Analysis of paddy productivity using normalized difference vegetation index value of sentinel-2 and UAV multispectral imagery in the rainy season

Liyantono¹,², A Sianjaya², IK Sari²

¹ Department of Mechanical and Biosystem Engineering, IPB University, Darmaga, Bogor, Indonesia.
² Undergraduate students, Department of Mechanical and Biosystem Engineering, IPB University, Darmaga, Bogor, Indonesia.
³ Corresponding author, Email: liyantono@apps.ipb.ac.id

Abstract. Rice is one of the primary food sources for people worldwide, especially in Indonesia. In 2018, Indonesia produced approximately 56.5 million tons of unhusked dry rice ready for milling with a total area of around 10.9 million hectares. Until these days, the estimation of rice productivity was done by the Indonesia Statistics Agency (BPS) and the Ministry of Agriculture. Aerial Images and Normalized Difference Vegetation Index (NDVI) were viewed as a tool that helps in monitoring and observing the crops. NDVI was used to analyze the paddy growth from planting until harvesting. Usage of Satellite Sentinel-2 and UAV was meant to compare the accuracy between them. Analysis of paddy productivity was used during the vegetative, reproduction, and maturity phases. The used methods are regression and correlation analysis. The model from the analysis was used to estimate the productivity of paddy. The result shows that the models with the highest correlation from both Sentinel-2 and UAV were obtained from analysis on the vegetative phase. This means the implemented NDVI value was from planting until the peak value in one cultivating season.

1. Introduction
Rice is one of the primary food sources for people worldwide, especially in Indonesia. Most Indonesian people use rice as a staple food in addition to other food sources such as sago and other types of tubers [1]. Rice is known to be divided into two subspecies, namely japonica, which grows in subtropical areas and indica, which is planted in Indonesia [2]. Based on BPS data in 2018 [3], Indonesia produced approximately 56.5 million tons of unhusked dry rice ready for milling with a total harvested area of 10.9 million hectares. Until now, the estimation of rice productivity was done by the Indonesia Statistics Agency (BPS) and the Ministry of Agriculture. The method used by BPS to estimate productivity are sampling in the rice field, the area of sampling is 2.5 m x 2.5 m. The utilization of manual method is too much time and energy. According to Nelson et al. [4], rice grows within 90-150 days, depending on the variety. Rice growth phase is divided into three, namely: vegetative (from planting to the appearance of panicles), reproductive (from the appearance of panicles to flowering), and maturation (from flowering to mature seeds).

Predicting the production of rice is important. Forecasting enabling local government to be able to import rice in case of deficiency or exporting in case of surplus. Because of the importance of that, monitoring of crop development, early productivity forecasting, and harvested area are important.

The appliance of remote sensing during data acquisition was meant to get more accurate data. Aerial Photography usage to calculate rice field potential was one of the steps to identify harvested area and productivity of the rice field [5]. From 2018, BPS use area sampling frames method to estimate the amount the rice production in Indonesia. This method using the area of the field as an enumeration unit with the basis of Geographical Information System (GIS). This method using satellite image as the basis for the estimation.
Aerial photography using UAV was one of the alternate technology to get more detail, real-time, fast, and easy [6]. Remote sensing using satellite and UAV has been widely used to analyze and identify agricultural crops. To analyze the phenological event of paddy, time-series data was used. Some research use time-series data from satellite-like Sentinel-2, MODIS, and Landsat-8 to analyze the productivity of paddy. Satellite imagery has its drawbacks, namely resolution per pixel and time of repeat visits [7]. Productivity estimation was done with Multispectral imagery. The vegetation index is the parameter used in this research. The vegetation index is one of the forms of spectral transformation that was applied to a multichannel image to highlight the aspect of leaf density or any other aspect regarding density, such as biomass, Leaf Area Index (LAI), chlorophyll concentration, and so on [1]. One of the algorithms used is NDVI (Normalized Difference Vegetation Index). NDVI algorithm utilizes NIR and RED spectral bands from the sensor.

The goals of this study were to analyze the characteristic of paddy vegetation index based on Sentinel-2 and UAV multispectral imagery, to compare the NDVI value of Sentinel-2 and UAV multispectral imagery to see if the Sentinel-2 can be used as a substitute of UAV, and to analyze the model of paddy productivity estimation with Sentinel-2 and UAV based on the correlation between paddy productivity and NDVI value.

2. Material and Method
The research was conducted in the rainy season, where water supply was sufficient along with planting until the harvesting stage. The research was divided into three stages, i.e. image capturing, field survey for data collecting, and image processing. The research procedure is shown in Figure 1.

![Figure 1. Research procedure](image)

2.1. Image Capturing
The image used for this research was obtained from UAV and satellite Sentinel-2. The images were captured from February until May 2019 in three different locations in Bogor, i.e. Situgede, Laladon, and Margajaya. The image from UAV was captured every 16 days. Aerial photography with UAV was using drone DJI Phantom 4 equipped with multispectral camera Parrot Sequoia. The flight route of the drone was made using PIX4D application with a flight height of 100 m. While image from satellite Sentinel-2 was captured every five days. According to ESA [8], Sentinel-2 is a twin...
polar-orbiting satellite in the same orbit, phased at 180° to each other. The sentinel-2 image was downloaded from the website by ESA that is https://scihub.copernicus.eu.

2.2. Field survey
The field survey was done to determine which rice fields will be observed. The surveyed data were planting date, harvest date, varieties, and production of each rice field. The data were obtained from local farmers in each area. The survey was done from February until May 2019.

2.3. Image Processing
The image captured from UAV was selected to choose the image that represented the flight route. After the image was selected, the image was stitched with software PIX4D to make one whole image of the rice field. Stitching using PIX4D was divided into three steps which are, (1) initial processing, (2) point cloud and mesh, (3) DSM, ortho-mosaic, and index. The equation to calculate NDVI value is as follows:

\[
NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}
\]  

where \( \rho_{NIR} \) is near-infrared reflectance and \( \rho_{RED} \) is red reflectance. The final result of this process is RGB image and NDVI image (Figure 2).

(a)                                              (b)

Figure 2. Stitching result using PIX4D software, (a) RGB image, (b) NDVI image

After the images were stitched, every rice field was digitized to morph the rice field into a vector image using Quantum GIS (QGIS) software. The digitization process was done by making polygons in the shape of the rice field. Every polygon was given an ID to differs each rice field.

NDVI image from Sentinel-2 was also processed with QGIS software. QGIS has a function called raster calculator to calculate the value from the NIR image and Red image. Equation (1) was used during the process. The next step is the digitization of each pixel in Sentinel-2 NDVI image using a grid feature in QGIS to make a grid layer based on each pixel. The grid layer then intersected with the rice fields polygon to find the weight of each pixel in every rice field.

The next step is NDVI value extraction. After the digitization process, the NDVI values were extracted using the zonal statistic function in QGIS. The NDVI values were varied for each rice field. This function was used for both UAV and satellite images. The NDVI value was exported into .xml format to further analyzed using MS Excel.

In MS Excel, the NDVI value will be sorted based on farmer names. Regression analysis is a statistical method that aims to investigate and model the relationship between productivity and NDVI variables [9]. The correlation between NDVI value and productivity analysis was done for each growth phase, which is vegetative, maximum vegetative, maturity, and one harvest season. There are three methods involved, which are linear regression method and multilinear method. In the linear regression method, the average NDVI value for each respective phase was plotted into a scatter graph to find the model for paddy productivity estimation. While in the multilinear regression method, curve fitting was done with plant age in days for the independent variable (x) and NDVI values as
dependent variable (y) for each farmer, which results in the linear and polynomial equation for each farmer. The variable in each equation was used as parameters to estimate paddy productivity. Figure 3 is the flowchart for the multilinear regression method used.

Figure 3. Multilinear regression procedure

3. Result and Discussion

3.1. NDVI value analysis

UAV image acquisition starts from January 30th, 2019 and ends on May 27th, 2019. The image was taken every 16 days. The condition for each field was different. There was some field already planted, and another filed was still in the process of soil tillage. An example of a phenological event of paddy is shown in Figure 4, 5, and 6.

Figure 4. Phenological event of paddy in Situgede
According to figure 4, 5, and 6, paddy phenological events have parabolic curve forms. NDVI value will increase accordingly from planting until it reaches the maximum value, and then it will decrease until it reaches the harvesting period. The NDVI value of Sentinel-2 is higher than UAV when it first planted. Reaching the maximum point, the NDVI value of UAV is higher than Sentinel-2. Approaching the harvesting period, the NDVI value of Sentinel-2 is higher than UAV. The difference in the phenological event is because the resolution of the Sentinel-2 is lower than UAV. Each field has different treatments by each farmer, resulting in a difference in growth and NDVI value. There are seven different varieties in Situgede location, which is, Ciherang, Inpari 33, Inpari 40, Inpari 43, Mekongga, Mentik wangi, and Cisantana. The majority of them are Ciherang variety, with a sample count of 19 fields. For Margajaya location, there are five different varieties which are, Inpari 30, Inpari 32, Inpari 42, GH, and IR 64. While in Laladon, there are three different varieties which are, Inpari 32, GH, and IR 42. While in Laladon, there are three different varieties which are, Inpari 32, GH, and IR 42.

3.2. Productivity analysis

Paddy was harvested using different methods and different tools. Paddy was cut in the rice panicles section, bottom section, or middle section. After the paddy was cut, it was threshed to separate the unhusked rice and the stem. After the unhusked rice was threshed, it was dried to reduce the moisture content of the rice. The productivity (Y) of the paddy was obtained using this equation as follows:

\[ Y = \frac{P}{A} \]  

(2)
where $P$ is production and $A$ is area.

The production of the paddy was obtained from field survey to local farmers in each respective location, while the area of the rice field was obtained from digitization with QGIS software. Table 1 shows the productivity value for each variety based on all location observed.

| Varieties   | Area (ha) | Productivity (t/ha) |
|-------------|-----------|---------------------|
| Cisantana   | 0.0337    | 4.451               |
| Inpari 30   | 2.1392    | 6.096               |
| Inpari 32   | 2.4514    | 4.959               |
| Inpari 33   | 1.6872    | 5.073               |
| Inpari 40   | 0.0887    | 4.047               |
| Inpari 42   | 0.7332    | 6.365               |
| Inpari 43   | 0.4889    | 4.252               |
| Mentik Wangi| 0.7134    | 2.690               |
| Mekongga    | 0.1263    | 3.167               |
| Ciherang    | 0.9557    | 4.244               |
| IR 42       | 1.6277    | 5.040               |
| IR 64       | 1.6922    | 4.022               |
| GH          | 0.4000    | 6.774               |

The highest productivity achieved by GH variety with the value of 6.774 t/ha while the lowest productivity achieved by mentik wangi variety with the value of 2.690 t/ha. The productivity of GH is still lower than the expected production, which is 9.3 t/ha. The productivity of mentik wangi is the lowest because it was only given organic fertilizer.

3.3 Estimation of paddy productivity

According to Wahyunto and Herwanto [9], the NDVI value in vegetative phase will increase until it reaches the maximum point (maximum vegetative phase) and then when the plants reaching generative phase, and the NDVI value will decrease because the grain will be filled and changing of leaf color when the paddy approaching harvesting time. Illustration of each growth phase can be seen in Figure 7.

The first method is the linear regression method; the average value of each phase was plotted into the graph. The equation obtained will be used as a model to estimate paddy productivity. The second method is multilinear regression, the first thing to do is plotting the NDVI value with the age of plants for each farmer, and then the coefficient of each obtained equation was extracted as input for the next linear regression. The result of regression and correlation analysis for each location with UAV imagery was shown in Table 2, while the result with Sentinel-2 imagery is shown in Table 3.
Figure 7. NDVI value of paddy in (a) reproductive phase, (b) one season phase, (c) vegetative phase, and (d) generative phase

Table 2. UAV paddy productivity estimation model

| Location  | Phase          | Age in weeks | Linear Data count (n) | Linear R² | Multilinear Data count (n) | Multilinear R² |
|-----------|----------------|--------------|-----------------------|-----------|---------------------------|---------------|
| Situgede  | Vegetative     | 0-13         | 17                    | 0.149     | 17                        | 0.325         |
|           | Maximum vegetative | 9-13        | 17                    | 0.208     | -                         | -             |
|           | Reproductive   | 4-14         | 8                     | 0.169     | 17                        | 0.096         |
|           | Generative     | 9-20         | 17                    | 0.0004    | 17                        | 0.096         |
|           | One season     | 0-20         | -                     | -         | 8                         | 0.623         |
| Margajaya | Vegetative     | 0-9          | 9                     | 0.247     | 7                         | 0.171         |
|           | Maximum vegetative | 5-9        | 12                    | 0.284     | -                         | -             |
|           | Reproductive   | 6-12         | 7                     | 0.051     | 7                         | 0.905         |
|           | Generative     | 8-20         | 11                    | 0.165     | 4                         | 0.690         |
| Laladon   | Vegetative     | 0-9          | 9                     | 0.127     | -                         | -             |
|           | Maximum vegetative | 7-9        | 9                     | 0.046     | -                         | -             |
|           | Reproductive   | 8-10         | 9                     | 0.107     | 9                         | 0.439         |
|           | Generative     | 7-20         | 9                     | 0.036     | 8                         | 0.551         |
Table 3. Sentinel-2 paddy productivity estimation model

| Location | Phase             | Age in weeks | Data count (n) | Linear  | R²   | Multilinear | R²   |
|----------|-------------------|--------------|----------------|---------|------|-------------|------|
|          |                   |              |                |         |      |             |      |
|          | Vegetative        | 0-13         | 18             | 0.008   | 11   | 0.529       |      |
|          | Maximum vegetative| 9-13         | 17             | 0.146   | -    | -           |      |
|          | Generative        | 9-20         | 17             | 0.002   | 20   | 0.104       |      |
|          | One season        | 0-20         | 10             | 0.168   | 11   | 0.684       |      |
| Situgede |                   |              |                |         |      |             |      |
|          | Vegetative        | 0-9          | 9              | 0.213   | 5    | 0.889       |      |
|          | Maximum vegetative| 5-9          | 13             | 0.056   | -    | -           |      |
|          | Generative        | 6-12         | 13             | 0.035   | 12   | 0.203       |      |
|          | One season        | 8-20         | 7              | 0.305   | 6    | 0.998       |      |
| Margajaya|                   |              |                |         |      |             |      |
|          | Vegetative        | 0-9          | 8              | 0.00001 | -    | -           |      |
|          | Maximum vegetative| 7-9          | 9              | 0.150   | -    | -           |      |
|          | Generative        | 8-10         | 9              | 0.041   | 9    | 0.324       |      |
|          | One season        | 7-20         | 3              | 0.831   | -    | -           |      |
| Laladon  |                   |              |                |         |      |             |      |

Table 2 and 3 show that the data count for each phase is not the same. It’s because in some phases, the data can’t be used. The reproductive phase in Sentinel-2 can’t be analyzed because of insufficient data. The R² in the linear is quite low; that means the dependent variable (productivity) has a low dependency on the independent variable (NDVI) or productivity has more high dependency on another variable. Another probability is the high variability of the dependent variable because data collected based on an interview with the farmer (not on direct measurement). Table 2 shows that the best model from UAV in Situgede was obtained from one season with multilinear regression with an R² value of 0.623. The equation from this analysis is \( y = 1266.2\alpha_0 + 197.9\alpha_1 + 21.02\alpha_2 - 21.94 \). From Margajaya, the best model obtained is from the reproductive phase with multilinear regression with an R² value of 0.905. The equation from this analysis is \( y = -6.390\alpha_0 - 66.65\alpha_1 - 8.48\alpha_2 + 17.23 \). From Laladon, the best model obtained is from the generative phase with multilinear regression with an R² value of 0.551. The equation from this analysis is \( y = 6.76\alpha_0 + 98.95\alpha_1 + 1.47 \).

Table 3 shows that the model from Sentinel-2. The best model obtained from Situgede was also obtained from one season phase with multilinear regression analysis. The R² value is 0.684. The equation from this analysis is \( y = 191.817\alpha_0 + 18.755\alpha_2 - 3.927 \). The best model obtained from Margajaya was obtained from one season phase with multilinear regression analysis. The R² value is 0.998. The equation from this analysis is \( y = -65409.208\alpha_0 - 1954.99\alpha_1 - 66.957\alpha_2 + 47.081 \). The best model obtained from Laladon was obtained from one season phase with linear regression analysis. The R² value is 0.831. The equation from this analysis is \( y = 31.456x - 11.8 \). These model still has limitation to the specific location, so the future research is needed to improve the model become a generic model.

4. Conclusion and suggestion

4.1. Conclusion

In this research, we know that multispectral imagery can be used to estimate paddy productivity. The phenological events of paddy formed a parabolic curve in which the growth has been divided into three phases. The first phase is the vegetative phase, the second is the reproductive phase, and the third is the generative or maturity phase. The NDVI value of Sentinel-2 is higher than UAV at the beginning of the vegetative phase. Approaching the reproductive phase, the NDVI value of UAV is higher than Sentinel-2 until the maximum point. The NDVI value of Sentinel-2 is higher than UAV when approaching the harvest time.

The best model obtained from UAV is from Margajaya during the reproductive phase with multilinear regression analysis, which has an R² of 0.905. The equation for this model is \( y = -6.390\alpha_0 - 66.65\alpha_1 - 8.48\alpha_2 + 17.23 \). The best model obtained from Sentinel-2 is from Laladon during one season phase with linear regression analysis, which has an R² of 0.831. The equation for this model is \( y = 31.456x - 11.8 \).
4.2. Suggestion
The amount of observed field should be increased because high data count leads to higher data variability that we can use. Image acquisition should have a fixed interval. Image acquisition of Sentinel-2 should be made in the area with lower cloud coverage, so the amount of data increased.

5. References
[1] Sudarsono NW, Surdarsono B, Wijaya AP. 2016. Analyzed of paddy growth phase using NDVI, EVI, SAVI, and LSWI on Landsat 8 Imagery (in Indonesian). Jurnal Geodesi Undip 5(1): 125-134.
[2] Purwono, Purnamawati H. 2007. Cultivation of 8 types of superior food plants (in Indonesian). Depok (ID): Penebar Swadaya.
[3] [BPS] Badan Pusat Statistik. 2018. Official news on the harvested area and rice production statistics in West Java 2018 (in Indonesian). [internet]. [downloaded 2019 September 20]. Available on: https://jabar.bps.go.id/pressrelease/2018/11/01/682/luas-panen-dan-produksi-padi-di-jawa-barat-2018.
[4] Nelson A, Setiyono T, Rala AB, Quicho ED, Raviz JV, Abonete PJ, Maunahan AA, Garcia CA, bhatti HZM, Villano LS. 2014. Towards an operational SAR-based rice monitoring system in Asia: examples from 13 demonstration Sites across Asia in the RICE Project. Remote Sens. doi:10.3390/rs61110773.
[5] Irsyad F. 2017. Application of aerial photos to predict the potential of paddy fields at Solok City using an unmanned aerial vehicle (in Indonesian). Jurnal Teknologi Pertanian Andalas. 21(2).
[6] Shofiyanti R. 2011. Unmanned aerial vehicle technology for mapping and monitoring crops and agricultural land (in Indonesian). Jurnal Informatika Pertanian. 20(2): 58–64.
[7] Jinru X, Baofeng S. 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. Journal of Sensors. 2017: 1-17.
[8] [ESA] European Space Agency. 2015. Sentinel-2 User Handbook. [internet] [downloaded 2019 March 14]. Available on: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook.
[9] Wahyunto W, Heryanto B. 2006. Estimating the productivity of paddy fields through satellite image analysis (in Indonesian). Warta Informatika Pertanian. 15(2):853-859.