

# ON THE JOINT EXPLOITATION OF OPTICAL AND SAR SATELLITE IMAGERY FOR GRASSLAND MONITORING

A. Garioud<sup>1,2,\*</sup>, S. Valero<sup>2</sup>, S. Giordano<sup>1</sup>, C. Mallet<sup>1</sup>

<sup>1</sup> Univ. Gustave Eiffel, IGN-ENSG, LaSTIG, Saint-Mande, FRANCE - [firstname.lastname@ign.fr](mailto:firstname.lastname@ign.fr)

<sup>2</sup> CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS, Toulouse, FRANCE - [silvia.valero@cesbio.cnes.fr](mailto:silvia.valero@cesbio.cnes.fr)

## Commission III, WG III/6

**KEY WORDS:** Time-series, Data fusion, Sentinels, Regression, NDVI, Recurrent Neural Networks, Gap filling, Ablation Study.

### ABSTRACT:

Time series of optical and Synthetic Aperture RADAR (SAR) images provide complementary knowledge about the cover and use of the Earth surface since they exhibit information of distinct physical nature. They have proved to be particularly relevant for monitoring large areas with high temporal dynamics and related to significant ecosystem services. Grasslands are such crucial surfaces, both in terms of economic and environmental issues and the automatic and frequent monitoring of their agricultural practices is required for many purposes. To address this problem, the deep-based SenDVI framework is presented. SenDVI proposes an object-based methodology to estimate NDVI values from Sentinel-1 SAR observations and contextual knowledge (weather, terrain). Values are regressed every 6 days for compliance with monitoring purposes. Very satisfactory results are obtained with this low-level multimodal fusion strategy ( $R^2=0.84$  on a Sentinel-2 tile). Finer analysis is however required to fully assess the relevance of each modality (Sentinel-1, Sentinel-2, weather, terrain) and feature sets and to propose the simplest conceivable framework. Results show that not all features are necessary and can be discarded while others have a mandatory contribution to the regression task. Moreover, experiments prove that accuracy can be improved by not saturating the network with non-essential information (among contextual knowledge in particular). This allows to move towards more operational solution.

## 1. INTRODUCTION AND MOTIVATION

### 1.1 Context

Accurate and regular monitoring of terrestrial biomass is today an essential need in environmental, economic, and social terms. Among vegetation, grasslands are crucial ecosystems: they are the second most represented land cover surfaces throughout the world after forests. They support many so-called *ecosystem services*, such as carbon storage, prevention of soil erosion, production of quality food, and preservation of biodiversity (Bengtsson et al., 2019, Buchmann et al., 2019). Remote sensing has proved to be a suitable tool to monitor in such a direction (Atzberger, 2013, Weiss et al., 2020). Initial efforts have explored the needs and challenges of regular monitoring of agricultural practices such as mowing, grazing or ploughing on grasslands (Garioud et al., 2019). Quantifying and determining the nature of these technical acts is important since they affect the quality of grassland ecosystem services. The key identified factor is the periodicity of monitoring. Many grasslands have a prominent part in production systems: they are intensively farmed. Grass regrowth can be favoured by nitrogen inputs or favourable climatic conditions. As a result, biomass values can return to their initial state after only about ten days following a technical act. Thus, reliable weekly measurements are required.

Satellite constellations allowing a temporal revisit of the order of a few days, coupled with a free-access policy have emerged in the recent years. Possibilities of Earth Observation applications in monitoring vegetation status have steadily increased (Frampton et al., 2013, Mercier et al., 2020). The Sentinel satellites of the European Copernicus program allow the joint exploitation of optical and SAR data with similar high spatial

resolution, with global coverage, and with a temporal revisit of 5/6 days at mid-latitudes.

### 1.2 Problem Statement

For vegetation monitoring, *e.g.*, grasslands, optical sensors are mostly preferred as recently enlightened in (Griffiths et al., 2020) and (Liu et al., 2020). These sensors capture the interaction of light with the chemical properties of the target, efficiently describing the vivacity and photosynthetic activity of the plants. However, optical data suffer from the recurring presence of clouds, resulting in numerous missing data, which can last from a few days to a few months. It prevents meeting the requirements of dense surveillance of grasslands (Kolecka et al., 2018). Most monitoring approaches integrating optical satellite data are based on the exploitation of derived vegetation indices (Ali et al., 2017, Pasqualotto et al., 2019, Tiscornia et al., 2019). Among these indices, NDVI is still widely used for its simplicity and ease of interpretation (Stendardi et al., 2019, Solano-Correa et al., 2020, Belgiu, Csillik, 2018). To deal with cloud coverage, most of the studies apply a gap filling step to reconstruct missing values in time-series with available cloudless images. Among traditional gap-filling methods, common practice relies on a linear (or splines) interpolation method between the cloudless date preceding and following the missing one. This approach has shown its effectiveness especially in classification tasks (Vuolo et al., 2017, Fauvel et al., 2020). Nevertheless, for a precise follow-up of the targets evolution and especially in the case of abrupt and rapid phenomena such as management practices on grasslands, these techniques do not allow for a proper handling of missing data.

The integration of various optical sensors from different satellites has been proposed in (Claverie et al., 2018, Li, Guo, 2018, Griffiths et al., 2019). This strategy however leads to problems

\* Corresponding author

of spatial and radiometric standardization between data sources. Despite an increase in the number of images, relying on optical imagery alone only diminishes the missing data issue. Synthetic Aperture Radar (SAR) data has therefore been increasingly exploited. As an active remote sensing source, it is capable of acquiring data day and night. The radio frequency transmission power of the SAR sensors also allows them to be cloud insensitive. On the ground, SAR sensors interact with the dielectric and geometric properties of the target. As a result, these data are affected by climatic events (rain, frost, humidity), as well as by the effects of ground relief. Although monitoring of grasslands with SAR data only shows satisfactory results (Buckley, Smith, 2010, Siegmund et al., 2016, Tamm et al., 2016, Zalite et al., 2016, Chiboub et al., 2019), most of studies are either relying on full polarimetry (giving access to the Radar Vegetation Index) or on Very High Resolution (VHR) data. Both are difficult to access and not compliant with dense monitoring requirements.

### 1.3 Motivation and contribution

The joint exploitation of optical and SAR data is necessary to monitor grasslands with high regular temporal sampling. These techniques mainly focuses on independent processing of optical and SAR data with a late fusion strategy. High-level fusion schemes are based on complex ad-hoc decision rules defined on restricted areas. They often rely on empirically tuned parameters. Few works propose a low-level fusion of optical and SAR data, mainly for *image-to-image* translations tasks (Merkle et al., 2017, He, Yokoya, 2018, Hughes et al., 2019). In a similar direction, supervised regression of optical and SAR data for vegetation parameter retrieval has been proposed in (Baghdadi et al., 2016) using Neural Networks or with Gaussian Processes in (Pipia et al., 2019), both for LAI estimation. In (Scarpa et al., 2018) Convolutional Neural Networks are adopted for NDVI S1/S2 reconstruction but focused on short time series, discarding the temporal information. In parallel, a first low-level fusion approach has been suggested in (Garioud et al., 2019) enabling the regression of S1 SAR observations to S2 NDVI values. This Recurrent Neural Network framework integrates contextual data (weather/terrain) provided from external sources. It allows the regression task to refine the satellite measurements. Nevertheless, the model relies on a number of important and sometimes difficult to access sources of information. In this paper, we target to:

- Explore the relationship between optical, SAR and contextual data by proposing a deep-based network for a regression task. The analysis is two-fold: on the basis of physical understanding and computational constraints.
- Quantify, for various feature sets, for scalability and automation purposes, gains and losses in accuracy. We aim to select the required input data for satisfactory regression results.

## 2. PROPOSED REGRESSION FRAMEWORK

### 2.1 Data

The study area is located on the South-East of France covering a Sentinel-2 tile (T31TFM). The French Land-Parcel Identification System (LPIS) is used to retrieve 23,850 parcels of grasslands from the studied area. Agricultural season from October 2016 to October 2017 is considered. For each grassland parcel,

the two types of features described in the following have been considered :

**Sentinel derived features** consist of 53 optical Level-2A S2 images from which NDVI is calculated. The pre-processing is carried out by the MAJA chain which integrates cloud and shadow masks. 60 ascending and 60 descending orbit images from S1 are available during the agricultural season.  $\sigma_0$  values (VV, VH, and VV/VH polarization), and 6-day InSAR coherence (VV and VH polarization) are derived. In order to take into account incidence angles and acquisition times, the ascending and descending orbits of Sentinel-1 are independently processed. For each parcel, mean, median, and standard deviation of the optical and SAR satellite bands are computed. Additionally, for each SAR feature, slope between each date and the previous one is given for individual parcels as well as the mean slope for parcels in a spatial neighborhood (*e.g.*, 2km) around this parcel.

**Contextual information** are derived from ancillary data and from acquisition metadata. 25 climatic variables (*e.g.* rain, temperature, frost, etc.) are extracted from the Météo France SAFRAN-ISBA dataset since weather has a strong effect on S1 signals and grasslands (Moreira et al., 2019). These variables are daily provided and computed for each parcel from the nearest weather station. 5 topographic descriptors (*e.g.* height, exposition, slope, etc.) are given with a 5 m resolution Digital Surface Model provided by IGN-France. Temporal context (*e.g.*, week number, month, season) is given to each input date, as well as time shift separating Sentinel-1 and Sentinel-2 acquisitions. The LPIS agronomic type of the grassland is also provided (22 grassland classes on the area).

Because of the prerequisites of the method described below, Sentinel-1 and contextual data are all standardized by channel. With less Sentinel-2 data available, a common temporal grid between S2 and S1 data is required for the regression task. S2 data is interpolated by the nearest neighbor method on the S1 descending acquisition grid.

### 2.2 Method

The Sentinels-NDVI (SenDVI) network is based on Recurrent Neural Networks (RNN) and follows the preliminary works described in (Garioud et al., 2019). Refined with a two-branch approach in order to integrate contextual information more accurately, the network is assessed by its capacity to reconstruct NDVI time-series on a high temporal sampling. A deep learning approach was chosen for the task since it generally offers better results than traditional machine learning approaches such as Random Forest or Support Vector Machine (Ienco et al., 2017, Reichstein et al., 2019, Kamilaris, Prenafeta-Boldú, 2018). This method allows to alleviate the issue of selecting input data, permits to explore voluminous data and solve complex regression configurations.

SenDVI aims to learn the statistical relationships among multi-sensor multi-variate time-series. Sentinel-1 features are considered as *inputs* and Sentinel-2 cloud-free NDVI values as *labels*. Cloud mask associated to each S2 date is used to select the cloudy dates among the time-series for each parcel. These dates are not considered for computing the loss (*Mean Squared Error loss function*) during the training phase since they may not be reliable labels.

The network is divided into three blocks as shown in Figure 2. The first block consists of two separate branches, one for ascending and one for descending SAR acquisitions. This block is

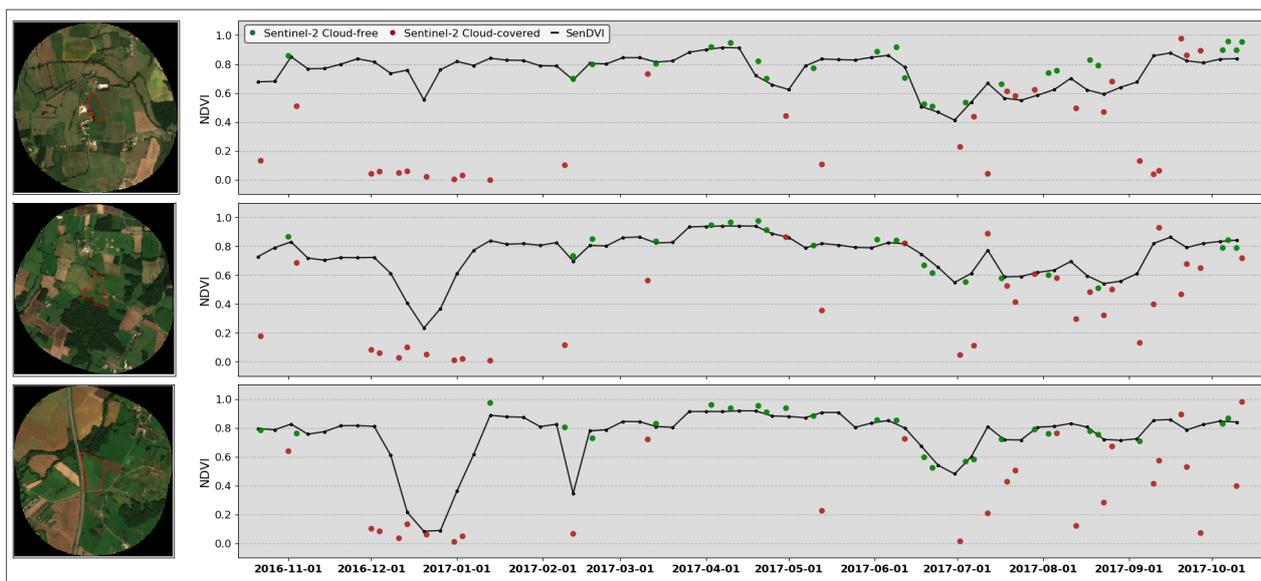


Figure 1. Comparison between predicted NDVI time-series obtained by the SenDVI network and actual Sentinel-2 NDVI values for three grassland parcels (left). Green dots indicates valid non-cloudy Sentinel-2 values while red dots are cloudy days (missing data). Black line shows the SenDVI network prediction with high temporal sampling. It follows correctly the non-cloudy values and keep high frequencies of the signal.

used to encode both SAR data and contextual data with Multi-Layer Perceptrons (MLPs) composed of successive linear regression layers, normalization, non-linear activation (*i.e.* Rectified Linear Units layer), and dropout. The outputs of each branch are merged by element-wise multiplication (Hadamard product) in order to apply contextual measurements to SAR observations. Finally, ascending and descending branches are concatenated in a single weighted feature vector for each date corresponding to the input data of the second block.

The second block performs the recurrent learning of the SAR-to-optical regression parameters for each date. The Recurrent Networks have the advantage of retaining a memory of past observations through gating mechanisms. This capability is particularly effective for time series and is suitable for vegetation measurements with periodicity. A bi-directional Gated Recurrent Unit (GRU) network was chosen. It exhibits similar results to other types of RNNs for fewer parameters (Ndikumana et al., 2018, Zhao et al., 2019).

Lastly, the third block decodes the outputs of the second block with an in-depth funnel-shaped MLP. It results in a single value of NDVI for each date.

### 2.3 Validation of SenDVI framework

The input data used to train and validate SenDVI is composed of 25 Sentinel-1 values and 92 contextual values per branch. Performance evaluation of the SenDVI network is measured with three standard regression quality metrics: coefficient of determination ( $R^2$ ), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Metrics are only calculated on dates where Sentinel-2 data was not affected by clouds.

The hyper-parameters of the model are empirically defined. A learning rate of  $10^{-3}$  is used with a batch size of 128 and the Adam optimizer is selected. The encoding MLPs are 4-layer deep and the decoding MLPs are 6-layer deep. The dropout during encoding is set to 0.2 while the dropout during decoding starts at 0.4 and is reduced for each subsequent layer. The GRU hidden size is set to 256 and is dropout-exempt. The result model has about 1.5 million parameters.

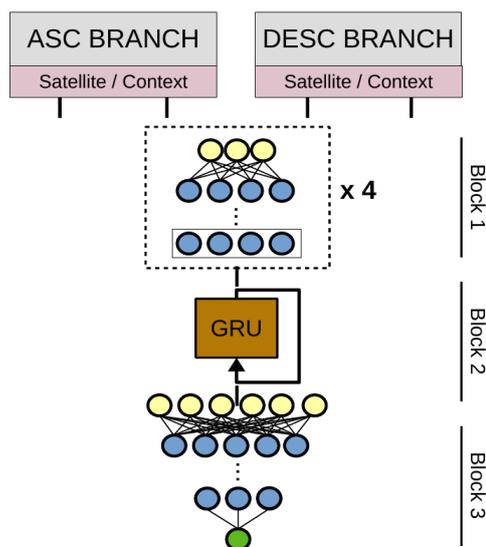


Figure 2. Schematic overview of the proposed SenDVI Network.

All SenDVI results provided hereafter are the average results obtained of 5 runs of a model configuration. Each model has a 5-fold cross-validation with 3/5 of the dataset retained for training (14,310 parcels), 1/5 for validation during training and the remaining 1/5 for testing (each of the 4,470 parcels). During each fold, the test parcels on which the metrics are calculated are different. With the use of one GPU NVIDIA RTX 2080, one run of a 5-fold cross-validation model takes about 1 hour and 20 minutes in total for 150 epochs per fold.

Although Sentinel-1 and Sentinel-2 sensors do not measure the same physical phenomena, nor under the same conditions, the SenDVI regression architecture shows a satisfactory quality for time-series reconstruction. Between S2 cloudless NDVI values and SenDVI-derived NDVI values, a mean  $R^2$  of **0.84**, MAE of **0.038** and RMSE of **0.058** were obtained for the five runs (Table 1).

	Mean	Median	Std
$R^2$	0.8410	0.8430	0.0039
MAE	0.0388	0.0385	0.0007
RMSE	0.0586	0.0583	0.0009

Table 1. SenDVI regression scores over grasslands.

Qualitative assessment can be performed through visual inspection of reconstructed time-series by looking at Figure 1. Results show very satisfactory reconstructed NDVI time-series. Our method is able to (i) reconstruct the signal despite significant temporal gaps and (ii) preserve main breakpoints, which are crucial for detecting human activity. In average for the whole dataset, only 26 images are exploitable among the 53 available. More than half of the image set over a year is obscured by clouds and therefore do not provide information on the parcel vegetation evolution. With the help of the low-level fusion from SenDVI, even when significant gaps exists (*e.g.*, >1 month, see left part of the signals in Figure 1), 60 reliable values are ensured every year. Our results are therefore compliant with the requirements of weekly measurements needed in an accurate monitoring of grasslands.

### 3. OPTICAL, SAR AND ANCILLARY DATA : BEHAVIOR AMONG REGRESSION TASK

The reconstructed NDVI time-series show satisfactory statistical as well as visual results. However, the acquisition and pre-processing steps for the various features can be tedious, cost-intensive, and therefore not in line with operational constraints. A first question arises about the need to process the large amount of input data. The second question is the assessment of which input and contextual information are mostly essential for this regression task. Therefore, the qualitative and quantitative analysis of the contribution of each feature is studied by an ablation study. Considering the two different types of inputs described in 2.1, each feature within the two categories will be removed one by one for training a new model. The same metrics as in 2.3 are computed allowing to compare the regression accuracy of the model without the feature of interest against the full model. The rank of the model among the configurations is also given according to the  $R^2$  score of the configuration (the highest being rank 1). The so-called full model consists of 5 runs of 5-fold cross-validated models with the full set of data as presented in 2.3 and is our baseline. The splitting of the dataset (train, validation, test) as well as the networks hyper-parameters and architecture (*e.g.* encoder depth and sizes, learning rate and hidden layer size of RNN and decoder depth and sizes) are equally fixed for all the experiments.

#### 3.1 Ablation study

We call "ablation" the relevance assessment of the various features. We do not here evaluate the potential simplification or complexity of the deep-based architecture.

We consider 16 distinct configurations of feature sets. They are decomposed into 3 families (see the respective subsections): 6 context features, 5 features derived from Sentinel-1, and 5 more generic sets (called *global* features) will be separately removed. Main results are shown in Figure 3. For the 16 configurations, the median  $R^2$  value is **0.8227** and the standard deviation is **0.064**. Five configurations exhibit very close scores, with a  $R^2$  value around **0.84**: the full model, the model without the features related to topography, temporal context, agronomic type,

and without the VV/VH ratio band. Results shown that the best performances are not obtained by the full model, which is ranked in 4<sup>th</sup> position. We can however notice a limited difference.

The best results are obtained by removing the information related to the temporal context. This information is in fact induced by the stable sequencing (*e.g.* every 6 days) of the time-series. Somehow duplicating this information is therefore not necessary and does not provide relevant additional information. The use of a GRU rather than an Long short-term memory (LSTM) network, which has a dedicated long-term memory vector, may also explain the low contribution of this feature. Nevertheless, tests done by replacing the GRU with an LSTM do not seem to affect the importance of this feature. The model without the topographic information obtains also equivalent results, even if this information seems important at first glance. Indeed, the topographic feature provides information on the topography related to the area but also on the exposition of each parcels that could be related to the aiming specifications of the sensor. However without this information the standard deviation between the five runs is the lowest among the tested configurations. Agronomic type of the grassland parcel also does not seem to influence the results. Although 22 agronomic classes are present in the area, 61.78% of the parcels are represented by a single class and 95.9% of the total are spread over 3 classes (long rotation grassland of 6 years or more, permanent grassland of 5 years minimum, temporary grassland of 5 years or less). Moreover, these 3 classes are mainly discriminated by the inter-annual duration of the grassland cover and do not necessarily distinguish between different agronomic types. VV/VH ratio band often appears to be the most correlated with NDVI (Fieuzal et al., 2013, Veloso et al., 2017, El Hajj et al., 2019), albeit hiding this information to the network input has little impact on the regression capability. Although, the standard deviation between runs is more significant. The MLPs used for encoding in the first block of the network (Section 2.2) are certainly efficient to compute similar simple information such as a ratio of two values. Their preliminary calculation therefore appears to be negligible.

**3.1.1 Contextual features.** Among the contextual features, the time shift ( $\delta$ ) between the acquisition date of Sentinel-2 and the interpolated Sentinel-1, the climate information as well as the information from neighboring parcels are the most important ones. For these three features, their removal results in a larger standard deviation of the accuracy scores. Time- $\delta$  information is visibly more expressive for the network than temporal context (*e.g.* moment within the year, etc.). Climatic conditions have a direct impact on the state of the vegetation. This information therefore seems adequate to regress NDVI values accurately. Nevertheless, it is derived from the nearest meteorological station and is therefore similar for a large number of parcels. It seems certain that if such information with a finer spatial resolution was integrated with the contextual information, the results would be improved. The spatial component of contextual information (*e.g.* neighborhood slope) is the most significant. Indeed, knowing the average evolution of neighbouring parcels for the network makes it possible to frame the evolution of the parcel of interest within a range. The suppression of this information thus leads to a drop in the regression score to **0.8227**. Since the information given is only a simple mean, more elaborate metrics on the evolution of neighboring parcels such as spatial entropy (Gao et al., 2014) or Moran's Index (Das, Ghosh, 2017) could improve the results.

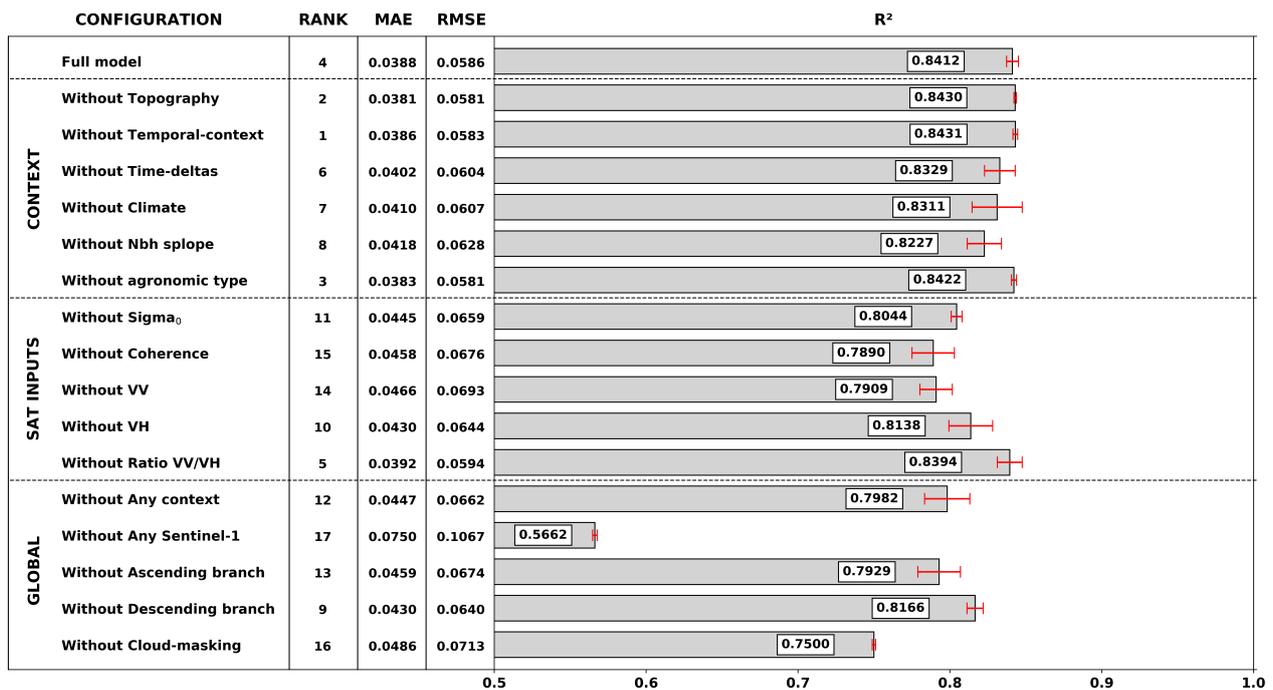


Figure 3. Ablation study of the SenDVI Sentinel-1 to Sentinel-2 NDVI regression task. Mean MSE, RMSE and  $R^2$  as well as standard deviation of  $R^2$  (red bar) for five runs of each configuration are provided. Ranks are given in terms of  $R^2$  scores.

**3.1.2 Satellite features.** Regarding the five Sentinel-1 features, their removal leads to a greater drop in scores as for the contextual features. This is not true for the ratio band (see above). Comparing  $\sigma_0$  features (which measure the proportion of signal backscattered by the target), and coherence features (which calculates the interferometric stability of the target between two dates), we assess that the sigma bands have less relevance than that the coherence information ( $R^2=0.8044$  and  $0.7890$ , respectively). Since two acquisition dates are embedded with coherence, the later is particularly less affected by climatic effects. Moreover, as a proxy for change between two dates, coherence tends to introduce less variance between measurements. As a result, coherence information can be considered important in time-series analysis. Regarding both VV and VH polarization channels, the removal of VV has more impact than VH ( $0.7909$  against  $0.8138$ ). Nevertheless, physically, the VH polarization should be more sensitive to vegetation growth since vertical structures cause depolarization of the wave and thus increased its response sensitivity.

**3.1.3 Global features.** Finally, full families of attributes are considered for ablation. Removing one by one context features does not result in a significant loss of regression quality. Nevertheless, when removing all of the context features, the regression score drops to  $0.7982$  (-0.05). This result demonstrates the relevance of multimodality and ancillary data for such a regression task. Then, Sentinel-1 attributes are removed from the network (and thus also contextual data related to the neighborhood). This configuration gives the lowest regression score ( $0.5662$ ). The network, fed only with the climate, topography, temporal context, and agronomic type features of the parcel, does not correctly estimate NDVI values over grasslands. An example of differences in regression on a parcel is given in Figure 4. It is apparent that without the Sentinel-1 data the time-series shows very little dynamics, based mainly on the evolution of contextual features. The model without the contextual features seems very close to the full model, especially for

the dates when the Sentinel-2 values were available. For the cloudy dates mainly at the beginning of the season, however, a distinction is present which by visual analysis reflects less the evolution of the plot. Another global family considered are the ascending and descending acquisitions of Sentinel-1. Indeed, the integration of the two orbits requires twice as much storage and processing. The ablation of all the descending acquisitions allows the network to obtain a score of  $0.8166$  compared to  $0.7929$  for the descending acquisitions. On the study area, significant relief is present on the eastern part. Many of the grasslands parcel are thus not visible by SAR acquisitions in descending orbit, explaining the difference in these results. This distinction could also be caused by the difference in time-deltas between Sentinel-2 measurements and ascending and descending acquisitions. Figure 5 illustrates the inconsistency in  $R^2$  results among global features removed. Overall, parcels decrease in regression accuracy as more information is removed. For few parcels however the full model shows lower results. Finally, the cloud mask wasn't used for loss calculation. This suppression results in a considerable decrease in accuracy with a  $R^2$  of  $0.75$ . Despite intrinsic errors of the mask (omission and commission of clouds and shadows), its usefulness remains significant.

### 3.2 Computational and storage efficiency

Operational mapping requires network optimization and simplicity while keeping high performances (Mallet, Le Bris, 2020). The first ablation study is informative in terms of regression quality. Nevertheless, in order to reduce storage and preprocessing requirements, other configurations have been tested. All the features were first iteratively removed: the interdependence of these features must be analyzed for the purpose of proposing a satisfactory network with a minimal number of inputs.

Among the contextual information, the topography and climatic features often require the acquisition of proprietary data with

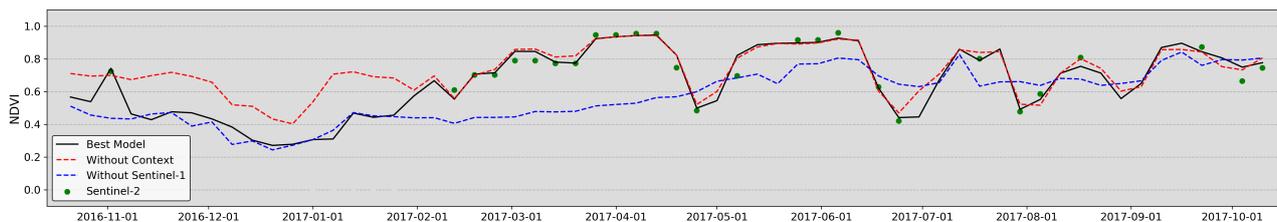


Figure 4. Results of different reconstructions of NDVI time series with various feature sets for one grassland parcel. Only non-cloudy S2 values are shown. It shows that SenDVI can restore the signal despite clouds (S1 features) while keeping only significant high frequencies (context features).

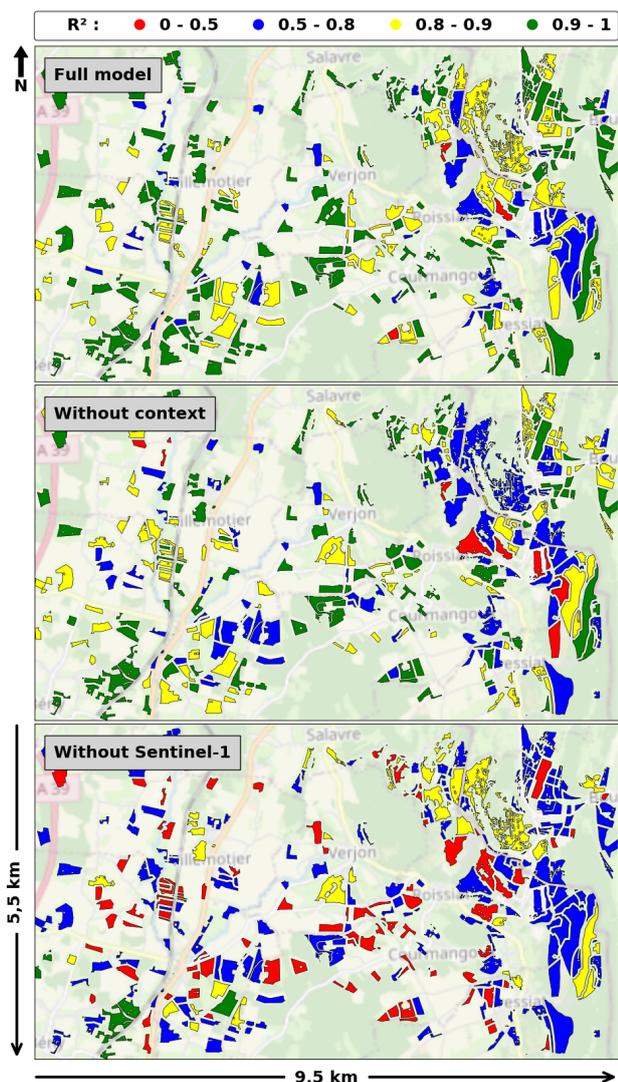


Figure 5. Visual assessment of  $R^2$  values for the full model against two ablation studies removing global features (without contextual features and without Sentinel-1 features).

spatial interpolation. Topographic information do not seem to provide crucial information to the network. Digital Terrain Models with higher spatial resolution may favorably impact the accuracy of SenDVI. However, they are bound not to be available at large scales.

Topographic features were thus first removed from the network entries and the remaining contextual features were suppressed one by one. The configuration with the topographic information and the additional contextual information removed having

the minimal impact on the score is retained and the operation is repeated from this configuration. We consider satisfactory a regression score close to 0.84. Results are provided in Table 2. At a certain point, regression scores are increasing. One possible explanation is that essential data are kept but with approximately same number of parameters, the network has further opportunity to explore this inputs and therefore achieves better high-level representation. Indeed, by removing the topography, time shift between acquisitions (deltas), and temporal context, the scores of  $R^2$  reaches 0.8438. Similarly these three features had little impact on the scores obtained in the ablation study. In the same way, meteorological and agronomic information about grasslands can be neglected. This conclusion is in agreement with the ablation study above on the agronomic type but the results following the ablation of the climatic information are more surprising. Indeed, removing only the climate information scores lower than when it is removed with a set of features.

Add. Removed feature	$R^2$	MAE	RMSE
Topography	0.8430	0.0381	0.0586
- Time-deltas	0.8429	0.0384	0.0583
- Temporal context	0.8438	0.0380	0.0576
- Climate	0.8428	0.0383	0.0579
- Agronomic type	0.8411	0.0393	0.0586

Table 2. Features removed from the SenDVI network that preserve a satisfactory regression score, starting with the topographic set. Removing other attributes would negatively impact the regression quality.

Removing all of the features (Table 2) provides the best results (*i.e.* above  $R^2=0.84$ ) of all configurations while requiring less data. Removing  $\sigma_0$  ratio band VV/VH in this case makes the score drop at 0.8355. For the other features (Sentinel-1 or global ones), in all cases, the scores were significantly worse.

#### 4. DISCUSSION

The network integrates numerous multi-modal data: both satellite data and data aimed at contextualizing the measurements (*i.e.* topographic, climatic data). It was appropriate to assess which features were most contributing to the regression. The results presented above allow to draw preliminary conclusions but are certainly not generally applicable: a single Sentinel-2 tile is not sufficient.

Contextual features are derived from external sources, which may be acquired or exploited differently. Specifically, topographic and climatic data can be handled in different ways. Topographic data have not been standardized by the orientation of the Sentinel-1 acquisition plan. Fluctuations found in the data are likely not to reflect actual variations from one parcel to another.

Climate measurements are (i) daily averaged, and (ii) provided with a coarse spatial resolution (no dense sampling), with significant disparities depending on the location. For ascending and descending orbits, the acquisition times are not the same (beginning of the morning and end of the afternoon, respectively). This difference is therefore not taken into account in the contextualisation effort. In addition, there are a limited number of weather stations in the area, whose values are interpolated. This results in (i) a highly variable accuracy of the measurements depending on the location of the plots and (ii) in similar values for neighboring parcels.

The agronomic type is also subject to uncertainty. In addition to the fact that such classes are highly imbalanced, it includes a declarative bias due to the financial subsidies granted by the French Paying Agency and the constraints imposed on certain classes.

This analysis determines that contextual data provided a gain in regression accuracy. Counter-intuitively, however, the ablation study scores show that the most important features are all derived from Sentinel-1 data (*i.e.*, neighborhood as most important context feature). In order to improve these results, it would therefore be advisable to extract from external sources, features more specifically tailored in relation to Sentinel-1 acquisitions and the parcels of interest. The improvement of the neighborhood feature, through a graph-based approach for example, would support a complete data-driven approach that could also surely lead to at least similar results.

Several points are nevertheless conclusive: (i) a simple approach integrating only contextual data types does not provide satisfactory regression scores; (ii) InSAR coherence although requiring costly calculations is a significant information, and (iii) "the more data, the better" approach often associated with deep learning does not necessarily apply despite a dedicated solution (two branches, attention mechanisms).

Areas for improvement in the network architecture are related to the strategy proposed to integrate the contextual data into the SenDVI architecture. The common Hadamard product used here could be improved. The use of more dedicated networks, such as a Transformer (Vaswani et al., 2017, Sainte Fare Garnot et al., 2020), could improve the results. The hypothesis of significantly increasing the size of the network could also potentially provide better results. However, for most uses, and with a perspective of automation and deployment over large areas, this option is not feasible.

## 5. CONCLUSION

The SenDVI neural network based regression model proposes a multi-sensor approach allowing to obtain time series of high temporal sampling of vegetation indices that do not suffer from the traditional missing data due to clouds. Resulting highly sampled time series showed satisfactory profiles and regression scores. The ablation study made it possible to select the features needed for this purpose and to propose a strategy integrating less data and therefore less pre-processing and storage requirements.

These time series allow weekly monitoring of the evolution of the vegetation, objective identified for the detection of management practices on grasslands. The development of a change point detection framework will be based on these reconstructed time series. Since the proposed regression method is based solely on Sentinel-1 data, both an on-line and off-line method can be implemented. A multi-task learning approach based on SenDVI architecture seems to be a promising direction.

## ACKNOWLEDGEMENTS

This research has been funded by the Agence pour le Développement Et la Maîtrise de l'Énergie (ADEME) and the Centre National d'Études Spatiales (CNES).

## REFERENCES

- Ali, I., Cawkwell, F., Dwyer, E., Green, S., 2017. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7), 3254–3264.
- Atzberger, C., 2013. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sensing*, 5(2), 949–981.
- Baghdadi, N., El Hajj, M., Zribi, M., Fayad, I., 2016. Coupling SAR C-Band and Optical Data for Soil Moisture and Leaf Area Index Retrieval Over Irrigated Grasslands. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(3), 1229–1243.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204, 509 – 523.
- Bengtsson, J. et al., 2019. Grasslands—more important for ecosystem services than you might think. *Ecosphere*, 10(2).
- Buchmann, N., Fuchs, K., Feigenwinter, I., Gilgen, A. K., 2019. Multifunctionality of permanent grasslands: ecosystem services and resilience to climate change. *Improving sown grasslands through breeding and management*, 24, European Grassland Federation EGF, 19 – 26.
- Buckley, J. R., Smith, A. M., 2010. Monitoring grasslands with Radarsat 2 quad-pol imagery. *IGARSS*, 3090–3093.
- Chiboub, O., Kallel, A., Frison, P.-L., Lopes, M., 2019. Monitoring of grasslands management practices using interferometric products Sentinel-1. *Advances in Remote Sensing and Geo Informatics Applications*, 239–242.
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145 – 161.
- Das, M., Ghosh, S. K., 2017. Measuring Moran's I in a Cost-Efficient Manner to Describe a Land-Cover Change Pattern in Large-Scale Remote Sensing Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6), 2631–2639.
- El Hajj, M., Baghdadi, N., Wigneron, J.-P., Zribi, M., Albergel, C., Calvet, J.-C., Fayad, I., 2019. First Vegetation Optical Depth Mapping from Sentinel-1 C-band SAR Data over Crop Fields. *Remote Sensing*, 11(23).
- Fauvel, M., Lopes, M., Dubo, T., Rivers-Moore, J., Frison, P.-L., Gross, N., Ouin, A., 2020. Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series. *Remote Sensing of Environment*, 237, 111536.

- Fieuzal, R., Baup, F., Marais-Sicre, C., 2013. Monitoring Wheat and Rapeseed by Using Synchronous Optical and Radar Satellite Data—From Temporal Signatures to Crop Parameters Estimation. *Advances in Remote Sensing*, 2(2), 162-180.
- Frampton, W. J., Dash, J., Watmough, G., Milton, E. J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83 - 92.
- Gao, S., Zhu, Z., Liu, S., Jin, R., Yang, G., Tan, L., 2014. Estimating the spatial distribution of soil moisture based on Bayesian maximum entropy method with auxiliary data from remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 32, 54 - 66.
- Garioud, A., Giordano, S., Valero, S., Mallet, C., 2019. Challenges in Grassland Mowing Event Detection with Multimodal Sentinel Images. *MultiTemp*.
- Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*, 220, 135-151.
- Griffiths, P., Nendel, C., Pickert, J., Hostert, P., 2020. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sensing of Environment*, 238, 111124.
- He, W., Yokoya, N., 2018. Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation. *ISPRS International Journal of Geo-Information*, 7(10).
- Hughes, L. H., Merkle, N., Bürgmann, T., Auer, S., Schmitt, M., 2019. Deep learning for SAR-optical image matching. *IGARSS*, 4877-4880.
- Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1685-1689.
- Kamilaris, A., Prenafeta-Boldú, F., 2018. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70 - 90.
- Kolecka, N., Ginzler, C., Pazur, R., Price, B., Verburg, P., 2018. Regional Scale Mapping of Grassland Mowing Frequency with Sentinel-2 Time Series. *Remote Sensing*, 10(8), 1221.
- Li, Z., Guo, X., 2018. Non-photosynthetic vegetation biomass estimation in semiarid Canadian mixed grasslands using ground hyperspectral data, Landsat 8 OLI, and Sentinel-2 images. *International Journal of Remote Sensing*, 39(20), 6893-6913.
- Liu, J., Atzberger, C., Huang, X., Shen, K., Liu, Y., Wang, L., 2020. Modeling grass yields in Qinghai Province, China, based on MODIS NDVI data—an empirical comparison. *Frontiers of Earth Science*.
- Mallet, C., Le Bris, A., 2020. Current challenges in operational very high resolution land-cover mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020.
- Mercier, A., Betbeder, J., Baudry, J., Roux, V. L., Spicher, F., Lacoux, J., Roger, D., Hubert-Moy, L., 2020. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 231 - 256.
- Merkle, N., Fischer, P., Auer, S., Müller, R., 2017. On the possibility of conditional adversarial networks for multi-sensor image matching. *IGARSS*, 2633-2636.
- Moreira, A., Fontana, D. C., Kuplich, T. M., 2019. Wavelet approach applied to EVI/MODIS time series and meteorological data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, 335 - 344.
- Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., Hossard, L., 2018. Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, 10(8).
- Pasqualotto, N. et al., 2019. Retrieval of Evapotranspiration from Sentinel-2: Comparison of Vegetation Indices, Semi-Empirical Models and SNAP Biophysical Processor Approach. *Agronomy*, 9(10).
- Pipia, L., Muñoz-Marí, J., Amin, E., Belda, S., Camps-Valls, G., Verrelst, J., 2019. Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes. *Remote Sensing of Environment*, 235, 111452.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204.
- Sainte Fare Garnot, V., Landrieu, L., Giordano, S., Chehata, N., 2020. Satellite image time series classification with pixel-set encoders and temporal self-attention. *CVPR*.
- Scarpa, G., Gargiulo, M., Mazza, A., Gaetano, R., 2018. A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. *Remote Sensing*, 10(2), 236.
- Siegmund, R., Grant, K., Wagner, M., Hartmann, S., 2016. Satellite-based monitoring of grassland: assessment of harvest dates and frequency using SAR. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII*, 9998.
- Solano-Correa, Y. T., Bovolo, F., Bruzzone, L., Fernández-Prieto, D., 2020. A Method for the Analysis of Small Crop Fields in Sentinel-2 Dense Time Series. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3), 2150-2164.
- Stendardi, L., Karlsen, S. R., Niedrist, G., Gerdol, R., Zebisch, M., Rossi, M., Notarnicola, C., 2019. Exploiting Time Series of Sentinel-1 and Sentinel-2 Imagery to Detect Meadow Phenology in Mountain Regions. *Remote Sensing*, 11(5).
- Tamm, T., Zalite, K., Voormansik, K., Talgre, L., 2016. Relating Sentinel-1 Interferometric Coherence to Mowing Events on Grasslands. *Remote Sensing*, 8(10), 802.
- Tiscornia, G., Baethgen, W., Ruggia, A., Do Carmo, M., Ceccato, P., 2019. Can we Monitor Height of Native Grasslands in Uruguay with Earth Observation? *Remote Sensing*, 11(15).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. *NIPS*.
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, 199, 415-426.
- Vuolo, F., Ng, W.-T., Atzberger, C., 2017. Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data. *International Journal of Applied Earth Observation and Geoinformation*, 57, 202 - 213.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402.
- Zalite, K., Antropov, O., Praks, J., Voormansik, K., Noorma, M., 2016. Monitoring of Agricultural Grasslands With Time Series of X-Band Repeat-Pass Interferometric SAR. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8), 3687-3697.
- Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., Feng, M., 2019. Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang, China. *Remote Sensing*, 11(22).