,067	384	96
5	95	21,111	7,600	1,900	1,056	380	95
10	90	10,000	3,600	900	1,000	360	90
15	85	6,296	2,267	567	944	340	85
20	80	4,444	1,600	400	889	320	80
30	70	2,593	933	233	778	280	70
40	60	1,667	600	150	667	240	60
50	50	1,111	400	100	556	200	50
60	40	741	267	67	444	160	40
70	30	476	171	43	333	120	30
80	20	278	100	25	222	80	20

$$n = \frac{pq}{S_p^2} \qquad n' = \frac{pq}{S_p^2}p$$

Actual sample size can be reduced dramatically !!

Theoretical sample size is decided by two factor, "**share" and "aimed precision"**
Actual sample size is decided [two factor, "**share" and "aimed precision"] and ["share"]**
Therefore, the **actual sample size** for the field survey is smaller than the sample size in the stage of preparatory survey which is decided considering precision.

4. Conclusion

1. We have established the "**Dot Sampling Method for Area Estimation".**
2. It is connected **Attribute method** with **Excel** and **Google Earth**.
3. The method have **resolved various issues** which were difficult.
4. The **advantages** and **achieved techniques** are shown as follow.

Items		Advantages/Achieved techniques
Attribute method Excel Google Earth		Simple statistical methodology. No measuring (Category survey), Non sampling errors hardly occur.
		Sampling without list.
		Accurate Web maps. (Location, Category)
Procedure	Sampling	on Excel Sheet, Google Earth.
	Preparatory	on Google Earth. Survey item: One (category).
	Field survey	on the Field, No-contact with farmers. Resolved complicated land, rare crops, etc.
Results (Estimates)		Simple Manual/Questionnaire. Speedy, reliable outputs.

Simple

Easy

Reliable

Cost effective

5. **We hop**e that the method is **tested** and **used** in **Asia and Pacific regions .**
6. **Let's learn the method together.**

5. Acknowledgement

Thank you very much for your attention

At the same time, we appreciate their cooperation[1] in MAFF of Japan, JICA, MAFC of Tanzania, Africa Rice Center and FAO who helped us to establish and to spread the method.

[1] **Experts:**
MAFF of Japan: Mr. Kenji Kamikura, Mr. Nobunori Kuga, Mr. Yasuhiro Miyake, Ms. Emiko Morimoto. Mr. Ryuki Ikeda . [Retired]: Mr. Takejirou Endo, Mr. Akira Kato.
JICA Tanzania: Mr. Minoru Homma.
JICA M&E Project: Dr. Fuminori Arai, Dr. Michio Watanabe, Ms. Kyoko Akasaka, Mr. Hakan Yuksel.
MAFC of Tanzania: Mr. Oswald Ruboha, Mr. Alli Kisusu.
AfricaRice: Dr. Aliou Diagne, Dr.Toure Ali, Dr. Alioune Dieng.
FAO: Dr. Naman Keita, Dr. Elisabetta Carfagna, Dr. Mukesh Srivastava.

Reference

FAO, THE WORLD BANK. (2010) GLOBAL STRATEGY TO IMPROVE AGRICULTURAL AND RURAL STATICTICS

FAO, THE WORLD BANK. (2012) ACTION PLAN of the global strategy to improve agricultural and rural statistics. 45-46

FAPRAP, (Feb 2014), Concept Note: Expert Meeting on Crop Monitoring for Improved Food Security

Frank Yates, (1949), SAMPLING METHODS for CENSUS AND SURVEYS 35, 82

Google Earth on the Internet web site

Improving Food Security Information in Africa. Africa Rice Center, Ministry of Agriculture , Forestry and Fisheries, Japan.(July 2013), Rice Production Survey using the Dot Sampling Method and Google Earth. July 2013

Issei Jinguji. Agricultural statistics. JICA expert. (Dec 2012), How to Develop Master Sampling Frames using Dot Sampling Method and Google Earth

Kenji Kamikura Senior Statistician, Statistics Department Ministry of Agriculture, Forestry and Fisheries of Japan. (FAO APCAS 24 Document, October 2012) Estimation of Planted Area using the Dot Sampling Method 1-2,4

Naman Keita Manager CountrySTAT Statistics Division Food and Agriculture Organization of the United Nations (28 November 2011). GLOBAL STRATEGY TO IMPROVE AGRICULTURAL AND RURAL STATISTICS A FRAMEWORK FOR CAPACITY DEVELOPMENT AND TECHNICAL ASSISTANCE

Nobunori Kuga Senior Statistician, Statistics Department Ministry of Agriculture, Forestry and Fisheries of Japan. (January 2014) Comparison of the random sampling and the systematic sampling for the dot sampling using Monte Carlo method.

Theresa "Terry" Holland USDA National Agricultural Statistics Service. (21 June 2012) Improving Agricultural Statistics in Tanzania: Results of USDA and FAO-USDA Assessments

William G. Cochran, New York, (1977), Sampling Techniques (Third Edition) 50-55