Improving Paddy Statistics in Indonesia: Developing Crop Cutting Survey Using Area Sampling Frame

Amalia, Ratna Rizki and Kadir, Kadir

Statistics Indonesia, Statistics Indonesia

19 June 2019

Improving Paddy Statistics in Indonesia: Developing Crop Cutting Survey Using Area Sampling Frame

Ratna Rizki Amalia[1]
ratna.amalia@bps.go.id

Kadir[2]
kadirsst@bps.go.id

Abstract

Although the official figure of production data that was fully based on objective measurements was successfully released in 2018 for the first time, the productivity data is still obtained through a crop-cutting survey that is conducted based on a list frame (households) instead of an area frame. The use of a list frame has several limitations such as a high non-response rate that affects the accuracy of estimation. In other words, the crop cutting survey also is going to make use of an area frame instead of a list frame. By doing so, there is no need to conduct a household listing for the sampling frame, which is time-consuming and costly. Therefore, the implementation of the crop cutting survey will become more efficient and faster than before. The aim of this paper is to give an explanation about the technicalities that are going to be implemented regarding the integration. The defence would be focused on the methodology that is going to be applied and the business process of the data collection.
Keywords: productivity, area sampling frame, crop cutting survey, list frame

1. Introduction

For many years paddy production figure in Indonesia has been a source of policy and political debate, especially when it comes to importation of rice to stabilize the domestic price. Amid a strong commitment towards self-sufficiency, Indonesia is still one of the main rice importers in the world. It has continued to happen although the production figure pointed out that there was no need for importation due to plenty of production surplus. In fact, dissociation between the development of rice price in the market and the production figure happened very frequently. Soaring in price, which indicates a supply shortage, still happened when the official figure recorded a substantial amount of rice surplus.

Many parties have blamed the accuracy of the official production figure as the main cause of that such inconsistency. It has been suspected to suffer from overestimation because of the use of subjective measurements in estimating it. Technically, the production figure of paddy is calculated by multiplying two variables, namely the harvested area and productivity (yield per hectare). The former variable has been suspected as the main source of an upward bias in production data since it was collected through subjective measurements, mainly "eye-estimate" method.

In improving the accuracy of paddy statistics, especially harvested area, BPS-Statistics Indonesia in collaboration with the Agency for Assessment and Implementation of

Technology (BPPT) has developed an objective measurement based on area sampling frame (ASF) technique to estimate the harvested area of paddy since 2015. Although the official figure of production data that was fully based on objective measurements was successfully released for the first time in 2018, the productivity data is still obtained through a crop-cutting survey that is conducted based on a list frame (households) instead of an area frame.

The use of a list frame suffers from some considerable limitations. The main limitation is due to a high probability of non-response because enumerators miss observing selected household targets. This case happens quite frequent since a household that was selected as a sample did not report its harvest activities to the enumerator as supposed. In some cases, the information given by farmers about harvest plan is not accurate. As the consequence, the number of sample targets is failing to meet, which affects the accuracy of estimation. Another limitation is the practical issue. The updating of the list frame for a selected census block is costly and time-consuming. In practice, an enumerator has to visit (door-to-door) each household in the census block to collect information about their existence and the state of their cultivated paddy crops. This practice, of course, needs a substantial effort regarding the operational cost, especially when it comes to a geographical issue. Sometimes, the location of a selected census block is not easy to reach since it is situated in a remote area and far away from the town.

Therefore, this paper suggests the use of crop cutting surveys with an area frame approach for productivity data collection in Indonesia. This work is part of a national project carried out by BPS-Statistics Indonesia in improving food crops data in Indonesia. The aim of this paper is to give an explanation about the technicalities that are going to be implemented regarding the integration. The explanation would be focused on the methodology that is going to be applied and the business process of the data collection. The main contribution of this paper is to enrich the knowledge and discussion about the methodology of producing paddy statistics, especially productivity, by making use of ASF.

2. Current Rice Data Collection in Indonesia

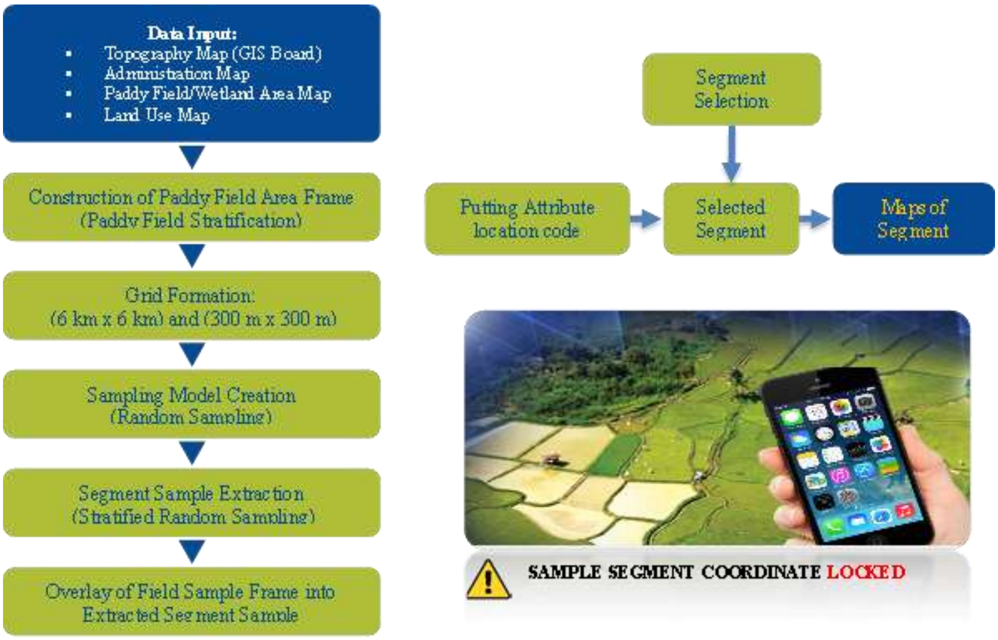
As mentioned earlier, the production figure of paddy in Indonesia is estimated by multiplying harvested area and productivity. Since 1973, data for the former variable has been collected by subjective measurements (BPS, 2015). Some approximations frequently used to collect the data, among others, are seed use, water use for irrigation, information from farmers and village officers and eye-estimate. In practices, the last approximation is the most frequently implemented.

The use of those subjective measurements in estimating harvested area has been suspected as the main contributors of inaccuracy in the production figure. Some studies pointed out that the harvested area measured by subjective measurements, especially eye-estimate, suffered from overestimation. In 1997, a study conducted by BPS-Statistics Indonesia and Japan International Cooperation Agency (JICA) found that the harvested area collected by subjective measurements was overestimated by around 17 per cent (Maksum and Sastrotaruno, 1997). Since then, some measures to find out a better method of estimation based on an objective measurement have been done. Thanks to the availability of paddy area data developed from satellite imagery, BPS-Statistics Indonesia and BPPT started to develop an objective measurement to estimate the harvested area by making use of area sampling frame (ASF) technique in 2015. The official figure from the new method was released for the first time in 2018 ending the long debate about the accuracy of the production figure of paddy in Indonesia.

An area frame can be defined as a list of land parcels, sometimes called segments, which have distinct and identifiable boundaries. In a census or survey, those segments are treated as statistical units to be observed (Muchlis, 2018). Meanwhile, Wigton and Bormann (1978) defined ASF as a statistical design based on a partition of an area into N smaller parts (sub-areas) so-called sampling units. From those sampling units, n units are selected randomly as enumeration areas and called segment samples. The use of ASF for agricultural statistics actually has been started since the 1950s in the US. For almost 70 years, the National Agricultural Statistics Services (NASS) has been implementing ASF to produce agricultural statistics such as grain production, crop acreage, and livestock inventories (Davies, 2009). In Indonesia, the implementation of ASF to obtain agricultural statistics dated back in 1979. At that time, with the funding of the US government, ASF was introduced for the first time to obtain statistics for rubber, coconut and paddy. A pilot survey was conducted in Lampung Province in 1980 (Willet, 1981). However, there was no a follow-up of the pilot survey being done.

As a breakthrough in estimating the harvested area of paddy, the implementation of ASF in Indonesia combined paddy filed data based on remote sensing technology (satellite imagery) as a sampling frame and the utilization of Android device for field observation. It provides rich information about the development of paddy plants on a monthly basis. The information then can be used to obtain an estimation of the paddy area in each growth phase including the harvested area. The implementation of ASF for the harvested area estimation is summarized in Figure 1.

Figure 1. Implementation of ASF for harvested area estimation



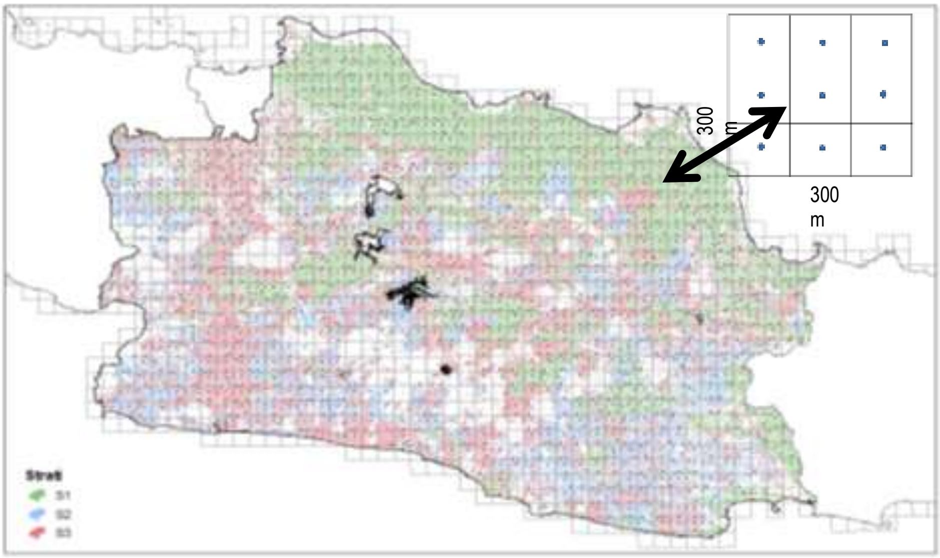
Source: BPS-Statistics Indonesia (2018)

The implementation of ASF is started with the development of a sampling frame by making use of some spatial data, namely administrative map, paddy field map, land cover map, and topography map. The frame is constructed by overlaying those maps altogether. The frame then is stratified into four strata as follows.

a. Stratum-0 (S-0) which contains polygons of non-arable land such as forests, plantations, ponds, bodies of water, and settlements. This stratum will be excluded from sample selection.
b. Stratum-1 (S-1) which contains polygons of irrigated rice fields, either cultivated once a year, twice or more. Segments in this stratum will be chosen as samples.
c. Stratum-2 (S-2) which contains polygons of non-irrigated or rain fed rice fields. Segments in this stratum will also be chosen as samples.
d. Stratum-3 (S-3) which contains polygons of possible rice fields that in practice these are actually dry land polygons.

Once a stratified area frame is ready, it is divided into grids and sub-grids sized 6 km x 6 km and 300 m x 300 m respectively. Random sampling then is applied to obtain sample segments. The selected sample segments equipped with geo-reference information and ID information (province, regency, district, and randomizations codes) then are observed on a regular basis (monthly) by surveyors. The illustration of segment and sub-segment are presented in Figure 2. The survey can be considered as a panel study since the same segment samples are going to be observed every month without any sample replacement.

Figure 2. Distribution of segments by stratum



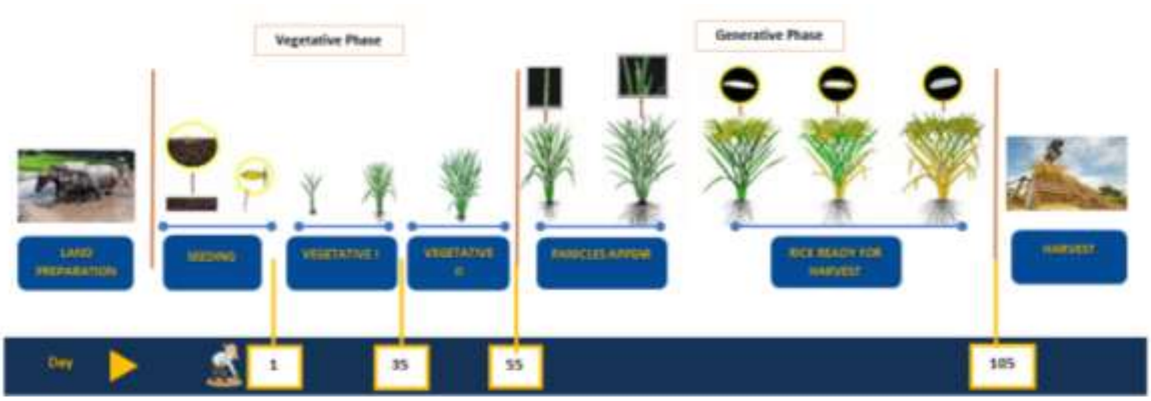
Source: Mubekti and Sumargana (2016)

In conducting the field observation, the surveyors will make use of their smartphone equipped with an Android application, which is specifically developed for ASF, the surveyor observed growth phase and take a picture on the centre of all subsegments in a selected segment. There are nine sub-segments in each segment sized 100 m x 100 m to be observed by a surveyor. The growth phase information and the picture obtained from each sub-segment then are sent to a server. This procedure can get rid off the subjectivity in identifying the growth phase of paddy crops and minimizing the human errors. The final output is the estimation of paddy field area according to the growth phases, namely land preparation, vegetative, generative, and harvest.

The output of ASF allows us to make predictions (forecasts) of the harvested area for the next three months. The prediction for one month ahead can be obtained from the total area of paddy that is in the generative phase. The forecasts for two and three months ahead can be determined from the total area of vegetative II and vegetative I phases respectively. This procedure can be applied since paddy crops in the generative phase would be harvested in one month ahead while those in vegetative II and I would need 2 and 3 months respectively to be ready to harvest. For example, from observation in January, the forecast of harvested area for February is simply obtained from the area of generative phase in January while the forecast for March and April are just the area of vegetative II and I in phases January respectively. The same pattern applies for other months in a successive way. This pattern basically follows the growing phase calendar of paddy crops presented in Figure 4.

Figure 3. Rice Growth Phase Calender

Rice Growth Phase by Month											
January	February	March	April	May	June	July	August	Sep	Oct	Nov	Dec
V1	V1	V1	V1	V1	V1	V1	V1	V1	V1	V1	V1
V2	V2	V2	V2	V2	V2	V2	V2	V2	V2	V2	V2
G	G	G	G	G	G	G	G	G	G	G	G
H	H	H	H	H	H	H	H	H	H	H	H

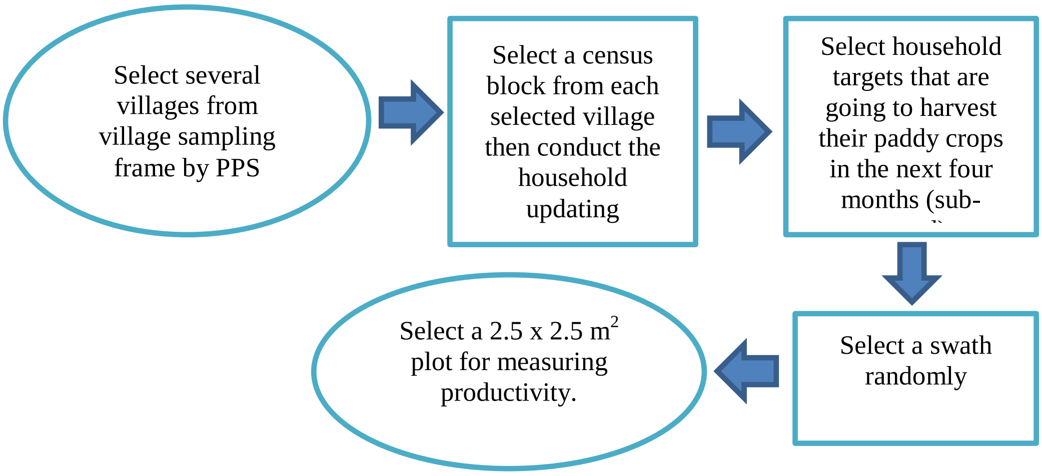


Source: BPS-Statistics Indonesia (2018)

Meanwhile, productivity is obtained from an objective measurement, namely the crop-cutting survey or *Ubinan*. The survey design applies a multi-stage random sampling. In the first stage, several villages are proportionally chosen by Probability Proportional to Size (PPS) with the number of food crops farmer as the size. In the second stage, a census block is proportionally selected from each selected village by using the food crops farmer as the size. In every selected census block, the household updating is then conducted. This stage makes use of a list of household targets in a selected census block as a sampling frame. From the updating process, the household targets that are going to harvest their paddy crops in the next four months (sub-round) can be identified as well their harvest schedule. In each selected household, a swath of paddy field is selected by simple random sampling (SRS). The ultimate sampling unit is a plot sized 2.5 m x 2.5 m selected randomly from a selected field. The final step of the survey is the measurement of the yield in the plot to be extrapolated to obtain estimation for yield per hectare (productivity). In summary, the description of the current sampling design can be seen in Figure 4.

The survey is conducted annually in three sub-rounds, namely the first sub-round covering the period of January-April, the second sub-round (May-August), and the third sub-round (September-December). Meanwhile, the updating process to obtain a list frame is conducted in the last month of the previous sub-round. For example, the frame for the second sub-round is updated in April.

Figure 4. Sampling Design of Current Crop Cutting Survey



3. Crop-cutting Survey Using an Area Frame

From the discussion in the previous section, it can be seen that although the production figure released in 2018 were fully based on objective measurements, the crop-cutting survey still makes use a list frame to estimate productivity. As mentioned earlier, the use of a list frame suffers from some considerable limitations. Those limitations have motivated BPS-Statistics Indonesia to integrate the measurement of harvested area and productivity by using an area frame. In general, the sampling frame for the crop-cutting survey will be obtained from the result of the field observation of ASF for estimating harvested area, i.e, all subsegments with the growth phase of paddy crops on it is vegetative II. This section is going to give a more detail explanation about the implementation of this idea covering both the methodology and the business process.

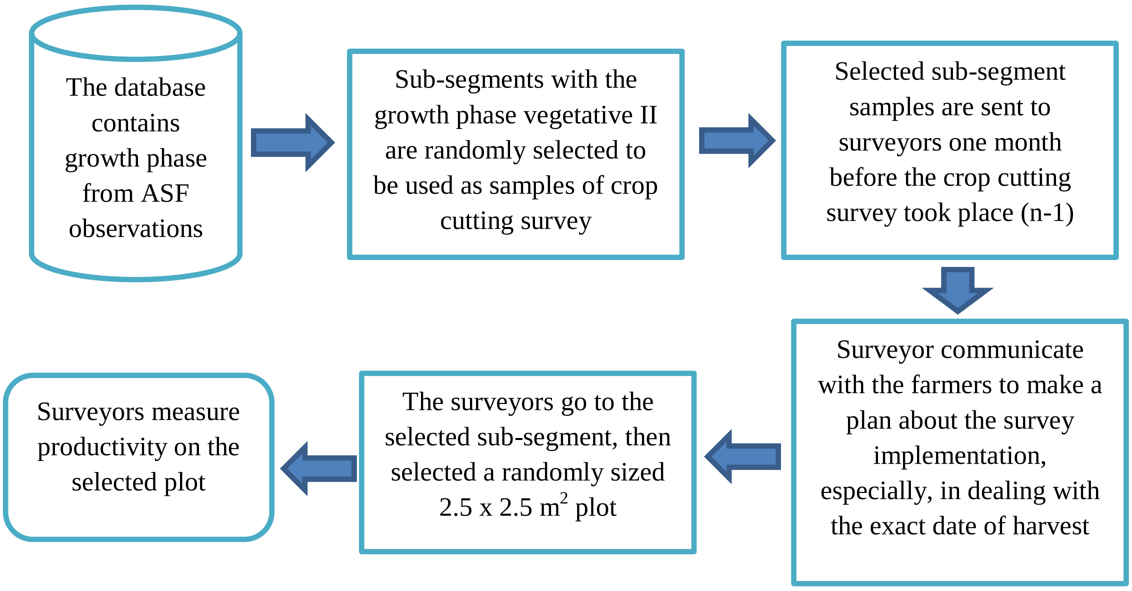
As highlighted earlier, in implementing ASF for estimating harvested area, surveyors will make a visit to the selected sample segments to observe and take pictures of the growth phases of paddy crops in all sub-segments. The growth phases that will be recorded are land preparation, vegetative I, vegetative II, generative, harvest, wetland cultivated non-paddy, damaged paddy crops, and dry land.

Sample Selection

The implementation of the crop-cutting survey would make use the database of subsegment samples belonged to vegetative II phase as a base for sampling selection. This scenario is chosen since paddy crops in the vegetative II phase possibly will be harvested in the next three months. From the database, a set of sub-segment sample will be selected randomly by SRS. The next step is the selection of paddy swath from each selected sub-segments by SRS. It should be done as in each sub-segment may consist of more than one swaths. A selected swath is determined based on the coordinate of the selected sub-segment by which a swath located in the centre of the coordinate will be the sampling unit. This point is basically a place where the observation of the growth phase was done. The final stage is the measurement of productivity a plot sized 2.5 m x 2.5 m that is selected randomly from each selected swath.

By this setting, it can be seen that a list of sub-segment sample will be available for surveyors in two months before the survey will be conducted. Therefore, there is plenty of time for them to communicate with the farmers to make a plan about the survey implementation, especially, in dealing with the exact date of harvest. By doing so, the issue of non-response that frequently take place can be addressed and the number of samples realisation can be improved substantially. Moreover, a better efficiency regarding the cost and time can be gained since there is no need for updating in a selected census block to obtain a list frame. The business process of implementing the new framework is summarized in Figure 5.

Figure 5. The Business Process of Implementing Crop Cutting Survey Using ASF



Finding the Farmers

One of the main challenges in implementing the new framework is how to find out the farmers. This issue arises as we do not have information about to whom a selected sub-segment belongs to. The only information that we have is the location of it, which is the coordinate and ID information. Based on our pilot survey, this issue can be addressed by finding out the farmers once the list of sub-segment samples was already in hand. Therefore, the surveyors can identify the owner of a selected sub-segment at most one month prior to the survey. This information then can be used for the next survey when the sub-segment is re-elected. Once the farmers of selected-sub segments are identified, the surveyors can make an appointment regarding the exact date of harvest. Therefore, they can make a visit at the right time.

Our experience shows that finding the farmers is not a big issue for some areas. Sometimes, it depends on the time of visit. In general, it is quite easy to find the farmers in the field when the visit to a selected sub-segment was made at around 08.00-11.00 am or at about 01.00-04.00 pm, especially during the planting and harvest seasons. For some areas, like in Bogor Regency, West Java Province, finding the farmers is even easier since most of them usually set a semi-permanent building made from bamboo or tent around their paddy field and stay in there during the planting and harvest seasons to oversee their paddy crops.

However, for some cases, finding farmers was quite difficult. To overcome this problem, we had to ask help from other farmers in the same area and village officers. In general, our pilot project pointed out that finding the farmers was not difficult although it needed effort and was a little bit time-consuming. Moreover, finding the farmers may be an issue in the first time, but it can be addressed when the implementation of ASF has run for several times as long as the information about the field ownership is well recorded in each implementation of ASF. Another strategy is to build a database containing information on the ownership of all sub-segments in the frame through a special enumeration. This strategy, of course, will need time and effort to be realised. Therefore, information about a sub-segment is not only about the location but also the farmers belong to it.

4. Conclusion

This paper discusses the implementation of the crop-cutting survey using an area frame to improve the production statistic of paddy in Indonesia. The survey has been making use of a list frame for years, which is time-consuming and costly to develop. The survey also becomes prone to a substantial non-response due to poor awareness of farmers in reporting their harvest activity that can affect the accuracy of estimation. The use of an area frame obtaining from ASF that has been already applied since 2018 to estimate the harvested area of paddy could address this issue. Therefore, it may result in better efficiency and accuracy. One of the main challenges in implementing the new framework is finding and identifying the farmer to whom a selected sub-segment belongs to. One of the measures that can be done to address this issue is to develop a database consists the information of field ownership of all sub-segments observed regularly in ASF to estimate the harvested area.

5. References

BPS-Statistics Indonesia. (2018). Pedoman Teknis Pendataan Statistik Pertanian Tanaman Pangan Terintegrasi dengan Metode Kerangka Sampel Area (KSA) 2018. Jakarta: Badan Pusat Statistik.

Davies, C. (2009). Area Frame Design for Agricultural Surveys. Washington DC, USA: USDA, 2009.

Willet, J.W. (1981). Area Sampling Frames For Agriculture in Developing Countries. The US Department of Agriculture.

Muchlis. (2018). Area Sampling Frame: A New Approach to Reform Agricultural Data Collection. *Asia–Pacific Economic Statistics Week*.

Mubekti and Sumargana, L. (2016). Pendekatan Kerangka Sampel Area Untuk Estimasi dan Peramalan Produksi Padi. *Pangan*, vol. 25, pp. 71-82.

Rosner, L. Peter and McCulloch, Neil. (2008). A Note on Rice Production, Consumption and Import Data In Indonesia. *Bulletin of Indonesian Economic Studies*, vol. 44, No. 1, pp. 81-91, 2008

Wigton, W.H. and Bormann, P. (1978). A Guide to Area Sampling Frame Construction Utilizing Satellite Imagery. *2nd Int'l Training Course in Remote Sensing Applications for Agriculture*: Crop Statistics and Agricultural Census. UN Outer Space Affairs Division,.