# Dense NDVI Time Series by Fusion of Optical and SAR-Derived Data

Thomas Roßberg, *Graduate Student Member, IEEE*, and Michael Schmitt, *Senior Member, IEEE*

***Abstract*—Gaps in normalized difference vegetation index (NDVI) time series resulting from frequent cloud cover pose significant challenges in remote sensing for various applications, such as agricultural monitoring or forest disturbance detection. This study introduces a novel method to generate dense NDVI time series without these gaps, enhancing the reliability and application range of NDVI time series. We combine Sentinel-2 NDVI time series containing cloud-induced gaps with NDVI time series derived from the Sentinel-1 synthetic aperture radar sensor using a gated recurrent unit, a variant of recurrent neural networks. To train and evaluate the model, we use data from 1206 regions around the world, comprising approximately 283 000 Sentinel-1 and Sentinel-2 images, collected between September 2019 and April 2021. The proposed approach demonstrates excellent performance with a very low mean absolute error of 0.0478, effectively filling even long-lasting gaps while being applicable globally. Thus, our method holds significant promise for improving the efficiency of numerous downstream applications previously limited by cloud-induced gaps.**

***Index Terms*—Cloud removal, data fusion, deep learning, gap filling, recurrent neural network (RNN), vegetation monitoring.**

## I. Introduction

The normalized difference vegetation index (NDVI) [1] derived from multispectral optical data is an important tool for vegetation monitoring. It serves as a valuable indicator of vegetation health and vitality, enabling insights into various vegetation-related phenomena. NDVI time series offer an opportunity to analyze temporal trends in vegetation dynamics and, therefore, are used for a multitude of applications, such as crop classification [2], yield estimations [3], or forest disturbance mapping [4]. However, the effective utilization of NDVI time series faces challenges, particularly in regions with frequent cloud cover such as tropical or subtropical areas, which impedes consistent monitoring efforts.

A variety of strategies have been investigated to mitigate the impact of cloud cover when analyzing NDVI time series. A simple strategy is the reconstruction of NDVI time series [5], for example, by utilizing multiyear long time series and filling gaps using long-term trends or by using linear interpolation and a Savitzky–Golay filter [6]. Both the methods can result in decent results, if long-term developments are studied or the gaps are short, but are not optimal when sudden changes are frequent and irregular, for example, for cropland areas.

Another straightforward method is the use of multiple optical multispectral sensors with comparable resolutions and characteristics. After matching and harmonizing their reflectance values to ensure consistent spectral information, the temporal frequency is increased, decreasing the gap length. An example of this method is the Harmonized Landsat and Sentinel-2 product [7].

Another approach is spatiotemporal fusion, where frequent coarse-resolution imagery and less frequent but high-resolution data are combined to enhance temporal resolution while maintaining spatial resolution. Techniques such as STARFM [8] and the fusion of Sentinel-2 and Sentinel-3 data [9] are examples of this approach. Even though all the three aforementioned approaches can reduce the number of gaps in optical time-series data, their reliance solely on optical data limits their effectiveness in regions with persistent cloud cover.

To overcome the limitations posed by cloud cover, synthetic aperture radar (SAR) sensors can be a solution, offering distinct advantages. As active sensors, they employ radar waves capable of penetrating cloud cover and operating independently of sunlight. In addition, the longer wavelengths compared to optical light enable them to look into the volume below a surface, which is especially helpful when studying forests or crops. There, SAR data allow one to look below the top tree canopy or get information on the whole crop and not only the top surface. Therefore, several studies could showcase the potential of SAR data, for example, for crop monitoring [10], [11], [12] or forest biomass retrieval [13], [14]. Despite these benefits, the widespread adoption of SAR data in vegetation monitoring remains limited and can be attributed to various factors: one primary challenge is the need to redevelop algorithms and applications originally designed for NDVI or other optical data to accommodate SAR data. Furthermore, SAR data processing is more complex, and its interpretation is less intuitive compared to traditional optical data. Finally, long-time series data only exist for optical data, for example, in the Landsat archive, but not for SAR data, which limits the use of SAR data for long-term studies.

A way to bridge this gap is to translate images or time series of SAR data into NDVI data. This enables the use of existing applications while being unaffected by cloud coverage. One

Manuscript received 1 December 2023; revised 18 January 2024; accepted 16 March 2024. Date of publication 20 March 2024; date of current version 10 April 2024. This work was supported in part by the German Federal Ministry for Economic Affairs and Climate Action through the project Dense Satellite Time Series for Agricultural Monitoring under Grant FKZ 50EE2018 and in part by University of the Bundeswehr Munich. *(Corresponding author: Thomas Roßberg.)*

The authors are with the Department of Aerospace Engineering, University of the Bundeswehr Munich, 85577 Neubiberg, Germany (e-mail: thomas.rossberg@unibw.de).

Digital Object Identifier 10.1109/JSTARS.2024.3379838

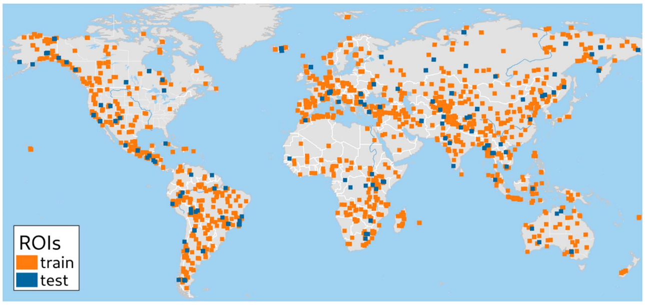


Fig. 1. Location of the ROIs used for training (orange) and testing (blue).

study utilized a combination of convolutional neural networks and long short-term memory networks to transform pixelwise Sentinel-1 time series into Sentinel-2-derived NDVI time series [15]. However, this research was limited to a 10 km × 10 km area, predominantly featuring croplands. Another significant limitation of this approach is its lack of demonstrated transferability to other regions or landscapes, necessitating the training of a new model for each different area. This requirement poses substantial challenges for broader applications. Moreover, the proximity of training and test pixels in such a confined area may lead to a high correlation, potentially resulting in an overestimated test performance. The study's demand for a fixed time-series length and temporal resampling further complicates its adaptability and practical application in diverse settings.

Another study adopted a sequence-to-sequence model for this SAR-to-NDVI time-series translation with aggregated time series from field polygons, achieving a low mean absolute error (MAE) of 0.04 [16]. With approximately 20 000 km$^2$, the study encompasses a larger area than that in [15], yet it remains confined to two regions in France. Also, it demands a common temporal grid, the preexistence of polygons to aggregate time-series data, and the employed random split that was neither spatial nor temporal giving an insufficient understanding of the model's performance.

We propose a worldwide applicable and flexible solution to generate dense NDVI time series by fusing accurate optical NDVI values with denser yet less accurate SAR-derived NDVI values. This involves employing a recurrent neural network (RNN), a deep learning architecture adapted for handling time series. The irregularity of accurate optical NDVI data and the regularity of SAR-derived NDVI values lead to the creation of dense and accurate NDVI time series. Utilizing RNNs brings the key advantage of accommodating missing values and sequence lengths that vary [17], a feature critical for effective fusion of heterogeneous data.

The novelty of our approach lies first in the fusion of earth observation time-series data of different modalities to predict dense NDVI values and second on its global applicability. Our overall approach is split into two main stages: initially, we perform a single-image SAR-to-NDVI translation, which has been extensively discussed and evaluated in our previous work [18]. This is followed by the fusion of the SAR-derived NDVI time series with optical data, a process detailed in the current publication. This strategic separation not only permits a thorough evaluation of each step but also enhances the method's adaptability and versatility, as this method is not limited to the NDVI but can be extended to other vegetation indices or time series of different modalities.

The rest of this article is organized as follows. We present a description of the dataset creation in Section II. Sections III and IV summarize the RNN model and the experiments conducted, respectively. Subsequently, we present the quantitative and qualitative results in Section V and discuss them in Section VI. Finally, the conclusion in Section VII completes this article.

## II. Creation of the Time-Series Dataset

Our approach integrates optical NDVI and SAR-derived NDVI data, necessitating the creation of a comprehensive dataset for training and evaluation. We selected data from the Sentinel-1 SAR and Sentinel-2 optical sensors, motivated by their global coverage, high revisit frequency (at the equator five days for Sentinel-2 and six days for Sentinel-1), and free data access. Our regions of interest (ROIs) are derived from the 1206 globally distributed ROIs of the SEN12TP dataset [18]. These ROIs are chosen for their balanced representation in terms of land cover, climate, and global distribution and are displayed in Fig. 1.

The SEN12TP dataset [19], with its ROIs each measuring 20 km× 20 km, amounts to a size of 222 GB. Extending this to include images at many different times from each region would significantly increase the data volume. To limit the required storage, we utilize only the central 2.56 km× 2.56 km area of each ROI, but download all available imagery between September 1, 2019 and March 30, 2021 to form our image time series. The size of 2.56 km is chosen, so that the image has a size of 256 px× 256 px at 10-m resolution.

Overall, 157 050 Sentinel-1 and 125 860 Sentinel-2 images were downloaded, encompassing 252 GB. Subsequent sections detail the necessary data and preprocessing steps for both the optical NDVI and SAR-derived NDVI values. Fig. 2 provides an overview of all these steps. For preprocessing and image download, Google Earth Engine (GEE) [20] was used.

### A. Optical Data and Processing

As the source of optical data and to calculate the NDVI, data of the Sentinel-2 multispectral sensors are used, which are masked for cloud and cloud shadows and coreferenced onto each other. Thereby, atmospherically corrected Level-2A of the GEE collection `COPERNICUS/S2_SR_HARMONIZED` is used. To mask all the clouds, the cloud probabilities contained in the GEE collection `COPERNICUS/S2_CLOUD_PROBABILITY` created using the Sentinel Hub's cloud detector [21] are used with a threshold set to 40%. To clean the borders of the detected clouds, the cloud probabilities are first convolved with a circle-shaped filter with a radius of 40 m, thresholded and dilated again with 20 m large circular filter kernel. These parameters are taken from the s2cloudless example script [22]. Cloud shadows were masked using a geometrical method by projecting the previously calculated cloud masks onto the ground [23]. The alternative approach, which is not employed in this study, is the spectral method that utilizes the different multispectral bands for shadow detection. Potential cloud

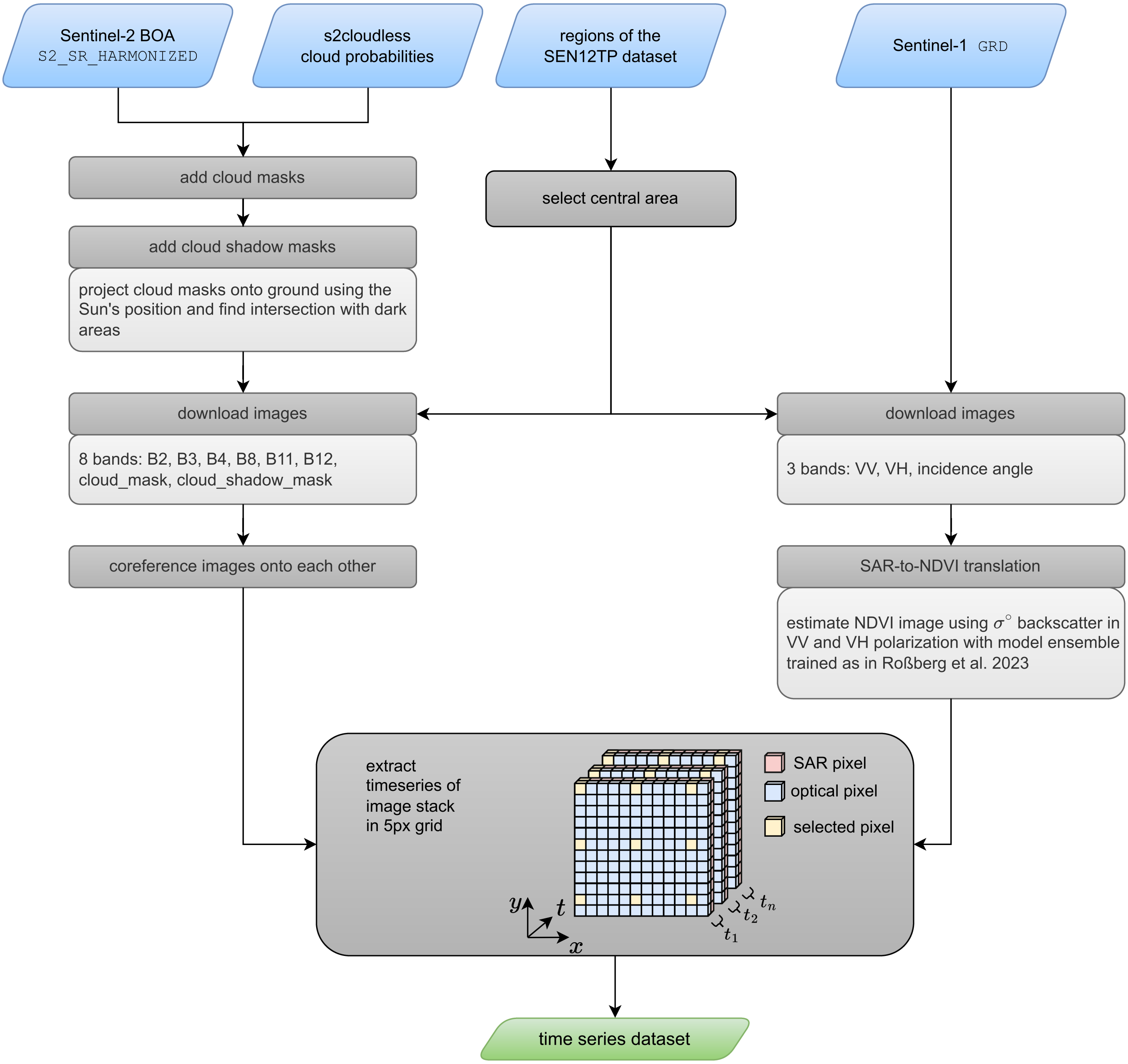

Sentinel-2 BOA
S2_SR_HARMONIZED
s2cloudless
cloud probabilities
regions of the
SEN12TP dataset
Sentinel-1 GRD
add cloud masks
select central area
add cloud shadow masks
project cloud masks onto ground using the Sun's position and find intersection with dark areas
download images
8 bands: B2, B3, B4, B8, B11, B12, cloud_mask, cloud_shadow_mask
download images
3 bands: VV, VH, incidence angle
coreference images onto each other
SAR-to-NDVI translation
estimate NDVI image using σ° backscatter in VV and VH polarization with model ensemble trained as in Roßberg et al. 2023
extract timeseries of image stack in 5px grid
SAR pixel
optical pixel
selected pixel
y
t
x
t1
t2
tn
time series dataset

Fig. 2. Overview of the creation of the time series used for training and testing with the required preprocessing steps. Sentinel-1 and -2 image series are processed and retrieved from GEE for the regions of the SEN12TP dataset. From the image stacks, pixelwise time series are retrieved in a regular 5 px grid to form the dataset.

shadows are all cloud mask pixels projected along the Sun's azimuth angle for 2 km. They are intersected with dark areas, which are all areas with a reflectance sum of the bands B8, B11, and B12 smaller than 0.3, which are not water according to scene classification layer (SCL) band. Again, to clean up the mask, a morphological opening is applied with an erosion (40 m circle) followed by a dilation (100 m circle) [23]. The bands B2, B3, B4, B8, B11, and B12 together with the cloud and cloud shadow masks are downloaded for all the optical images, which are at least 10% cloud and cloud shadow-free.

Upon download of the images, the need for coregistering each optical image stack became apparent, as the geolocalization of the Sentinel-2 imagery is not pixel perfect which is also noted in [24]. We use a method similar to the one in *eolearn* Python package [25]: the template image upon which each image is registered on the gradient of the temporal mean of all the images after cloud masking and conversion to gray scale. To find the best matching position, a translation-only motion model with the enhanced correlation coefficient is used [26].

Finally, the NDVI is calculated using the red and infrared Sentinel-2 bands B8 and B4, respectively [1]

$$\mathrm{NDVI} = \frac{\text{infrared} - \text{red}}{\text{infrared} + \text{red}} = \frac{\mathrm{B8} - \mathrm{B4}}{\mathrm{B8} + \mathrm{B4}}. \quad (1)$$

### B. SAR Data

The SAR data as the second data modality of our approach are acquired by the Sentinel-1 sensors and sourced from the GEE collection `COPERNICUS/S1_GRD`. From this collection, data acquired in the interferometric wide swath mode with two polarizations i.e., VV and VH, are used. For these ground range detected (GRD) data, geometric terrain correction is applied, and the data are transformed to a logarithmic scale. This process yields sigma naught ($\sigma^{\circ}$) backscatter values, expressed in decibels, to be used in our approach.

### C. Retrieval of SAR-Estimated NDVI Values

SAR data are a valuable tool for remote sensing due to their ability to acquire images of the earth's surface, even through cloud cover or at night. This capability is enabled by the sending and receiving of microwave radiation. However, the side-looking geometry required for SAR and the different wavelengths it uses change the data processing and analysis compared to optical sensors. Therefore, creating optical-like data from SAR or, in our case, estimating NDVI values from SAR data is highly desirable.

Estimating NDVI values from SAR data is possible, despite the fact that they sense the earth in different parts of the electromagnetic spectrum. This is possible due to their relationship to each other through different surface parameters, such as land cover [27], [28] and plant or vegetation parameters [10], which are measurable through both NDVI and SAR data. The relationship between SAR and NDVI data has been demonstrated in [29], [30], and [31], with various studies showing effective pixelwise retrieval of NDVI values from SAR data in small regions [32], [33], [34]. In addition, our prior work has shown that translating single Sentinel-1 SAR scenes to NDVI images is feasible on a global level [18]. We employ this method to estimate NDVI values and utilize these estimations to improve the reconstruction of NDVI time series.

To estimate NDVI values from SAR data, we use the deep learning model, as described in [18]. There, a U-Net [35] is used to transform an SAR backscatter image into an NDVI image. As this requires a large dataset with SAR and optical images of the same locations and time, the SEN12TP dataset was created and presented in that study. It consists of approximately 220 GB of timely paired Sentinel-1 and Sentinel-2 imagery with a maximal time discrepancy of less than 12 h. Using this dataset, a good performance and low MAE of 0.122 could be achieved using only $\sigma^{\circ}$ backscatter data.

In this study, we used an ensemble of five trained models as described above. Taking the mean from the outputs of five models ensures superior model performance and reliability. In the end, an estimated NDVI image is created for each location and SAR acquisition.

### D. Time-Series Extraction and the Final Dataset

Time series are extracted from the optical, SAR, and SAR-derived NDVI images. Only every fifth column and every fifth row are used to form a regular 5 px grid. This reduces the data volume to a manageable level and additionally reduces unneeded redundancy due to the high similarity among the time series of neighboring pixels. Next to the image data, the acquisition dates are extracted for each time step. Overall 330 GB of image data are used (consisting of 72 GB optical, 181 GB SAR, and 78 GB SAR-estimated NDVI imagery) resulting in 124 GB of extracted time-series data.

Two splits of the dataset are created to assess spatial and temporal generalization performance. The spatial split dataset contains the same train, validation, and test scenes as those found in the SEN12TP dataset, while the temporal split dataset uses the first 12 months of all the scenes for training and the remaining six months of all the scenes for testing. This allows performance evaluation on unseen data from different scenes and future dates.

## III. RNN-Based Time-Series Fusion

To fuse the optical and SAR-derived NDVI time series, we employ a gated recurrent unit (GRU) [36], a common RNN variant. We selected RNNs because of their suitability for handling variable sequence lengths and accommodating missing data [17]. Utilizing a many-to-many architecture, our chosen GRU model generates an output prediction for each step of the input sequence.

The model input is first transformed using a single fully connected layer with 128 neurons and rectified linear unit (ReLU) activation with shared weights for each time step. Then, the GRU consisting of three bidirectional layers, each with a hidden size of 256, and a dropout layer with a dropout rate of 0.3 fuses the time series while learning temporal patterns. Finally, the GRU outputs are fed to two consecutive linear layers, again with shared weights for each time step. The first linear layer has 128 neurons and a ReLU activation function, while the second layer, which serves as the final output, uses a sigmoid activation function. The model is depicted in Fig. 3. The hidden state of the GRU is initialized for each batch with random weights taken from a standard normal distribution. The model is trained using the Adam optimizer [37] with a learning rate of $5 \times 10^{-4}$, a batch size of 128, and the mean squared error loss.

The extracted time series (cf. Section II) undergo further processing before being passed to the RNN model. The optical NDVI time series serve as both the label and the model input. For each time series, a subset of values is randomly selected as the label, while the remaining values are used as input. We use $66\% = 2/3$ of the values as labels and the remaining ones as model inputs to facilitate model learning. This simulates frequent cloud cover and showed a good performance in preliminary tests compared to using only 33% or 50% of the values as labels. For each training epoch, the selection as label or model input is different to make the model more robust and avoid overfitting.

From the 18-month long time series, we extract sequences with a length between one and six months to train a versatile model, which can be used for a variety of use cases. The extracted sequence length is a balance between data demands and practicality: extremely short sequences are avoided as they lack sufficient usable data, while excessively long sequences

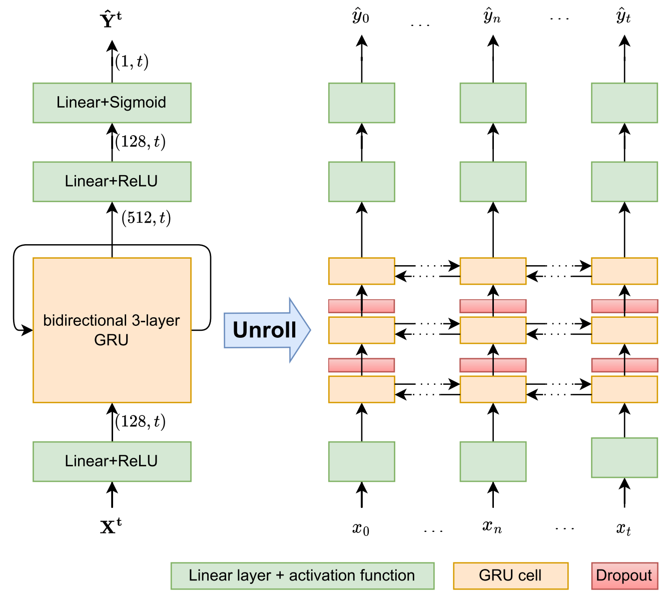

$\hat{\mathbf{Y}}^{\mathbf{t}}$
$(1, t)$
Linear+Sigmoid
$(128, t)$
Linear+ReLU
$(512, t)$
bidirectional 3-layer GRU
Unroll
$(128, t)$
Linear+ReLU
$\mathbf{X}^{\mathbf{t}}$
$\hat{y}_0$ ... $\hat{y}_n$ ... $\hat{y}_t$
$x_0$ ... $x_n$ ... $x_t$
Linear layer + activation function
GRU cell
Dropout

Fig. 3. Overview of the RNN model architecture. To the right is the unrolled version. $\mathbf{X}^t$ denotes the input time series, and $\hat{\mathbf{Y}}^{\mathbf{t}}$ denotes the predicted or rather the fused time series. The numbers denote the feature dimensions with the variable sequence length $t$.

demand impractical data resources impeding easy applicability when deploying the model.

From these sequences, all the steps that contain no values for all the input features or the label are removed. This leads to shorter sequence lengths, which improved performance in preliminary experiments.

All the model inputs are normalized as well as the optical NDVI labels. Model inputs undergo $Z$-standardization using the mean and standard deviation from the training set. Initially, the labels are clipped to the range $[-1, 1]$. Subsequently, they undergo min–max normalization to ensure that their values fall within the interval $[0.2, 0.8]$, which corresponds to the linear region of the sigmoid activation function. The day of the year (DOY) is not put directly into the model, but $\sin(\text{DOY})$ and $\cos(\text{DOY})$ are used to avoid sudden jumps at the end of the year and to capture cyclic patterns of the time series more efficiently.

## IV. Experiments

To evaluate our approach, we run multiple experiments: we compare the model performance on the spatially split test set for different input modalities, compare it to baseline models, and evaluate the transferability to different times by using the temporal split of the dataset.

We train three GRU models with our time-series dataset to compare different input data.

1) *NDVI-Fusion:* This is our proposed method and model relying on the optical NDVI and SAR-derived NDVI as well as the DOY.
2) *Optical-Only:* This is a model relying only on the optical NDVI and DOY to compare the effect of adding the SAR-estimated NDVI to the optical one. This model can only learn temporal characteristics of the time series and is, therefore, expected to be worse, especially for long gaps in the NDVI time series.
3) *NDVI-Backscatter:* This is a model that uses the $\sigma°$ SAR backscatter instead of the SAR-derived NDVI to evaluate whether the single scene SAR-to-NDVI translation is beneficial or can be omitted.

To assess the efficacy of our proposed model, we benchmark it against two baselines: first, a simple linear interpolation, and second, a histogram-based gradient boosting regression tree (hGBRT) model.

The linear interpolation approach uses the daily values of the time-series dataset (cf. Section II). As such, clouds are simulated by masking two-thirds of the available optical NDVI values, and only one-third of the values are used for the actual interpolation. The unmasked values are interpolated and compared with masked values to calculate the performance metrics.

The hGBRT model is a more complex machine learning approach similar to the lightGBM model [38]. We have opted for hGBRT because of its inherent capability to handle missing values, its enhanced computational efficiency compared to standard gradient boosting regression trees for large datasets, and its performance comparable to more complex deep learning models [39]. The model is tasked to predict the current NDVI value utilizing the previous optical and SAR-derived NDVI values, along with the DOY.

One drawback of the hGBRT model is its requirement for constant sequence lengths and a uniform temporal grid. Consequently, we select weekly values to mitigate the issue of excessive missing data that daily values would present. These values are extracted from intervals spanning six months, selected through a sliding window approach applied to the entire 1.5-year-long sequences. We chose six-month intervals under the assumption that a model with access to more extensive data is likely to achieve better performance compared with a model trained only with one or three months of data.

For each week, the available data from each source are averaged, or the values are left empty in the absence of data. This approach, however, leads to imperfect modeling due to reduced temporal accuracy and the inability to accurately reflect sharp data changes. Furthermore, a breakdown of the performance for differently long gaps in the data during evaluation is not accurately possible because of the lower temporal resolution of the aggregated weekly values. For instance, a gap of one to six days may exist between two aggregated weekly values or a missing weekly value could indicate a gap ranging from 8 to 20 days in the daily time series.

For performance evaluation, we use two dataset splits. First, a spatial split for general performance evaluation and assessing how well the model is able to generalize to different areas. To this end, we use the same test regions as the SEN12TP dataset but the full 1.5-year-long time series. Second, we implement a temporal split for our data: the first 12 months (October 2019 to September 2020) are used to train the model, encompassing the temporal characteristics of an entire year. The subsequent six months (October 2020 to March 2021) serve as the testing period. This approach mirrors a real-world scenario, where a

TABLE I
COMPARISON OF THE PERFORMANCE OF DIFFERENT MODELS USING THE SPATIALLY SPLIT TEST SET DATA

| Method | MAE | RMSE | R$^2$ |
|---|---|---|---|
| linear interpolation | 0.0482 | 0.0843 | 0.9190 |
| hGBRT baseline | 0.0675 | 0.1081 | 0.8680 |
| Optical-Only | 0.0513 | 0.0882 | 0.9115 |
| NDVI-Backscatter | 0.0517 | 0.0874 | 0.9130 |
| NDVI-Fusion | 0.0478 | 0.0806 | 0.9261 |

model is trained on currently available data and then applied to future data posttraining. Identical to training, 66% of the optical NDVI values of each time series are used as labels to evaluate the performance and the remaining 33% as model inputs.

For all the experiments, we report the MAE, the root-mean-square error (RMSE), and the coefficient of determination $R^2$.

## V. Results

To evaluate our fusion approach, we first evaluate it numerically on the globally distributed test scenes (see Section V-A). Then, we show the resulting fused time series for two example areas in detail (see Section V-B).

### A. *Quantitative Results*

To show the good performance of our approach, we apply the trained fusion model on the test set data and calculate the error between optical NDVI and the fused NDVI. The test data are comprised of time series of 124 globally distributed regions, the same as the test regions of the SEN12TP dataset [19]. As a baseline to our model, we use linear interpolation of the optical data as well as an hGBRT, a machine learning model able to handle missing data (cf. Section IV).

Our fusion approach *NDVI-Fusion* achieves a very low error, with an MAE of 0.0478. Using our fusion model only with the optical NDVI time series or the optical NDVI together with the $\sigma^\circ$ SAR backscatter results in a slightly higher MAE of 0.0513 and 0.0517, respectively. For the two baselines, we have a higher MAE of 0.0675 of the hGBRT model, and a slightly higher MAE of 0.0482 using linear interpolation. The relative order of the MAE of the compared models is the same for the RMSE and $R^2$ score, as shown in Table I.

To demonstrate that our fusion approach can generalize from one year to another and, therefore, be used in a real-world scenario, we train a model using the temporally split dataset. For this, we train the model using data from all the areas of the first year and then test it using data from the remaining six months. This ensures that all the testing data follow the training data temporally, thereby simulating a real-world scenario where the model, trained on current data, is subsequently applied to future data. Compared to the model trained with spatially split data, we see a slight performance decrease; the MAE increases from 0.0478 to 0.0498, as illustrated in Table II.

TABLE II
PERFORMANCE OF *NDVI-FUSION* MODEL ON TEMPORALLY SPLIT DATA

| data split | MAE | RMSE | R$^2$ |
|---|---|---|---|
| spatial | 0.0478 | 0.0806 | 0.9261 |
| temporal | 0.0498 | 0.0855 | 0.9017 |

For comparison, the performance using the spatial split from Table I is reported.

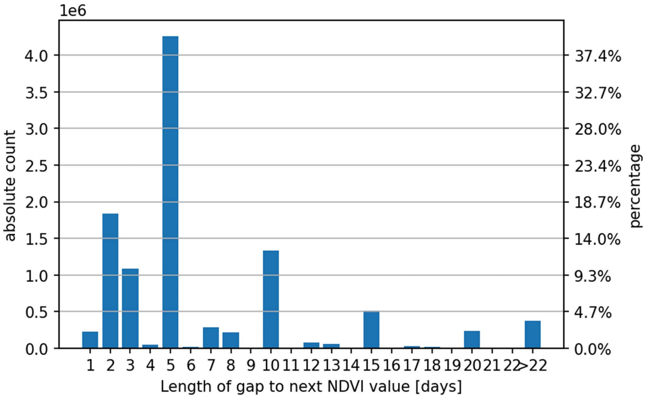


Fig. 4. Distribution of the gap lengths of the optical NDVI time series in the test set used for performance calculation. Gaps with a length of 5, 10, 15, and 20 days are more common because of the five-day revisit of the Sentinel-2 constellation.

The distribution of gap lengths is unbalanced: most gaps are rather short with 70% of the gaps being shorter than six days, as shown in Fig. 4. The peaks for gap lengths with a duration as multiple of five days are due to the five-day revisit of the Sentinel-2 constellation [40]. Because of this imbalanced distribution and the importance of filling especially longer gaps, we determine the predicted error in conjunction with the length of the optical gap or rather how near the next optical NDVI value is. For this, we calculate for each optical NDVI used as label the gap length as the minimum distance to the next or previous optical NDVI used as model input. We do not include the hGBRT model performance here, because the aggregation step from daily to weekly NDVI values only allows rather broad gap length classes: the weekly values can have a distance of one to six days gaps between them and having no value for a week could signify an NDVI data gap between 7 and 20 days.

The evaluation shows that the error of all the models increases with an increasing gap length. The increase, however, is different for the different models: linear interpolation is becoming significantly worse for long gaps ($\geq$ 20 days long), whereas the fusion approach only shows a mild increase, as listed in Table III.

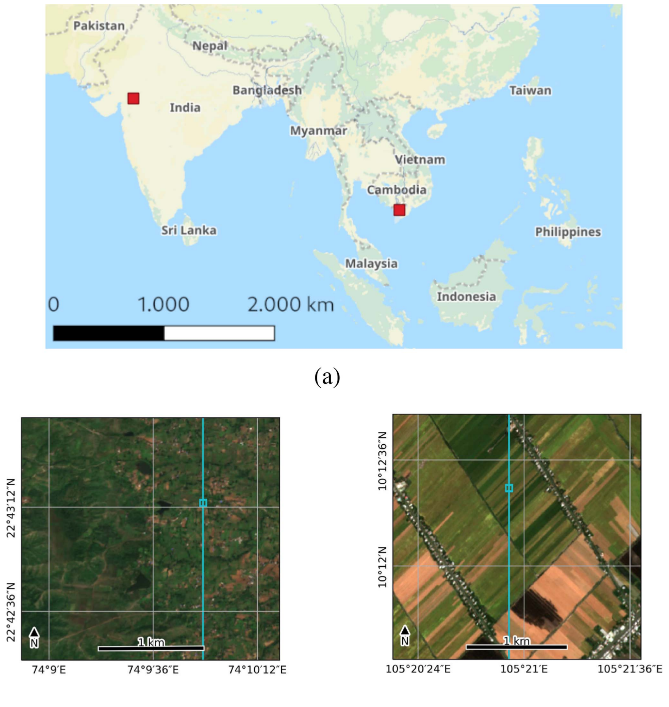
### B. *Qualitative Results*

To demonstrate our fusion approach visually and present qualitative results, we pick two example areas of 124 areas of the test set. They are chosen to demonstrate the lower and upper performance bounds: one area located in the south of Vietnam has one of the worst test performances and highest errors, whereas the other area has a very low error and is located in India. In the subsequent text, these are referred to as the *Vietnam* and *India* example areas, respectively. The location of both the areas, along with an RGB image, is presented in Fig. 5.

TABLE III
COMPARISON OF THE MODEL'S PERFORMANCES FOR DIFFERENTLY LONG GAPS IN THE OPTICAL DATA FOR THE SPATIALLY SPLIT TEST SET DATA

|  |  | gap length [days] |  |  |  |  |
|---|---|---|---|---|---|---|
| Metric | Method | $< 5$ | 5–9 | 10–14 | 15–19 | $\geq 20$ |
| MAE | Interpolation | 0.035 | 0.040 | 0.047 | 0.056 | 0.088 |
|  | Optical-Only | 0.034 | 0.040 | 0.048 | 0.058 | 0.091 |
|  | NDVI-Fusion | 0.034 | 0.040 | 0.046 | 0.054 | 0.076 |
|  | temporal split | 0.035 | 0.042 | 0.050 | 0.060 | 0.082 |
| RMSE | Interpolation | 0.063 | 0.070 | 0.080 | 0.091 | 0.137 |
|  | Optical-Only | 0.059 | 0.067 | 0.080 | 0.092 | 0.139 |
|  | NDVI-Fusion | 0.059 | 0.066 | 0.077 | 0.087 | 0.116 |
|  | temporal split | 0.062 | 0.072 | 0.084 | 0.097 | 0.126 |
| $R^2$ | Interpolation | 0.949 | 0.941 | 0.924 | 0.906 | 0.809 |
|  | Optical-Only | 0.955 | 0.945 | 0.925 | 0.904 | 0.798 |
|  | NDVI-Fusion | 0.955 | 0.946 | 0.930 | 0.914 | 0.860 |
|  | temporal split | 0.919 | 0.919 | 0.900 | 0.876 | 0.823 |

Performance of the NDVI-Fusion model is also reported for the temporally split test data.



(a)

(b)

(c)

Fig. 5. Location of two example areas used for demonstrating our time-series fusion approach marked red on a world map (top) and Sentinel-2 true-color images of these two locations (bottom). The vertical line and the box in cyan denote which data are used to display exemplary time-series progression. (a) Location of the two example areas. (b) *India*, 2019-10-12. (c) *Vietnam*, 2020-02-25.

All the plots are created by masking all the optical NDVI values in a sliding window approach: going over each NDVI value of the time series, the NDVI is masked, the fusion model is run, and the prediction of this date is saved. This ensures that for each predicted value, the optical NDVI of that date was not seen by the model.

Analyzing the time series of a single pixel, whose location is shown as cyan square in Fig. 5, shows that the SAR-derived NDVI values have a similar behavior, but applying the fusion approach and augmenting the SAR derived with the optical NDVI values result in a high agreement between optical NDVI values of a date and the prediction at that day. The complex and quickly changing NDVI course of the *Vietnam* example is modeled very well with low errors, even though the optical NDVI values are often not dense and have long gaps. The *India* example area has very dense optical values most of the years; only around August to October, does the NDVI increase and become rather sparse. This is very well reflected in the fused NDVI. Both examples are shown in Fig. 6.

To visualize the performance, we plot the NDVI of not only a single selected pixel but also a whole image row, denoted as a cyan vertical line in Fig. 5. The optical data have many gaps due to clouds and cloud shadows in the data, which also results in gaps in the calculated error. In contrast, the fused NDVI is dense without long gaps and results in a low error for almost all the dates. For the *India* example area, the NDVI is sufficiently dense for most of the year but between June and October, almost no cloud-free pixels could be acquired. Our fusion approach can estimate reasonable NDVI values and achieves a low error for the few cloud-free optical dates. Only for the *Vietnam* example area, a few dates in December have a higher error, which coincides with a drastic increase of the NDVI at these dates. Visualizations for both the areas are depicted in Fig. 7.

Finally, we compare our fusion approach for a series of images. For 15 partly cloud-free images of the *Vietnam example area* between December 2019 and February 2020, we calculate the fused NDVI and compare it to the optical one. We achieve for almost all the scenes a low error but can fill all the gaps due to clouds, as shown in Fig. 8.

All these figures show the high performance of our approach and the high accordance with the optical NDVI.

## VI. DISCUSSION

We presented a fusion approach to fuse time series of different sources to construct dense and accurate time series. This approach integrates sparse but accurate optical NDVI time series with denser, albeit less precise, SAR-estimated NDVI time series, effectively addressing gaps caused by clouds and cloud shadows. Both quantitatively and qualitatively, we show the high performance and low error of the fused NDVI time series. Gaps can be filled, and the resulting time series have a high similarity with the optical values.

Our findings reveal that prefusion NDVI estimates from individual SAR scenes are preferable to the direct use of $\sigma^\circ$ SAR backscatter (cf. Table I). This efficacy stems primarily from two factors: first, the SAR-to-NDVI translation leverages spatial neighborhood information of the SAR data, enhancing NDVI estimation accuracy by considering each pixel's spatial context; second, SAR-derived NDVI values are less impacted by speckle noise, a prevalent issue in SAR backscatter time series, resulting in a cleaner dataset for fusion.

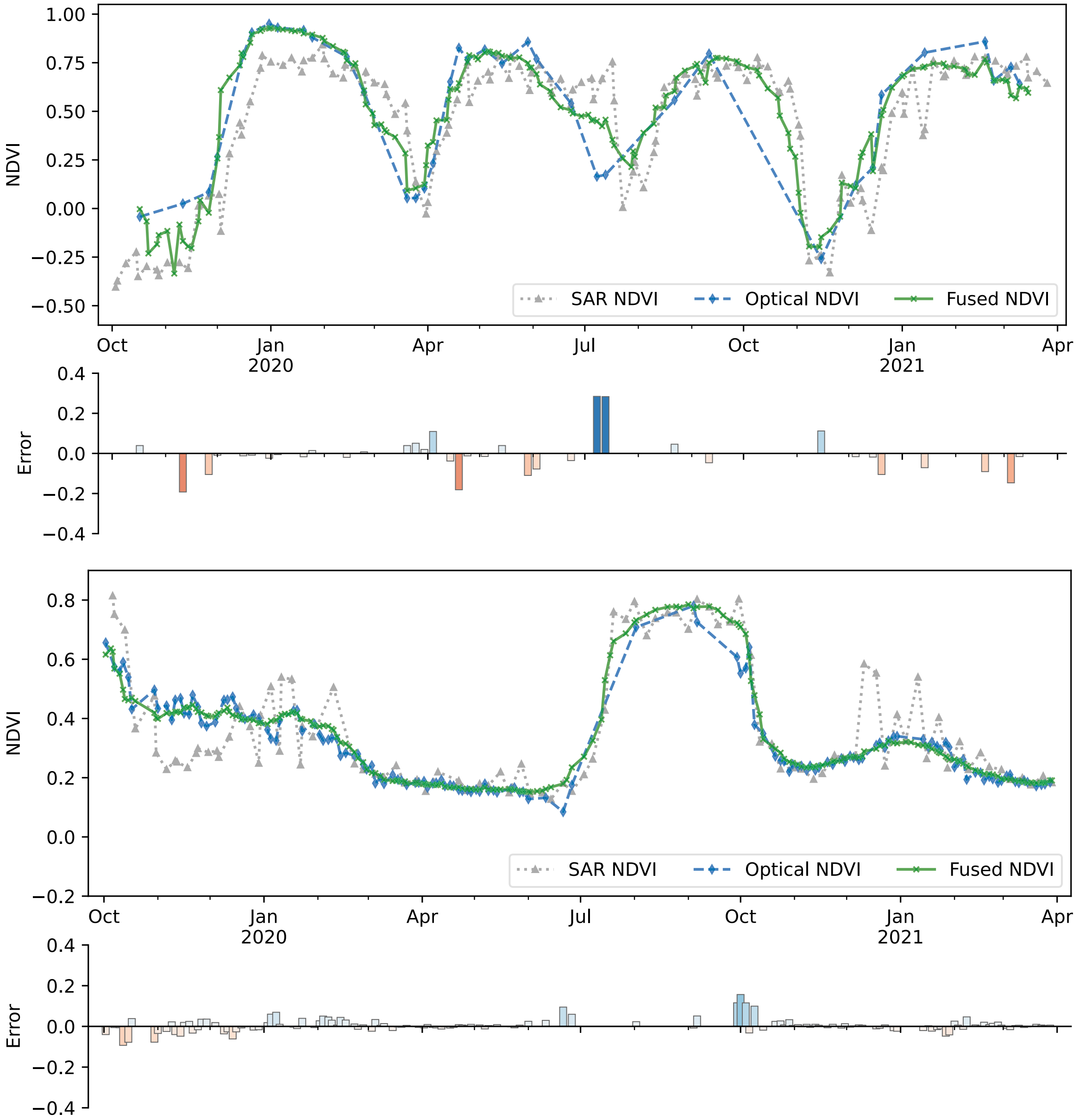
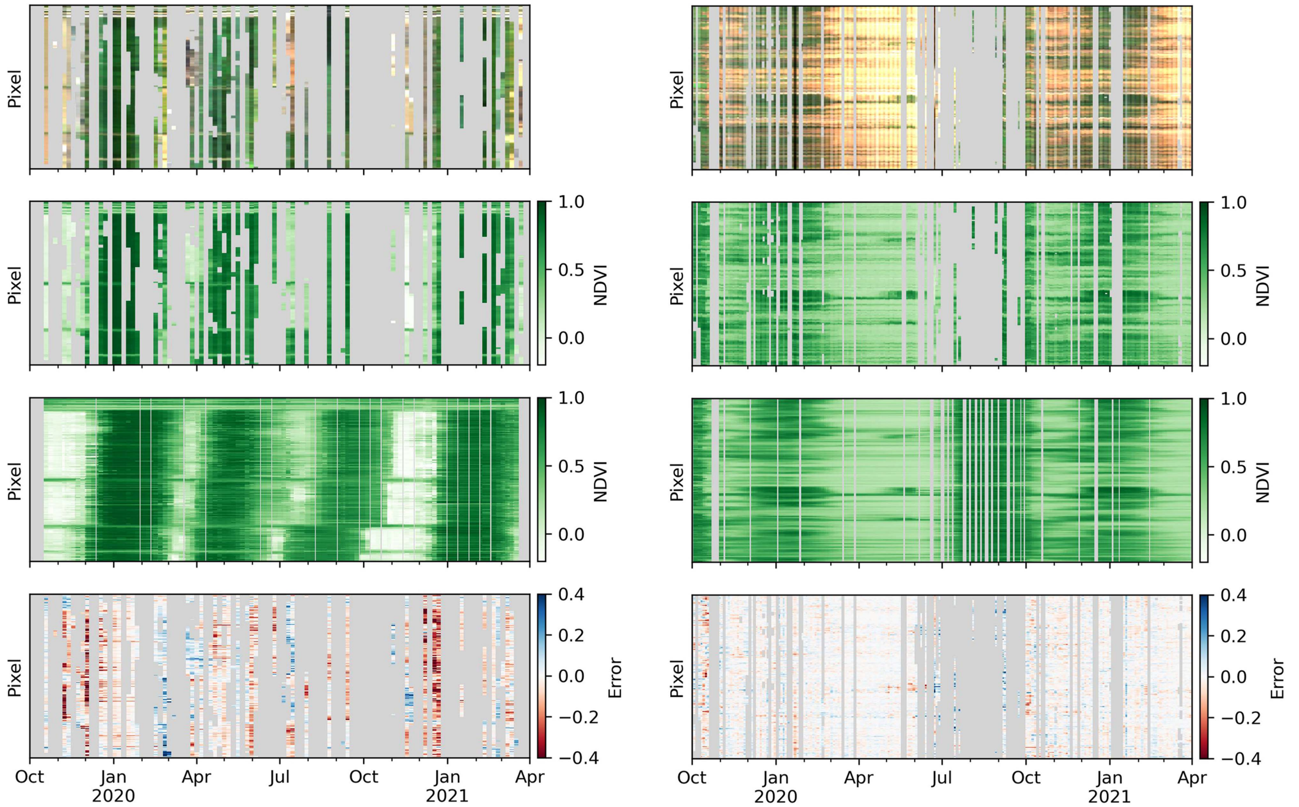


Fig. 6. Comparison of the NDVI time series and associated errors for optical, SAR-derived, and fused NDVI data for a cropland pixel within the *Vietnam* (top) and *India* (bottom) example areas. The location of the used pixels is given in Fig. 5. The fused NDVI time series (green) is dense and is closely aligned with the optical NDVI measurements, illustrating the effectiveness of the fusion approach. In contrast, the optical NDVI time series (blue) suffers from irregular acquisition intervals due to cloud cover. Meanwhile, the SAR-estimated NDVI (gray) is consistent in temporal coverage but displays a higher level of uncertainty and noise.

As shown numerically in Table III, a low error is achieved for short as well as for long gaps in the optical data. Nonetheless, there are instances showing limitations of our approach. First, our model struggles to accurately predict abrupt changes in the time series due to agricultural practices or vegetation burns. This can be evidenced in the *India* example area, where higher errors were observed during rapid NDVI fluctuations in July and October 2020, as depicted in Fig. 6. Similarly, for the *Vietnam* example area, on multiple occasions, a sudden change results in an increased error, for example, in April or July 2020.

These inaccuracies can be attributed to two main reasons: first, the necessity for high temporal precision, where even slight timing discrepancies can lead to significant errors, and second, a lack of data with rapid vegetation transitions. As most of the vegetation is rather stable, only very little optical data are available with sudden changes in the NDVI. To mitigate this, we could expand our dataset to include more instances with high and quick vegetation and NDVI changes. However, it might be computationally expensive to acquire a sufficient amount of suitable time series, as there has to be a rapid vegetation change and coinciding available optical data.

Another case in which the method has a lower accuracy is in flooded areas or for water bodies due to the interaction of microwave signals with water surfaces. The reflection of microwaves off water results in minimal backscatter signals and,



(a) (b)

Fig. 7. Comparison of the NDVI of one image row over time for the example areas. The RGB (top row) and the NDVI derived from optical data (second row) have many gaps due to cloud coverage. In contrast, the fused NDVI (third row) is almost gap-free and is closely aligned with optical NDVI as evidenced by the low error between fused and optical NDVI (bottom row). The location of the row is shown in Fig. 5. Missing data are displayed in gray due to missing image retrievals or masked clouds and cloud shadows. (a) *Vietnam* example area. (b) *India* example area.

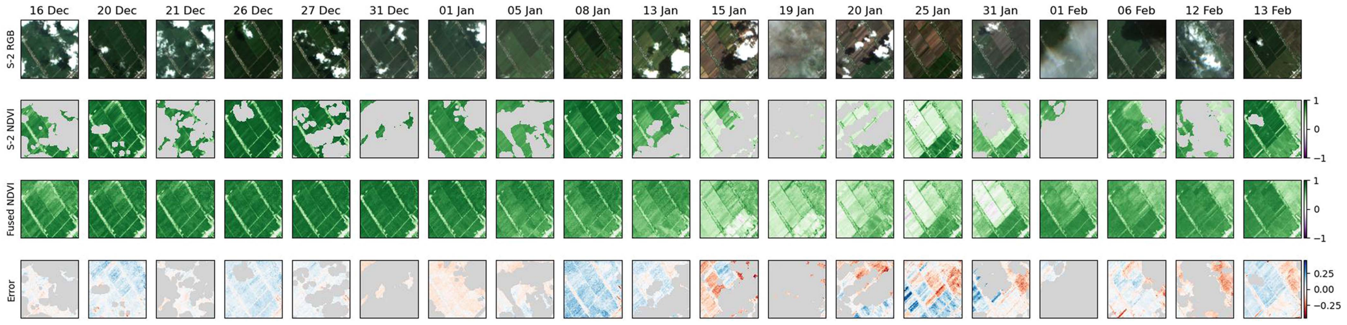


Fig. 8. Overview over the RGB and NDVI images between December 2019 and February 2020 for the *Vietnam* example area. The top row shows the optical images, and the second row shows the NDVI images masked for clouds and cloud shadows. The fused images (third row) are only shown for all the dates, where S-2 data were captured. The bottom row shows the error between optical and fused NDVI images. Gray areas are masked because of detected clouds or cloud shadows.

consequently, a lack of usable information for NDVI translation in the received signal. Without information about the earth's surface in the SAR data, the translation to NDVI values is not accurately possible, which results in an increased error of the SAR-derived NDVI time series and, therefore, also an increased error in the fused NDVI series in such environments. This phenomenon likely explains the elevated errors observed on January 25 in Fig. 8, as the rice fields in the studied area undergo regular flooding [41].

We also acknowledge that the optical NDVI, our standard of truth, has some inherent uncertainty and inaccuracy, arising from factors such as atmospheric correction methodologies, topographic, solar, and viewing angles, and adjacency effects. We did not apply corrections for these effects to remain as close as possible to the original data. Moreover, our cloud and shadow masking, while effective, is not infallible, with occurrences of both false positives and false negatives. A manual annotation of clouds is infeasible, due to the sheer volume of images. However, employing a more sophisticated temporal cloud detection algorithm, such as in [42], could improve cloud detection accuracy, albeit at the cost of increased computational demand and more complex data processing.

## VII. Conclusion

To acquire dense NDVI time series despite frequent cloud cover, we propose a novel method, which combines sparse optical NDVI time series and denser but less accurate SAR-derived NDVI time series using a deep learning GRU. A large and globally distributed dataset is used, comprised of 283 000 images from 1206 locations taken between September 2019 and April 2021.

The proposed method yields a low MAE of 0.0478 outperforming approaches using only optical data or additionally SAR backscatter $\sigma^\circ$ time series. It demonstrates consistent accuracy, even with extended gaps in the data, and generalizes well across different regions and times. Visual assessments confirm a high similarity to optical values across various spatial scales, supporting the reliability of denser NDVI time series for downstream applications despite cloud cover challenges.

## References

[1] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, “Monitoring vegetation systems in the great plains with ERTS,” in *Proc. 3rd Earth Resour. Technol. Satell. Symp.*, 1974, pp. 309–317.

[2] J. Xiong et al., “Automated cropland mapping of continental Africa using Google Earth Engine cloud computing,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 126, pp. 225–244, 2017.

[3] M. Schwarz et al., “Satellite-based multi-annual yield models for major food crops at the household field level for nutrition and health research: A case study from the Nouna HDSS, Burkina Faso,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 117, 2023, Art. no. 103203.

[4] S. Hislop, S. Jones, M. Soto-Berelov, A. Skidmore, A. Haywood, and T. H. Nguyen, “A fusion approach to forest disturbance mapping using time series ensemble techniques,” *Remote Sens. Environ.*, vol. 221, pp. 188–197, 2019.

[5] S. Li, L. Xu, Y. Jing, H. Yin, X. Li, and X. Guan, “High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, 2021, Art. no. 102640.

[6] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, “A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter,” *Remote Sens. Environ.*, vol. 91, no. 3–4, pp. 332–344, 2004.

[7] M. Claverie et al., “The harmonized Landsat and Sentinel-2 surface reflectance data set,” *Remote Sens. Environ.*, vol. 219, pp. 145–161, 2018.

[8] F. Gao, J. Masek, M. Schwaller, and F. Hall, “On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.

[9] Q. Wang and P. M. Atkinson, “Spatio-temporal fusion for daily Sentinel-2 images,” *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.

[10] H. McNairn and J. Shang, “A review of multitemporal synthetic aperture radar (SAR) for crop monitoring,” in *Multitemporal Remote Sensing: Methods and Applications*, vol. 20, Berlin, Germany: Springer, 2016, pp. 317–340.

[11] C.-a Liu, Z.-x Chen, Y. Shao, J.-s Chen, T. Hasi, and H.-z. Pan, “Research advances of SAR remote sensing for agriculture applications: A review,” *J. Integrative Agriculture*, vol. 18, no. 3, pp. 506–525, 2019.

[12] M. Hosseini, H. McNairn, S. Mitchell, L. D. Robertson, A. Davidson, and S. Homayouni, “Synthetic aperture radar and optical satellite data for estimating the biomass of corn,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 83, 2019, Art. no. 101933.

[13] S. Mermoz, M. Réjou-Méchain, L. Villard, T. Le Toan, V. Rossi, and S. Gourlet-Fleury, “Decrease of L-band SAR backscatter with biomass of dense forests,” *Remote Sens. Environ.*, vol. 159, pp. 307–317, 2015.

[14] M. Pichierri, I. Hajnsek, S. Zwieback, and B. Rabus, “On the potential of polarimetric SAR interferometry to characterize the biomass, moisture and structure of agricultural crops at L-, C- and X-bands,” *Remote Sens. Environ.*, vol. 204, pp. 596–616, 2018.

[15] W. Zhao, Y. Qu, J. Chen, and Z. Yuan, “Deeply synergistic optical and SAR time series for crop dynamic monitoring,” *Remote Sens. Environ.*, vol. 247, 2020, Art. no. 111952.

[16] A. Garioud, S. Valero, S. Giordano, and C. Mallet, “Recurrent-based regression of Sentinel time series for continuous vegetation monitoring,” *Remote Sens. Environ.*, vol. 263, 2021, Art. no. 112419.

[17] P. B. Weerakody, K. W. Wong, G. Wang, and W. Ela, “A review of irregular time series data handling with gated recurrent neural networks,” *Neurocomputing*, vol. 441, pp. 161–178, 2021.

[18] T. Roßberg and M. Schmitt, “A globally applicable method for NDVI estimation from Sentinel-1 SAR backscatter using a deep neural network and the SEN12TP dataset,” *J. Photogrammetry Remote Sens. Geoinf. Sci.*, vol. 91, pp. 171–188, 2023.

[19] T. Roßberg and M. Schmitt, “SEN12TP—Sentinel-1 and -2 images, timely paired,” 2023. [Online]. Available: https://zenodo.org/records/7342060

[20] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google Earth Engine: Planetary-scale geospatial analysis for everyone,” *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017.

[21] A. Zupanc, “Improving cloud detection with machine learning,” 2017. [Online]. Available: https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13

[22] M. Aleksandrov, A. Zupanc, M. Lubej, and Ž Lukšič, “Sentinel hub’s Sentinel-2 cloud detector,” 2023. [Online]. Available: https://github.com/sentinel-hub/sentinel2-cloud-detector/blob/master/examples/sentinel2-cloud-detector-example.ipynb

[23] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, “Aggregating cloud-free Sentinel-2 images with Google Earth Engine,” *ISPRS Ann. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W7, pp. 145–152, 2019.

[24] C. Requena-Mesa, V. Benson, M. Reichstein, J. Runge, and J. Denzler, “EarthNet2021: A large-scale dataset and challenge for Earth surface forecasting as a guided video prediction task,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2021, pp. 1132–1142.

[25] D. Peressutti, “How to co-register temporal stacks of satellite images,” 2023. [Online]. Available: https://medium.com/sentinel-hub/how-to-co-register-temporal-stacks-of-satellite-images-5167713b3e0b

[26] G. D. Evangelidis and E. Z. Psarakis, “Parametric image alignment using enhanced correlation coefficient maximization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1858–1865, Oct. 2008.

[27] M. C. Hansen, R. S. Defries, J. R. G. Townshend, and R. Sohlberg, “Global land cover classification at 1 km spatial resolution using a classification tree approach,” *Int. J. Remote Sens.*, vol. 21, no. 6–7, pp. 1331–1364, 2000.

[28] H. Balzter, B. Cole, C. Thiel, and C. Schmullius, “Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests,” *Remote Sens.*, vol. 7, no. 11, pp. 14876–14898, 2015.

[29] J. Alvarez-Mozos, J. Villanueva, M. Arias, and M. Gonzalez-Audicana, “Correlation between NDVI and Sentinel-1 derived features for maize,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 6773–6776.

[30] A. -K. Holtgrave, N. Röder, A. Ackermann, S. Erasmi, and B. Kleinschmit, “Comparing Sentinel-1 and -2 data and indices for agricultural land use monitoring,” *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 2919.

[31] P. -L. Frison et al., “Potential of Sentinel-1 data for monitoring temperate mixed forest phenology,” *Remote Sens.*, vol. 10, no. 12, 2018, Art. no. 2049.

[32] G. Scarpa, M. Gargiulo, A. Mazza, and R. Gaetano, “A CNN-based fusion method for feature extraction from Sentinel data,” *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 236.

[33] R. Filgueiras, E. C. Mantovani, D. Althoff, E. I. F. Filho, and F. F. da Cunha, “Crop NDVI monitoring based on Sentinel 1,” *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1441.

[34] E. P. dos Santos, D. D. da Silva, C. H. do Amaral, E. I. Fernandes-Filho, and R. L. S. Dias, “A machine learning approach to reconstruct cloudy affected vegetation indices imagery via data fusion from Sentinel-1 and Landsat 8,” *Comput. Electron. Agriculture*, vol. 194, 2022, Art. no. 106753.

[35] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

[36] K. Cho et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” 2014, doi: 10.3115/v1/D14-1179.

[37] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, *arXiv:1412.6980.*

[38] G. Q. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” in *Proc. Int. Conf. Neural Inf. Process. Syst.* 2017, pp. 3146–3154.

[39] S. Elsayed, D. Thyssens, A. Rashed, H. S. Jomaa, and L. Schmidt-Thieme, “Do we really need deep learning models for time series forecasting?” 2021, *arXiv:2101.02118.*

[40] *Sentinel-2 User Handbook*, Eur. Space Agency, Paris, France, 2015.

[41] V.-H. Nguyen et al., "An assessment of irrigated rice cultivation with different crop establishment practices in Vietnam," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 401.

[42] Z. Zhu and C. E. Woodcock, "Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change," *Remote Sens. Environ.*, vol. 152, pp. 217–234, 2014.

**Thomas Roßberg** (Graduate Student Member, IEEE) received the Dipl.-Ing. degree in information systems engineering from the Dresden University of Technology, Dresden, Germany, in 2020. He is currently working toward Ph.D. degree in remote sensing with the University of the Bundeswehr Munich, Neubiberg, Germany.

He was a Research Assistant with the University of Applied Sciences, Munich, under the supervision of Prof. Michael Schmitt. His research interests include the use of synthetic aperture radar data to reduce the negative impact of clouds on normalized difference vegetation index images and time series.

**Michael Schmitt** (Senior Member, IEEE) received the Dipl.-Ing. (Univ.) degree in geodesy and geoinformation, the Dr.-Ing. degree in remote sensing, and the Habilitation degree in data fusion from the Technical University of Munich (TUM), Munich, Germany, in 2009, 2014, and 2018, respectively.

Since 2021, he has been the Chair for Earth Observation with the Department of Aerospace Engineering, University of the Bundeswehr Munich, Neubiberg, Germany. Before that, he was a Professor of Applied Geodesy and Remote Sensing with the Department of Geoinformatics, Munich University of Applied Sciences, Munich. From 2015 to 2020, he was a Senior Researcher and the Deputy Head with the Professorship for Data Science in Earth Observation, TUM. In 2019, he was an Adjunct Teaching Professor with the Department of Aerospace and Geodesy, TUM. In 2016, he was a Guest Scientist with the University of Massachusetts, Amherst, MA, USA. His research interests include image analysis and machine learning applied to the extraction of information from multimodal remote sensing observations, in particular, remote sensing data fusion with a focus on synthetic aperture radar and optical data.

Dr. Schmitt is a Co-Chair of the Working Group "Active Microwave Remote Sensing" of the International Society for Photogrammetry and Remote Sensing and an Active Member of the Image Analysis and Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society. He is a Reviewer for a number of renowned international journals and conferences and has received several best reviewer awards. He is an Associate Editor for *IEEE Geoscience and Remote Sensing Magazine*.