

Crop Phenology Estimation in Rice Fields Using Sentinel-1 GRD SAR Data and Machine Learning-Aided Particle Filtering Approach

Jie Yang¹, Hongtao Shi², Qinghua Xie^{1,*}, Juan M. Lopez-Sanchez³, Xing Peng¹, Jianghao Yu¹, Lei Chen¹

¹ The School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China - (yangj28, xieqh, pengxing, yu721, lei.chen) @cug.edu.cn

² The School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China - hongtao.shi@cumt.edu.cn

³ The Institute for Computer Research (IUII), University of Alicante, E-03080 Alicante, Spain - (juanma.lopez) @ua.es

Keywords: Crop Phenology, Particle Filtering, Sentinel-1, Real-time, Machine-learning.

Abstract

Monitoring crop phenology is essential for managing field disasters, protecting the environment, and making decisions about agricultural productivity. Because of its high timeliness, high resolution, great penetration, and sensitivity to specific structural elements, synthetic aperture radar (SAR) is a valuable technique for crop phenology estimation. Particle filtering (PF) belongs to the family of dynamical approach and has the ability to predict crop phenology with SAR data in real time. The observation equation is a key factor affecting the accuracy of particle filtering estimation and depends on fitting. Compared to the common polynomial fitting (POLY), machine learning methods can automatically learn features and handle complex data structures, offering greater flexibility and generalization capabilities. Therefore, incorporating two ensemble learning algorithms consisting of support vector machine regression (SVR), random forest regression (RFR), respectively, we proposed two machine learning-aided particle filtering approaches (PF-SVR, PF-RFR) to estimate crop phenology. One year of time-series Sentinel-1 GRD SAR data in 2017 covering rice fields in Sevilla region in Spain was used for establishing the observation and prediction equations, and the other year of data in 2018 was used for validating the prediction accuracy of PF methods. Four polarization features (VV, VH, VH/VV and Radar Vegetation Index (RVI)) were exploited as the observations in modeling. Experimental results reveals that the machine learning-aided methods are superior than the PF-POLY method. The PF-SVR exhibited better performance than the PF-RFR and PF-POLY methods. The optimal outcome from PF-SVR yielded a root-mean-square error (RMSE) of 7.79, compared to 7.94 for PF-RFR and 9.1 for PF-POLY. Moreover, the results suggest that the RVI is generally more sensitive than other features to crop phenology and the performance of polarization features presented consistent among all methods, i.e., $RVI > VV > VH > VH/VV$. Our findings offer valuable references for real-time crop phenology monitoring with SAR data.

1. Introduction

Crop phenology is a significant biophysical feature of crops. For crop growth monitoring and yield estimation, access to crop phenology data can be very helpful. Crop phenology has the ability to accurately reflect important time periods, such as sowing and harvesting. This precise timing information is then utilized to implement irrigation and fertilizer applications, which in turn encourages more effective planting practices. Therefore, crop phenology monitoring is crucial for controlling field dangers, safeguarding the environment, and determining agricultural productivity.

Rice is the staple food for more than half of the world's population. Accurate crop phenology information can provide decision support for rice yield and irrigation. Traditional monitoring of crop phenology information is mostly obtained through field observations, which is time-consuming and labor-intensive. With the development of modern technology, remote sensing technology makes it possible to monitor crop phenology information on a large scale (Khanal et al., 2014). And the study area can be sampled several times in a short period of time, hence it can be used for dynamic monitoring of crop phenology information.

Optical data and synthetic aperture radar (SAR) data are widely employed in phenology monitoring. The advantage of optical data in phenology monitoring primarily arises from the varying sensitivities of different spectra to the canopy structure of different crops. At the culmination of the crop growth cycle, vegetation indices are amalgamated to identify key phenological periods, such as the start of the growing season (SOS) and the end of the growing season (EOS) (Yang et al., 2017; Sisheber et al., 2023; Liao et al., 2023). By utilizing various techniques, such as Savitzky-Golay filtering, Fourier transform, and wavelet transform, to smooth the time series of vegetation indices, the optical remote sensing enables the extraction of pivotal nodes such as maxima, minima, and inflection points, thus pinpointing critical time points within the crop growth cycle. Nevertheless, this method is limited to monitoring specific crop growth stages and remains susceptible to external environmental factors such as soil conditions, temperature, and climate.

The primary advantage of SAR technology lies in its high penetration capability, allowing for all-weather, all-day acquisition of surface information. Additionally, different SAR feature parameters exhibit high sensitivity to the morphological structure of crops (Liu et al., 2013). Research indicates significant potential for SAR data's feature information in monitoring crop growth (Wiseman et al., 2014). Crop phenology estimation

* Corresponding author.

problems are typically treated as classification problems, with researchers employing methods such as decision trees and K-means clustering. To enhance accuracy, machine learning has been applied to phenology classification, achieving promising results when combined with algorithms like Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) (Mascolo et al., 2016; Taskin et al., 2016; Yuzugullu et al., 2015). To achieve continuous estimation of crop lifecycles, the principles of dynamic system theory have been introduced into phenology estimation research. Based on this theory, Kalman Filtering (KF) and Extended Kalman Filtering (EKF) have gradually been used in SAR data for rice crop phenology studies (Mascolo et al., 2021). However, due to KF's applicability to linear problems and EKF's suitability for linearizing nonlinear problems (Mascolo et al., 2021). Particle Filtering (PF) is not constrained by model limitations, has begun to be applied in crop phenology estimation. When constructing predictive and observational models, curve fitting methods are commonly employed, including polynomial fitting (for state equations) and Logistic growth model fitting (for observations or state equations) (De Bernardis et al., 2016;), with few alternatives considered for model construction.

Therefore, this paper combines dynamic systems theory with machine learning regression algorithms to provide novel insights in crop phenology estimation. The application of machine learning combined with dynamic systems theory makes it possible to estimate crop phenology and growth in real time. Compared with the common polynomial fitting (POLY), this paper proposes two machine learning-assisted particle filtering methods (PF-SVR, PF-RFR) for crop phenology estimation by combining two integrated learning algorithms consisting of Support Vector Machine Regression (SVR) and Random Forest Regression (RFR), respectively. Four common polarization features including VV polarization, VH polarization, VH/VV polarization and Radar vegetation index (RVI) were integrated as observational information for model construction. Confirming that RVI is usually more sensitive to crop phenology than other features in the study area, and that the polarization features perform consistently across all methods, our results provide a valuable reference for real-time monitoring of crop phenology using SAR data.

The paper content is structured as follows: In Section 2, the study area and the data used are introduced. Section 3 describes the principles of dynamical systems theory, particle filtering and the main machine learning algorithms used. Section 4 specifically analyses the experimental results obtained. Section 5 summarizes the paper and draws conclusions.

2. Study site and Data set

2.1 Study site

This paper uses Sentinel-1 GRD SAR data to monitor the phenology of rice fields through the experimental technique. The research site is located in a rice crop cultivation region close to the mouth of the Guadalquivir River, Sevilla, southwest Spain. The planting area measures roughly 30 km by 30 km, and rice is grown there from May to October every year. The present work centers on the experimental examination of seven rice parcels, the distribution of which is depicted in Figure 1. The rice growth season in this location is approximately 135-150 days.



Figure. 1. Google–Earth picture of rice fields in Sevilla, Spain

2.2 Remote sensing image data

In this study, we downloaded all the images of the Sentinel-1 A/B constellation covering the rice cultivation time in the study area during 2017 and 2018 via the ESA official website, thus ensuring a revisit time of 6 days with a dual-polarized VV-VH image in Ground Range Detected (GRD) format. The images were processed using the SNAP software provided by ESA (<http://step.esa.int/main/download/snap-download/>), which includes orbit correction, thermal noise removal, radiometric calibration, speckle filter, conversion from linear to dB and terrain correction.

2.3 Ground Campaign Data

Our ground-truthing data come from the local group of rice farmers (*Federacion de Arroceros de Sevilla*), where each year throughout the rice-growing season, meticulous phenology data are measured in the field on four to seven rice-growing plots. These statistics represent the key phenological time using the worldwide standard definition of Biologische Bundesanstalt, Bundessortenamt and Chemical industry (BBCH).

3. Methods

3.1 Dynamic Systems Theory

A mathematical framework for examining how complex systems behave over time is called dynamic systems theory. According to the idea, a system is usually described as a set of variables and the interactions between them. The crop growth cycle can be viewed as a continuous dynamic process that changes over time. Among these, the collection of states X_N ($X_N = [x_1, x_2, x_3 \dots x_N]$) in the system across time can be thought of as the crop phenology information. It can be known that the state equation of the system, also known as the prediction equation, is as shown in equation (1). This is in accordance with the first-order Markov assumption, which states that the current moment state X_k of the system is only determined by the previous moment state X_{k-1} .

$$X_k = F(X_{k-1}) + v_k, \quad (1)$$

where X_k = current moment
 X_{k-1} = previous moment
 v_k = random noise
 F = corresponding functional relationship

The observation equation is as shown in equation (2), which mainly reflects the relationship between the observed value Y_k and the current moment state X_k .

$$Y_k = H(X_k) + w_k, \quad (2)$$

where Y_k = observed value
 X_k = current moment
 w_k = random noise
 H = corresponding functional relationship

Hence, precise formulation of state and observation equations within dynamic system theory is paramount for achieving optimal estimation. PF and KF stand out as commonly employed parameter estimation methods in dynamic system theory.

3.2 Particle Filtering

In general, obtaining direct crop phenology state X_k is challenging and is primarily achieved through field observations. In cases where field observation is impractical, indirect parameter estimation relies on other observational information Y_k , such as vegetation index, climate data, SAR polarization characteristics, etc.

Particle filtering is a Bayesian filtering method for state estimation of dynamic systems, which is particularly suitable for nonlinear and non-Gaussian systems (Arulampalam et al., 2002). For example, crop phenology over time dynamics is a nonlinear process of change, and thus is well suited for corresponding crop phenology estimation by PF algorithms. It represents the probability distribution of the state of the system by a set of randomly sampled particles, and updates and estimates it according to the observed data. PF algorithm is an "observation-update" form of parameter estimation, which roughly consists of five parts: (1) particle initialization; (2) prediction; (3) update; (4) updating the particle weights; and (5) particle resampling. The details of each of these parts are as follows:

(1) Particle initialization

During the particle initialization phase, a random set of particles, denoted by the number N , is drawn from the a priori probability distribution to represent the potential distribution of the system state. Typically, the initial positions of these particles can be determined based on the prior knowledge of the system or existing observations. Each particle is assigned a weight, which is initially set to an equal value of $\frac{1}{N}$.

(2) Prediction

During the prediction phase, the state of each particle is forecasted based on a dynamic model of the system. This is accomplished by applying the state transition function of the system, typically incorporating a model of system dynamics noise to account for uncertainty. The formula for predicting the state is as shown in equation (3).

$$X_k^{[i]} = F(X_{k-1}^{[i]}) + v_k^{[i]}, \quad (3)$$

where $X_k^{[i]}$ = predicted state of the i -th particle at time k
 $X_{k-1}^{[i]}$ = predicted state of the i -th particle at time $k-1$
 $v_k^{[i]}$ = random noise of the i -th particle
 F = corresponding functional relationship

(3) Update

During the state update stage, the state of each particle is corrected based on the observed data. This is achieved by

considering the difference between the observed value and the actual observation for each particle. The formula for updating the state is as shown in equation (4).

$$w_k^{[i]} = p(Y_k | X_k^{[i]}), \quad (4)$$

where $w_k^{[i]}$ = the weight of the i -th particle at moment k
 $p(Y_k | X_k^{[i]})$ = conditional probability

(4) Updating particle weights

During the stage of updating particle weights, new weights for each particle are computed based on the state and observation data of each particle. Typically, the weights of the particles are calculated using normalized observation probabilities to ensure that the sum of the weights equals 1. The formula for updating the particle weights is as shown in equation (5).

$$w_k^{[i]} = p(Y_k | X_k^{[i]}) / \sum_{j=1}^N p(Y_k | X_k^{[j]}), \quad (5)$$

where $w_k^{[i]}$ = the weight of the i -th particle at the moment k
 $p(Y_k | X_k^{[i]})$ = conditional probability
 N = total number of particles
 $\sum_{j=1}^N p(Y_k | X_k^{[j]})$ = sum of observation probabilities

(5) Particle resampling

Resampling is based on the weight of the particles, so that the high-weight particles are more repeated, so as to maintain the diversity of the particle population and solve the problem of particle degradation. In this paper, residual resampling is mainly used.

3.3 Equation construction process

3.3.1 Prediction Equation: The state equation in this study chiefly delineates the correlation between crop phenology (state variables) over time, typically depicted by a first-order time-varying differential equation. Therefore, predetermining the change pattern is imperative, commonly achieved through a crop growth model. Considering the diverse growth models across different crop types, this paper predominantly adopts a straightforward empirical model to formulate the state equation. Hence, a polynomial fitting function serves as the functional relationship between the state variable and time, succinctly captured as follows:

$$X_k = a_0 + a_1 k + a_2 k^2 + \dots + a_n k^n, \quad (6)$$

where $a_0, a_1, a_2, \dots, a_n$ = polynomial fitting factors

Ordinary Least Square Fitting approach can be used to solve the parameters. The polynomial order of the fitted crop biophysical parameters is typically not greater than 5 in order to avoid overfitting, which can result in decreased accuracy.

3.3.2 Observation Equation: In this instance, the link between the system state X_k at a given moment H and the SAR feature or vegetation index acquisition Y_k will be represented by the observation equation. The observation equation's formulation is akin to that of the state equation, and the measured data can be used to empirically fit the observation equation between the crop phenology and the polarized SAR observation feature.

Three primary methods of fitting were used in this paper: (1) polynomial fitting; (2) support vector machine regression; (3)

random forest regression. Polynomial fitting is a straightforward, intuitive technique that is easy to comprehend and apply, which is why the first strategy is employed. Polynomials can be tailored to fit a range of nonlinear relationships and data forms and patterns.

In addition, this paper also adopts two other algorithms for machine learning: SVR, RFR. Firstly, SVR has a strong ability to fit non-linear datasets, and secondly, SVR can achieve the fitting of non-linear relationship through kernel function, which is suitable for complex data patterns. At the same time, SVR has less effect on noise and local extremes, and has strong robustness. The decision function of SVR can be represented by support vectors, which is easy to explain and understand. RFR can process data efficiently without feature selection, and at the same time, it has relatively less effect on outliers and can deal with data with more noise.

4. Results

In this study, we used this method to estimate the phenological information (BBCH) of crops. Based on the data in 2017, we used the polynomial fitting method to construct a prediction model, and used polynomial fitting, SVR and RFR algorithms to construct an observation model, aiming to compare the impact of different algorithms on the parameter estimation results. Based on the established prediction model and observation model, we made prediction estimates of the phenological information in 2018, and compared the real phenological information in 2018 to verify the accuracy of the results.

4.1 Predictive equation construction

In this paper, the prediction equation describes the relationship between rice crop phenology and time (DoS). Due to the inconsistency of the sowing date of rice in different plots, we converted the original time coordinate system with the day of year (DoY) as the time coordinate system to the DoS time coordinate system with the sowing date starting from 0.

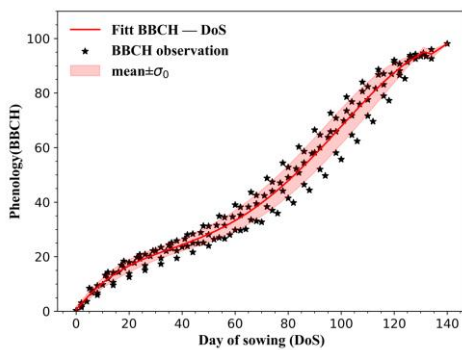


Figure 2. Predictive equations to construct a plot of results.

The equation is mainly completed by polynomial fitting, and after many fittings, we select the optimal fitting polynomial order of 5, and the specific expression is as shown in equation (8).

$$y = ax^5 + bx^4 + cx^3 + dx^2 + ex + f, \quad (8)$$

where $a=7.88316233e-10$
 $b=-1.56373028e-6$
 $c=3.86594439e-4$
 $d=-2.97796674e-2$
 $e=1.27101659$

$$f=3.25316695e-1$$

As shown in Figure 2, the fitting results of the prediction equation constructed from the measured data in 2017 are presented, where the red line represents the fitting result and the black star dot represents the actual ground observation. It can be seen from the figure that in the middle and late stages of rice growth, the dispersion degree of BBCH information is higher than that in the early stage, and the growth trend of rice is basically the same among different plots.

4.2 Observational equation construction

In this paper, the observation equation describes the relationship between SAR observation feature information and rice crop phenology, and three curve fitting algorithms are mainly used for the construction of the observation equation, including polynomial fitting, SVR, and RFR. As shown in Figure 3,4,5, where the black line represents the actual observed value and the red line represents the fitting curve, and it can be seen from the observation equation that the four polarization characteristics are more sensitive in the early stage of rice growth, and the sensitivity is lower in the late growth stage.

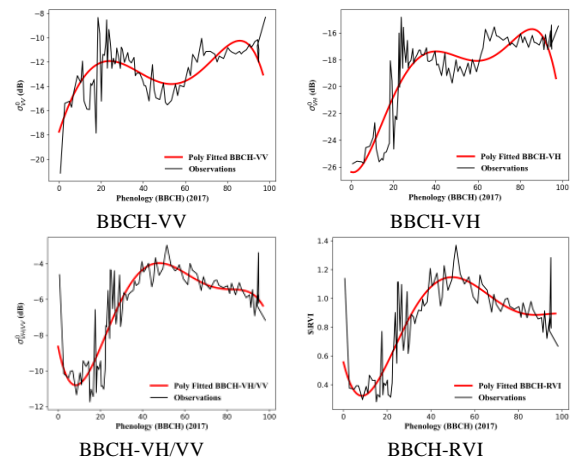


Figure 3. Observation equation based on POLY.

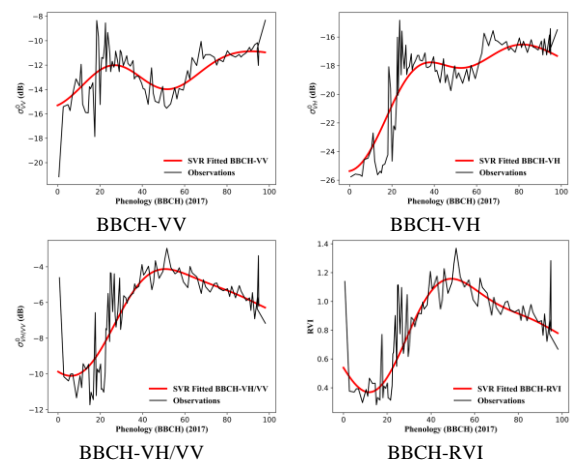


Figure 4. Observation equation based on SVR.

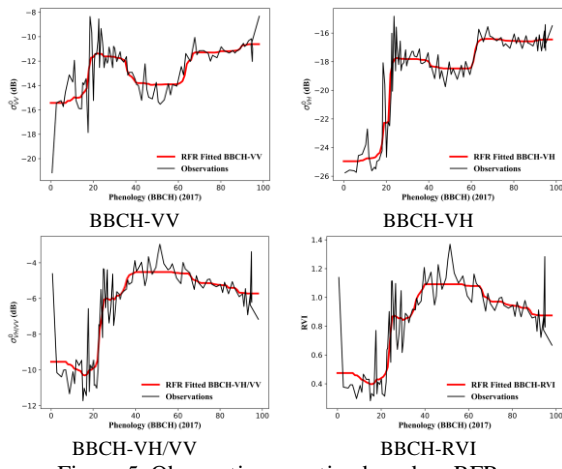


Figure 5. Observation equation based on RFR.

4.3 Estimation of phenology

By constructing the prediction equation and observation equation based on the 2017 data, combined with the polarization characteristics of the Sentinel-1 GRD data, we can use the particle filter algorithm to estimate the phenology of rice crops in 2018. Different observation equation fitting algorithms were used to estimate different phenological parameters, as shown in Figure 2. It can be observed that among the phenological estimation information obtained by the observation equations constructed by the three algorithms, the Root Mean Square Error (RMSE) values are as follows: $RVI < VV < VH < VH/VV$. In addition, the observation equation constructed by the SVR algorithm obtained the best accuracy of the results, with an RMSE of 7.79, and the results obtained by RFR also showed good accuracy, with an RMSE of 7.94. Both algorithms outperformed the polynomial fitting results, with an RMSE of 9.1. However, it is worth noting that in the later stages of rice crops, there is a certain reduction in estimation accuracy. This may be due to the fact that the polarization feature does not exhibit a high degree of sensitivity at all stages of the crop growth cycle.

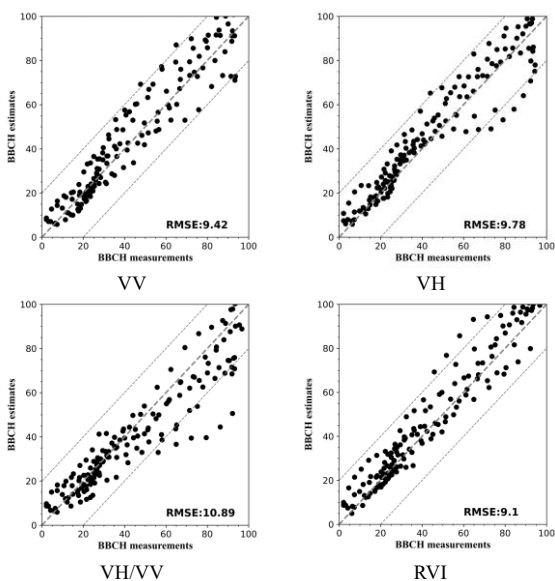


Figure 6. BBCH estimation results based on POLY.

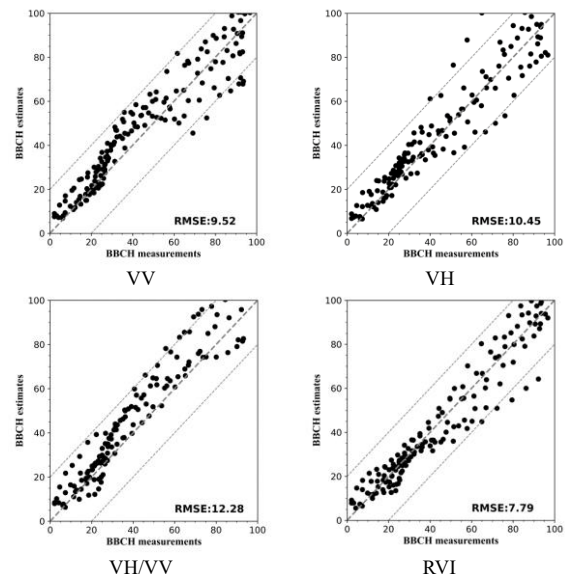


Figure 7. BBCH estimation results based on SVR.

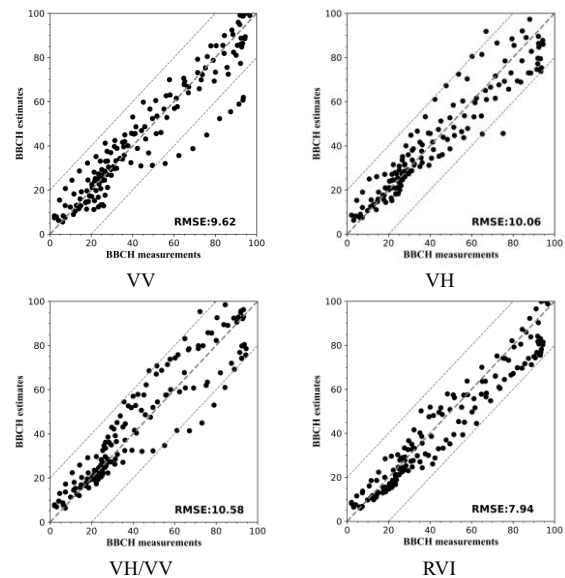


Figure 8. BBCH estimation results based on RFR.

5. Conclusions

In this paper, we propose a method for estimating rice phenological parameters by integrating machine learning and dynamical systems theory. By comparing the effects of two typical machine learning algorithms (SVR and RFR) in the process of constructing observation equations, and combining PF algorithm and SAR polarization features for rice phenology estimation. The following main conclusions were drawn from this study:

- (1) The rice phenology estimates obtained from constructing observation equations using the SVR algorithm had an optimal RMSE of 7.79 over the full growth cycle. In comparison, the results obtained using the RFR algorithm had an RMSE of 7.94, while polynomial fitting yielded a result of 9.1. This demonstrates the superiority of machine learning algorithms in the construction of observation equations.

(2) For different model building algorithms, the results of the four polarization features in rice crop phenology estimation showed the following results: $RVI > VV > VH > VH/VV$. This indicates that the RVI feature has a higher dominance in rice phenology estimation and is more sensitive than the other features over the whole growth cycle of rice.

In future, we will consider introducing more polarization features, such as feature parameters after polarization decomposition. In addition, since rice grows in water for a long time, the backscattering coefficient features are susceptible to the influence of the water body and soil, etc. Therefore, we will consider adding the corresponding feature information to attenuate these influences and improve the accuracy of rice phenology estimation. This will provide a more reliable reference value for rice phenology estimation, thus supporting applications such as crop yield estimation and crop growth monitoring.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Grant No. 42171387, 41820104005, 42101400) and the Spanish Ministry of Science and Innovation (AEI Grant No. PID2020-117303GB-C22/AEI/10.13039/501100011033). All Sentinel-1 data used in this study were graciously provided free of charge by ESA. The Ground-truthing datasets were generously supplied by the Federacion de Arroceros de Sevilla.

References

- Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 50(2), 174-188.
- De Bernardism, C., Vicente-Guijalba, F., Martinez-Marin, T., Lopez-Sanchez, J.M., 2016. Particle filter approach for real-time estimation of crop phenological states using time series of NDVI images. *Remote Sens.*, 8(7), 610.
- Khanal, S., Kushal, K.C., Fulton, J.P., Shearer, S., Ozkan, E., 2020. Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sens.*, 12(22), 3783.
- Liao, C., Wang, J., Shan, B., Shang, J., Dong, T., He, Y., 2023. Near real-time detection and forecasting of within-field phenology of winter wheat and corn using Sentinel-2 time-series data. *ISPRS J. Photogramm.*, 196, 105-119.
- Liu, C., Shang, J., Vachon, P.W., McNaim, H., 2013. Multiyear crop monitoring using Polarimetric RADARSAT-2 Data. *IEEE Trans. Geosci.*, 51(4), 2227–2240
- Mascolo, L., Lopez-Sanchez, J.M., Vicente-Guijalba, F., Nunziata, F.A., Migliaccio, M., Mazzarella, G., 2016. Complete Procedure for Crop Phenology Estimation With PolSAR Data Based on the Complex Wishart Classifier. *IEEE Trans. Geosci.*, 54(11), 6505-6515.
- Mascolo, L., Martinez-Marin, T., Lopez-Sanchez, J.M., 2021. Optimal grid-based filtering for crop phenology estimation with sentinel-1 SAR data. *Remote Sens.*, 13(21), 4332.
- Sisheber, B., Marshall, M., Mengistu, D., Nelson, A., 2023. Detecting the long-term spatiotemporal crop phenology changes in a highly fragmented agricultural landscape. *Agric. For. Meteorol.*, 340, 109601.
- Kucuk, C., Taskin, G., Erten, E., 2016. Paddy-Rice Phenology Classification Based on Machine-Learning Methods Using Multitemporal Co-Polar X-Band SAR Images. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 9(6), 2509-2519.
- Wiseman, G., McNaim, H., Homayouni, S., Shang, J., 2014. RADARSAT-2 Polarimetric SAR response to crop biomass for agricultural production monitoring. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 7(11), 4461–4471.
- Yang, Z., Shao, Y., Li, K., Liu, Q., Liu, L., Brisco, B., 2017. An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data. *Remote Sens. Environ.*, 195, 184-201.
- Yuzugullu, O., Erten, E., Hajnsek, I., 2015. Rice Growth Monitoring by Means of X-Band Co-polar SAR: Feature Clustering and BBCH Scale. *IEEE Geosci. Remote Sens. Lett.*, 12(6), 1218-1222.