

# SEMI-SUPERVISED PHENOLOGY ESTIMATION IN COTTON PARCELS WITH SENTINEL-2 TIME-SERIES

Vasileios Sitokonstantinou<sup>1,2</sup>, Alkiviadis Koukos<sup>1</sup>, Charalampos Kontoes<sup>1</sup>,  
Nikolaos S. Bartsotas<sup>1</sup>, Vassilia Karathanassi<sup>2</sup>

<sup>1</sup>Institute for Space Applications and Remote Sensing, National Observatory of Athens, Penteli, Greece

<sup>2</sup>Laboratory of Remote Sensing, National Technical University of Athens, Zografou, Greece

## ABSTRACT

This study presents a dynamic phenology stage estimation methodology for cotton towards early warning and mitigation advice against natural disasters. First, a time-series comparison algorithm, based on Earth Observation (EO) data, is used to assign pseudo-labels to approximately 1,000 parcels. For this, we employ only a limited number of ground truth samples. The pseudo-labels are then used to train Random Forest (RF) regression models for phenology stage estimation. The pseudo-labeling process is used to augment the annotated dataset and allow for modelling the growth of cotton. The models are applied and evaluated on two different test sites in Greece; for which field campaigns were carried out to collect the labels. The results are satisfactory and showcase the successful generalization of the models to other areas. The dynamic predictions for cotton growth and extreme weather events, from numerical weather prediction (NWP) models, are invaluable information for decision-making relevant to agricultural insurance schemes and farm management.

**Index Terms**— cotton phenology, agricultural insurance, semi-supervised learning, early warning, natural disasters

## 1. INTRODUCTION

Crop yield is exposed to a number of risks that result in volatile farm profit and hence an unstable income for the farmers [1]. The timely knowledge of the occurrence and severity of upcoming weather perils is significant information for the development of risk management tools, but also the optimization of farm management and the control of inputs. In this context, farmers, agricultural consultants and agricultural insurance companies could benefit greatly [2]. With respect to agricultural insurance schemes, the usually applied indemnity-based insurance is subject to asymmetries in information, and for this reason index-based insurance has been used as an alternative [3]. However, unlike traditional compensation schemes, index-based insurance is not based on the actual farm losses but rather on exceeding an index threshold [4]. Given that the overall goal is the reduction of the farm losses and the increase of the transparency in

the compensation process, indices should be able to describe the reality of the individual farm. This discrepancy between indices and actual farm losses is known as basis risk. The knowledge of phenology significantly decreases the basis risk and the expected utility for the farmers [4].

The yield losses caused by adverse weather events can vary significantly, as there is high dependence on the phenological stage of the plant at the instance of the event. Therefore, instead of having fixed time windows for the index determination, phenology estimation can play a key role in the adjustment of the index results based on the expected impact of the disastrous event on yield. Big Earth Observation (EO) data that cover large areas with high frequency and at high spatial resolution have introduced new opportunities for the large-scale monitoring of phenology, without the need for costly and time-consuming manual observations [5, 6].

Additionally, proactive actions and mitigation measures against imminent adverse weather events, but also the management and scheduling of farm practices, can be assisted by the timely and accurate knowledge of the growth stage of the crop. For instance, soil fertilization can be applied earlier or foliar fertilization can be delayed based on the knowledge of an upcoming precipitation event. Another example is the early picking of cotton bolls, during the boll opening phase, in the case of an approaching extreme weather event.

In this study, we have implemented a dynamic phenology estimation methodology for cotton. To overcome the scarcity of labelled data for training the Machine Learning (ML) models, we have developed a pseudo-labeling technique that is based on a comparison analysis of EO data time-series from the Sentinel-2 mission. The pseudo-labels are used to train a Random Forest (RF) regression model and its performance is evaluated on 27 parcels for which ground truth data was collected through field campaigns.

## 2. METHODOLOGY

### 2.1. Early warning for weather events

In order to predict the occurrence of an imminent extreme weather event, atmospheric parameters from numerical weather

prediction (NWP) models are exploited. Currently, information both from a long-term/coarse-grid-spacing global model (15-days/0.25-degrees respectively) as well as a short-term/high-grid-spacing regional model (2.5 days/0.02-degrees) are available. The latter refers to an in-house configuration of WRF-ARW on a spatial configuration of 6-km grid spacing over Europe and 2-km over Greece. The model configuration in terms of resolution, as well as the microphysical schemes that are used, allow for an explicit resolution of complex processes such as the initiation of deep convection without the need for parameterization schemes. This also benefits the estimation of difficult to estimate processes such as hail growth, that are known to challenge the reliability of any NWP model. Forecasts from both sources are updated daily. The model outputs that are used to identify approaching extreme weather phenomena include 2m temperature and soil temperature at different depths (0-10, 10-40 and 40-100cm), the wind speed at 10 m (gale risk), accumulated precipitation (flood-inducing heavy precipitation event risk) and Convective Available Potential Energy - CAPE (hailstorm risk).

## 2.2. Phenology stage prediction

The phenology stage prediction methodology is comprised of two steps. Initially, a time-series comparison algorithm is used to generate predictions based on a limited amount of reference parcels, for which on-the-spot acquired growth stage timestamps are available (Section 2.2.1). The second step involves utilizing the predictions from the first step as pseudo-labels to then train RF regression models (Section 2.2.2). The reference data refer to the collection of phenological stage timestamps for 10 parcels in Rodopi, Greece; as collected through field visits in 2018 and 2019. The growth stages that concern this study include the root establishment (germination and emergence), the leaf development, the squaring, the flowering, the boll development and the boll opening.

Growth stage	Cont. scale	DoY range	Days
Seeding	100	110-125	-
Root Establishment	100 - 199	110-150	15-25
Leaf Development	200 - 299	130-190	25-40
Squaring	300 - 399	160-215	20-25
Flowering	400 - 499	180-250	35-45
Boll Development	500 - 599	220-270	20-25
Boll Opening	600 - 699	240-315	25-45
Harvest	700	265-315	-

**Table 1.** Phenological stages with reference to this study's continuous scale, the possible DoY ranges in which a stage can occur and the expected ranges of duration

Table 1 lists the seven growth stages, along with their corresponding values in i) a continuous growth scale that is used for the model predictions, ii) the range in DoY that stages

can occur and iii) the range of their expected duration according to literature [7]. The continuous scale ranges from 100 to 700, where 100 refers to the seeding day and 700 refers to the completion of the boll opening stage.

### 2.2.1. Pseudo-labeling

The methodology is based on the 1vAll comparisons of the time-series of an incoming parcel with the ones from the reference dataset. The time-series feature space comprises of Sentinel-2 images, including all spectral bands - except B01, B09 and B10 - and the vegetation indices Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index and Boll Opening Rate Index (BORI). Since phenology is dependent on the relative temporal progression of features and not their absolute values, the final parameter space is comprised of the slopes ( $y'$ ) of the time-series with a step of  $s$  ( $=5$ ) days, as shown in Algorithm 1.

---

#### Algorithm 1: Pseudo-labeling

---

```

Result: Current phenological stage
Input: data, ref_data, DoP, tw, s
foreach feature in data do
  y = data.subset(feature);
  for i=DoP-tw to DoP do
    |  $y'(i) = [y(i) - y(i - s)] / s$ ;
    possible_stages = get_stages(DoP);
    foreach stage in possible_stages do
      foreach ref_parcel in ref_data do
        start = start_day(stage, ref_parcel);
        end = end_day(stage, ref_parcel);
        y_ref = ref_data.subset(feature, ref_parcel);
        for i=start to end do
          for j=i-tw to i do
            |  $y'_{ref}(j) = [y_{ref}(j) - y_{ref}(j - s)] / s$ ;
            end_pos = (i - start) / (end - start);
            pred_stage = stage + 100 * end_pos;
            errors.add(MAE( $y'$ ,  $y'_{ref}$ ));
            preds.add(pred_stage);
          sort(preds, by = errors);
          results.add(preds[:3]);
        final_stage = median(results);

```

---

In order to predict the phenological stage for a parcel, a time window ( $tw$ ) that refers to the number of examined days prior to the Day of Prediction (DoP) must be defined. This way, we generate the feature subspace for the given DoP. For this study,  $tw$  was set to 75 days. In other words, each time-series segment represents the last 75 days prior to the DoP. Then, based on the literature derived DoY ranges, as given in Table 1, we record all possible phenological stages for any given DoP. The  $tw$ -long feature subspace of the examined parcel

is compared with multiple equivalent segments of the reference parcels using the Mean Absolute Error (MAE). These segments stem from sliding in time, from the start of the first (*start*) to the end of the last possible stage (*end*), for a DoP. The comparisons are made for each feature individually, and the three smallest errors for each are recorded. The prediction targets refer to the continuous phenology scale, as defined in Table 1. Each error corresponds to a particular segment from the reference parcels and in turn to a specific value in the continuous scale. For instance, the prediction 510 refers to the 5th phenological stage (Boll development), with a 10% completion percentage. The three segments with the smallest MAE are recorded for each parameter. The median value of the respective predictions for those recorded segments is the final growth stage prediction at a given DoP.

### 2.2.2. RF regression model using pseudo-labels

The reference parcels are confined and the extracted knowledge from those can only be fully representative for parcels of the same region and of similar agro-climatic conditions. Therefore, the time-series comparison method was applied to 994 cotton parcels in close proximity to the reference parcels. Due to the scarcity of ground truth data and the need to generalize the methodology to be applicable in other areas, we examined a pseudo-labeling approach. The pseudo-labels refer to those 994 predictions that are then used to train an RF regression model.

Since the phenology prediction methodology needs to be dynamic and capable to execute at any time instance, multiple RF models are trained for every 5 days throughout the growing cycle of cotton. Therefore, predictions are made every 5 days, using different RF models of increasingly larger feature spaces. The feature spaces comprise of images from seeding to the DoP. Each model has been fine-tuned individually.

## 3. EXPERIMENTAL RESULTS

The validation data, based on which the performance of the two phenology estimation techniques were evaluated (Sections 2.2.1 and 2.2.2), has been collected through a field campaign on 16 parcels in Rodopi, Greece and 11 parcels in Thessaly, Greece. Two experts visited each of the fields 3 or 4 times from the beginning of August 2020 until their harvest. On each visit, the experts recorded the prevailing phenological stage of the parcel in the BBCH scale, which was then translated to the continuous scale of this study.

Table 2 shows the percentiles of predictions with respect to error ranges in the continuous scale. The results are given for the two validations datasets, namely of Thessaly and Rodopi, separately. It should be noted that the Thessaly region is situated far from both the reference data and the pseudo-labels; and is characterized by different agro-climatic conditions. For both regions, but particularly for Thessaly, the

RF performs better than the time-series comparison method (pseudo-labeling). Even though the validated samples are limited, it could be argued that the model generalizes well, providing satisfactory results when transferred to different regions.

Error	Thessaly		Rodopi	
	RF	PL	RF	PL
<10	29%	31%	24%	31%
<20	76%	51%	54%	51%
<30	100%	74%	63%	64%
<40	-	86%	80%	82%
<50	-	96%	92%	91%

**Table 2.** Percentiles of predictions of RF and pseudo-labeling (PL) for different ranges of error in the continuous scale

The MAE of the two methods has been averaged over all predictions for the 97 field visits and was computed in both the continuous scale and in days. The pseudo-labeling method resulted in a MAE of 23.98 in the continuous scale and 6.88 in days, while the RF regression gave a MAE of 20.33 and 5.82 days. The results are satisfactory, particularly given the inherent ambiguity of the target. The different plants in a parcel do not all grow exactly at the same pace. Therefore, both the reference and validation data, which were given with a single BBCH description for a single DoY timestamp, are subject to observation errors. The experts have quoted that even though they are confident of their aggregated decision, they have witnessed an intra-parcel deviation for the growth stage of up to 4 days. The results are within the limits of this aggregation error and thus estimation errors should not be interpreted based on their absolute value but rather their relative importance. Having said that, both approaches appear to perform well.

## 4. DISCUSSION

The prediction of extreme weather events combined with the timely knowledge of the current growth stage is invaluable information for pertinent decision making. An indicative application would be an alerting mechanism that results in i) prevention and mitigation actions and ii) evidence-based insurance processes.

**Ideal conditions for seeding based on temperature predictions.** The germination of cotton requires soil temperatures (0-20cm) larger than 16 °C for 10 consecutive days [8]. Also, for non irrigated parcels an indicative average rainfall of 50mm should be recorded prior to seeding [8].

**Predicted heatwaves and intensification of irrigation.** This is relevant particularly during the flowering and end of flowering stages. Temperatures larger than 32 °C during flowering, but also during boll development can be thought as threshold temperatures [9]. Nevertheless, the knowledge of an immi-

ment heatwave is important for every stage of the cultivation. **Interruption of irrigation based on the estimation of the current phenological stage and expected rainfall.** Irrigation can be interrupted on the onset of boll opening [10], in order to stop the continuous growth of cotton and for the photosynthetic carbohydrates to start contributing to the development of bolls and not the development of leaves and flowers. On the other hand, the interruption of irrigation at the physiological cutout could be compromised due to an upcoming rainfall, with a significant cost to the yield.

**Application of fertilizers depending on rainfall predictions.** There are many fertilization methods, i.e. application of standard fertilization prior and during seeding and application of superficial fertilization at the first stages of the plant [11]. The knowledge of an imminent rainfall can assist in rushing the fertilization prior to the event in order to better integrate the fertilizer. Furthermore, foliar fertilization could be postponed based on the knowledge of an imminent rainfall, avoiding rinsing prior to its absorption.

**Prediction of adverse weather event and early harvest.** The knowledge of imminent gale, frost, hail and flooding risk near the end of the cultivation can trigger an earlier harvest of cotton. If the hazard and the consequent damage are expected to be severe, then the early harvesting is justified even in cases of merely 30% of bolls are open.

## 5. CONCLUSIONS

In this study we implemented a dynamic phenology prediction pipeline that is based on generating pseudo-labels to then train RF regression models. The available ground truth data is scarce and limited to a single region. Therefore the proposed semi-supervised methodology can provide a scalable and geographically transferable solution. The time-series comparison and the RF regression methods provided satisfactory and comparable results. However, within the context of large scale applications, the RF solution is more computationally efficient than the exhaustive time-series comparison equivalent and with greater potential for generalization.

The results presented showcase the performance of the methods only for the last stages of cotton growth. In the near future, the methodology will be tested for the full growth cycle in order to inspect and understand the potential performance differences among the various growth stages. Nonetheless, the results clearly illustrate the potential of dynamically identifying the growth stages over large areas with only minimal ground truth information. This in turn can have great impact towards a more resilient agricultural sector, both from the farm management and agricultural insurance perspective.

## 6. ACKNOWLEDGEMENTS

This work has been supported by the e-shape project, which has been funded by the European Union's Horizon 2020 in-

novation programme under grant agreement 820852. The authors acknowledge NEUROPUBLIC S.A. for providing the reference data and thank Dr. Leonidas Toullos and Dr. Dimitra A. Loka, Researchers at the Institute of Industrial and Forage Crops in Greece, for their valuable consultations.

## 7. REFERENCES

- [1] J. B. Hardaker, *Coping with risk in agriculture*, Cabi, 2004.
- [2] J. W. Glauber, "The growth of the federal crop insurance program, 1990–2011," *American Journal of Agricultural Economics*, vol. 95, no. 2, pp. 482–488, 2013.
- [3] C. G. Turvey, "Weather derivatives for specific event risks in agriculture," *Applied Economic Perspectives and Policy*, vol. 23, no. 2, pp. 333–351, 2001.
- [4] T. Dalhaus and R. Finger, "Can gridded precipitation data and phenological observations reduce basis risk of weather index-based insurance?," *Weather, Climate, and Society*, vol. 8, no. 4, pp. 409–419, 2016.
- [5] T. Jianwu, C. Körner, H. Muraoka, S. Piao, M. Shen, S. J. Thackeray, and X. Yang, "Emerging opportunities and challenges in phenology: A review," *Ecosphere*, vol. 7, no. 8, pp. 1–17, 2016.
- [6] V. Sitokonstantinou, A. Koutroumpas, T. Drivas, Karathanassi V. Koukos, A., C. Kontoes, and I. Papoutsis, "A sentinel based agriculture monitoring scheme for the control of the cap and food security," in *RSCy2020*. International Society for Optics and Photonics, 2020, vol. 11524, p. 1152407.
- [7] D. M. Oosterhuis, "Growth and development of a cotton plant," *Nitrogen nutrition of cotton: Practical issues*, pp. 1–24, 1990.
- [8] J. M. Bradow and P. J. Bauer, *Germination and Seedling Development*, pp. 48–56, Springer Netherlands, 2010.
- [9] M. Ashraf, M. M. Saeed, and M. J. Qureshi, "Tolerance to high temperature in cotton (*Gossypium hirsutum* L.) at initial growth stages," *Environmental and Experimental Botany*, vol. 34, no. 3, pp. 275–283, 1994.
- [10] D. M. Oosterhuis and F. M. Bourland, *Physiological Cutout*, University of Arkansas, Division of Agriculture, Agricultural Experiment Station Fayetteville, 2008.
- [11] J. S. McConnell and M. Mozaffari, "Yield, petiole nitrate, and node development responses of cotton to early season nitrogen fertilization," *Journal of Plant Nutrition*, vol. 27, no. 7, pp. 1183–1197, 2005.