

An Approach for Improvement of Goodness of Fit on the Estimation of Paddy Rice Yield Using Satellite(MODIS) Images

Bae-Sung Kim¹, Jae-Hwan Kim² and Seong-Bo Ko^{1*}

¹Dept. of Industrial and Applied Economics in JeJu National University,
Research Institute for Subtropical Agriculture and Animal Biotechnology

²World Institute of Kimchi

MODIS 영상을 이용한 논벼 생산량 추정모형의 적합도 개선을 위한 연구

김배성¹, 김재환², 고성보^{1*}

¹제주대학교 산업응용경제학과 · 아열대농업생명과학연구소

²세계김치연구

Abstract This research was performed in order to improve the goodness of fit of paddy rice production forecasting using MODIS images and to find out appropriate explanatory variables in the forecasting model. The aim of this paper is to review the use of satellite images for the survey of paddy rice production in Korea. Many developed countries, including the United States, Australia, and Japan, have been using satellite images to produce agricultural statistics such as crop production, cultivated acreage, etc. The survey accuracy of crop production by using satellite images, however, is not satisfied in practical use. In this paper, we reviewed several methods to increase the survey accuracy of rice production statistics, gained from satellite images. Rice was selected for this study because its cultivated area and production amount could be more easily identified than other crops by using satellite images. The MODIS images were used because they involved more appropriate images to estimate and analyze rice production. This study estimated yield functions by using the NDVIs, gained from paddy rice yields and annual average isothermal lines, and the meteorological variables such as sunshine hours, rainfall, and temperature during ripening stage. As a result of yield function estimation, the goodness of fit(R-squared) for the models was shown from 0.768 to 0.891. In this study, it is noteworthy academically and practically that vegetation index(NDVIs) identified by annual average isothermal lines and meteorological variables are very useful for estimating yield functions.

요 약 본 논문은 MODIS 위성 영상을 이용하여 논벼 생산량을 추정하는 모형의 적합도 개선 및 추정모형내 적절한 설명변수를 탐색하고자 수행되었다. 또한 이 연구는 한국에서 논벼 생산량 조사를 위해 위성 영상을 사용하는 방안을 검토하기 위해 수행되었다. 미국, 호주, 일본 등 많은 선진국들은 재배면적 및 생산량 조사와 같은 농업 통계를 산출하기 위해 위성 영상을 이용하고 있다. 그러나 위성 영상을 이용한 작물 생산량 조사의 정확성은 아직 충분치 않은 수준이다. 본 연구는 위성 영상을 이용한 논벼 생산량 조사의 정확도를 증대시키기 위한 몇 가지 방법을 검토하고 있다. 많은 작물 중 논벼를 연구대상으로 선정한 이유는 논벼가 다른 작물 보다 재배면적과 작황의 영상 분석이 용이하였기 때문이고, 다양한 위성 영상 중 MODIS 영상을 이용한 것은 한국 논벼 생산량 조사 연구를 위해 보다 적절한 영상을 다수 포함하고 있었기 때문이다. 이 연구에서 등온선에 의해 구분된 논벼로부터 도출된 NDVI지수, 논벼 등숙기의 일조시간, 강우량, 온도 등 기상변수를 이용하여 단수함수가 추정되었다. 단수함수 추정결과, 모형의 적합도(R-squared)는 0.768-0.891를 보였다. 이 연구는 연평균 등온선에 의해 구분된 NDVI지수와(등숙기) 기상변수가 단수함수 추정에 매우 유용하게 이용될 수 있음을 보이고 있다.

Key Words : Satellite Images, Rice Yield, Isothermal line, NDVI, MODIS, Meteorological variables

*Corresponding Author : Seong-Bo Ko(JeJu National Univ.)

Tel: +82-64-753-3351 email: ksb5262@jejunu.ac.kr

Received July 9, 2013

Revised (1st August 9, 2013, 2nd August 28, 2013)

Accepted November 7, 2013

1. Introduction

Developed countries, such as the United States, Japan, etc, have been using satellite images for investigating crop cultivated area and crop production because a wide range of area can be easily analyzed within a short time.

In Korea, satellite images have been applied to various fields since high resolution satellite images, obtained from Kompsat-2 and Kompsat-3(launched in 2006 and in 2012, respectively), were provided at affordable prices. So, demonstration researches using the images have been recently conducted in the field of agriculture in Korea.

The first attempt, using remote sensing techniques in agricultural field in Korea, was performed by Rural Development Administration in the middle of 1990. Korea National Statistical Office has been implementing a demonstration project to introduce the remote sensing techniques into agricultural statistics and setting up execution plans for mid-long terms. In fisheries industry from 2004, under the leading of Korea Maritime Institute and Fisheries Outlook Center, satellite images have been used for recognizing the quantity of aquaculture farming facilities and for estimating the aquaculture products. The feasibility review for the introduction of remote sensing techniques into agricultural field has been studied by Korea Rural Economic Institute from 2005 to 2010[7].

In Korea, studies about the estimation of cultivated areas or production areas by using satellite images have been fulfilled[2,4,6,7]. However, it seemed the investigation of crop production was performed only by Hong et al.[3,5]. Hong et al.[3] performed the regression analysis with the NDVIs, obtained from MODIS data about North Korea between 2000 and 2008, and the quantitative data, presented by Korea National Statistical Office; as a result, they showed that the R-squared, meaning the goodness of fit of estimated rice production in North Korea, was 0.591. Also, in 2012, they performed the regression analysis with the NDVIs, obtained from MODIS satellite images of paddy rice area in South Korea between 2002 and 2010, and the meteorological data such as sunshine hours, solar radiation, and rainfall during ripening periods; as a result, they showed that the R-squared was between 0.70 and 0.82[5].

This study suggests more effective methods for increasing the goodness of fit of paddy rice production

estimation by using MODIS images in Korea from 2008 to 2009.

This study was revised and supplemented from the excerpt written by B.S. Kim and J.W. Kim in KREI research report[1].

2. MODIS Images Collection and NDVIs Calculation

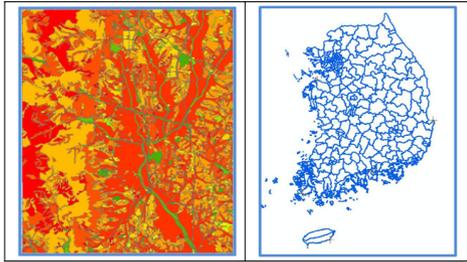
The purpose of this study is to calculate NDVIs of paddy rice by using MODIS images and to develop a forecasting model of paddy rice yield through the NDVIs. The MODIS (Moderate-resolution Imaging Spectroradiometer) is a payload scientific instrument launched into Earth orbit by NASA in 1999 on board the Terra (EOS AM) Satellite, and in 2002 on board the Aqua (EOS PM) satellite, and the NDVI(Normalized Difference Vegetation Index) is a graphical indicator that can be used to analyze remote sensing measurements and assess whether the target being observed contains live green vegetation or not(Wikipedia). So, MODIS images during August and September, when the NDVIs of paddy rice are maximized, were searched and collected. Collecting and analyzing MODIS images and calculating NDVIs were performed by Sea & River Consultant INC., specialized in satellite image processing.

A total of ten legible MODIS images were finally selected for the period from 2008 to 2010, and their cloud volumes were below 10 percent. This study used three images in 2008, five in 2009, and two in 2010.

Also, the land use data, provided by the Ministry of Environment, were used for calculating the cultivated area that is essential to extraction of vegetation index of paddy rice. The production data of paddy rice, provided by Korea National Statistical Office, were employed for estimating yield functions and for analyzing correlations between NDVIs and paddy rice yields. Administrative boundary data were utilized for correcting locations on MODIS images and for classifying vegetation index by administrative districts.

The collected MODIS images are geometrically distorted due to filming location and origination; so, relevant correction on the images is necessary. For

correcting location problems, layers of administrative district boundaries were summed as one layer and then applied. The location correction of MODIS images was performed through the processes such as satellite image input, GCP collection, and reprocessing.

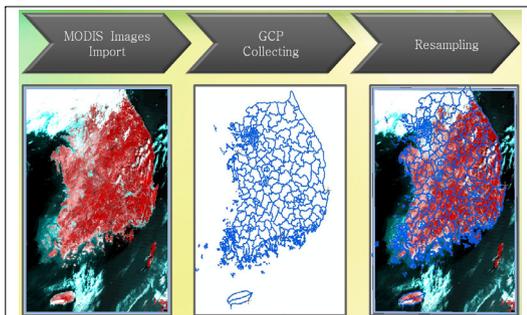


[Fig. 1] Land Use(left) and the Boundary of Province(right)

Indices about plants growth are necessary for the estimation of paddy rice production by using MODIS images. This study used the NDVIs (Normalized Difference Vegetation Index), widely used for recognizing plants growth through satellite images. The range of NDVI is between 0 and 1, and the greater in number the better in growth.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

The resulted range of NDVI calculation through MODIS images were from -1 to 1, but only the range of between 0 and 1 was used because the objective plant was paddy rice in this study.



[Fig. 2] Data Preprocessing of MODIS Images

In this study, NDVIs of paddy rice area for administrative boundaries were calculated, and then the

average NDVI for cities and counties was used. This study did not include any vegetation index which was not significant due to cloud volume and so on; so, the satellite images only from 2008 to 2009 were used.

3. Correlation Analysis between Paddy Rice Yield and NDVI

Prior to specification of a paddy rice yield model, the correlation analysis between paddy rice yields and NDVIs was performed. This was fulfilled to develop a yield function with better goodness of fit and higher forecasting ability.

The correlation coefficients between paddy rice yields and NDVIs, obtained from satellite images from May to September in 2008 and 2009, were measured. The following cities were not included in the calculation because their paddy rice area was relatively less than others. They were Seoul Metropolitan City, Busan Metropolitan city, Daegu Metropolitan City, Incheon Metropolitan City, Gwangju Metropolitan City, Daejeon Metropolitan City, Ulsan Metropolitan City, and Jeju Special Self-Governing Province.

As shown in the Table 1, the range of correlation coefficients between paddy rice yields and nationwide NDVIs are from 0.04 to 0.39 or below 0.

[Table 1] Correlation Coefficient bet. Paddy Rice Yild and NDVI of the Whole Country

date	'08.5.3	'08.8.7	'08.9.7	'09.5.7
Coefficient	-0.0113	0.1527	-0.0716	0.0363
date	'09.8.15	'09.9. 6 /	'09.9.10	'09.10.3
Coefficient	0.3874	0.3928		-0.0090

The NDVIs for whole country were not available for estimating a yield model because its correlation coefficients were very low; so, we alternatively reviewed the correlation analysis between yields and NDVIs for the 8 provinces as follows: Gyeonggi Province, Gangwon Province, Chungcheongbuk Province, Chungcheongnam Province, Jeollabuk Province, Jeollanam Province, Gyeongsangbuk Province, and Gyeongsangnam Province. Then, the NDVI for each province was revised, and the correlation between a local paddy rice yield and a

corresponding NDVI was reviewed.

As indicated in the Table 2, the resulted coefficients among provinces were higher than the nationwide but still low.

The regions with higher correlation coefficients appeared on isothermal lines; as a result, the correlation coefficients in this study were reviewed by calculating NDVIs for each isothermal line.

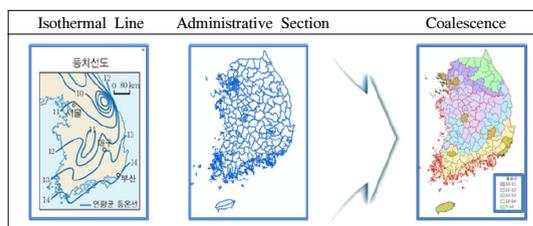
[Table 2] Correlation Coefficient bet. Paddy Rice Yield and NDVI by Province

date	'08.5.3	'08.8.7	'08.9.7	'09.5.7	'09.8.15	'09.9.6 '09.9.10	'09.10.3
Gyeonggi	-0.418	0.100	-0.121	0.226	0.325	-0.078	-0.212
Gangwon	0.411	0.030	0.085	-0.370	0.103	0.450	-0.270
Chungbuk	-0.482	-0.303	-0.285	-0.406	-0.550	-0.282	-0.446
Chungnam	-0.379	-0.122	0.273	-0.524	0.415	-0.345	-0.461
Chonbuk	0.034	0.217	0.313	-0.177	0.530	0.096	-0.424
Chonnam	0.225	0.176	0.153	0.432	0.613	0.502	0.465
Gyeongbuk	0.053	0.060	-0.287	-0.070	-0.050	0.048	-0.150
Gyeongnam	0.196	0.334	-0.239	-0.047	0.077	0.040	-0.271

So, the NDVIs for corresponding regions were recalculated by combining the annual average of isothermal lines and the administrative district boundaries.

Paddy rice area in South Korea was divided in 5 groups by annual average isothermal lines. The group 1 is characterized by annual average temperature between 13 and 14 degrees Celsius, and it consists of 38 cities and counties in Jeollanam Province, Gyeongsangnam Province, and Gyeongsangbuk Province. The group 2 is characterized by annual average temperature between 12 and 13 degrees Celsius, and it consists of 35 cities and counties in Gangwon Province, Chungcheongnam Province, Jeollabuk Province, Jeollanam Province, and Gyeongsangbuk Province.

The group 3 is characterized by annual average temperature between 11 and 12 degrees Celsius, and it consists of 50 cities and counties in Gyeonggi Province, Chungcheongbuk Province, Chungcheongnam Province, Jeollabuk Province, and Gyeongsangbuk Province. The group 4 is characterized by annual average temperature between 10 and 11 degrees Celsius, and it consists of 11 cities and counties, and it consists of Gyeonggi Province and Gangwon Province.



[Fig. 3] Process of using annual average of isothermal lines

The group 5 is characterized by annual average temperature below 10 degrees Celsius and consists of 11 cities and counties in Gangwon Province.

[Table 3] Correlation Coefficient bet. Paddy Rice Yield and NDVI by Isothermal Line(Annual Average)

	'08.8.7	'08.9.7	'09.8.15	'09.9.6 '09.9.10	'09.9.10
Group 1	0.814	0.778	0.755	0.709	0.732
Group 2	-0.764	-0.840	0.819	-0.842	0.787
Group 3	0.728	0.798	0.849	0.843	0.761
Group 4	0.764	-0.752	0.681	-0.838	0.929
Group 5	0.821	0.745	0.817	0.893	0.776

As mentioned in the Table 3, the correlation coefficients between paddy rice yields and NDVIs, based on annual average isothermal lines, were shown higher than the results in other methods. The negative coefficients could be turned out due to the size difference between actual area and sampling area for calculating vegetation indices.

4. Development to Paddy Rice Yield Forecasting Model

Based on the examined results so far, paddy rice area was divided into 5 groups by annual average isothermal lines, and the yield forecasting model for paddy rice was developed as follows.

$$Y = f(NDVI, TEMP, SUNH, RAIN, CLOUD) \quad (2)$$

Here, Y is the paddy rice yield data, presented by Korea National Statistical Office, and NDVI is the average of cities and counties from MODIS images. SUNH is the cumulative sunshine hours, which are the

sum of sunshine hours during ripening stage. RAIN is the amount of cumulative rainfall, meaning the sum of rainfall during ripening stage. CLOUD is the cloud volume of MODIS filming zones.

This study applied the data in 2009 to the yield function, which was the equation 2, and applied to group 1, 2, and 3. The group 4 and 5 were not included due to the small number of observations.

[Table 4] Estimation Results of Paddy Rice Functions by Group

date	Sept. 6	Sept. 10	Sept. 6
Group	1	2	3
NDVI	0.362***	0.286***	0.462***
TEMP	0.003	-	-
SUNH	-	-	0.336***
RAIN	-0.008**	-	-0.007**
CLOUD	-	-	-
Intercept	6.353***	6.411***	4.695***
R-squared	0.891	0.873	0.768
Functional form	log-log	log-log	log-log

NOTE: ***, **, and * are 1%, 5%, and 10% level of significance, respectively.

As shown in the Table 4, the goodness of fit for the yield model, ranged from 0.768 to 0.891, was relatively high. Hong et al. (2012) reported that the R-squared among paddy rice yields, NDVIs, and sunshine hours, was 0.70 and that the R-squared among paddy rice yields, NDVIs, and rainfalls, was 0.82. In their researches, the reason for the high R-squared values might be because they calculated regional NDVIs on the basis of isothermal lines, and because they used a log-log functional form.

As mentioned in the estimation results, NDVIs for all models were significant under 1% level of significance; thus, the use of satellite images is useful for estimating yield functions. Also, the meteorological variables such as sunshine hours and rainfall during the paddy rice ripening stage were significant under 5% and 10% levels of significance, respectively. In our research, it was resulted that the introduction of meteorological variables could improve the goodness of fit for models. Therefore, in order to estimate more appropriate yield models, more accurate vegetation indices and meteorological variables should be made up.

5. Conclusion

This study was performed in order to improve the goodness of fit and predictability for a paddy rice yield model and to find appropriate explanatory variables.

In this research, NDVIs were calculated from MODIS images, and the images were provided by the Satellites, Terra and Aqua, during 2008 and 2010. Also, the meteorological variables, such as sunshine hours, rainfall, and temperature during ripening stage, were collected and applied.

According to the correlation analysis between paddy rice yields and NDVIs, the resulted coefficients for NDVIs on the basis of annual average isothermal lines were shown higher than the results for NDVIs based on nationwide scale or provinces.

This study estimated yield functions by using the NDVIs, gained from paddy rice yields and annual average isothermal lines, and the meteorological variables such as sunshine hours, rainfall, and temperature during ripening stage. The images, filmed on September 6th and 10th in 2009, were selected for NDVI calculation.

As a result of yield function estimation, the goodness of fit for the models was shown from 0.768 to 0.891, which was higher than the results given by Hong et al. in 2009.

In this study, it is noteworthy academically and practically that vegetation index(NDVIs) identified by annual average isothermal lines and meteorological variables are very useful for estimating yield functions. Also, establishment of more accurate vegetation index and meteorological variables is necessary for more appropriate yield functions.

It is expected that the accuracy of NDVIs and goodness of fit of the paddy rice yield model will be improved if the images from KOMPSAT-3(Korea Multi-purpose Satellite-3) or other radar satellite images are applied.

Reference

- [1] Han, Suk-Ho, Bae-Sung Kim, Kung-Pil Kim, Jae-Hwan Kim, and Marl-Jing Kim, 「A Pilot Study on Estimating Rice Yield with Satellite Images」, Korea

Rural Economic Institute(KREI), Research Report D308, 2011.

- [2] Hong, Suk-Young, Sang-Kyu Rim, Kyu-Sung Lee, In-Sang Jo and Kil-Ung Kim, "Estimation of Rice-Planted Area using Landsat TM Imagery in Dangjin-gun area", Korean Journal of Agricultural and Forest Meteorology, Vol. 3 No. 1, pp. 5~15, 2001.
- [3] _____, Eun-Young Choe, Gun-Yeob Kim, Sin-Kyu Kang, Yi-Hyun Kim, and Yong-Seon Zhang, "Estimating Rice Yield of North Korea using a Satellite Vegetation Index", Korean Society of Soil Science and Fertilizer, Vol. 2009, P. 64, 2009.
- [4] _____, Byoung-Keol Min, Jee Min Lee, Yi Hyun Kim, and Kyung Do Lee, "Estimation of Paddy Field Area in North Korea Using RapidEye Images", Vol.45 No.6, pp. 1194-1202, 2012a.
- [5] _____, Ji-Na Hui, Joong-Bae Ahn, Jee-Min Lee, Byoung-Keol Min, Chung-Kuen Lee, Yi-Hyun Kim, Kyung-Do Lee, Sun-Hwa Kim, Gun-Yeob Kim, and Kyo-Moon Shim, "Estimating Rice Yield Using MODIS NDVI and Meteorological Data in Korea", Korean Journal of Remote Sensing, Vol. 28 No.5, pp. 509-520, 2012b.
DOI: <http://dx.doi.org/10.7780/kjrs.2012.28.5.4>
- [6] Kim, Chung-Sil and Jae-Hwa Park, "A Study on the Effective Direction Applying Remote Sensing Technology as a Tool for Agricultural Statistics", Journal of Rural Development, Vol.29 No.2, pp. 161-176, 2006.
- [7] Park, Mee-Sung, Bo-Young Ham, Byoung-keol Min, and Bae-Sung Kim, "Estimation of Paddy acreage in Dangjin district using Komsat-2 Images", Journal of Rural Development, Vol. 33 No. 3, pp. 1-17, 2010.

Bae-Sung Kim

[Regular member]



- June 1999 : Korea Univ. Ph.D. in Economics
- Feb. 2003 ~ Feb. 2012 : Korea Rural Economic Institute, Senior Fellow.
- Mar. 2012 ~ current : Jeju National Univ., Dept. of Industrial and Applied Economics, Professor

<Research Interests>

Production Economics, Applied Econometrics, Agricultural Outlook

Jae-Hwan Kim

[Regular member]



- Feb. 2003 : Seoul National Univ. Master. D. in Agri - Economics
- Feb. 2003 ~ Jul. 2011 : Korea Rural Economic Institute, Research associate .
- Aug. 2011 ~ current : World institute of Kimchi, senior researcher

<Research Interests>

Food Industry, Kimchi Globalization, Agricultural Outlook

Seong-Bo Ko

[Life member]



- Feb. 1995 : Korea Univ. Ph.D. in Agricultural Economics
- May 1997 ~ Aug. 2004 : Jeju Development Institute, Head of Research Division
- Sep. 2004 ~ current : Jeju National Univ., Professor

<Research Interests>

Agricultural Policy, Agricultural Outlook, Regional Industry Analysis.