

# Rice Yield Estimation Based on SAR and Meteorological Parameters

Digvijaya Paudel<sup>1</sup>, Bikash Sherchan<sup>1</sup>, Krishna Prasad Bhandari<sup>2</sup><sup>1</sup>Department of Geomatics Engineering, IOE, Pashchimanchal Campus, Tribhuvan University, Nepal<sup>2</sup>Center for Space Science and Geomatics Studies, IOE, Pashchimanchal Campus, Tribhuvan University, Nepal

(Manuscript Received: 15/08/23; Revised: 18/09/2023; Accepted: 02/10/2023)

## Abstract

Estimating food production, demand, and distribution allows for timely yield prediction, crucial for managing food security. In-depth research has been published in the literature on applying vegetation indices discovered using optical remote sensing observation for yield estimation. The fundamental limitation of optical remote sensing is that it cannot penetrate cloud cover, which may contaminate the data. Therefore, a novel approach has been studied in this study employing SAR images from Sentinel-1 to construct a regression model combining vegetation index derived from SAR images and climatic variables over a 6-year time series (2017 to 2022). It is evident from the findings that the predictor variables have a non-linear relationship with the yield, which a straightforward linear regression model cannot describe. Other regression models, such as Random Forest, could be more useful in explaining such a complicated and non-linear connection. When the Multiple Linear regression mode was used for testing, it was found that the  $R^2$  value was to 0.918 and the MSE was 0.513 Mt/ha. When the RF regression mode was used for testing, it was found that the  $R^2$  value increased to 0.918 and the MSE improved to 0.513 Mt/ha. Furthermore, observation showed a prediction error of around 0.353 Mt/ha when employing the Spatial Error model. Therefore, rice yield estimation has considerably improved when employed in a spatial model.

**Keywords:** Estimation; Regression; Rice; SAR; Yield

## 1. Introduction

In agriculture, remote sensing is used to analyze data from sensors to determine the crop developmental stage, including emergence, vegetative growth, reproductive growth, and maturity. Farmers and researchers can decide when to use inputs such as fertilizer, insecticides, and other chemicals and when to harvest the crop by keeping track of the crop's growth stage. Relative to agriculture, yield prediction is a significant application of remote sensing. Before a crop is harvested, the yield can be predicted using data from remote sensors. It is possible to employ yield prediction models to give farmers a head start on prospective production shortages or surpluses by developing them using a mix of crop growth data, meteorological data, and soil data.

Most studies on remote sensing data for agricultural yield estimation have used visible and infrared sensors, such as Landsat or sophisticated high-resolution radiometers. More recently, high temporal and spatial resolution data from the Copernicus program's Sentinel-1 and Sentinel-2 constellations of the European Space Agency (ESA) are freely available. These data

formats provide a fresh window of opportunity for high-resolution crop monitoring activity [1]. Due to its accessibility regardless of weather and ability to supplement optical data in inclement weather, SAR data offer special advantages in the remote prediction of agricultural yield [2]. As an active sensing technique, spaceborne synthetic aperture radar (S-SAR) is able to record signal amplitude and phase information as well as the target's polarization scattering data.

The use of vegetation indices obtained from optical remote sensing observation for yield estimation has been extensively documented in the literature. However, the main drawback of optical remote sensing is that it cannot see through cloud cover, which may somehow taint the data. Additionally, Nepal experiences monsoon rains during the rice crop cycle's Peak Season (PoS). As a result, a large portion of the country, particularly hilly areas, is covered with a sizable portion of cloud, making it challenging to obtain cloud-free satellite images that can be used in various studies.

There are various varieties of rice cycles according to the growth season, including Summer Autumn (SA), Autumn-Winter (AW), and Winter-Spring (WS). There are various stages in the rice growth cycle, including sowing, transplanting, tilling, jointing, head-

\*Corresponding author. Tel.: +977- 9846880994,  
E-mail address: digbijay223@gmail.com

ing, and ripening. Various stages respond to micro-waves differently. At various growth phases and combination settings, optical and SAR indices are suitable to varying degrees. According to observations, optical indices for vegetation cover at the growth stages and interferometric coherence show a different tendency [3].

According to the study by [4], strong nonlinear changes were evident in the several phenological phases of rice backscatter. Advanced Synthetic Aperture Radar (ASAR), a multi-polarization comprised of C-band S-SAR, allows for the research of rice monitoring and yield computation using multi-polarization data. When the agrometeorological model could not be employed, [5] suggested a statistical yield estimation model and used this approach to pinpoint rice-growing zones. They fitted the three-time phase HH/VV ratio based on the statistical model to estimate the rice yield in the research area. It was challenging to choose the right moment to collect the data, though. In order to identify rice by inverting the Leaf Area Index based on the water cloud model, [6] used the ASAR alternating polarization precision mode and the HH/VV polarization ratio threshold. With parameter optimization of the assimilation approach, based on the associated backscattering coefficients, the yield estimation was achieved, improving the universality of the yield estimation scheme with an accuracy of nearly 85%. Consequently, it is clear that multi-polarization and multi-temporal SAR data can guarantee the accuracy of rice yield estimation. In order to map rice in the study region, [7] determined that the HH/VV combination was the optimum data source and thus ratio change detecting technology was employed using this ratio combination.

The environment must be warm and humid for the rice production. It works best when there is a lot of humidity, a lot of sunlight, and a reliable water supply. For the duration of the crop's life, an average temperature between 21°C and 37°C is needed. In Nepal, rice is grown between 60 meters above mean sea level to 3050 meters above mean sea level in moderate temperature zones. Rainfall, humidity and terrain characteristics are the major factors influencing the rice yield [8].

In the Nepalese context, the application of multi-temporal Sentinel-1 data for rice yield estimation is still not explored. As a result, this study investigated

the utilization of ratios of two polarizations, i.e., VV and VH of Sentinel-1 as a novel and experimental approach for estimating rice production through a regression approach that takes into account meteorological characteristics.

## 2. Materials and Method

### 2.1 Data Collection

In the Hilly area of Nepal, the Start of Season (SoS), Peak of Season (PoS) & End of Season (EoS) for rice correspond to mid-July, mid-September, and mid-November, respectively. However, for better analysis S-SAR image product within the C band of Sentinel 1 were acquired corresponding to the whole crop cycle for the period of 2017 - 2022. These images were used to calculate the ratio between the two available polarizations (i.e., VV/VH) alternative to NDVI for rice crop cycle of 6 years in the Google Earth Engine (GEE) platform.

### 2.2 Mask to Rice Crop Area

It was crucial to keep the study's scope restricted to the zone where rice is grown in order to improve its accuracy and authenticity. Therefore, a LULC map was utilized as an initial mask. Additionally, the map created by ICIMOD using sophisticated algorithms was adopted in light of the possibility that manually prepared LULC could have some limitations. Crop calendar was used to further mask the area where rice was grown.

### 2.3 Interpolation of Meteorological Parameters

For a six-year study period, daily measurements of the climatic parameters were taken for the crop cycle. Following processing, a monthly average for each parameter was obtained, put into the GIS platform, and interpolated to create a raster format. The widely used kriging method was selected for the interpolation task. During the interpolation, it was done in such a way to match the spatial resolution of sentinel images, i.e., 10 m.

### 2.4 Preparing Yield Raster

Six years of rice yield data were gathered from the MoALD's official documentation of agriculture statistics. The yield data was collected at the district level and was measured in metric tons per hectare. To create the raster, this scalar value was then spread to the cropped area that was being masked in proportion to the size of each individual pixel. Many literary works have utilized this workflow. Similar methodology was adopted in the Terai districts of Nepal for rice yield estimation [9].

### 2.5 Regression Analysis

In this work, three regression models, i.e., Multiple Linear Regression (MLR), Random Forest (RF), and Spatial Error Regression (SE) were investigated to estimate the yield of rice. MLR and RF, as a result, are non-spatial regression models that do not explicitly describe spatial auto-correlation or spatial heterogeneity and do not take into consideration the geographical correlations between data points. As a result, this work is unique in that it explores the spatial relation and creates a model for yield estimate based on a variety of predictor factors.

Prior to creating regression models for yield estimation, a preliminary correlation analysis was performed to ascertain the relationship between yield and various other variables and to better understand whether there is a potential relationship between these variables and whether they may be useful in predicting the outcome variable. Each model's dependent variable was chosen to be the yield for the year 2020, with the remaining data serving as the dependent variables for the other models. Each model was also trained and tested according to the same protocol, with 70% of the data utilized for training and the remaining 30% for testing. The yield for the year 2022 was then estimated using the model with the highest level of accuracy, and a model was created. Regarding the Spatial Regression Model, a spatial weight matrix was created before the Spatial Error model could be run, and each variable's spatial association was then investigated using this matrix. Following that, the Spatial Error model was used to perform spatial regression.

### 2.6 Study Area

The Tanahun district, which is a part of Gandaki Province, lies almost in the center of all of Nepal. The district's cropping season has changed, which has resulted in a shorter rice harvesting season. Rice productivity has fallen over the last few years, according to statistics from the Tanahun District's 17-

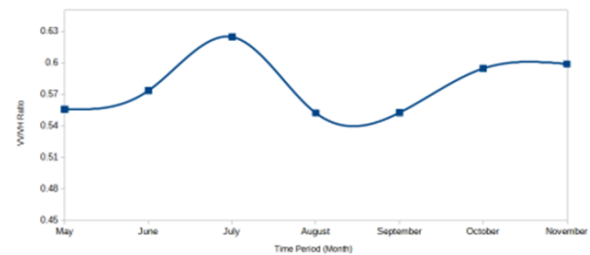


Figure 1 Sample Plot of VV/VH over Crop Cycle for 2017

year crop output history. Untimely rainfall in an excessive amount could be to blame for this lower productivity.

An untimely downpour during pollination, flowering or grain filling combined with a dry spell caused by a drought, could reduce crop yield. Intense rainfall during the monsoon season has an impact on rice and causes climatic dangers like water logging in the district [10].

## 3. Results and Discussion

### 3.1 VV/VH Pattern over Crop Cycle

In contrast to the typical pattern of the NDVI, it was found that the pattern of the ratio decreased relative to other times during the rice crop cycle. In other words, the crop cycle frequently had higher NDVI measurements than the other months. On the other hand, it was demonstrated that the ratio decreased during the crop cycle as compared to previous times.

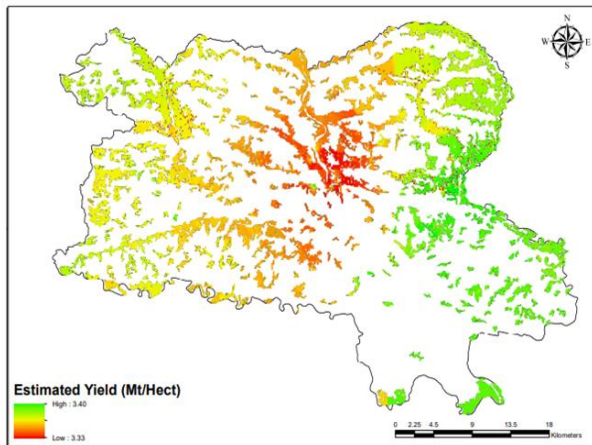
### 3.2 Relationship Between NDVI and VV/VH Ratio

While examining the association, it was also taken into account that optical satellite pictures for the POS of the rice crop cycle were not available. The six months (the first four and last two of each year) were therefore used for the analysis. For the rice crop cycle, observations showed that the indices had a negative correlation with one another. This connection was reliable enough to support the reported pattern by the ratio of the two selected polarizations, VV/VH.

### 3.3 Multiple Regression Model

For each set of predictor variables, an MLR model was created in order to assess how those variables will affect the estimate model. It was discovered that when variables are included in the regression model, the accuracy of the model steadily improves over time, with the prediction error falling from 1.31 Mt/ha to 0.810 Mt/ha. This indicates that using all the variables in the estimating model would allow us to

reach a higher level of accuracy. The entire result,



however indicated that the  $R^2$  value was 0.293.

Figure 2 Yield Estimation by Random Forest Regression Model

This observation's fundamental cause may be the nonlinear connection between the yield and its predictor factors, which the linear regression model fails to account for.

### 3.4 Random Forest Model

When the RF regression mode was used for testing, it was found that the  $R^2$  value increased to 0.910 and the MSE improved to 0.513 Mt/ha. Moreover, there was a yield discrepancy of around 513 kg per hectare between the estimated and actual yield. We came to the conclusion that there is a non-linear relationship between the yield and its predictor variables, which this model more effectively captures than other linear regression models due to the enhanced result obtained using the random forest model. The model was created using data spanning five years, from 2017 to 2021, and was ultimately reviewed for the year 2022, after testing for the year 2020.

### 3.5 Spatial Error Model

It was observed that all the variables exhibit significant spatial correlation and therefore spatial regression model was employed to have yield estimation by incorporating all the predictor variables. When the SE model was employed by incorporating all the variables, the results seemed to improve than by incorporating other non-spatial regression models. Unlike statistical models, efficiency of spatial regression models is not fully represented by the value of  $R^2$ . Observation showed the prediction error of around 0.353 Mt/ha. Therefore, the estimation of rice yield has considerably improved when employed spatial

model. From this, we may infer that the variables are

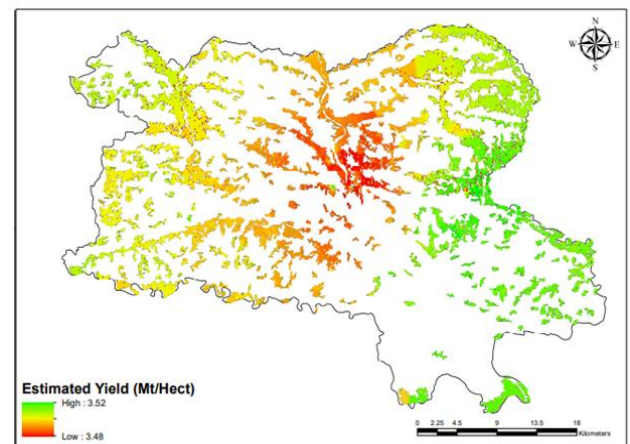


Figure 3 Yield Estimation by Spatial Error Regression Model

spatially auto correlated with each other and thus a spatial error model could provide better estimation than others.

### 3.5 Discussion

The study's initial discovery was that, in contrast to the NDVI trend, the VV/VH ratio decreased during the crop growth cycle. There are a number of causes for the decreasing value of the polarization ratio. The radar signal travels vertically during VV polarization and travels vertically upon transmission and reception. Because of the increased attenuation of the signals, backscatter levels drop as plant development increases. VV backscatter is the primary example of this, but VH backscatter is less noticeable. The direct contribution of soil and vegetation to VV backscatter predominates, and the signal is becoming more and more attenuated by plants' growing vertical structures, i.e. stems [11]. This may serve as an explanation for the negative association between VV/VH backscatter ratio and yield.

Similar to how the rice crop expands, we may see an increase in canopy height. It is evident that when plant height rises, the VV radar signal intensity decreases [12]. The biomass of rice plants tends to be higher and the moisture content to be lower during the mature stage than it was during earlier stages. As radar signals travel farther into the canopy with higher biomass levels, they may be attenuated weakly, especially in VV polarization. Reduced scattering of radar signals in both VV and VH polarizations can also be a result of lower moisture content. As rice



reaches its mature stage of growth, these elements may result in a decreased VV/VH ratio.

Three regression models have been looked at in this work to determine the rice yield based on the polarization ratio derived from Sentinel's SAR image and climatic factors. It is clear from the findings that the predictor variables have a non-linear relationship with the yield, which cannot be described by a straightforward linear regression model. Other regression models, such as Random Forest, could be more useful in explaining such a complicated and non-linear connection. With the implementation of RF model, we can conclude that the Humidity could be the first factor to distinguish the amount of yield for a crop calendar followed by Minimum Temperature, Maximum Temperature, Rainfall and VV/VH polarization ratio. Regarding the spatial regression model, the parameters used here exhibit a significant degree of spatial correlation. As a result, the model's efficiency has been increased by the use of spatial regression models.

Furthermore, as can be seen in the figure, the yield value varies across many geographical regions of the study area. The topographical conditions within the low and highland of the area affects the water availability throughout the growing season and also other soil factors. This difference can also lead to the different value in the yield of rice. Therefore, in order to have a precise yield estimation, geographical factors should also be considered in further works.

#### 4. Conclusions

It has been determined as a result of this study that SAR data can be used to examine the rice crop cycle and, ultimately, in the estimation of yield. The spectral features of rice that can be explained by NDVI may also be understood by using the VV and VH polarization from SAR images. For future research, additional primary data could be used in smaller geographic areas. Additionally, the effects of agricultural and irrigation methods might be considered, and a comparison between the two scenarios could be done. Additionally, the optical index might be added to the SAR values used here to conduct a comparable study.

#### Acknowledgment

The authors would like to acknowledge Department of Hydrology and Meteorology and Agriculture Knowledge Center, Tanahun for their assistance to provide meteorological and yield data respectively. Also, we would like to acknowledge ICIMOD for the LULC data employed here in the study.

#### References

- [1] Aschbacher, J., & Milagro-Pérez, M. P. The Europe an Earth monitoring (GMES) programme: Status and perspectives, *Remote Sensing of Environment*, 120 (2012) 3–8.
- [2] Alebele, Y. and et al. Estimation of Canopy Biomass Components in Paddy Rice from Combined Optical and SAR Data Using Multi-Target Gaussian Regressor Stacking, *Remote Sensing*, 12 (16) (2020) 2564.
- [3] Phi Phung, Hoang & Lam, Nguyen & Nguyen-Huy, Thong & Le Toan, Thuy & Apan, Armando. Monitoring rice growth status in the Mekong Delta, Vietnam using multitemporal Sentinel-1 data, *Journal of Applied Remote Sensing*, 14 (1) (2020) 014518.
- [4] De Bernardis, C. G., Vicente-Guijalba, F., Martinez-Marin, T., & Lopez-Sanchez, J. M. Estimation of Key Dates and Stages in Rice Crops Using Dual-Polarization SAR Time Series and a Particle Filtering Approach, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (2014) 1–11.
- [5] Nguyen, Lam-Dao., Armando, Apan., Frank, R., Young., Trung, Le-Van., Thuy, Le-Toan., Alexandre, Bouvet. Rice monitoring using ENVISAT-ASAR data: preliminary results of a case study in the Mekong River Delta, Vietnam, (2007).
- [6] Shen, S., Yang, S., Li, B., Tan, B., Li, Z., & Le Toan, T. A scheme for regional rice yield estimation using ENVISAT ASAR data, *Science in China Series D: Earth Sciences*, 52(8) (2009) 1183–1194.
- [7] Yang, S., Shen, S., & Zhao, X. Assessment of RADARSAT-2 quad-polarization SAR data in rice crop mapping and yield estimation, *Remote Sensing and Modeling of Ecosystems for Sustainability*, 9 (2012). DOI:10.1117/12.929252
- [8] Joshi, N.P., Maharjan, K.L., & Piya, L. Effect of Climate Variables on Yield of Major Food-crops in Nepal: A Time-series Analysis, *ASAE 7th International Conference*, (2011).
- [9] Fernandez-Beltran, R., Baidar, T., Kang, J., & Pla, F. Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal, *Remote Sensing*, 13(7) (2021).
- [10] P. Koirala. Impact of Climate Change on Rice Production in Tanahun, Nepal, (2019).
- [11] K. Harfenmeister, D. Spengler, and C. Weltzien. Analyzing Temporal and Spatial Characteristics of Crop Parameters Using Sentinel-1 Backscatter Data, *Remote Sensing*, 11 (13) (2019). DOI: 10.3390/rs11131569.
- [12] Bousbih, S., Zribi, M., Lili-Chabaane, Z., Baghdadi, N., El Hajj, M., Gao, Q., & Mougnot, B. Potential of Sentinel-1 Radar Data for the Assessment of Soil and Cereal Cover Parameters, *Sensors*, 17 (11) (2017). DOI: 10.3390/s17112617.