Comparison of Gap-filling Algorithms of Sentinel-2A/B NDVI Images for Monitoring in Rice Paddy Fields

Jeong Eun[1], Sun-Hwa Kim[2*], Tae Ho Kim[3]

[1]Principal Researcher, Corporate Research Institute, PerPixel Inc., eun2@perpixel.co.kr

[2]CEO, Perpixel Inc., sunhk@perpixel.co.kr

[3]Director, Underwater Survey Technology 21 Corp., thkim@ust21.co.kr

*sunhk@perpixel.co.kr

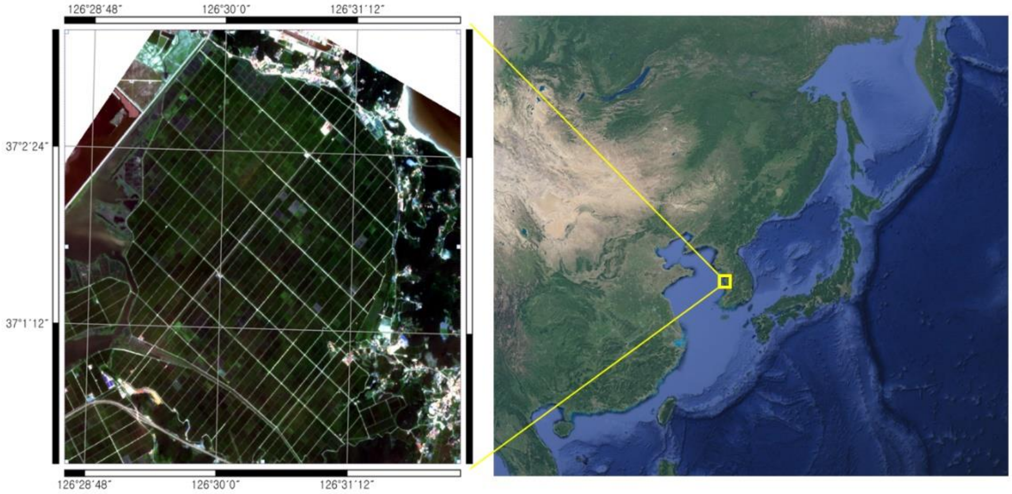
1. Introduction

The CAS500-4 satellite, which is set to launch in 2025, is an optical satellite designed for agriculture and forestry observation. It will be in a polar orbit and have a spatial resolution of 5 meters and an observation width of 120 kilometers. Its purpose is to improve the prediction and management of agricultural and environmental changes in the Republic of Korea. One important application of optical satellite images is crop growth monitoring, but there are limitations due to gaps in the data caused by cloud cover. This makes it challenging to monitor crop growth using optical satellite images. However, machine learning algorithms, particularly Gaussian Process Regression (GPR), have emerged as powerful tools for processing time-series data and filling in these gaps. GPR has been proven effective for seasonal crop monitoring and filling gaps in optical images caused by clouds. In this study, we aim to evaluate the suitability of the GPR correction technique for providing high-quality information by applying it to areas with missing data for crop monitoring. We will compare various gap-filling techniques and propose the most suitable one for the CAS500-4 based on field data.

2. MATERIALS AND METHODS

2.1 Study Area

The study was conducted in the rice paddy area located in Dangjin-si, Chungcheongnam-do, Republic of Korea, with coordinates 37.0331°N 126.5062°E. Field data were collected in this area to analyze the effectiveness of the gap-filling technique. The study area covers 1,042 hectares and is mainly dedicated to the cultivation of rice, which is a major crop in the Republic of Korea. The area receives an average annual precipitation of approximately 1,066 mm, with the majority of rainfall occurring during the monsoon season and typhoon periods from July to September. The average monthly temperature is around 13°C, and there are distinct seasonal temperature variations. Rice has a growth period of about 150 days, with planting typically starting in May to June and harvesting taking place in September to October. In the study area, harvesting generally occurs in October, although in different regions, it can be as late as November due to variations in temperature sensitivity.



Source: Reference for Goole Earth and satellite Image

Figure 1: Location of the study area (rice paddy).

2.2 Dataset used

The optical satellite images used in this study were obtained from Sentinel-2A/B satellites. These satellites provide free data with a spatial resolution of 10 meters and a revisit period of 5 days. We specifically used Level-2A images, which are atmospherically corrected reflectance images. The Sentinel-2A/B L2A images were sourced from the Copernicus Open Access Hub (https://dataspace.copernicus.eu/). For our study, we utilized satellite data from May to October 2021, which aligns with the growth period of rice. To verify the accuracy of the satellite data, we collected spectral reflectance data of rice through three field observations using an ASD (Analytical Spectral Devices) instrument. The specific instrument used was the Field Spectrometer 3, which measures reflectance information in the 350-2500 nm wavelength range.

Table 1: The dataset used in this study.

Satellite/ Data	Dataset and specifications
Sentinel-2A/B L2A	05/07/21 ~ 10/24/21
Field data(ASD)	08/05/21, 09/09/21, 10/13/21 (ASD Fieldspec3)

Figure 4: Time-series ASD measurement photos for the study area.

2.3 Methods

In this study, NDVI (Normalized Difference Vegetation Index) images were generated using Sentinel-2A/B L2A satellite imagery of the Dangjin area in 2021. MaxNDVI composite images were created at 15-day intervals by combining daily NDVI images. For the analysis, DATimeS (Decomposition and Analysis of Time Series Software) was utilized. DATimeS is a component of the ATRMO (Automated Radiative Transfer Models Operator) software package developed by ESA Sentiflex. It supports research on satellite-based vegetation production monitoring and offers the application of 30 different machine learning algorithms using optical imagery. DATimeS also provides the capability to derive time series correction results, enabling the assessment of seasonal crop growth conditions. In this study, MaxNDVI composite images were used as input data, and representative gap-filling techniques including Gaussian Process Regression (GPR), Linear Interpolation, and Nearest Neighbor Interpolation were applied using DATimeS. The performance of these gap-filling techniques was analyzed by comparing the results with the field reflectance data that was collected.

3. Results

The results of applying the time series vegetation index gap-filling technique using 34 Sentinel-2A/B NDVI images from May to October 2021 are as follows. It was challenging to obtain clear images without clouds during the rainy and typhoon seasons in Korea, not only in the study area but also due to the seasonal characteristics of the country. While the gap-filled images exhibited similar time-series patterns to the original images, limitations were observed as cloud effects still persisted in the gap-filled images. This suggests that the influence of clouds could not be completely eliminated by using the composite NDVI image, which still retained cloud effects, as input data for the gap-filling technique.

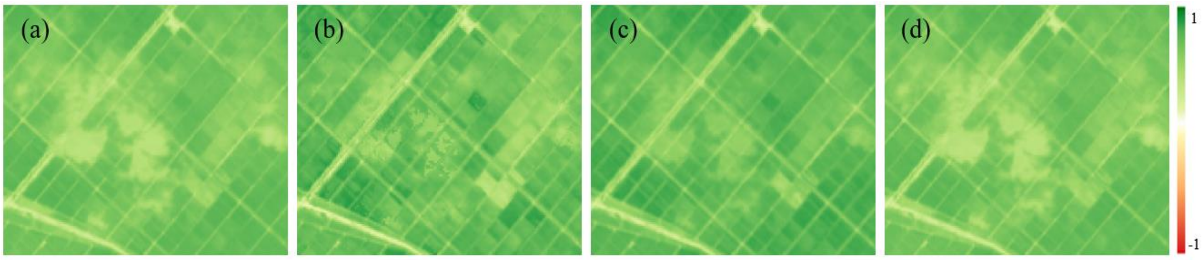


Figure 6: Residual cloud effect after applying gap-filling techniques (Example: Composite image from August 30 to September 9, 2021). 15-days MaxNDVI composited image (a), image with GPR technique applied (b), image with linear interpolation technique applied (c), and image with nearest neighbor interpolation technique applied (d).

To evaluate the effectiveness of the gap-filled techniques, RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) indices were used to compare the images obtained from the applied correction techniques with the field data. The three correction techniques (Gaussian Process Regression, linear interpolation, nearest neighbor interpolation) yielded similar results, but with slight differences. Among these techniques, the Gaussian Process Regression (GPR) technique exhibited relatively lower values for both RMSE and MAE, indicating better performance in reducing the differences between the corrected images and the field data.

Table 2: Accuracy of three gap-filled technique results estimated based on field measured data.

Measure	**Gaussian process regression**	**Linear interpolation**	**Nearest neighbor interpolation**
RMSE	0.17289	0.17297	0.17474
MAE	0.13119	0.13879	0.14235

4. Conclusion and Recommendation

Based on existing research examples and considering the significance and effectiveness of correction techniques, the GPR technique is considered the most effective for time-series correction of agricultural and forestry satellites. Current research in this field is focused on applying step-by-step time-series correction techniques that differentiate between pixels at gap boundaries and other pixels. In the future, it is anticipated that the GPR technique will be utilized to estimate missing data, resulting in the production of high-quality time-series data. This data will have valuable applications in areas such as crop monitoring, including the detection of growth changes and disaster response.

Acknowledgements

This study was conducted with the support of Republic of Korea the Rural Development Administration's research project(task number: PJ016234).

References

Dario De Caro, Matteo Ippolito, Marcella Cannarozzo, Giuseppe Provenzano and Giuseppe Ciraolo, (2023). Assessing the performance of the Gaussian Process Regression algorithm to fill gaps in the time-series of daily actual evapotranspiration of different crops in temperate and continental zones using ground and remotely sensed data. *Agricultural Water Management*. https://doi.org/10.1016/j.agwat.2023.108596

Kim Sun-Hwa & Jeong Eun, (2021). Development of Cloud and Shadow Detection Algorithm for Periodic Composite of Sentinel-2A/B Satellite Images. *Korean Journal of Remote Sensing*, Vol.37, No.5-1, 2021, pp.989~998

Kim Sun-Hwa & Jeong Eun, (2022). Development of Score-based Vegetation Index Composite Algorithm for Crop Monitoring. *Korean Journal of Remote Sensing*, Vol.38, No.6-1, 2022, pp.1343~1356

Luca Pipia, Eatidal Amin, Santiago Belda, Matías Salinero-Delgado and Jochem Verrelst, (2021). Green LAI Mapping and Cloud Gap-Filling Using Gaussian Process Regression in Google Earth Engine. *Remote Sensing*. 2021, 13, 403. https://doi.org/10.3390/rs13030403

Santiago Belda, Luca Pipia, Pablo Morcillo-Pallarés and Jochem Verrelst, (2020). Optimizing Gaussian Process Regression for Image Time Series Gap-Filling and Crop Monitoring. *Agronomy 2020*, 10, 618; doi:10.3390/agronomy10050618

Copernicus Open Access Hub Website, Copernicus Browser, Satellite Download form https://dataspace.copernicus.eu/

Korea Meteorological Administration from https://data.kma.go.kr/cmmn/main.do

Malvern Panalytical Website, ASD FieldSpec products information from https://www.malvernpanalytical.com/en