Research Progress and Challenges in Remote Sensing Monitoring of Oilseed Crop Yield: Postprint

Authors: Ma Yujing, Wu Shangrong, Yang Peng, Cao Hong, Tan Jieyang, Zhao Rongkun

Date: 2023-12-04T00:00:00+00:00

Abstract

[Purpose/Significance] Oil crops constitute an important component of both food and non-food supplies, and represent a crucial source of edible vegetable oil and plant protein. Real-time, dynamic, and large-scale monitoring of oil crop growth is of great significance for guiding agricultural production, maintaining stability in grain and oil markets, and ensuring national health and well-being. Remote sensing technology, due to its advantages of wide coverage, timely information acquisition, and rapid processing, has been widely applied in research and applications for regional crop yield monitoring. [Progress] This paper first introduces the relevant background of oil crop yield estimation using remote sensing technology; secondly, it reviews the current research status of oil crop monitoring based on remote sensing technology from three aspects: remote sensing parameter retrieval, area monitoring, and yield estimation research, pointing out that data assimilation technology has great potential in oil crop yield estimation, and elaborates in detail on both assimilation methods and grid selection. [Conclusion/Outlook] It identifies opportunities for remote sensing technology in oil crop monitoring, proposes some problems and challenges in oil crop yield estimation based on remote sensing technology regarding crop feature selection, spatial scale determination, and remote sensing data selection, and provides an outlook on future development trends in oil crop yield estimation research. This paper can provide reference and guidance for in-depth research on regional yield estimation and growth monitoring of oil crops.

Full Text

Preamble

Research Progress and Challenges of Oil Crop Yield Monitoring by Remote Sensing

MA Yujing[1], WU Shangrong[2*], YANG Peng[2], CAO Hong[2], TAN Jieyang[3], ZHAO Rongkun[2]

[1]School of Information and Communication Engineering, North University of China, Taiyuan 030051, China
[2]State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China (Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences), Beijing 100081, China
[3]Institute of Agricultural Economy and Agricultural Regionalization, Hunan Academy of Agricultural Sciences, Changsha 410125, China

Abstract:
[Significance] Oil crops constitute a vital component of both food and non-food supplies, serving as an important source of edible vegetable oils and plant proteins. Real-time, dynamic, and large-scale monitoring of oil crop growth is of great significance for guiding agricultural production, maintaining stability in grain and oil markets, and ensuring national health. Remote sensing technology has been widely applied in regional crop yield monitoring research and applications due to its advantages of broad coverage, timely information acquisition, and rapid processing. [Progress] This paper first introduces the background of using remote sensing technology for oil crop yield estimation. Second, it reviews the current status of remote sensing-based oil crop monitoring from three perspectives: remote sensing parameter inversion, area monitoring, and yield estimation research. It points out that data assimilation technology has great potential for oil crop yield estimation, which is elaborated in detail from both assimilation methods and grid selection perspectives. [Conclusions/Prospects] The paper identifies opportunities for remote sensing technology in oil crop monitoring, proposes several problems and challenges in crop feature selection, spatial scale determination, and remote sensing data selection for oil crop yield estimation, and forecasts future development trends in oil crop yield research. This study can provide valuable reference for in-depth research on regional yield estimation and growth monitoring of oil crops.

Keywords: remote sensing; yield simulation; data assimilation; oil crops; yield monitoring; parameter inversion

Foundation items: Hunan Provincial Natural Science Foundation Project (2021JJ40286); National Natural Science Foundation of China (42271374); Fundamental Research Funds for Central Nonprofit Scientific Institutions (1610132021009); Youth Innovation Program of Chinese Academy of Agricultural Sciences (Y2023QC18)

2 Current Status of Remote Sensing Monitoring for Oil Crops

Remote sensing technology enables growth monitoring and yield estimation of oil crops, with parameter inversion and area monitoring serving as critical steps in this process.

2.1 Parameter Inversion for Oil Crops

During remote sensing-based crop monitoring, crop parameters such as biomass and Leaf Area Index (LAI) can effectively describe crop growth status and are key indicators for yield research. However, direct measurement of these growth parameters is destructive and costly. Therefore, conducting inversion research based on remote sensing technology to obtain observation information of oil crops is of great importance.

Currently, parameter inversion for oil crops primarily involves inferring crop biophysical parameters through empirical models and physical models to support yield estimation research [?]. Compared to microwave radar data, optical remote sensing is more widely applied in oil crop inversion studies.

2.1.1 Empirical Models Empirical models typically calculate characteristic parameters from canopy remote sensing data that show strong correlation with known crop parameters. These models establish relationships between characteristic parameters and crop parameters through statistical regression or machine learning methods, representing a widely applied approach [?]. Statistical regression methods such as linear and exponential functions offer simplicity and strong interpretability, making them commonly used for crop parameter estimation.

Research has demonstrated good correlation between vegetation indices (VI) and crop parameters. Based on the advantages of statistical regression methods, researchers have investigated the correlation between VI and oil crop growth parameters for parameter inversion studies. Zhang et al. [?] and Sun et al. [?] established LAI inversion models for rapeseed using the statistical relationship between Normalized Difference Vegetation Index (NDVI) and LAI, finding that NDVI correlates well with LAI and can be used for crop parameter inversion. Qiu et al. [?] constructed rapeseed LAI inversion models based on exponential functions using NDVI and its derivative parameters, proving that NDVI series in vegetation indices have good inversion effects on LAI. Qi et al. [?] used Normalized Difference Spectral Index (NDSI), Ratio Spectral Index (RSI), and other indices to build Simple Regression (SR) models for inverting chlorophyll content (CC) in peanut leaves, indicating that spectral indices can effectively retrieve peanut leaf chlorophyll content. Additionally, studies have shown that Synthetic Aperture Radar (SAR) polarization information can also be used for oil crop parameter inversion. Zhang et al. [?] explored rapeseed growth parameter inversion using linear, logarithmic, and quadratic statistical regression methods, finding that polarization information is highly sensitive to rapeseed crop parameters and suitable for growth parameter inversion. These studies prove that parameter inversion based on statistical regression methods using correlations between remote sensing data and crop parameters is feasible. However, when data exhibit complex nonlinear relationships, simple spectral feature extraction and regression models may not accurately invert crop parameters, necessitating machine learning or physical model approaches to improve inversion accuracy.

In addition to linear, exponential, and logarithmic regression methods, machine learning has been employed to construct inversion models, offering greater advantages in practical applications. Machine learning establishes complex nonlinear relationships between remote sensing data and crop observation data, enabling more accurate oil crop yield estimation through model training and prediction using large datasets. Yuan et al. [?] used Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Partial Least Squares Regression (PLSR) to construct models for soybean LAI inversion, finding ANN more suitable for single-growth-stage soybean LAI estimation while RF performed better across the entire growth period. Qi et al. [?] used 12 vegetation indices related to LAI, including Ratio Vegetation Index and NDVI, to build prediction models for peanut LAI using SR and Back Propagation Neural Network (BPNN) methods, discovering that BPNN achieved higher prediction accuracy. Wei et al. [?] constructed inversion models for rapeseed LAI using both linear fitting and RF methods, concluding that machine learning provides higher inversion accuracy. Furthermore, studies have used machine learning to explore complex relationships between SAR remote sensing data and oil crop parameters. Mercier et al. [?] used Gaussian Process Regression (GPR) to invert rapeseed biomass, finding that polarization indices VH and VV correlate well with wet and dry weight, and that radar data at different bands show potential for rapeseed parameter inversion. Ghosh et al. [?] used GPR methods based on C-band SAR data to construct models estimating Plant Area Index (PAI), Water Content (WC), wet weight, and other plant parameters for rapeseed and soybean, demonstrating that both full and dual polarization can be used for inverting these crop parameters.

These studies demonstrate that machine learning offers greater advantages than statistical regression for oil crop parameter inversion. However, oil crop canopies have complex structures, and parameter inversion based on canopy reflectance and crop biophysical parameters is susceptible to canopy structure effects [?]. Empirical model methods alone cannot provide detailed explanations of the physical mechanisms between surface characteristics and spectral responses, limiting parameter inversion research for oil crops.

2.1.2 Physical Models Physical model methods are based on light transmission and interaction processes, establishing mapping relationships between remote sensing data and physical models to infer crop optical parameters and optimize models using observation data [?, ?]. The geometric shape and biochemical characteristics of oil crop canopies change during growth, altering radiation transfer within the canopy [?]. Physical models can describe how canopy reflectance varies with canopy, leaf, and soil background characteristics [?], making them suitable for oil crop parameter inversion. The PROSAIL model is a classic model in optical remote sensing inversion, applied to retrieve biochemical and structural variables of oil crops, with its coupling principle shown in Figure 1 [Figure 1: see original paper]. Optimization methods mainly include iterative optimization, Lookup Table (LUT), and neural networks. Compared with empirical models, physical inversion methods are applicable under various conditions and more suitable for regional-scale parameter inversion [?].

Numerous studies based on physical inversion methods have been conducted on oil crops, achieving good parameter inversion results. Li et al. [?] used the PROSAIL model and LUT method to invert crop LAI and canopy CC from UAV multispectral images of rapeseed and sunflower, finding that LUT-inverted LAI correlated well with ground-measured biomass and proving that CC can serve as an indicator for canopy nitrogen content estimation in oil crops. Similarly, Duan et al. [?] used the PROSAIL model based on LUT to invert sunflower LAI, evaluating its performance and demonstrating the method's suitability with good accuracy. Tomícek et al. [?] proposed using PROSAIL model, LUT, and ANN to invert leaf CC and LAI for rapeseed, achieving high inversion accuracy and finding potential in combining optical data with PROSAIL radiative transfer models. Studies have shown that physical inversion methods perform better than empirical methods for crop parameter inversion [?]. Nandan et al. [?] used both empirical and PROSAIL physical models to invert soybean LAI, noting that LUT inversion performed better and was more suitable for regional-scale LAI estimation. Thus, physical model-based parameter inversion methods show greater potential for oil crop monitoring. However, physical models require biophysical, biochemical, and soil characteristic parameters, some of which are difficult to obtain, presenting certain limitations [?].

In summary, selecting appropriate inversion methods and data sources is crucial for oil crop parameter inversion research. Empirical models establish empirical relationships between remote sensing data and surface parameters, while physical models provide physical explanations for parameter estimation based on surface physical properties and light transmission principles. For data sources, optical remote sensing data are commonly used, offering rich spectral information but being susceptible to weather conditions, while radar data have strong penetration capabilities and sensitivity to surface structure. Therefore, considering oil crop characteristics, appropriate inversion methods and data sources must be selected to obtain more accurate and comprehensive parameter estimates.

2.2 Area Monitoring of Oil Crops

Remote sensing technology development provides effective means for rapid and accurate acquisition of crop planting information. Crop sowing area monitoring is a key step in yield estimation and has become an important component of crop remote sensing monitoring [?].

Current methods for extracting oil crop area using remote sensing technology mainly include classification methods such as RF, Decision Tree (DT), and SVM. Data sources include optical and microwave remote sensing data, each with distinct advantages.

2.2.1 Optical Remote Sensing Data Different ground objects have different spectral characteristics. Based on remote sensing images from different periods, spectral information from different crops at various growth stages can be obtained. Through spectral features and vegetation indices, different types of oil crops and other vegetation can be distinguished to achieve oil crop classification and monitoring [?].

Optical data have demonstrated great potential for oil crop area monitoring. Song et al. [?] used NDVI from Landsat and MODIS data to estimate soybean area in the United States, noting that time-series indicators can effectively characterize vegetation biophysical properties and be applied to oil crop classification and area monitoring. Shangguan et al. [?] used RF methods based on Landsat data to extract national soybean planting area in Argentina, finding that features calculated from NDVI and near-infrared bands played important roles in distinguishing soybean from non-soybean areas. Li et al. [?] used RF methods with Landsat data to estimate soybean planting area in Heilongjiang Province, achieving an overall accuracy of 84%. Due to complex crop planting patterns, distinguishing oil crops from other crops and interfering ground objects based solely on spectral features presents certain difficulties. Yang et al. [?] used DT methods with Enhanced Vegetation Index (EVI) from MODIS to extract rapeseed planting area, showing that 250 m resolution is suitable for large-scale plain area monitoring, but higher-resolution imagery should be considered for small-area regions with complex planting structures and fragmented plots. Jiang et al. [?] conducted area monitoring of soybean, rapeseed, and other crops in Northeast China, North China, and the middle-lower Yangtze Plain based on Sentinel-2 optical data, achieving high monitoring accuracy. The study noted that rapeseed fields in the middle-lower Yangtze Plain are scattered, and factors such as clouds and mountains interfere with crop area monitoring, resulting in relatively lower accuracy, indicating that optical data face challenges in mountainous crop monitoring.

These studies show that although optical remote sensing offers many advantages for crop monitoring, it encounters problems when monitoring oil crops in mountainous areas. China's oil crop planting structure is dispersed, with some oil crops distributed in fragmented hilly and mountainous regions. In these areas, terrain complexity, mixed cropping, and frequent cloud cover make crop monitoring data acquisition difficult, affecting monitoring accuracy.

2.2.2 Microwave Remote Sensing SAR offers important application potential for oil crop monitoring due to its strong penetration capability, all-weather observation capability, multi-temporal observation, and surface monitoring capabilities.

Scholars have conducted area monitoring research on oil crops based on SAR. Jiao et al. [?] found that phenological stages of rapeseed could be estimated by observing its polarization response in soybean and rapeseed monitoring studies. Valcarce-diñeiro et al. [?] monitored rapeseed area using RADARSAT-2 and Sentinel-1 polarimetric data, achieving high classification accuracy. Additionally, combining microwave and optical data has proven effective for high-precision area monitoring of oil crops [?]. Ren et al. [?] used Sentinel-1/2 to classify and map soybean, rapeseed, and other crops in mountainous areas, showing that Sentinel-1 performed well in classifying rapeseed from other crops across different feature combinations, while Sentinel-2 spectral information was suitable for soybean classification mapping. The study noted that optical and microwave combinations show potential for mountainous crop monitoring. Although SAR offers advantages in crop classification and identification, it also presents challenges. SAR image processing and correction analysis are complex, requiring specialized microwave processing expertise and technical skills.

In practical applications of oil crop area monitoring, it is necessary to consider crop structure and distribution characteristics, comprehensively evaluate the advantages and limitations of SAR and optical data, and combine observation data and monitoring methods to improve identification accuracy of oil crops in mountainous areas.

2.3 Remote Sensing Yield Estimation for Oil Crops

The rapid development of modern remote sensing technology provides effective technical support for regional oil crop yield estimation. Remote sensing technology can obtain crop information in real-time, rapidly, and over large areas, making it the primary means of current crop yield estimation. Unlike other crops, oil crops such as soybean and rapeseed have complex canopy structures and special physiological characteristics. The selection of phenological periods, data sources, and modeling methods all affect yield estimation results. A review of grain and oil yield estimation research from the past five years shows that researchers have achieved good results in regional staple crop yield simulation for rice, wheat, and corn using remote sensing methods. However, yield simulation research for regional oil crops such as soybean, rapeseed, peanut, sesame, castor, and sunflower accounts for only about 26% of studies. This is due to the complexity of oil crop planting layouts and structural characteristics, which present difficulties in crop feature selection, spatial scale determination, and remote sensing data selection.

2.3.1 Yield Estimation Phenological Periods Phenological periods are specific stages in crop growth and development that significantly impact yield estimation. The selection of phenological periods affects remote sensing yield estimation results, with models based on different periods showing substantial differences in accuracy and effectiveness [?].

Different crops use different growth stages for yield estimation. Table 1 shows phenological period selections for some oil crops. In remote sensing yield estimation research for oil crops, phenological period selection differs from staple crops due to the unique plant structure. Richetti et al. [?] used MODIS data from different phenological periods to calculate EVI for regional soybean yield estimation, obtaining accurate results by selecting appropriate phenological information. They found that soybean remote sensing yield estimation is often based on seed initiation and pod stages. Li et al. [?] estimated regional soybean yield using UAV hyperspectral data and explored yield estimation differences across growth stages, finding high accuracy during seed initiation, pod, and grain-filling stages. The flowering and silique stages are critical periods determining rapeseed yield [?], with most rapeseed yield estimation research based on these two stages. Han et al. [?] noted that flowering and silique stages have greater impact on rapeseed biomass accuracy validation than other periods. Ma et al. [?] constructed linear and nonlinear regression models to estimate rapeseed biomass based on canopy hyperspectral data at different growth stages, finding high correlation between NDVI and biomass during flowering and silique stages. Fan et al. [?] simulated regional rapeseed yield based on remote sensing vegetation indices, showing that LAI during rapeseed flowering stage has potential for early yield prediction. Gong et al. [?] established yield estimation models based on rapeseed flowering and silique stages. For other oil crops, the seed initiation stage is optimal for peanut yield research [?], while the inflorescence emergence stage before flowering shows highest correlation between vegetation indices and yield for sunflower [?].

These studies show that oil crops have unique structures and growth cycles. Phenological periods such as flowering, silique, and pod stages are critical growth and development phases closely related to yield. Therefore, when conducting oil crop yield estimation research, it is essential to consider the particularities of oil crops and select appropriate phenological periods.

2.3.2 Remote Sensing Data Sources Different remote sensing data sources have different sensor characteristics and spatiotemporal resolutions, making appropriate data source selection crucial for accurate yield estimation results.

Optical remote sensing data provide rich spectral information reflecting crop growth status. Current crop yield estimation research primarily uses optical remote sensing data such as MODIS [?, ?], Landsat [?, ?], Sentinel-2 [?, ?], and HJ-1 [?]. However, remote sensing sensors cannot directly sense crop yield; they must use remote sensing bands or vegetation indices calculated from bands to invert canopy or crop parameters closely related to yield, thereby establishing relationships between remote sensing features and crop parameters for yield estimation [?, ?]. Vegetation indices such as NDVI and EVI are widely used in regional yield estimation for corn [?] and wheat [?, ?], achieving good progress. In addition to staple crops, studies have proven that optical vegetation indices are also suitable for oil crop yield estimation research. Song et al. [?] noted that NDVI during the growth season peak is one of the most important variables in soybean simulation. Zamani-noor et al. [?] used NDVI calculated from multispectral data to monitor rapeseed growth status, finding NDVI values more reliable for full-growth-period monitoring.

Furthermore, some studies have found that oil crop yield estimation based on remote sensing technology must consider photosynthetic organ parameters such as soybean pods and rapeseed siliques. He et al. [?] noted that soybean pods, as homologous organs of leaves, play a decisive role in soybean grain yield. Peng et al. [?] simulated rapeseed yield based on ground hyperspectral data and evaluated LAI's predictive capacity for rapeseed yield at different growth stages, noting that rapeseed silique wall area is closely related to yield. Gong et al. [?] found that silique effects must be considered when simulating rapeseed yield using remote sensing technology. Therefore, when studying oil crops, it is necessary to combine parameters from pods and silique walls with leaf parameters for yield estimation research.

However, crops such as rapeseed and soybean have diverse optical organs and complex canopy structures. Studies have proven that relying solely on spectral data for regional oil crop yield estimation has certain limitations. Liu et al. [?] and Bognár et al. [?] estimated regional rapeseed yield by combining multispectral remote sensing data and machine learning algorithms, finding that direct prediction using optical remote sensing data had low accuracy. Yoosefzadeh-najafabadi et al. [?] and Sun et al. [?] used optical remote sensing data to estimate regional soybean biomass and yield, noting low simulation accuracy. SAR can penetrate vegetation to obtain deeper crop structural information, offering advantages in crop structure monitoring. Therefore, SAR data with strong penetration capability has been considered for oil crop yield estimation. Yang et al. [?] and Nguyen et al. [?] used fully polarimetric SAR remote sensing data to estimate rapeseed aboveground biomass and yield, showing that SAR data can improve regional rapeseed yield prediction capability and accuracy.

2.3.3 Modeling Methods Crop remote sensing yield modeling methods mainly include statistical regression models and machine learning methods.

Statistical regression model methods primarily establish empirical regression relationships between remote sensing bands or vegetation indices and yield, building mathematical models between remote sensing data and yield, including Least Squares (LS) and Linear Regression (LR) methods. Statistical regression models are simple, intuitive, fast, efficient, and widely applicable, suitable for obtaining large-scale crop yield information. Li et al. [?] used PLSR to establish yield prediction models for regional soybean yield, validating significant correlations between LAI, biomass, and soybean yield. Although simple, regression modeling has limitations, including insufficient model robustness [?]. Additionally, crop yield formation has nonlinear characteristics, and using simple regression models for complex nonlinear relationships may affect estimation accuracy [?].

Machine learning methods are data-driven modeling approaches that train on large datasets to learn complex relationships between remote sensing data and yield information [?], thereby establishing crop yield estimation models. These mainly include Support Vector Regression (SVR) and BPNN. Machine learning can automatically learn features from multiple data levels and is widely applied due to its high accuracy, fast training, and ability to model with small samples. Mateo-sanchis et al. [?] used machine learning to combine optical EVI and microwave Vegetation Optical Depth (VOD) complete time series, constructing yield models to predict soybean yield with high accuracy (R^2 up to 0.9). Krupavathi et al. [?] noted that ANN methods for constructing yield prediction models are relatively stable, capable of capturing complex relationships between crop yield and remote sensing parameters, offering advantages in yield prediction. Therefore, compared with statistical regression models, machine learning methods show greater application potential in oil crop yield estimation.

Domestic and international scholars have explored yield estimation for different oil crops using remote sensing technology, demonstrating the superiority of machine learning methods. Pejak et al. [?] used multispectral vegetation indices and various machine learning algorithms including SVM, RF, XGB, and SGD to predict soybean yield, finding SGD performed best (MAE = 0.436 t/ha, correlation coefficient = 0.83%). Schwalbert et al. [?] integrated optical remote sensing EVI and NDVI with Long Short-Term Memory (LSTM) neural networks to construct regional soybean yield prediction models, finding LSTM had lower MAE and RMSE than RF, indicating better performance. Sun et al. [?] used optical remote sensing reflectance with Convolutional Neural Network (CNN) and LSTM algorithms to propose a CNN-LSTM deep learning model for county-level soybean yield prediction, showing better performance than standalone CNN or LSTM ($R^2 = 0.78$). Abbaszadeh et al. [?] integrated 3DCNN and ConvLSTM deep neural networks to build yield models using MODIS data for multiple U.S. counties, demonstrating superior soybean yield estimation performance compared to individual models. Zhou et al. [?] developed a hybrid CNN model based on multispectral soybean data and CNN architecture, achieving high yield estimation accuracy ($R^2 = 0.78$) and indicating deep learning's potential for soybean yield prediction. Teodoro et al. [?] compared deep learning networks with shallow learning models (RF, SVM, LR) for soybean yield prediction, finding deep learning models achieved highest accuracy. Reisi-gahrouei et al. [?] used L-band airborne SAR data with Multiple Linear Regression (MLR) and ANN to estimate biomass of rapeseed and soybean, showing ANN-based models provided more accurate biomass estimates. Yu and Shang [?] combined NDVI from HJ-1A/1B with phenological features using RF methods to estimate regional sunflower yield, proving RF models can accurately predict yield (RMSE = 0.4 t/ha, relative error = 10.1%). Zeng et al. [?] used PLSR and ANN for sunflower yield prediction, showing advantages of combining remote sensing data with ANN models. Amankulova et al. [?] used vegetation indices from Sentinel-2 multispectral data with MLR, RF, and SVM to predict sunflower yield, finding RF was the best method for field-scale crop yield prediction. Numerous studies have confirmed the superiority of machine learning methods in oil crop remote sensing yield estimation.

Compared with regression models, machine learning algorithms typically require more computational resources and time for model optimization. Therefore, despite better performance, appropriate modeling methods should be selected based on actual conditions and research objectives, considering the advantages and limitations of each approach.

2.3.4 Data Assimilation for Yield Estimation Since the 1960s, crop model research has developed rapidly with advances in agricultural science, computer technology, and understanding of crop growth mechanisms [?, ?]. Crop models can effectively simulate crop growth and development processes at point scales but are not suitable for regional-scale oil crop yield simulation [?]. When extending crop yield simulation from point to regional scales, increased spatial scales introduce surface and near-surface environmental heterogeneity, creating difficulties in parameter acquisition and regionalization [?, ?]. Satellite remote sensing offers unique advantages in broad coverage and spatiotemporal resolution for ground crop information acquisition but can only obtain limited, discrete crop growth observations that cannot effectively support dynamic studies of crop growth, development, and yield formation. Oil crops have special growth structures with photosynthetic organ succession during growth, creating significant temporal differences. Compared with staple crops, oil crop planting patterns also differ, typically appearing as small, scattered fields, thus showing diversity in spatial distribution. These particularities may challenge remote sensing-based oil crop monitoring. Therefore, introducing remote sensing information into crop models through data assimilation can improve crop growth simulation and yield estimation capabilities, achieving spatiotemporal extension of remote sensing inversion and crop models. Assimilation algorithms and grids are important components of data assimilation yield estimation, and selecting appropriate algorithms and grids is particularly important when combined with oil crop characteristics.

(1) Assimilation Algorithms. Assimilation algorithms are the most important components of crop assimilation yield estimation systems. Parameter optimization algorithms based on cost functions and ensemble filtering algorithms based on estimation theory are currently the most widely used assimilation algorithms. The former mainly includes the Simplex search algorithm and Shuffled Complex Evolution (SCE-UA), with cost functions including Root Mean Square Error and four-dimensional variational methods. The latter primarily includes Ensemble Kalman Filter (EnKF) and Particle Filter (PF).

Researchers have studied oil crops using data assimilation technology, confirming its potential for oil crop yield estimation. Trépos et al. [?] used EnKF to assimilate LAI into the SUNFLO crop model for sunflower yield prediction, comparing direct simulation with different LAI assimilation algorithms. They found that assimilating LAI into crop models improved yield prediction results compared to simulation alone (RMSE decreased from 9.88 to 7.49 q/ha), and EnKF further improved accuracy compared to Least Square Estimator (RMSE decreased from 7.92 to 7.49 q/ha), significantly improving sunflower yield prediction. Studies show that data assimilation technology can compensate for the complementarity between remote sensing and field observation data, thereby improving regional oil crop yield estimation accuracy.

(2) Assimilation Grids. Assimilation grids divide the geographic space involved in crop models and remote sensing observations into different unit areas or grids for data assimilation and model updating. In regional crop yield simulation based on data assimilation, assimilation grid selection is closely related to final yield assimilation accuracy [?].

Assimilation grid size depends not only on satellite remote sensing resolution but also on crop model input parameter resolution (meteorological elements, crop and soil parameters, and field management measures). Gaso et al. [?] assimilated remote sensing-inverted LAI information into soybean growth models for yield prediction, finding spatial variation in LAI across different soybean fields and that remote sensing spatial resolution directly affected assimilation accuracy. Gaso et al. [?] used variational assimilation to incorporate observed LAI into crop growth models for field-scale soybean yield prediction, showing that within-field LAI variability introduced uncertainty in yield prediction, affecting accuracy. Tang et al. [?] conducted assimilation yield estimation for rapeseed using LAI inverted from MODIS data, demonstrating that assimilation results were significantly influenced by LAI curves and that correcting LAI bias could improve accuracy. Thus, remote sensing spatial resolution and model input parameter resolution both affect assimilation yield estimation results.

With improved satellite remote sensing spatial resolution, finer assimilation grids can be obtained, making spatial differences in yield simulation more significant. However, continuously reducing assimilation grid size does not always improve assimilation accuracy; instead, an optimal assimilation unit exists, closely related to field plot size [?]. Reasonable assimilation grid selection is key to achieving accurate regional crop yield simulation. Therefore, considering assimilation grid size and conducting oil crop yield estimation research based on data assimilation technology represents an important future research direction.

3 Challenges and Prospects for Oil Crop Yield Monitoring

Compared with traditional ground survey monitoring methods, remote sensing technology provides technical support for oil crop growth monitoring and yield estimation, representing an important application of remote sensing technology in agricultural production [?].

In recent years, remote sensing technology has made significant progress in oil crop monitoring and gained widespread global attention and application. In addition to commonly used foreign remote sensing satellites such as the Sentinel series, Landsat series, MODIS, and RADARSAT, China has also launched resource series, Gaofen series, and environmental satellites, achieving important accomplishments in optical and microwave research. Using remote sensing satellite data for oil crop yield estimation research has become a future development trend. With continuous remote sensing technology development, its application in oil crop monitoring will receive increasing attention from researchers worldwide.

3.1 Main Challenges

Remote sensing technology provides comprehensive data support for oil crop monitoring and yield estimation, making such research highly significant. However, due to the complexity of oil crop planting layouts and structural characteristics, regional oil crop yield simulation accuracy remains low with considerable technical difficulty. Therefore, applying remote sensing yield estimation technology to Chinese oil crop research faces challenges in crop structure, distribution, and remote sensing data source selection.

3.1.1 Crop Feature Selection In assimilation yield estimation systems, assimilation variables bridging crop models and remote sensing observations directly affect assimilation efficiency and accuracy. LAI is closely related to crop yield [?] and is commonly used as an assimilation variable in staple crop yield estimation [?, ?]. However, research shows that using only LAI for yield prediction in oil crops such as soybean and rapeseed results in low estimation accuracy.

Analysis from botany and crop science perspectives reveals that low regional oil crop yield simulation accuracy stems from the presence of photosynthetically active non-leaf organs such as pods or siliques in oil crops like soybean and rapeseed [?]. Taking rapeseed as an example, leaves are the main canopy component from seedling to flowering stage, responsible for most photosynthesis. After flowering, siliques grow rapidly, with surface area increasing quickly, and both leaves and siliques serve as canopy components for photosynthesis. During the silique stage, siliques develop and mature while leaves decline, and rapeseed mainly relies on silique wall photosynthesis to fill seeds, making siliques the primary canopy component [?]. Both rapeseed siliques and soybean pods participate in photosynthesis and contribute to yield formation, with 50%-70% of rapeseed grain yield coming from silique walls, while pods contribute second only to leaves [?, ?]. Thus, organs such as siliques and pods are unique structures in oil crops like soybean and rapeseed, closely related to yield. These special growth structures and physiological indicators significantly impact oil crop yield estimation. Therefore, using only canopy LAI as the assimilation state variable for yield estimation in crops with active photosynthesis in siliques or pods may cause serious yield underestimation, affecting regional crop yield simulation accuracy.

3.1.2 Spatial Scale Determination Oil crop planting distribution is a factor causing low regional yield estimation accuracy. Unlike staple crops, Chinese oil crop fields are generally small, especially rapeseed, which is mainly planted in southern hilly and mountainous areas with complex landscape patterns and fragmented plots. Meanwhile, crop planting structures are chaotic, with multiple crop types in fields and buildings such as factories constructed in fields in disorderly fashion. These conditions increase mixed pixels, reducing rapeseed information extraction within unit grids and affecting assimilation accuracy. From a remote sensing perspective, using regularized yield simulation grids for regional oil crop yield estimation may confuse numerous background ground objects, introducing additional errors and affecting oil crop yield assimilation accuracy. Therefore, paying attention to crop planting area characteristics is scientifically significant for regional oil crop yield estimation.

3.1.3 Remote Sensing Data Selection Optical remote sensing is widely used in crop yield estimation. However, for oil crops such as rapeseed and soybean, the shape, arrangement, and distribution of siliques and pods differ from leaves, making it difficult for optical remote sensing reflectance characteristics alone to support remote sensing inversion of these three-dimensional photosynthetic organ parameters. Additionally, considering Chinese oil crop planting distribution and spatial layout characteristics, sufficient optical remote sensing data may be difficult to obtain due to meteorological conditions. China's key soybean growth stages occur in summer and autumn with frequent clouds and rain, causing meteorological interference with optical remote sensing data. Rapeseed main production areas are located in southern hilly and mountainous regions with year-round cloudy and rainy conditions, making optical remote sensing data acquisition difficult during key growth stages. Therefore, using only optical data affects regional oil crop yield estimation results. SAR signals are unaffected by meteorological conditions such as clouds and rain, with strong penetration capability to obtain surface information under crop canopies. However, due to strong reflection and scattering, SAR signals are affected by oil crop canopy structure, causing signal confusion. Additionally, interactions between SAR data and ground objects produce multiple scattering patterns, leading to complex scattering mechanisms and increased difficulty in data interpretation and processing, which also presents challenges for yield estimation research. Thus, remote sensing data source selection also affects oil crop yield estimation research.

3.2 Prospects

Remote sensing technology faces challenges and difficulties in crop structure, distribution, and multi-source data synergy when applied to oil crop yield estimation. With continuous remote sensing technology development and improvement, it will play an increasingly important role in oil crop yield estimation, providing more accurate and efficient information support for agricultural production and resource management.

First, regarding crop feature selection, it is necessary to comprehensively consider crop plant characteristics and the agronomic mechanisms of yield formation through siliques and pods. The photosynthesis of rapeseed siliques and legume pods is highly active and crucial for yield formation in rapeseed and soybean. Therefore, when determining bridging parameters and optimizing assimilation algorithms in crop assimilation yield estimation systems, it is necessary to comprehensively consider the combined effects of crop plant characteristics and silique or pod wall area index on yield formation. Correcting LAI by incorporating silique or pod wall area index to improve crop yield estimation accuracy represents an important development direction for oil crop yield monitoring.

Second, regarding spatial scale determination, using regularized yield simulation grids may confuse numerous background ground objects, introducing additional errors and affecting yield assimilation accuracy. Therefore, referencing remote sensing mapping research methods, spatial heterogeneity discriminant functions can be established based on field management, crop varieties, soil characteristics, and assimilation variable spatial distributions for spatial clustering, or semantic segmentation based on object-oriented knowledge (such as boundary-based, region-based, or machine learning approaches) can be used to divide yield estimation grids or units. This represents an important development direction for oil crop yield monitoring.

Finally, regarding remote sensing data selection, optical data are susceptible to climate conditions, and using only optical remote sensing to obtain oil crop structural information has limitations. SAR data not only provides all-day, all-weather observation capability unaffected by meteorological conditions but its side-looking slant-range projection imaging capability is also highly sensitive to the three-dimensional structures of rapeseed siliques and soybean pods. Particularly, C-band radar microwaves can penetrate crop canopies, scattering multiple times among stems, leaves, siliques, or pods, containing canopy information that compensates for limitations in optical remote sensing data perception capabilities and coverage. Therefore, combining optical data with radar off-nadir remote sensing measurement technology advantages to perceive and derive response relationships between crop leaves, siliques, pods, and remote sensing data parameters, bridging crop characteristics and remote sensing information for crop yield simulation, is also an important development direction for oil crop yield monitoring.

In summary, remote sensing technology is widely applied and advantageous in oil crop monitoring and yield estimation but still faces challenges requiring further research and exploration in crop structure, distribution, and multi-source data synergy. With continuous remote sensing technology development and improvement, future remote sensing technology will play an increasingly important role in oil crop yield estimation, providing more accurate and efficient information support for agricultural production and resource management.

Conflict of Interest Statement: This study has no conflicts of interest among researchers or with publicly disclosed research findings.

References: (The references section contains 111 citations that are preserved in the original format but not translated here for brevity.)

Note: Figure translations are in progress. See original paper for figures.