Integrating Synthetic Aperture Radar Data into a Monitoring Tool for Sugarcane in the Cauca Valley, Colombia

Alejandro Mestre-Quereda[a], Juan M. Lopez-Sanchez[a], Tomas Martinez-Marin[a], Victor Cazcarra-Bes[a], Carlos Mosquera[b], Juliana Lozano[b], Jhonatan Quinones[b], Henk Pelgrum[c], Antonia Skarli[c], Corné van der Sande[c], Andrea Idrovo[c], Dirk Hoekman[d], and Martin Vissers[d]

[a]Institute for Computer Research, University of Alicante, P.O. Box 99, 03080 Alicante, Spain

[b]AgroAp, C.I. AGROAP SAS, Ka 100 #11-60, ofi. 819. Cali, Valle, Colombia

[c]eLEAF, Stadsring 65B, 3811 HN Amersfoort, The Netherlands

[d]SarVision, Agro Business Park 10, 6708 PW Wageningen, The Netherlands

Abstract

A methodology to integrate SAR data with optical imagery to implement a reliable monitoring system for sugarcane under all weather conditions has been developed in the framework of the EU-funded DINOSAR project. An extensive field campaign was conducted to collect key biophysical parameters, e.g., stem count, plant height, and fresh biomass, for developing a model that establishes the relationship between these sugarcane traits and radar observables (backscatter and coherence). Building upon this model and the available models based on optical imagery, an integrated algorithm based on dynamical systems theory has been formulated and tested. This method synergistically combines SAR and optical data streams to generate a robust and accurate information source for sugarcane decision-support tools. Results of vegetation biomass estimation along a whole year demonstrate that such an integration approach outperforms methods based only on regressions (machine learning) or on only one type of input data (SAR or optical).

1 Introduction

Current techniques for agricultural crop monitoring are largely based on the exploitation of satellite images acquired in the optical part of the spectrum, thanks to the sensitivity of these frequency bands to the plant conditions (greenness, health, etc.). By using dedicated indices, like the normalised difference vegetation index (NDVI) computed from the red and near-infrared bands, agricultural fields are monitored along the cultivation season. In precision farming, crop monitoring is devoted to ensure the right growth of plants, triggering farming activities, like irrigation and fertilisation, when required.

The application of optical images and NDVI to sugarcane monitoring has been demonstrated in the literature [1]. Its potential for operational purposes in the Cauca Valley (Colombia) was shown by the ESA-funded CostCutting4Sugarcane project [2]. Unfortunately, optical satellite sensors cannot provide usable images in the presence of clouds. This is a strongly limiting factor in areas like the Cauca Valley, which are characterised by long rainy seasons and very frequent cloud coverage (with only 20-25% cloud-free images per year on average). Such a lack of data is especially relevant for sugarcane because its early growing season (i.e., the first 3-4 months) must be carefully monitored as it is the most relevant from the viewpoint of cultivation practices with impact in production. In that context, sensors operating in the microwave spectrum, and especially synthetic aperture radar (SAR), offer a useful complement to optical sensors. SAR satellites provide a consistent and reliable acquisition scheme, and the collected data are sensitive to vegetation structure, biomass, and wetness in the scene, i.e., aspects complementary to their optical equivalent.

The integration of optical and radar data has been attempted in particular study cases for sugarcane (e.g. [3, 4]). However, all approaches found are either very site-specific or purely based in machine learning, and hence without an easy transferability to other conditions (sites, cultivation practices, etc.). Another common approach to integrating optical and radar data for crop monitoring involves synthesizing optical vegetation indices, like NDVI, from radar images. That methodology uses radar as an NDVI proxy to fill gaps in optical time-series caused by cloud cover [5, 6]. However, this technique has limited performance. The inaccuracy stems from inherent NDVI limitations (e.g., saturation for large biomass values) and the challenge of accurately deriving NDVI from radar, as the two data types respond to different crop biophysical properties. This discrepancy is particularly evident in crops like sugarcane, where NDVI and radar backscatter exhibit divergent behaviours during key phenological stages [7].

To solve the mentioned issues, the EU-funded DINOSAR project [8] employs a novel integration technique based on state-space models and Bayesian filtering. This framework optimally merges satellite observations with auxiliary data (e.g., temperature, in-situ measurements) and agronomic crop models [9]. A Kalman filter is used to update the crop model-based predictions with the observational data, explicitly accounting for the uncertainty in each model and data type to achieve an optimal estimate. In this work, we present the results obtained in the project by integrating Sentinel-1 (radar) and Sentinel-2 (optical) images. In particular, the algorithm is designed to predict biomass values of the sugarcane plants throughout a complete growing season in the Cauca Valley (Colombia), thereby providing a continuous and reliable characterization of crop development.

2 Test site and Data

The test site is located in the Cauca Valley, Colombia, in which 237,000 ha are dedicated to its cultivation. The conditions in the Cauca Valley allow for year-round cultivation of sugarcane, with no distinct harvest seasons.

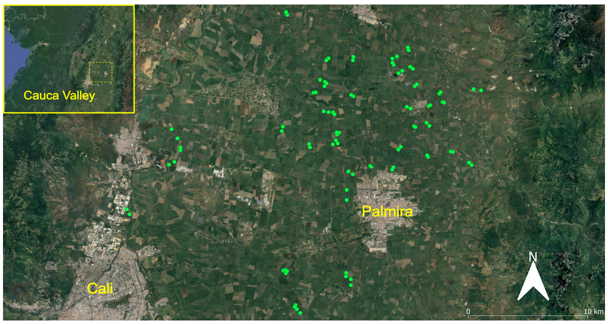

Figure 1 Image of the Cauca Valley with the location of the measurement points (green dots) employed in the in-situ measurement campaign.

2.1 Field campaign

The specific study area selected for this project (shown in **Figure 1**) covers an area of 560 km^2 and comprises 34 fields. A total of 70 measurement points were selected to conduct a field campaign. The in-situ data collection campaign spanned from July 2024 to August 2025, covering a complete cultivation season for all fields. During the first two months of the season, points were visited weekly. Afterwards, all points were visited every two weeks. For each point the following physical variables were measured:

- Stem height (cm).
- Stem diameter (cm).
- Number of stems per metre (along each row).
- Biomass of the stems (kg).
- Biomass of the leaves (kg).

Figure 2 shows all measurements of total biomass (obtained as the sum of the biomass of stems and leaves) gathered during the whole season, represented as a function of the date from the beginning of the season of each field (i.e., Day of Season, abbreviated as DoS). Biomass values are low during the first three months and then quickly increase. Notably, biomass exhibits considerable fluctuations, with sudden drops and increases observed all along the season. There is a noticeable variability among points for the same DoS because some plants grow faster or provide more biomass than others. This variability occurs even within the same field, as a consequence of different cultivation conditions and many external factors (soil texture, management, etc.).

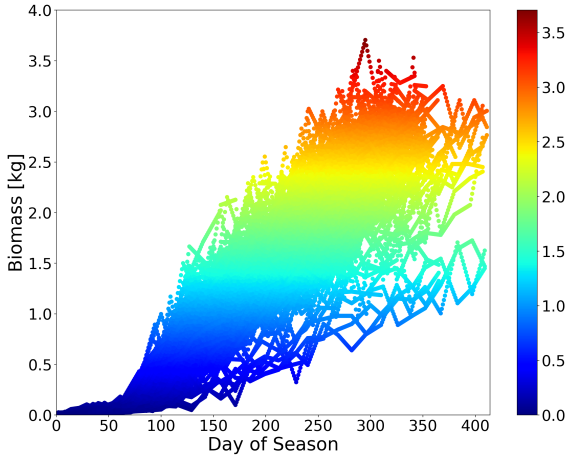

Figure 2 Total biomass measurements (from stems and leaves) for all points selected for the in-situ measurement campaign. For a better illustration of the biomass temporal evolution, measurements are linearly interpolated to obtain daily values.

2.2 Satellite images

Sentinel-1A images for the study area, with a 12-day revisit cycle, were processed as follows. Single Look Complex (SLC) products were radiometrically calibrated using GAMMA software to extract the backscattering coefficient γ_0 at both VV and VH polarimetric channels. Relevant bursts and subswaths were selected and coregistered to a common primary SLC. Subsequently, a multi-looked intensity image was generated from the coregistered SLC stack. Sequences of six consecutive intensity images were processed using a multitemporal speckle filter. The outputs from this filter were orthorectified using the Copernicus Digital Elevation Model (DEM).

In parallel, the coregistered SLCs were used to generate interferometric coherence products for consecutive 12-day pairs of images. These coherence images (for VV and VH channels) were also orthorectified to align with the geometry of the single-date intensity images.

Consequently, the final input datasets for the monitoring algorithm comprise orthorectified maps of the following variables:

- Backscattering coefficient for VV and VH channels.
- Coherence for VV and VH channels.

For optical data, Level-2A Sentinel-2 imagery, providing geometrically and atmospherically corrected surface reflectance at 10-m resolution, was used as the source for NDVI computation.

3 Integration Algorithm

The objective of the integration algorithm is to generate optimal estimates of sugarcane biomass over the growing season by fusing a model of its expected temporal evolution with satellite-derived observations. This paradigm of combining a system model with measurements is formally described by Dynamical Systems Theory [9, 10].

Within this framework, a crop is treated as a dynamical system. Its state, defined here by biomass, is modelled using a state-space approach. The system is characterized by two core components:

Evolution model A mathematical function or model that predicts the temporal progression of the state.

Observation model A mathematical function or model which employs a set of observations to provide indirect information about the state. In this work, the observations are the various satellite data products.

The fusion of the model's prediction (the prior) with the incoming measurements is performed using the Kalman Filter (KF) [11]. The KF provides an optimal recursive solution for such state estimation problems and forms the core of the proposed integration algorithm.

The KF operation is composed of two steps: prediction and update. The prediction step provides a prior estimation of the system state by projecting its expected evolution ahead, i.e. using the evolution model, whereas the update step refines the estimation by including (satellite) observations.

3.1 Evolution model

Empirical expressions for the evolution model can be derived through polynomial curve-fitting or machine learning regression. These techniques typically capture a smoothed, average temporal response, which fails to represent the rapid changes and significant variability in biomass observed at measurement points (see **Figure 2**). Consequently, their application would introduce an estimation bias, as predictions would be skewed towards the mean of the training data.

To address this limitation, we model the temporal evolution of biomass as a first-order, discrete-time Markov process using transition matrices [12]. In this formulation, the state space is discretized into distinct biomass levels, and the system dynamics is characterized by the probabilities of transitioning between these states over time. For practical purposes, we decided to generate weekly transitions so that each transition matrix captures how likely it is for biomass to change from one state to another between the previous week and the current one.

3.2 Observation model

The observation model maps the satellite observations to estimates of the crop's physical variable (biomass). As previously stated, this study utilizes time series from both Sentinel-1 and Sentinel-2. We implemented two separate models—one for SAR and one for optical data—using the Random Forest Regressor (RFR) from Python's scikit-learn library [13].

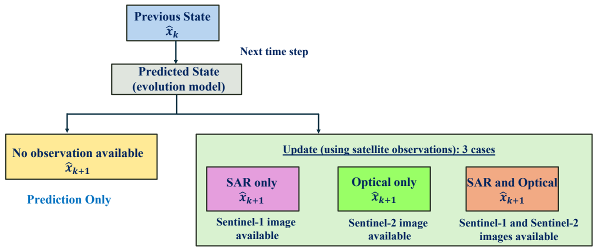

Figure 3 Workflow of the integration algorithm.

The SAR-based model can incorporate multiple features derived from the original data, such as backscatter coefficients and interferometric coherence across different polarimetric channels. A dedicated study was carried out to quantify the contribution of individual SAR features. In the specific implementation illustrated in this work, the model uses four features: backscatter and coherence for both VV and VH channels. The optical-based model uses the NDVI as its only input. In addition, to better capture phenological dynamics, the DoS was included as an input feature for both models. Both observation models were trained on 70% of the available data, with the remaining 30% reserved for testing. The RFR was configured with 300 trees.

3.3 Algorithm Workflow

Figure 3 shows how the complete integration algorithm proceeds to provide biomass estimates throughout the season. As described in Subsection 3.1, the transition matrices are defined on a weekly basis, enabling the evolution model to generate an estimate (prediction step of the KF) at weekly intervals. Then, different scenarios are considered depending on the availability of satellite observations acquired during the week for which the biomass value is to be predicted:

1. **Case 1: No observation** - The final estimate is provided by the evolution model (there is no update step in the KF).
2. **Case 2: Sentinel-1 observation** - When a Sentinel-1 image is available, but there is no Sentinel-2 image, the estimate is updated using the SAR observation model.
3. **Case 3: Sentinel-2 observation** - When a Sentinel-2 image is available, but there is no Sentinel-1 image, the estimate is updated using the optical observation model.
4. **Case 4: Sentinel-1 and Sentinel-2 observations** - When images from both satellites are available, the estimate is updated using both observation models.

3.4 Performance Evaluation

Model performance was assessed by comparing biomass estimates with the in-situ data collected during the field campaign. The quantitative metrics employed are the root mean squared error (RMSE) and the mean absolute error (MAE) between estimated and true (in-field measured) biomass values.
To correctly evaluate the methodology, we start by making a random selection of the measured points (references) to build both the evolution and observation models. Specifically, we employ 90% of the points (63 points out of 70) to build the models and the remaining 10% (7 points out of 70) are reserved for testing. In this way, all data associated with the testing points are not 'seen' by the models. The whole procedure is repeated 100 times, so the overall performance metrics represent the averages computed over 100 realisations.

4 Results

Figure 4 shows the weekly biomass estimates obtained at four measurement points over an entire season. It is important to note that all the data associated with these points (i.e., in-situ measured biomass as well as SAR and NDVI features) were not employed in the construction of the evolution and the observation models, i.e., these data were excluded from model training. As shown in all examples of **Figure 4**, the estimated biomass values (blue dots) exhibit good overall agreement with the field measurements (green dots). The resulting RMSE and MAE values are low, around one hundred grams (0.1 kg) for plants producing between 1.5 kg and 3 kg of biomass. Notably, nearly all biomass values lie within the interval formed by ± one standard deviation of the model estimates.
More importantly, the integration algorithm can correctly track abrupt biomass changes that happen across the season. A sudden drop is observed for point #11 between DoS 230 and 260, followed by an increase until the end of season. The proposed integrated model successfully reproduces this decrease and subsequent recovery in biomass. Such a predictive accuracy would be challenging to achieve using alternative approaches based on machine learning or mathematical regressions (curve fitting), which typically capture only general trends but not abrupt temporal changes.
Furthermore, if we compare points #11 and #12, which belong to the same field, they exhibit very different temporal biomass patterns, with point #11 producing more biomass than point #12 (approximately 1 more kilogram at the end of the season). However, the algorithm accurately captures these differences without introducing a bias, demonstrating that the model is well-adapted to the strong variability present in the scene, as shown in **Figure 2**.
The same approach was tested using only SAR or optical data, resulting in a partial integration of all available satellite observations. **Figure 5** shows biomass estimation results for the same points as the ones shown in **Figure 4**. The general trends of the biomass temporal evolution are correctly followed using only either SAR or optical im-

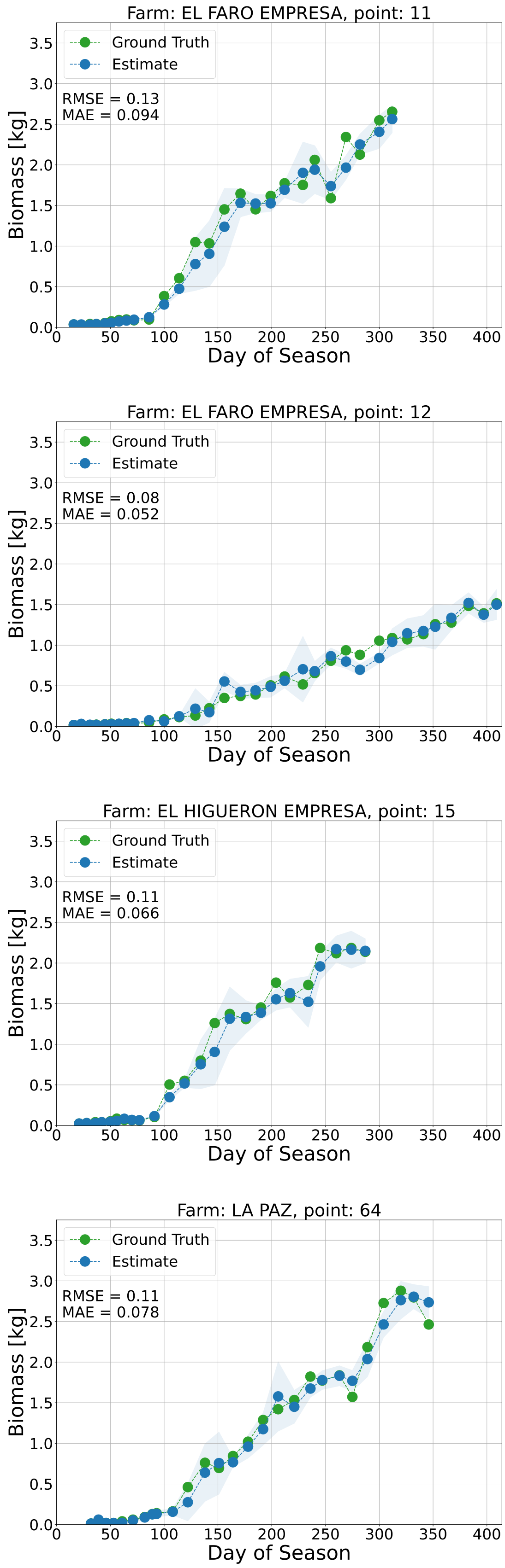

Figure 4 Temporal evolution of biomass estimates throughout a complete season for several points with field data. The shaded blue area represents ± one standard deviation of the model estimation.

ages (red and orange dots respectively). However, more fluctuations are observed in the estimation for these cases, resulting in higher RMSE and MAE values compared to the complete integration case. For instance, this is obvious between DoS 150 and 250 for point #12. Also note how the standard deviation of the estimates is consistently larger than when both types of observations (SAR and optical) are jointly considered, showing that the uncertainty of the model increases with only one satellite type.

Table 1 presents the global accuracy metrics for the proposed integration model and for models that rely on pure machine learning and satellite observations. The results demonstrate that the complete integration approach achieves the lowest RMSE and MAE, indicating superior estimation accuracy. Partial integration, i.e. using either SAR or NDVI data within the dynamical framework, yields better performance than observation-only machine learning (ML) models but is still outperformed by the full integration. ML models relying exclusively on satellite imagery (SAR-only, NDVI-only, or their combination) consistently show higher error. Notably, the NDVI-only method exhibits the lowest accuracy, which we attribute to significant data gaps from persistent cloud cover in the Cauca Valley. This finding underscores the critical importance of incorporating SAR data for reliable sugarcane monitoring in this region.

Table 1 Global accuracy metrics obtained with the proposed integration algorithm and with alternative approaches, after 100 realisations of random selection of points.

Method	RMSE [kg]	MAE [kg]
ML with SAR	0.153	0.099
ML with NDVI	0.189	0.101
ML with SAR+NDVI	0.150	0.098
Integration - SAR only	0.132	0.087
Integration - NDVI only	0.158	0.109
Integration - SAR+NDVI	**0.103**	**0.064**

5 Conclusions and Future Work

The results shown in this work suggest that a dynamical approach (starting with an initial estimate based on the expected evolution of the target variable, followed by updates using satellite images) is better suited for sugarcane monitoring than purely data-driven methods, particularly in situations where cloud cover reduces the number of available images (NDVI specifically).

This work has been carried out with Sentinel-1 data acquired every 12 days. The increased revisit rate (6 days) provided by two satellites (Sentinel-1A and 1C since April 2025, or Sentinel-1C and 1D in the next future) is expected to enhance the estimation performance of this method.

In addition, L-band SAR images from NISAR (expected from November 2025 onwards, after its commissioning phase is finished) could be easily integrated in this monitoring tool. L-band backscatter is expected to saturate at

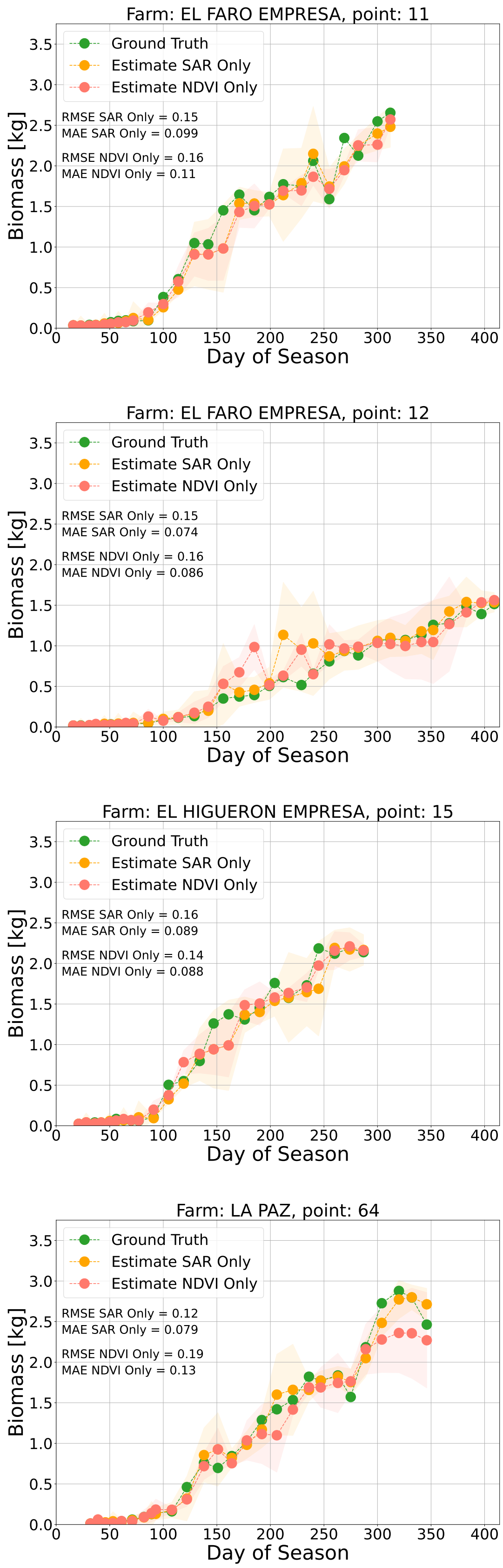

Figure 5 Temporal evolution of biomass estimates throughout a complete season for several points with field data using either only SAR or only optical imagery (partial integration). The shaded area represents ± one standard deviation of the estimates.

biomass values higher than at C-band, and repeat-pass interferometry will present a increased coherence. First results of the exploitation of NISAR data in this framework will be presented at the conference.

Acknowledgement

This work was funded by the European Union under project DINOSAR - Diagnostic tool that integrates optical, infrared and SAR data, Grant Agreement nr. 101129646.
The authors acknowledge the work of all people involved in the project activities.

6 Literature

[1] Som-ard J, Atzberger C, Izquierdo-Verdiguier E, Vuolo F, Immitzer M. Remote Sensing Applications in Sugarcane Cultivation: A Review. Remote Sensing. 2021;13(20).

[2] CostCutting4Sugarcane. ESA Space Solutions;. Available from: https://business.esa.int/projects/costcutting4sugarcane-0.

[3] Kavats O, Khramov D, Sergieieva K, Vasyliev V. Monitoring of Sugarcane Harvest in Brazil Based on Optical and SAR Data. Remote Sensing. 2020;12(24).

[4] Yeasin M, Haldar D, Kumar S, Paul RK, Ghosh S. Machine Learning Techniques for Phenology Assessment of Sugarcane Using Conjunctive SA and Optical Data. Remote Sensing. 2022;14(14).

[5] Pipia L, Muñoz-Marí J, Amin E, Belda S, Camps-Valls G, Verrelst J. Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes. Remote Sensing of Environment. 2019;235:111452.

[6] Roßberg T, Schmitt M. Dense NDVI Time Series by Fusion of Optical and SAR-Derived Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2024;17:7748-58.

[7] den Besten N, et al. Understanding Sentinel-1 backscatter response to sugarcane yield variability and waterlogging. Remote Sensing of Environment. 2023;290:113555.

[8] DINOSAR. Diagnostic that Integrates Optical, infrared and Synthetic Aperture Radar data;. Available from: https://www.dinosarproject.eu.

[9] Vicente-Guijalba F, Martinez-Marin T, Lopez-Sanchez JM. Dynamical Approach for Real-Time Monitoring of Agricultural Crops. IEEE Transactions on Geoscience and Remote Sensing. 2015;53(6):3278-93.

[10] Mascolo L, Martinez-Marin T, Lopez-Sanchez JM. Optimal Grid-Based Filtering for Crop Phenology Estimation with Sentinel-1 SAR Data. Remote Sensing. 2021;13(21).

[11] Kalman RE, et al. A new approach to linear filtering and prediction problems. Journal of Basic Engineering. 1960;82(1):35-45.

[12] Norris JR. Markov Chains. Cambridge, UK: Cambridge University Press; 1997.

[13] Pedregosa F, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825-30.