PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

LavenderVision: a dataset and methodology for lavender bloom detection using Sentinel-2 data

Andreas Stergioulas, Nikos Grammalidis, Dimitrios Kanelis, Vasilis Liolios, Chrisoula Tananaki

Andreas Stergioulas, Nikos Grammalidis, Dimitrios Kanelis, Vasilis Liolios, Chrisoula Tananaki, "LavenderVision: a dataset and methodology for lavender bloom detection using Sentinel-2 data," Proc. SPIE 12786, Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023), 127860B (21 September 2023); doi: 10.1117/12.2681856

SPIE.

Event: Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023), 2023, Ayia Napa, Cyprus

LavenderVision: A dataset and methodology for lavender bloom detection using Sentinel 2 data

Andreas Stergioulas[a], Nikos Grammalidis[a], Dimitrios Kanelis[b], Vasilis Liolios[b], and Chrisoula Tananaki[b]

[a]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
[b]Laboratory of Apiculture - Sericulture, Aristotle University of Thessaloniki, Thermi, Greece

ABSTRACT

The accelerated advancements in remote sensing technologies and the deployment of satellites offering freely accessible multispectral satellite imagery have facilitated the application of machine learning, particularly deep learning techniques, to tasks such as crop classification, yield estimation, and bloom detection. Additionally, several countries in the European Union have adopted the Land Parcel Identification System (LPIS), that obliges farmers to declare the exact area and crop type of their parcels each year while also making the LPIS data freely accessible to the public. For the purpose of the SmartBeeKeep research project, co-funded by EU and Greek funds, in this work we utilize the above with the objective to combine multispectral and multitemporal satellite data obtained from the Sentinel 2 satellite with the LPIS parcel maps in order to detect and classify the blooming period of the beekeeping plant lavender, with the use of automated deep machine learning methods. The specific plant type was selected as it is exhibits particular interest to the beekeeping community, which is the main focus group of the SmartbeeKeep project. For this task, a dataset was amassed and thoroughly sorted out, that comprises of approximately 15k individual parcels from the area of Southern France between January 2020 to December 2021. For each parcel, a study of its harmonized EVI index was carried out in order to roughly identify its blooming period temporal boundaries and with the help of experts, characteristic parcels for sub-regions of Marseille were selected to create more accurate temporal annotations. Additionally, freely available data from the EU Copernicus DIAS reference service WEkEO were utilized as an initial temporal estimation for the Start-of-Season (SOS) and End-of-Season (EOS) period. Two temporal deep learning methods were evaluated, namely a convolutional and a recurrent and model, so as to establish benchmark results on the created dataset. The dataset will be released upon publication.

Keywords: Remote Sensing, phenology monitoring, artificial intelligence, bloom detection

1. INTRODUCTION

Maintaining robust bee populations is vital for the global environment and economy. Bees are responsible for pollinating one-third of plant-based foods consumed by humans and contribute an estimated €265 billion[1] each year to the economy through their pollination services. Despite their environmental and economical importance, the ever changing climate conditions and the use of pesticides has resulted in a decline in bee population.[2] As such, systematic efforts are necessary to maintain a healthy bee population and aid the beekeeping community in such task.

One of the challenges that beekeepers face is the need of constant relocation of the beehives, since bees have a fly limit of 10 km radius from their beehive and that limits pollen gathering. Therefore, in order to increase production, beekeepers are in need of accurate information regarding blooming periods and locations of beekeeping plants in the general area that they operate. Recently, remote sensing tools have gained traction, providing non-intrusive ways for crop and phenology monitoring,[3] examinations of ecological changes over time,[4] soil degradation estimation[5] or yield production estimation.[6] Moreover, Satellite missions such as SENTINEL-1,

Further author information: (Send correspondence to A.S.)
A.S.: E-mail: andrster@iti.gr

Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023),
K. Themistocleous, D. G. Hadjimitsis, S. Michaelides, G. Papadavid, Eds., Proc. of SPIE Vol. 12786, 127860B
 · 0277-786X · doi: 10.1117/12.2681856

SENTINEL2-A/B and LANDSAT-7/8, with their high spectral and temporal resolution (13 spectral bands with a revisit rate of five days for SENTINEL2-A/B), provide an abundance of freely available data, that enables researchers to thoroughly study the field of remote sensing using popular methods such machine learning and deep learning, that have proven to be very effective, if provided with large volumes of data. These technological advancements have created the opportunity to incorporate automated methods in beekeeping, in order to provide such useful information and aid the beekeeping community in managing their bee hives.

Motivated by the above, in this work, our goal was to perform automated bloom detection for the bee keeping plant lavender. In order to accomplish this task, we amassed 15k lavender parcels across Southern France from January 2020 to December 2021. Given the unavailability of blooming annotations, we combined knowledge from experts in the field with automated, index-based data to create as close as possible target labels for the blooming period of lavender. We compared the performance of different deep learning methods, suitable to handle time series data and released the blooming dataset to be freely available for researchers in the field.

2. DATA ACQUISITION AND ANNOTATION CREATION

2.1 Study region

The study region selected encompasses a large area in southeastern France (Figure 1), covering 27,491.6 km2 and spanning across three administrative areas: Provence-Alpes-Côte d'Azur, Auvergne-Rhône-Alpes, and Occitanie. These specific areas were chosen as they contain a majority of France's lavender fields. The selection of this broader region was based on the availability of crop type and parcel border maps for France's metropolitan area, which are provided by the French government through the Land Parcel Identification System (LPIS). LPIS crop maps rely on the annual declarations from farmers who report the extent of their parcels and the type of crop grown. According to the French Government's on-the-spot checks, the accuracy of the marked parcels is 98%, with a relative error of 0,3%.

2.2 Sentinel-2 data

In order to acquire the Sentinel-2 multitemporal and multispectral in top-of-canopy reflectance images, a python script was prepared that uses Google Earth Engine to access and download the required data from COPERNI-

Figure 1. Selected area of study in southeast France.

CUS/S2_SR, an open access dataset from the Scihub[7] platform. For each parcel, the spatial borders provided by the LPIS system were utilized to download the lavender field's area and only parcels where lavender was cultivated during January 1st 2020 to December the 31th 2021 were selected. During this time period, the 10-day median images were acquired and images having more than 20% were discarded. Additional cloud masking was performed, utilizing the QA60 band of Sentinel-2 and more specifically the bits 10 and 11, that represent opaque and cirrus clouds. The exported images have 10m spatial resolution, with bands of different spatial resolutions upsampled by Google Earth Engine internal functions.

The dataset created in this work includes the provided 12 B bands provided from COPERNICUS/S2_SR: Aerosol (B1), Blue Green and Red (B2, B3 and B4), 4 Vegetation Red Edge bands (B5, B6, B7 and B8A), a near infrared band (B8), Water vapor (B9) and two shortwave infrared bands (B11 and B12). However bands B1 and B9 were not taken into consideration during the evaluation of different deep learning methods.

In order to enable the application of convolutional and non-convolutional methods, two dataset formats were considered, an image-based format and a vectorized format. For the image-based format, each parcel is comprised of a $T_p \times C \times H_p \times W_p$ image, where T_p is the total number of 10-day median observations with less than 20% percent cloud coverage, C is the number of Sentinel-2 bands and H_p, W_p represent the height and width respectively of a square patch containing the parcel's pixels. Pixels outside of a parcel's area are zero padded. In the vectorized format, each parcel is comprised of a $T_p \times C \times N_p$, where N_p is the total number of pixels in the parcel. As parcels observed from Sentinel-2 are relatively small, they can be treated as pixel vectors without considering their spatial extent. Following,[8] json files are provided containing a parcel's perimeter, pixel count, cover ratio perimeter to surface ratio.

2.3 Blooming annotation acquisition

Owing to the size of the selected study region, potential differences in lavender blooming periods from parcel to parcel may arise. Consequently, data from the EU Copernicus DIAS reference service WEkEO[9] were utilized for an initial estimation of the blooming period. Start-of-season and end-of-season data maps were employed to split the lavender parcels to $N_{cls} = 8$ clusters using the K-Means algorithm, creating initial groups of parcels exhibiting similar seasonal patterns. Parcels belonging to the same cluster were randomly selected and given to experts from the Aristotle University of Thessaloniki (AUTH), for accurate annotation of the start and end of their blooming period. In total, 24 randomly selected parcels were given to the experts.

For the purpose of obtaining the blooming period of each lavender crop, the annotation provided by the experts was combined with a study of the Enhanced Vegetation Index (EVI) of each parcel. EVI measures the greenness of vegetation while adjusting for atmospheric and soil noise and reducing saturation,[10] thus its was deemed as an appropriate index to study lavender condition during its blooming period. EVI is defined as:

$$\text{EVI} = G\frac{NIR - RED}{NIR + C_1 RED - C_2 BLUE + L} \tag{1}$$

Where NIR, RED and BLUE denote the reflectance values of bands B8, B3 and B1, G is a gain factor C1 and C2 are aerosol adjustment coefficients and L is the soil adjustment coefficient. Usually, $L = 1, C1 = 6, C2 = 7.5$, and $G = 2.5$. To take into account the scaling factor of Sentinel-2, the band reflectance values were divided by 10,000. For each individual lavender parcel, the mean EVI was calculated for every available observation in a span of 2 years, 01 January 2020 to 31 December 2021. The reasoning of examining EVIs across two years was that due to the periodic nature of plant phenology, a study of the temporal profile of EVI values would provide a rough estimate of lavender blooming periods.

One of the issues presented in utilizing EVI values as time series, is the existence of outliers that skew the time series wave form. To reduce the outlier values, previous works utilizing multitemporal satellite data have used harmonic regression to fit the time series,[11,12] a technique that proved to be very effective. Following these findings, the lavender parcels EVI time series were fitted to a harmonic function and the estimated harmonic coefficients were used to recreate a fitted EVI time series for the 2-year span. The harmonic function equation for a singular time step t was formulated as:

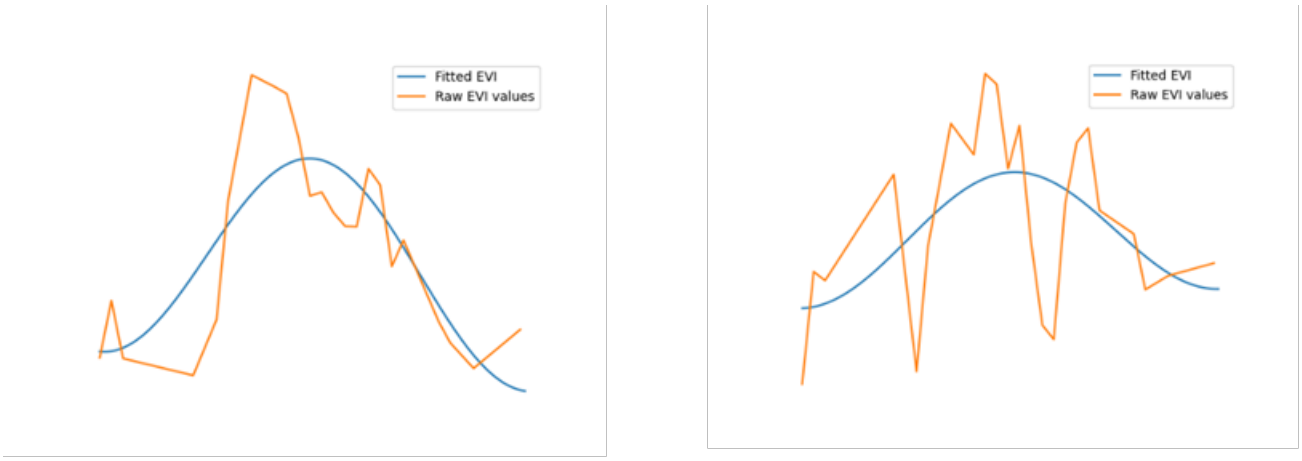

Figure 2. Example of 2 EVI time series for one year (orange) and their corresponding fitted harmonic (blue).

$$EVI_t = b_0 + b_1 t + b_2 \cos(\frac{2\pi}{365}t) + b_3 \sin(\frac{2\pi}{365}t) \tag{2}$$

Where b_0, b_1, b_2 and b_3 are the harmonic coefficients to be estimated and $\frac{2\pi}{365}$ is the angular frequency ω, with $T = 365$ days. The effectiveness of fitting raw EVI time series to a harmonic function in order to reduce spikes due to outliers is shown in Figure 2.

After the estimation of the mean EVI harmonised time series for each cluster, the phase shift from the annotated cluster was computed. The purpose of this process was to calculate the phase shift from the annotated cluster and shift the blooming period accordingly. This approach was adopted to alleviate the challenge of annotating each region individually. The phase, denoted by ϕ and measured in radians, was defined as follows:

$$\phi_{rad} = \arctan\left(\frac{b_3}{b_2}\right) \tag{3}$$

To estimate the shift in days, the radian phase ϕ_{rad} was converted in degrees and the following equation was applied:

$$\Delta\phi_{days,i} = T\frac{\phi^g_{degrees} - \phi^i_{degrees}}{360}, i = 1, ..., N_{cls} \tag{4}$$

With $\phi^g_{degrees}$ being the radian phase of the annotated cluster. Finally, the blooming period for a cluster i was calculated as:

$$Bloom_i = Bloom_g - \Delta\phi_{days,i}, Bloom_i \in [bloom_{start}, bloom_{end}] \tag{5}$$

All individual parcels in the same cluster i, were assigned the same blooming period $Bloom_i$, . Model training and testing were performed from April 2021 through September 2021.

3. IMPLEMENTED MODELS AND RESULTS

To assess the effectiveness of deep neural network (DNN) methods on the newly created dataset, firstly a model combining convolutional[13] and recurrent[14] layers was evaluated. The particular model, denoted as CNN-LSTM, takes a sequence $X \in \mathbf{R}^{T,C,H,W}$ as input and generates a label y_t for each timestep T, where y_t is a binary value that represents whether the timestep corresponds to a bloomed or non-bloomed time period. Additionally, another method has been evaluated where instead of handling the X as a time series of satellite images, a fixed number of S pixels was randomly sampled from the parcel. In this way, parcels are treated as pixel vectors

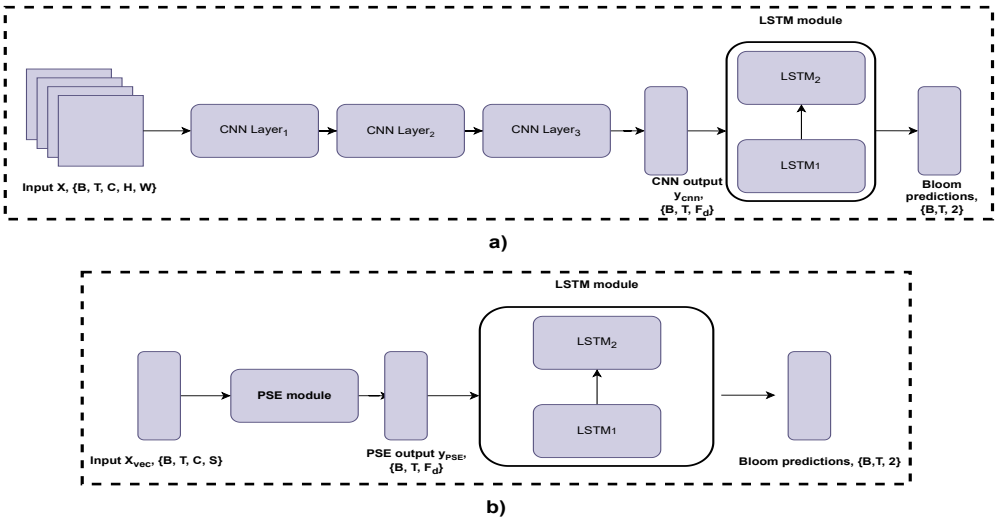

Figure 3. Architectures of a) CNN-LSTM and b) PSE-LSTM models

$X_{vec} \in \mathbf{R}^{T,C,S}$. A Pixel-Set encoder[8] (PSE), is utilised to process X_{vec} before feeding its output to an LSTM layer.

The CNN in the CNN-LSTM model consisted of three 2D CNN layers, with a kernel size of 3 for the first two layers and a kernel size of 4 for the last layer. Max pooling operations and ReLU non-linearities were used between each convolutional layer. The PSE module was designed based on the architecture proposed in,[8] comprised two bidirectional LSTM layers, with a hidden size of 128. A graphical illustration of the model architecture can be seen in Figure 3.

For training and testing, approximately 70% and 30% of the dataset were respectively used. The models were trained using the cross-entropy loss for 40 epochs, with a batch size of 32 and the Adam optimizer with a learning rate of $1e-3$. To augment the training dataset, random Gaussian noise with a standard deviation of $1e-2$ was added to the pixels. In the case of the CNN-LSTM evaluation, each parcel image was resized to 32×32 before being passed to the CNN-LSTM. The performance of both the CNN-LSTM and PSE-LSTM models is presented in Table 1.

In summary, the results show that both the CNN-LSTM and PSE-LSTM methods are effective for predicting blooming periods in lavender parcels, with their accuracy scores being close. These findings demonstrate the potential of using DNN methods for precision agriculture applications, particularly in the detection and prediction of crop cycles phenology.

Model	**Accuracy (%)**
CNN-LSTM	95.64
PSE-LSTM	91.15

Table 1. Lavender blooming detection results on the proposed dataset.

4. CONCLUSIONS

In this paper, a dataset enabling the training of deep learning methods on the task of bloom detection and more specifically for lavender, a plant of significant importance to bee keeping, was introduced. Potential future work could focus on further improving the annotations for the lavender blooming periods. One approach would be to utilize experts who could visit the lavender parcels in person and annotate the exact blooming periods. Additionally, there is potential to expand the annotations to cover all the production stages that lavenders may have, not just their blooming periods. Another avenue for exploration would be to examine other bee keeping plants, such as oilseed rape, that have more lengthy blooming periods, to further understand the relationship

between bee pollination and plant blooming. These areas of future work could provide deeper insights into the dynamics between bees and plants, and ultimately lead to more informed and effective beekeeping practices.

REFERENCES

[1] Khalifa, S. A., Elshafiey, E. H., Shetaia, A. A., El-Wahed, A. A. A., Algethami, A. F., Musharraf, S. G., AlAjmi, M. F., Zhao, C., Masry, S. H., Abdel-Daim, M. M., et al., "Overview of bee pollination and its economic value for crop production," *Insects* **12**(8), 688 (2021).

[2] Hendrikx, P., Chauzat, M.-P., Debin, M., Neuman, P., Fries, I., Ritter, W., Brown, M., Mutinelli, F., Le Conte, Y., and Gregorc, A., "Bee mortality and bee surveillance in europe," *EFSA Supporting Publications* **6**(9), 27E (2009).

[3] Stergioulas, A., Dimitropoulos, K., and Grammalidis, N., "Crop classification from satellite image sequences using a two-stream network with temporal self-attention," in [*2022 IEEE International Conference on Imaging Systems and Techniques (IST)*], 1–6, IEEE (2022).

[4] Lamba, A., Cassey, P., Segaran, R. R., and Koh, L. P., "Deep learning for environmental conservation," *Current Biology* **29**(19), R977–R982 (2019).

[5] Giuliani, G., Chatenoux, B., Benvenuti, A., Lacroix, P., Santoro, M., and Mazzetti, P., "Monitoring land degradation at national level using satellite earth observation time-series data to support sdg15–exploring the potential of data cube," *Big Earth Data* **4**(1), 3–22 (2020).

[6] Haghverdi, A., Washington-Allen, R. A., and Leib, B. G., "Prediction of cotton lint yield from phenology of crop indices using artificial neural networks," *Computers and Electronics in Agriculture* **152**, 186–197 (2018).

[7] "Copernicus open access hub," (2023). Accessed: 2023-03-10.

[8] Garnot, V. S. F., Landrieu, L., Giordano, S., and Chehata, N., "Satellite image time series classification with pixel-set encoders and temporal self-attention," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 12325–12334 (2020).

[9] "Wekeo: Copernicus and sentinel data at your fingertips," (2023). Accessed: 2023-03-10.

[10] Liu, H. Q. and Huete, A., "A feedback based modification of the ndvi to minimize canopy background and atmospheric noise," *IEEE transactions on geoscience and remote sensing* **33**(2), 457–465 (1995).

[11] Wilson, B. T., Knight, J. F., and McRoberts, R. E., "Harmonic regression of landsat time series for modeling attributes from national forest inventory data," *ISPRS journal of Photogrammetry and Remote Sensing* **137**, 29–46 (2018).

[12] Venkatappa, M., Sasaki, N., Shrestha, R. P., Tripathi, N. K., and Ma, H.-O., "Determination of vegetation thresholds for assessing land use and land use changes in cambodia using the google earth engine cloud-computing platform," *Remote sensing* **11**(13), 1514 (2019).

[13] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al., "Recent advances in convolutional neural networks," *Pattern recognition* **77**, 354–377 (2018).

[14] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," *Neural computation* **9**(8), 1735–1780 (1997).