Rice Productivity Prediction Model Design Based On Linear Regression of Spectral Value Using NDVI and LSWI Combination On Landsat-8 Imagery

Yudo Prasetyo 1*, Abdi Sukmono 1, Kurnia Wisnu Aziz 1 and Bernardinus Joko Prakosta Santu Aji 1

1Department of Geodetic Engineering, Faculty of Engineering, Diponegoro University Semarang, Indonesia

yudo.prasetyo@ft.undip.ac.id

Abstract. Rice is one of the most important agriculture crops in many countries, and it is a primary food source for more than three billion people worldwide. The global rice consumption is projected to be 873 million tonnes in 2030. Based on national socio-economic survey (susenas) data from Indonesia Statistic Agency (BPS), Indonesia's population of rice consumption 2016 reached 78,816 kg per capita per year. In the recent decades, two major issues like population growth (in particular in the major rice producing or consuming countries) and climate change put enormous pressure on the global food demand and its production. Almost all countries made the issue of food as a major issue. Now, Indonesian government wants to realize Indonesia became self-sufficient of rice in 2017. Some food security programs implemented to achieve it. One of them is predicting or estimation rice production and consumption. Prediction or estimation rice productivity before harvest is crucial, especially in regions characterised by climatic uncertainties. It is also enables governments to put in a good strategic contingency plans for the redistribution of food during times of famine. Satellite remote sensing has been widely applied and recognised as a powerful and effective tool for identifying agriculture crops. An important goal of agricultural remote sensing research is to spectrally estimate crop variables related to crop conditions, which can subsequently be entered into crop simulation and production models. In this research will be use a global and medium resolution of satellite remote sensing imagery (Landsat 8 ETM+) with 2016-2017 acquisition year in order to estimation of rice productivity. The rice productivity prediction modeling methods will be build in this research based on regression and spectral pattern analysis of NDVI (and LSWI (Land Surface Water Index in multi temporal satellite imagery. In addition, this prediction model will be tested against fact the level of productivity in the rice field. We will take a proportional field sampling test using cluster sampling test and random spectral value checking in order to measure accuracy level of rice productivity prediction model. The results of this research indicate that NDVI and LSWI algorithm is the best combination of linear regression in estimation of rice productivity level with coefficient of determination equal to 0.639. Validation of regression model equation to data of Agriculture Agency of Demak has difference or RMSE 8,394 Quintal/ Ha

1. Introduction

Rice is one of the most important agriculture crops in many countries, and it is a primary food source for more than three billion people worldwide [1]. The global rice consumption is projected to be 873
million tonnes in 2030[2]. In the recent decades, two major issues like population growth (in particular in the major rice producing/consuming countries) [3] and climate change put enormous pressure on the global food demand and its production [4]. Almost all countries made the issue of food as a major issue. Various food security programs promoted in order to maintain food sovereignty.

Indonesia is one of producer and consumer of rice. Based on national socio-economic survey (SUSENAS) data from Indonesia Statistic Agency (BPS), Indonesia’s population of rice consumption 2016 reached 78,816 kg per capita per year [5]. It is assessed by the Ministry of Agriculture is fairly high consumption. In fact, more than world rice consumption is only 60 kg per capita per year. In Asia, Indonesia's rice consumption is the highest. Indonesian rice production to meet domestic demand for rice in the range of 1980 to 2008. After that the needs of Indonesian rice deficit forced the government to import rice from abroad. Now, Indonesian government wants to realize Indonesia became self-sufficient of rice in 2017. Some food security programs implemented to achieve it. One of them is predicting or estimation rice production and consumption.

Prediction or estimation rice productivity before harvest is crucial, especially in regions characterised by climatic uncertainties. Forecasting enables planners and decision makers to predict how much to import in the case of a shortfall or, optionally, to export in the case of a surplus. Prediction and estimation also enables governments to put in place strategic contingency plans for the redistribution of food during times of famine. Therefore, the monitoring of crop development, crop growth and early productivity prediction are generally of great importance [6]. Indonesian territory is large enough. This makes an early estimate of rice production is very important.

Most common and widely used methods for estimating rice production are the use of agricultural statistical data acquired through field visits and interviewing the farmers. The methodology for mapping area under rice cultivation is basically done through annual/seasonal sample surveys based on a number of sample clusters that are constituted all over the country for measuring cultivated area during the crop growing season. Each cluster is visited many times and areas are recorded by the field staffs, checked, and then processed by regional statistical officers. Despite its invaluable ability for understanding historical trends in rice area, this method is extremely tedious, time-consuming, less precise, costly, inconsistent, too generalized and labour-intensive [7]. In most countries, the data become available too late for the appropriate actions to be taken to avert food shortage [6]. In addition, this method unable to provide information in timely manner as well as unable to delineate detailed geographical spatial distribution of areas under rice cultivation.

Satellite remote sensing has been widely applied and is recognised as a powerful and effective tool for identifying agriculture crops [8]. An important goal of agricultural remote sensing research is to spectrally estimate crop variables related to crop conditions, which can subsequently be entered into crop simulation and production models. Some research use global and medium resolution of satellite remote sensing imagery in estimation of rice productivity, such as MODIS, NOAA and Landsat ETM+. Nuarsa (2012) [9] use of remote sensing in rice productivity in Tabanan, Bali. In this research. This research develops estimation of rice productivity in the next generation of Landsat satellite imagery that Landsat-8 satellite imagery. Beside that, the modelling methods will be used in this research based on regression and spectral pattern analysis of NDVI and LSWI multi temporal

Based on the above introductory description as for the formulation of the problem of this research are:

1. How the right model for rice productivity estimation before harvesting through Landsat-8 satellite imagery?
2. How the productivity of rice in study area (Demak Regency) at 2017?

2. Methodology

2.1. Description of Study Area

The study was conducted in Demak of the Province of Central Java, Indonesia. Centred at latitude 6° 43’ 26"–7° 09’ 43" and longitude 110° 27’ 58"–110° 48’ 47" and is located approximately 25 km east of Semarang City (Figure 1). The longest distance from West to East is along 49 km and from North to South along the 41 km [5].
The Demak Regency was selected as the study area because Demak is the one of central area for rice production in Central Java. Based on BPS (Indonesian Central Statistics Agency) data 2016, rice production of Demak Regency in 2015 reached 653547 tons. It was 3rd rice production in Central Java after Sukoharjo Regency and Cilacap Regency. Beside that, Demak is located near from Semarang City. This makes Demak regency as food stock or food buffer zone for Semarang City [5]. Demak has twice the rice planting season, the rainy growing season and growing season during the dry season. Rice planting patterns simultaneously both for wetland paddy nor dry land. In Demak Almost 95 percent of paddy field are wetland paddy [10].

The harvested area of paddy of Demak in 2015 reached 98.618 hectares. If compared with previous year, it increased 2.01 percent. In 2015, paddy production was 653547 tons dried paddy. That increased 15.11 percent, compared with previous year. In 2015, the productivity of paddy was about 66.27 quintal per hectare, it increased 12.75 percent compared with 2014 as 58.73 quintal per hectare [10].

2.2. Data Collection

Data collection includes spatial data collection and data collection non spatial. Spatial data includes Landsat 8 satellite image data (obtained via download from glovis.usgs.gov), Administration Map of Demak (obtained from Development Planning Agency of Demak/BAPPEDA), Indonesian Topographical map (RBI) of Demak with scale 1:25.000 (obtained from Indonesian Spatial Agency/BIG). Non spatial data include rice productivity data (obtained from the Department of Agriculture), Biophysical data parameters (growth stage) obtained from rice planting calendar.

2.3. Method

Method of this study was followed the flowchart in Figure 2. It was divided in to four step, that were image preprocessing step, image processing and model development step, statistical testing and validation step, and model application step.

The principle method of this research was prediction of rice productivity using pattern value of NDVI, EVI and LSWI of rice plant before harvest in vegetative and generative growth stage. Because both in vegetative and generative growth stage were important phase of rice to growth. In this study, three images Landsat-8 multi temporal, i.e. the image on May 29, 2015 as time 1, the image on June 14, 2015, as time 2 and the image on June 30, 2015 as time 3 were used to make a Model rice productivity. The models acquired using linear regression between Index (NDVI, EVI, LSWI) multi temporal (time 1, time2 and time 3) and rice productivity. Then this model were tested using statistical testing to choose the best model. It was validated using independent rice productivity, with the requirement of RMSE is 10 quintal/ha. After that the model was applied to predict of rice productivity on rice planting II, 2017.
2.3.1 Preprocessing Image Method

a) Geometric Correction

Geometric correction geo-referenced image to real world coordinates. The geometric correction method of the image used image to map rectification with reference Indonesian Topographical map (RBI) of Demak, scale 1:25,000. Ground control point (GCP) on this research taken as many as 16 points of each image. Root mean square error (RMSE) in this process were required maximum 1 pixel.
b) Radiometric correction
This stage corrected image using the conversion of digital number to Top of Atmosphere (TOA) Reflectance. OLI band data of Landsat 8 can be converted to TOA planetary reflectance using reflectance rescaling coefficients provided in the product metadata file (MTL file). MTL is the format of metadata file for Landsat 8 imagery. The following equation is used to convert digital number (DN) or pixel values to TOA reflectance for operational land imager (OLI) data as follows [11]:

\[ \rho_{\lambda'} = M_{\rho} Q_{cal} + A_{\rho} \]

TOA reflectance with a correction for the sun angle is then [6]:

\[ \rho_{\lambda} = \frac{\rho_{\lambda'}}{\cos(\theta_{SZ})} = \frac{\rho_{\lambda'}}{\sin(\theta_{SE})} \]

where:
- \( \rho_{\lambda'} \) = TOA planetary reflectance, without correction for solar angle. Note that \( \rho_{\lambda'} \) does not contain a correction for the sun angle
- \( M_{\rho} \) = Band-specific multiplicative rescaling factor from the metadata (REFLECTANCE_MULT_BAND_x, where x is the band number)
- \( A_{\rho} \) = Band-specific additive rescaling factor from the metadata (REFLECTANCE_ADD_BAND_x, where x is the band number)
- \( Q_{cal} \) = Quantized and calibrated standard product pixel values (DN)
- \( \rho_{\lambda} \) = TOA planetary reflectance
- \( \theta_{SE} \) = Local sun elevation angle. The scene center sun elevation angle in degrees is provided in the metadata (SUN_ELEVATION)
- \( \theta_{SZ} \) = Local solar zenith angle; \( \theta_{SZ} = 90^\circ - \theta_{SE} \)

c) Highlight cloud
Clouds and shadows in the image is a problem that can’t be avoided. This is an obstacle in the use of medium resolution imagery such as Landsat 8, especially in equatorial regions. This causes the loss of some of the information contained in the images. Highlight Cloud is used to discriminate or distinguish clouds with land cover [12]. Because of the clouds in the application or interpretation of land use is considered as interference to other land cover reflectance values, so that should be eliminated in the analysis. In this study, the clouds separated by digitization after sharpening cloud value.

d) Image Cropping
Cropping images do reduce the file size of the image so that the processing of data to be lighter and faster. It are made in accordance with the required area or region of interest (ROI). Moreover, in this study uses data multi-temporal for classification process, so cropping of the image is needed in order to accelerate the image processing [13].

2.3.2 Image Processing and Model Development Method. In this research, the value of reflectance TOA are transformed to other spectral value using transformation spectral. Transformation method by using Normalized Difference Vegetation Index (NDVI) Enhanced Vegetation Index (EVI) and Land Surface Water Index (LSWI) will be used in each image. The results of image transformation at several different times are combined for a single image with three band (NDVI time1, NDVI time 2, NDVI time 3) and second image (LSWI time 1, LSWI time 2, LSWI time 3). The models acquired using linear regression between Index (NDVI, EVI, LSWI) multi temporal (time 1, time2 and time 3) and rice productivity in 50 point. The best model combination was chosen based on the highest value of coefficient determination in several models.
2.3.3 Statistical Testing and Validation Method. Several statistical testing were used to test the model. The statistical testing used were Classical assumption test (normality, multicollinearity, heteroscedasticity and autocorrelation) and F test. Beside that, models that have been formed need to be validated to determine the quality of the model. Validation the model using 15 point of rice productivity independent data. So it is known that the difference between the estimation results with the data of Agriculture Department of Demak Regency and RMSE.

3. Result and Discussion
3.1 Development Model of Rice Productivity Estimation
The model development of the relationship between independent variables (EVI, NDVI LSWI multi temporal before harvest) and the dependent variable of rice production using multiple linear regression method. Number after index vegetation showed time 1, time 2 and time 3 (example :NDVI 1 as NDVI time 1 ). In the regression process is also done a classical assumption tested to ensure the right statistical rules. After that the model was chosen by observing the highest coefficient of determination. This coefficient shows how strong the model is able to represent the bound variables of the independent variable. The model feasibility tested was used as a test tool to ensure that this model is truly feasible for estimating rice production using multi-temporal landsat-8 satellite image data.

Based on the relationship between rice productivity variables with NDVI transformation value, EVI and LSWI at multitemporal 3 time before harvest period can be made linear regression model as follows in Table 1.

| No | Parameters | Equation of Regression |
|----|------------|------------------------|
| 1  | NDVI (NDVI 1, NDVI 2, NDVI 3) | $Y = 99,030 - 236,143X_1 + 37,624X_2 - 56,722X_3$ |
| 2  | EVI (EVI 1, EVI 2, EVI 3) | $Y = 80,253 - 22,57X_1 + 22,684X_2 - 19,463X_3$ |
| 3  | LSWI (LSWI 1, LSWI 2, LSWI 3) | $Y = 68,461 + 8,133X_1 - 23,064X_2 + 28,567X_3$ |
| 4  | NDVI and LSWI (NDVI 1, NDVI 2, NDVI 3, LSWI 1, LSWI 2, LSWI 3) | $Y = 101,027 - 9,185X_1 + 16,334X_2 - 72,067X_3 - 38,408X_4 - 54,432X_5 + 45,716X_6$ |
| 5  | EVI and LSWI (EVI 1, EVI 2, EVI 3, LSWI 1, LSWI 2, LSWI 3) | $Y = 93,929 - 20,114X_1 + 16,867X_2 - 26,437X_3 + 1,135X_4 - 44,631X_5 + 24,95X_6$ |

Based on the regression result can be seen that the highest coefficient of determination resulted from the linear regression between rice productivity as the dependent variable and the value of NDVI and LSWI as independent variables. The coefficient of determination approach 0.639. This model has also tested the classical assumptions and required all assumptions. Therefore, in this research will be analyzed further is the relationship between rice productivity variables with the value of NDVI and LSWI. Based on regression, the equation model are:

$$Y = 101,027 - 9,185X_1 + 16,334X_2 - 72,067X_3 + 38,408X_4 - 54,432X_5 + 45,716X_6 \quad (3)$$

Where :

- $Y$ : The rice productivity (quintal/ha)
- $X_1$ : The value NDVI time 1
- $X_2$ : The value NDVI time 2
- $X_3$ : The value NDVI time 3
- $X_4$ : The value LSWI time 1
- $X_5$ : The value LSWI time 2
- $X_6$ : The value LSWI time 3
3.2 Statistical Testing Result

3.2.1 Classical Assumption Test Result. The classical assumption test is a test of statistical assumptions that must be required in multiple linear least squares regression (MLS). A classical assumption test is needed to determine whether the linear regression estimation results is done from the error. In this study the classical assumption test performed is the normality test, multicolinearity test, autocorrelation test and heteroscedasticity test. The linear regression estimation model must satisfy the test so that the resulting equation is of good quality. The result of classical assumption test can be seen in Table 2.

| Classical Assumption Test | Tested Method | Value | Requirement | Conclusion |
|---------------------------|---------------|-------|-------------|------------|
| Normality                | Komolgorov-Smirnov Method | Sig value parameters in range 0.087-0.2 | sig value > 0.05 | Required |
| Heteroscedasticity       | Glejer method | Sig value parameters in range 0.099-0.857 | sig value > 0.05 | Required |
| Multicolinearity         | Variance Inflection Factor (VIF) Method | VIF value parameters in range 1.692-4.236 | VIF value < 10 | Required |
| Autocorrelation          | Durbin Watson method | Durbin Watson value 2.152 | Durbin-Watson count is greater than 1.784 and smaller than 2.216 | Required |

3.2.2 F-Test Result. The model simultaneous test or model feasibility test or more popularly referred to as F test is a step to identify the estimated regression model is feasible or not. Worthy (reliable) here means a model that is feasible to be used to explain the effect of independent variables on the dependent variable. The test name is referred to as the F test, as it follows the following F distribution whose testing criteria are like One Way Anova.

| Model | Sum of Squares | Df | Mean Square | F    | Sig. |
|-------|---------------|----|-------------|------|------|
| 1     | Regression    | 1186,129 | 6 | 197,688   | 15,457 | .000b |
|       | Residual      | 549,961 | 43 | 12,790   |       |      |
|       | Total         | 1736,090 | 49 |          |      |      |

Taking conclusions on the F test is done by observing the sig value of the regression model made. In table 3, it can be seen that the model sig value is 0.000. It shows that the value of sig model meets the first requirement of sig <0.05. So Ho is rejected and Ha accepted. Therefore it can be concluded that the resulting linear regression model is feasible to be used to predict the dependent variable.

3.3 Validation

Multiple linear regression models generated to predict rice productivity need to be validated to find out how the quality of the model when applied. To perform such validation by entering the value of free variables are NDVI and LSWI on the model. Furthermore, based on the results of the calculation can be known the difference between the results of calculations with data from the Department of Agriculture Demak Regency. And can be known RMSE (Root Mean Square Error) from the results of the calculation data. The result calculation in 15 sample data there is total difference of rice productivity equal to 45,742 quintal / Ha between data from Agricultural Service and result of model calculation. In the 15 samples used it has an RMSE value of 8.394 quintal / Ha. RMSE value 8,394 quintal / Ha is still categorized as a medium error. The error rate is between 10-15%. This could be due to a sample in the form of secondary data that have no coordinates so that the location is not accurate. Therefore, further research is needed which can take samples of rice productivity directly in the field. The sampling must have the exact coordinates.
3.4 Application Model

The model was reliable to estimate the productivity because after compared with the rice productivity in 2016 from Department of Agriculture [14], it has different 10% error. The error can be caused from the sample from the secondary data statistic. The research [15] can result the best model until minimal error 4.2% with the same parameter (LSWI and EVI). It used MODIS data. [15] used rice productivity from in situ or primary data. Although this research has not yet reached the best model but with a 10% error can still be tolerated for initial estimation of rice productivity.

The resulting model is applied to the image in 2017. The objective of the implementation is to know the estimation of rice productivity in Demak Regency in 2017, planting season II (May-August). The productivity estimate will then be explained and visualized to obtain the distribution of the productivity value and the extent of each class.

Estimation of rice productivity in 2017 is done by entering the value of NDVI and LSWI (multi temporal three different time) in 2017 on the equation of multiple linear regression model that has been produced. The values of NDVI and LSWI are applied to the equations of multiple linear regression models that have been made. The result of model implementation resulted that the smallest rice productivity estimation in 2017 is 28,246 quintal/ha. While for estimation of the biggest rice productivity is 155,461 quintal/ha. The average on the estimate of rice productivity in 2017 amounted to 78,904 quintal/ha. Based on the result, a class can be used to determine the level of rice productivity. In this study, 3 classes were created to differentiate the level of rice productivity in Demak Regency by 2017. The calculation of the range for each class can be done by reducing the largest and smallest value in a data and dividing it by the desired number of classes. Range of rice productivity level in 2017 can be seen in table 4.

| Level of Rice Productivity | Range level productivity (Quintal/Ha) |
|---------------------------|--------------------------------------|
| Low Productivity          | 28,246 - 70,651                      |
| Medium Productivity       | 70,652 - 113,056                     |
| High Productivity         | 113,057 - 155,461                    |

Figure 3. Spatial Distribution of Rice Productivity Level in 2017
Based on Table 4, the distribution of productivity levels in 2017 can be visualized as shown in Figure 3. Based on figure 3 can be known the area of each class. In 2017, Demak District has a total area of rice fields of 47632.847 hectares of rice productivity estimates. The area consists of low rice productivity level of 8255,778 Ha, the level of medium rice productivity is 39010,413 Ha and high rice productivity level of 366,655 Ha.

4. Conclusion
This research resulted model of rice productivity estimation before harvesting using Landsat-8 with parameter NDVI and LSWI. The model was reliable to estimate the productivity of rice with determinant coefficient of 0.639 and average error 10%. So from the model that is able to show that the value of NDVI and LSWI affect the value of rice productivity as much as 63.9% while the rest of 36.1% is caused by other factors. Other factors could be the existence of pest and fertilizer application which is not discussed in this research. This model can be applied to predict the productivity of rice in Demak at 2017. The model could show that Demak Regency has a total area of rice fields of 47632.847 hectares of rice. The area consists of low rice productivity level of 8255,778 Ha, the level of medium rice productivity is 39010,413 Ha and high rice productivity level of 366,655 Ha.

References
[1] Khush G S 2005 What it will take to feed 5 billion rice consumers in 2030. International Journal of Molecular Biology, Vol 59, 1–6.
[2] Purevdorj M and Kubo M 2005 The future of rice production, consumption and seaborne trade: Synthetic prediction method. Journal of Food Distribution Research, 36 (1): 250-259.
[3] GeoHive—Historic, Current and Future Population: Asia. Available online: http://www.geohive.com/earth/his_proj_asia.aspx. (Accessed on 4 April 2017).
[4] Fifth Assessment Synthesis Report, IPCC: Geneva, Switzerland, 2014; p. 116. Available online: http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_LONGERREPORT.pdf. (Accessed on 4 April 2017).
[5] BPS (Center Statistic of Indonesia) 2014 Demak in Figure 2014. Demak: BPS.
[6] Sawasawa Haig L S 2003 Crop Yield Estimation: Integrating RS, GIS, and Management Factor. A case study of Birkooor and Kortigiri Mandals, Nizamabad District India. Thesis of Master Degree in ITC Enchede: Netherland.
[7] Prasad A K, Chai L, Singh R P and Kafatos M 2006 Crop yield estimation model for Iowa using remote sensing and surface parameters. International Journal of Applied earth observation and geoinformation, Vol 8: 26-33.
[8] Mekong Delta Using ENVISAT/ASAR Dual Polarization Data. IEEE Trans. Geoscience and Remote Sensing, 47, pp. 517 –526.
[9] Nuarsa I W, Nishio F and Hongo C 2012 Rice Yield Estimation Use Landsat ETM+ and Field Observation. Journal of Agriculture Science, Vol 4, No 3, 45-56.
[10] Department of Agriculture 2015 Conditions of Land Agriculture in Demak Th. 2014-2015. Demak: Department of Agriculture.
[11] USGS 2013 Using the USGS Landsat 8 Product. http://Landsat.usgs.gov. 10 April 2015.
[12] Putra E H 2011 Remote Sensing Using ERMapper. Yogyakarta: Graha Ilmu.
[13] Sukmono Abdi and Ardiansyah 2017 Identification of rice field using Multi-Temporal NDVI and PCA method on Landsat 8 (Case Study: Demak, Central Java). IOP Earth and Environmental Science Vol 54 paper 012001.
[14] Department of Agriculture 2017 Rice Productivity in 2016. Agricultural Agency: Demak.
[15] Wijesingha J S J, N L Deshapriya, L Samarakoon 2015 Rice Crop Monitoring And Yield Assessment With Modis 250m Gridded Vegetation Product : A Case Study in Sa Kaeo Province Thailand. The International Archives of The Photogrammetry, Remote Sensing and Spatial Information Science, Volume XL-7/W3.
Acknowledgments

We extend our thanks to remote sensing and photogrammetry laboratory Diponegoro University for all their help in data processing equipment and thanks to Agricultural Agency of Demak Regency for all their help in supporting data.