Science of the Total Environment 909 (2024) 168594

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Quantifying key vegetation parameters from Sentinel-3 and MODIS over the eastern Eurasian steppe with a Bayesian geostatistical model

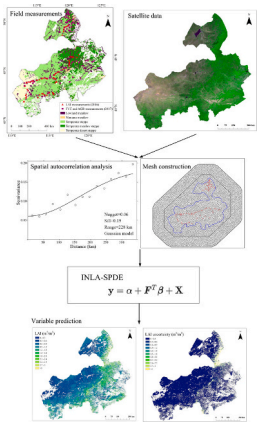
Zhenwang Li [a,b], Lei Ding [c], Beibei Shen [d], Jiquan Chen [e], Dawei Xu [d], Xu Wang [d], Wei Fang [f], Alim Pulatov [g], Maira Kussainova [h], Amartuvshin Amarjargal [i], Erkin Isaev [j], Tao Liu [a,b], Chengming Sun [a,b], Xiaoping Xin [d,*]

[a] Jiangsu Key Laboratory of Crop Genetics and Physiology, Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College, Yangzhou University, Yangzhou 225009, China
[b] Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou 225009, China
[c] College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China
[d] State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, National Hulunber Grassland Ecosystem Observation and Research Station, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[e] Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI 48824, USA
[f] Department of Biology, Pace University, New York, NY 10038, USA
[g] EcoGIS center, National Research University "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" (NRU-TIIAME), Tashkent 100000, Uzbekistan
[h] Sustainable Agriculture Center, Kazakh National Agrarian Research University, Almaty 050010, Kazakhstan
[i] University of the Humanities, Ulaanbaatar, Mongolia
[j] Mountain Societies Research Institute, University of Central Asia, Bishkek 720001, Kyrgyzstan

HIGHLIGHTS

- The potential of Sentinel-3 OLCI data was evaluated for estimating grassland parameters (LAI, FVC, and AGB).
- The Bayesian spatial model performed better than random forest and random forests kriging method.
- S3 OLCI data presented higher accuracy than MODIS data for estimating LAI, FVC, and AGB.
- The predicted results presented higher accuracy than PROBA-V and MODIS global land products.
- The employment of the red-edge bands improved the performance of vegetation variable estimation.

GRAPHICAL ABSTRACT



* Corresponding author.
E-mail address: xinxiaoping@caas.cn (X. Xin).

https://doi.org/10.1016/j.scitotenv.2023.168594
Received 15 February 2023; Received in revised form 5 November 2023; Accepted 13 November 2023
Available online 14 November 2023

ARTICLE INFO

Guest Editor: Jiaguo Qi

Keywords:
Biophysical parameter
Remote sensing
Bayesian modeling
Sentinel-3
MODIS
Grassland

ABSTRACT

Accurate estimation of grassland leaf area index (LAI), fractional vegetation cover (FVC), and aboveground biomass (AGB) is fundamental in grassland studies. The newly launched Ocean and Land Color Imager (OLCI) sensor onboard Sentinel-3 (S3) provides images with comparable spatial and spectral resolution with MODIS data. However, the use of S3 OLCI imageries for vegetation variable estimation is rarely evaluated. This study evaluated the potential of S3 OLCI and MODIS data for estimating grassland LAI, FVC, and AGB in the eastern Eurasian steppe. A Bayesian spatial model (Integrated Nested Laplace Approximation with Stochastic Partial Differential Equation, INLA-SPDE) was used to address spatial autocorrelation of in-situ observation data and to enhance our predictions. Our results showed that the models based on S3 OLCI data presented higher accuracy than models with MODIS data. The RMSEs decreased by 3.7–10.8 %, 3.7–7.5 %, and 1.6–14.2 % for LAI, FVC, and AGB predictions, respectively. Through combinations of multiple predictors, we confirmed the robustness of red edge bands for grassland variable estimation, the models employing red edge variables yielded 3.5 %, 3.2 %, and 0.4 % lower RMSEs than models with conventional visible and NIR bands for LAI, FVC, and AGB prediction, respectively. INLA-SPDE spatial model produced lower bias and higher prediction accuracy than random forest and random forests kriging method in most of the models; the INLA-SPDE predicted LAI and FVC maps also showed a better agreement with ground observations than MODIS and PROBA-V land products.

1. Introduction

Leaf area index (LAI), fractional vegetation cover (FVC), and aboveground biomass (AGB) are essential grassland biophysical parameters in modeling many ecosystem processes such as photosynthesis and transpiration (Baret, 2016; De Grave et al., 2020; Rocha et al., 2019). Conventional destructive or visual methods for estimating abovementioned variables are time-consuming and labor-intensive, which are applicable only for small-scale monitoring (Catchpole and Wheeler, 1992; Liang and Wang, 2020). Remote sensing (RS) meanwhile can provide an approach for large-scale land surface observations with high efficiency, including modeling these variables using multiple optical spectrum (Ali et al., 2016; De Grave et al., 2020).

The most common methods in RS estimation of these parameters include regression methods and physical based methods. In particular, the machine learning (ML) algorithms are increasingly employed because of their ability to combine different data structure features in a non-linear manner and to conform to the requirements of different tasks (Andreatta et al., 2022; Verrelst et al., 2015). However, current practice in model assessment has mainly focused on model fitting and overall accuracy, with a fundamental assumption that the ground observations are randomly distributed for calibrating RS products. Such an assumption of independent and random sampling is often violated because spatial autocorrelation is very common in both RS data (Rocha et al., 2019) and in-situ observations (Fang, 2005), which would significantly inflate the degree of freedom in statistical testing. Theoretically, plant traits from closer locations are more likely to be similar than those in distance (Tobler, 1970). When dealing with spatially autocorrelated datasets and RS imageries, conventional parametric statistics would not be appropriate (Rocha et al., 2019; Rocha et al., 2018).

Spatial models based on Bayesian inference have become more popular after powerful computational methods such as Markov Chain Monte Carlo (MCMC) became available (Banerjee and Fuentes, 2012; Gilks et al., 1995). Due to data subsampling and long burn-in periods, the use of MCMC methods is often computationally demanding, especially in the analysis of spatial data (Moraga et al., 2021). More recently, a novel Bayesian spatial model, the Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009) with the Stochastic Partial Differential Equation (SPDE) approach (Lindgren et al., 2011), was introduced by Poggio et al. (2016). Using a mesh to create a neighborhood structure and account for the spatial dependency for continuous surface, INLA-SPDE model is more computationally efficient than MCMC and has been successfully used in a large number of spatial problems including human health (Moraga et al., 2021), air quality (van Donkelaar et al., 2016), and soil mapping (Huang et al., 2017). However, very few studies applied INLA-SPDE model in vegetation variable estimation (Rocha et al., 2019).

Sentinel-3 (S3) is a new earth observation satellite launched by European Space Agency (Donlon et al., 2012). The Ocean and Land Color Instrument (OLCI) is a multispectral radiometer carried on board Sentinel-3A (launched in 2016) and Sentinel-3B (launched in 2018) with 21 bands in the 400–1020 nm spectral range at an approximately 300 m spatial resolution. The two satellites can provide a revisit time of less than two days at the equator, with higher overpass frequencies at higher latitudes. Compared to the widely used Moderate-resolution Imaging Spectroradiometer (MODIS) data, the S3 OLCI has 3 red-edge bands that are significant in terms of correlation with vegetation growth, which is expected to improve the accuracy of estimating vegetation biophysical variables. The S3 OLCI images have been employed for water clarity monitoring (Shen et al., 2020), water chlorophyll-a and optical properties retrieval (Kravitz et al., 2020; Xue et al., 2019). However, only a few studies explored the use of S3 OLCI imagery for vegetation variable estimation, and the comparison of its performance with widely used MODIS data has rarely been reported (De Grave et al., 2020; Reyes-Muñoz et al., 2022).

Grasslands are the biggest terrestrial ecosystems in the world and play a crucial role in soil and water conservation, biodiversity protection, and livestock production (Strömberg and Staver, 2022). Precisely retrieving of grassland biophysical variables at both regional and global scales is thus important for sustainable grassland management and understanding the responses and feedback of grass-lands to climatic change (Ali et al., 2016; Xu et al., 2021). Eurasia grassland is the most centralized and has the largest distribution globally, accounting for approximately 1/3 of the total global grassland area and 22 % of the total Eurasia area. In this context, our study objective was to apply the state-of-the-art INLA-SPDE spatial model to estimate three grassland essential biophysical variables (LAI, FVC, and AGB) from two main sources of RS data (S3 OLCI and MODIS) over the eastern Eurasian steppe. Specific objectives are to: 1) compare the performance of S3 OLCI and MODIS data in predicting three grassland essential variables (LAI, FVC, and AGB); 2) examine the prediction performance of INLA-SPDE model by comparing with random forests (RF) and random forests kriging (RFK); and 3) examine the significance of S3 red-edge bands and MODIS SWIR bands on LAI, FVC, and AGB prediction. The results of this study can support regional management of environmental and natural resources and promote the study of climate, biogeochemical cycles and vegetation dynamics.

2. Materials and methods

2.1. Study area

The study target is the grasslands in northern China (Fig. 1) that includes eastern part of Inner Mongolia and a small area of western Jilin

province. The area is also the most eastern region of Eurasian steppe, where grasslands are the basis for the animal husbandry. The average temperature in the study area decreases gradually from south to north, and the annual precipitation decreases gradually from east to west from 450 mm to 150 mm. The five dominant grassland types in the study area are temperate steppe (TS), temperate meadow steppe (TMS), temperate desert steppe (TDS), montane meadow (MM), and lowland meadow (LM). The detailed description of the five grassland types in northern

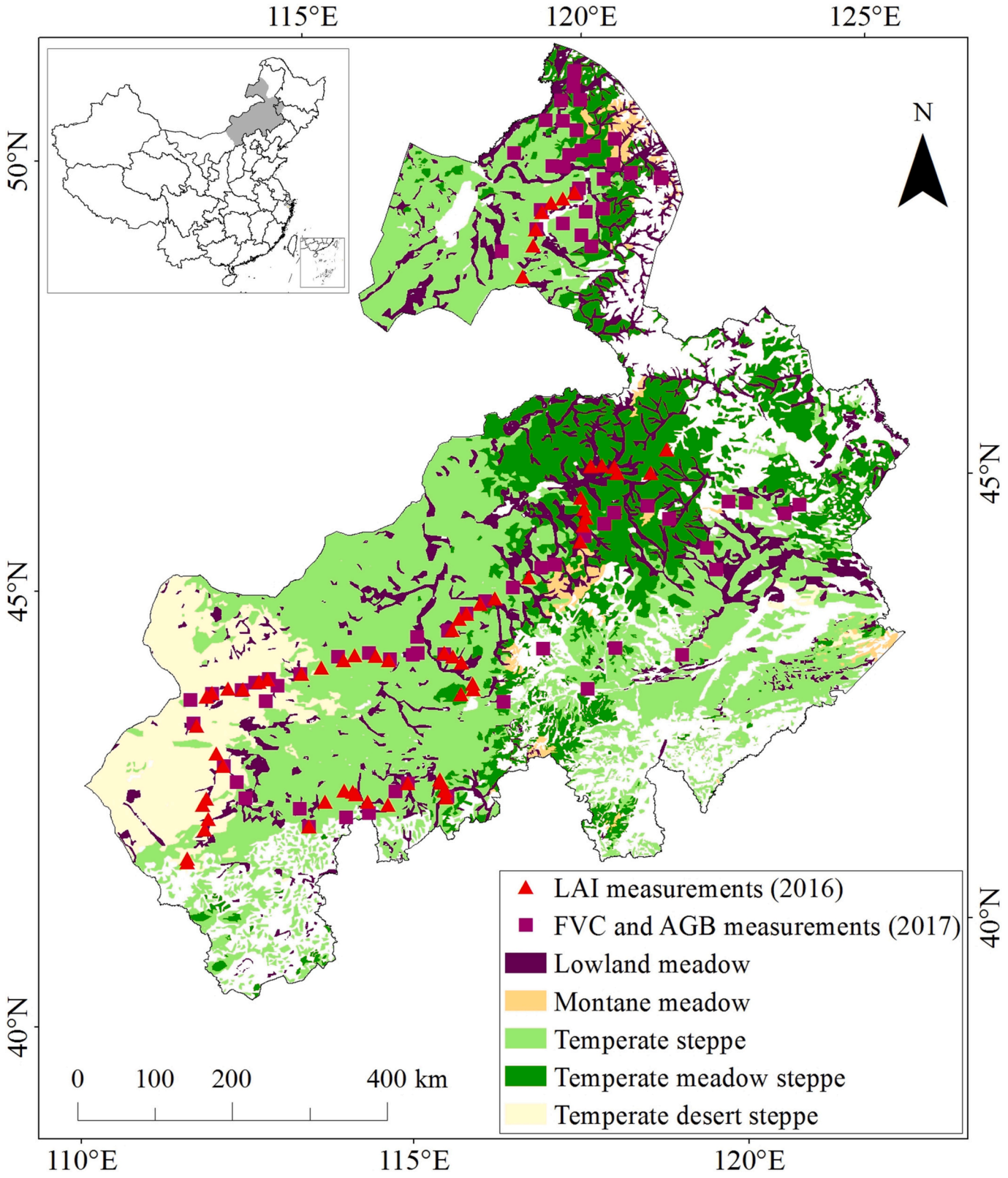


Fig. 1. The distribution of five dominant grassland types in the study region of northern China. Red triangles show the ground sampling plots of leaf area index (LAI) in 2016, purple squares show the ground sampling plots of fractional vegetation cover (FVC) and aboveground biomass (AGB) in 2017.

China can be found in Kang et al. (2007) and Li et al. (2020).

2.2. Ground data measurements

The field surveys were conducted during the peak of the growing season in July and August in 2016 and 2017. Two sampling transects were set up in the study area. One sampling transect was set along the West-East Transect (ranging from 111°E to 122°E), where the primary driving gradient is precipitation. Another sampling transect was set along the North–South Transect of the study region (ranging from 43°N to 51°N), where the primary driving gradient is the temperature (Xu et al., 2021). Taking grassland types, climate, topography condition, and reachability into account, the locations of the sampling plots were carefully selected to maintain their representativeness of the study area. In total, 68 LAI sampling plots were collected during July 29 to August 7 in 2016, 79 FVC and AGB sampling plots were collected during July 20 to August 5 in 2017. For each sampling plot, the area was 500 m × 500 m with homogeneous canopy and flat surface. Geographical coordinates of the central of the sampling plots were recorded using a Garmin GPS Map 62 s unit. At each plot, five 30 m × 30 m quadrats were laid out with one at the center of the plot and one each at the mid-point of the diagonal between the center and the corner. The grassland AGB was obtained using the harvest method: three 1 m × 1 m samples were randomly selected in each quadrat. Fresh grass in the samples was cut from ground with stubble no taller than 0.5 cm and brought to the laboratory. The fresh aboveground samples were dried at 85 °C for 48 h in the oven, and the weight was measured on an electronic scale with 0.01 g precision. The AGB of a sampling plot is the average weight of the five quadrats. The effective LAI was measured using an LAI-2200C plant canopy analyzer (Li-Cor, Lincoln, NE, USA). At each quadrat, one above-canopy and six below-canopy LAI-2200 measurements were recorded to obtain one effective LAI value before grass harvest, and then the five effective LAI values of each plot were averaged to calculate a mean LAI value. FVC was measured using a Canon 60D digital camera. 15–20 digital images were acquired along the two diagonals of each plot at the same canopy height. The modified excess green index (MExG), suggested by Tang et al. (2003) was then used to extract ground FVC. The MExG is defined as: MExG = 2 × G-R-B, where R, G, and B are red, green, and blue band of the digital images. The segmentation threshold was set to 30 to segment grass and soil. The ratio of vegetation (FVC) in the binary image was calculated, and FVC values in one plot were averaged to calculate a mean value.

2.3. Satellite data

2.3.1. Sentinel-3 data

The Sentinel-3 OLCI full resolution (EFR, 300 m spatial resolution) top-of-atmosphere (TOA) radiance products (Level-1) from Sentinel-3A were downloaded from the Copernicus Open Access Hub website. A total of 8 scenes for 2016 and 7 scenes for 2017 of S3A OLCI images covering the ground sampling periods and study site were collected (Appendix A). The S3 OLCI EFR Level-1 TOA product has undergone radiometric correction and geometric correction processing. Here we used the image correction for atmospheric effects (iCOR) plugin (De Keukelaere et al., 2018) of SNAP to conduct atmospheric correction to retrieve the surface reflectance products. Then a multi-day composition of the atmospherically corrected images was composed with the cloud contaminated pixels being removed. For each pixel, a value with the clearest condition identified by the cloud mask layer is selected from all the acquisitions within the composite period of each year, and a cloud free surface reflectance image for each year was generated. In this study, a total of 16 relevant S3 bands including Oa1-Oa12, Oa16-Oa18 and Oa21 were used in the predictive models (Appendix A); 11 vegetation indices (VIs) that derived from S3 OLCI surface reflectance (Appendix B), four solar and viewing angles (solar zenith angle (SZA), solar azimuth angle (SAA), view zenith angle (VZA), and view azimuth angle (VAA)) and geographic information (longitude and latitude) were also used in the grassland variable estimation. To decrease the effect of geolocation errors of ground measurements and S3 reflectance images, the average values of the pixel that cover the location of ground measurements and its surrounding 4 pixels were used to train and validate the models.

2.3.2. MODIS data

The MODIS nadir bidirectional reflectance distribution function (BRDF) adjusted reflectance (NBAR) product (MCD43A4 V0061) (Schaaf and Wang, 2021) was acquired from Earthdata (http://earthdata.nasa.gov). This product provides bands 1–7 at 500 m resolution in a 16-day time step, both Terra and Aqua data from a 16-day period were used to provide highest quality input data. The view angle effects are removed from the directional reflectance, resulting in a stable and consistent NBAR product. For each composite time interval in 2016 and 2017, three MODIS tiles (h25v03, h25v04, and h26v04) were mosaicked to cover the entire study region and then re-projected from original sinusoidal projection to geographic projection on the WGS84 datum through the MODIS reprojection tool (version 4.1). The 7 MODIS bands and 9 VIs derived from MODIS surface reflectance were used to estimate grassland variables (Appendix B). The S3 OLCI and MODIS images were then all resampled to 300 m spatial resolution using nearest neighbor method, and clipped to obtain images covering the study area.

2.3.3. Global land products

To further validate our predicted grassland LAI and FVC maps, two widely-used global land products were collected for intercomparisons. The MODIS LAI product (MCD15A2H v061) was collected from the NASA's Earthdata search client website (https://search.earthdata.nasa.gov/). The product has a spatial resolution of 500 m and a temporal interval of 8 days. Two algorithms were used by NASA to generate the products, the main algorithm is based on the use of Look Up Tables (LUTs) built for six different types of biomes, with simulations from a three-dimensional radiative transfer model (Knyazikhin et al., 1998). When the main algorithm fails, a backup algorithm is triggered to estimate the LAI based on the same radiative transfer model simulated LAI-NDVI relationships. In this study, only the main algorithm retrievals were considered.

The Copernicus Global Land Service (CGLS) PROBA-V LAI and FVC products were collected from the Copernicus Global Land portal (https://land.copernicus.eu). This product provides global time series of LAI and FVC data at a resolution of 300 m and a frequency of 10 days. The retrieval algorithm of the CGLS PROBA-V LAI and FVC products is based on a two-step process. Firstly, daily atmospherically corrected TOA reflectance in the blue, red and NIR PROBA-V spectral bands are used as inputs to the neural networks (NNTs), which were calibrated using four years of VEGETATION-2 reflectance data as input, to retrieve daily estimates of LAI and FVC. Then, the final 10-day product values are generated using a dedicated compositing scheme over a dissymmetric temporal window (Baret et al., 2013).

2.4. Prediction methods

2.4.1. INLA-SPDE model

The spatial models for grassland variable prediction were fitted using the INLA-SPDE approach available in the R package R-INLA (Lindgren and Rue, 2015). This model accounts for the spatial dependency using a mesh to represent the Matérn function and the default settings of priors on the hyper-parameters (Bakka et al., 2018). INLA uses a combination of analytical approximation and numerical integration to do approximate Bayesian inference in latent Gaussian models which includes a large class of models ranging from generalized linear mixed to spatial and spatio-temporal models (Rue et al., 2009). The spatial sparsity structure for the precision matrix is obtained by SPDE to link the Gaussian fields to the Gaussian Markov random fields (Lindgren et al.,

2011). Thus, the INLA-SPDE model is not solely built for discretely observed data location or for a grid, but approximate the entire processes defined on continuous domains (Rocha et al., 2019). A detailed description of the INLA-SPDE model is provided in the Supplementary Material. Fig. S1 shows the mesh selected and used in this study.

2.4.2. Random forests

The RF model is an ensemble learning method that can be used for either classification or regression (Breiman, 2001). The algorithm is a combination of tree predictors such that each tree depends on a collection of random variables sampled independently and then aggregates to produce accurate predictions. This method has showed better resistance to the over-fitting problem and to noise in the data compared with other regression methods (Prasad et al., 2006). In this study, n_{tree} of the RF model was set to 500 (after exploratory trials using the training data). For m_{try}, our previous studies showed that m_{try} optimization resulted in minimal improvements in RF predictions (Li et al., 2017), we therefore used the default m_{try} value (equals to one third of the predictive variable number) set by the RF. In this study, the analysis was accomplished using the "randomForest" package within the statistical software R 3.6.2 (Liaw and Wiener, 2007).

2.4.3. Random forests kriging

Although RF is a robust method that can improve prediction accuracy, this method ignores spatial autocorrelation information. Here a combination of RF and kriging was used to determine the spatial distribution of grassland LAI, FVC, and AGB. Firstly, an ordinary kriging (OK) technique was used to extract the components of the residuals obtained from the RF regression (Guo et al., 2015). The residual is the ground measurements minus the predicted values from RF (Eq. (1)); then the residual values of the sample points were modeled using OK to obtain the structure of the component in the residuals; finally, the final grassland variable estimates was obtained by adding the structure of the component to the RF-based predictions (Eq. (2)). By considering the spatial autocorrelation of the variable residuals through adding the extracted components to the RF-based predictions, a better prediction accuracy of grassland variables can be obtained. A detailed description of the implementation of OK model is provided in the Supplementary Material.

$$grass_{resd}(s_i) = grass_{obs}(s_i) - grass_{RF}(s_i) \quad (1)$$

$$grass_{RFK}(s_i) = grass_{RF}(s_i) + grass_{resd-OK}(s_i) \quad (2)$$

where $grass_{RFK}(s_i)$ is the predicted grassland variable at location s_i, $grass_{RF}(s_i)$ is the trend modeled by RF, and $grass_{resd-OK}(s_i)$ is the residual interpolated by OK, $grass_{obs}(s_i)$ is the observed value of the grassland variable, $grass_{resd}(s_i)$ is the grassland variable residual of the site s_i. This analysis was accomplished using the "randomForest" and "gstat" packages within the statistical software R 3.6.2.

2.5. Scenario development and variable important analysis

To investigate the spectral features causing the differences in the grassland variable prediction performance between S3 OLCI and MODIS, three scenarios for each satellite data were setup with different channel combinations (Table 1). Scenario 1 (SC1) and scenario 4 (SC4) included all the predictor variables for S3 OLCI and MODIS, respectively, scenario 2 (SC2) and scenario 5 (SC5) included optimized combination of predictor variables selected by the variable importance and correlation analysis for S3 OLCI and MODIS, respectively. To examine the significance of S3 red-edge bands and MODIS SWIR bands on the prediction performance of LAI, FVC, and AGB, scenario 3 (SC3) and scenario 6 (SC6) were also set, while SC3 included optimized S3 variables except the red-edge bands and related VIs, SC6 included optimized MODIS variables except the SWIR bands and related VIs.

To demonstrate the most important predictors, the relative importance of each predictor was calculated using the Boruta algorithm (Kursa and Rudnicki, 2010). In this study, the Boruta algorithm was run with 500 maximum runs of random forest, 0.99 confidence level, and z-scores of mean decrease accuracy measure to gather permutation importance. This analysis was accomplished using the "Boruta" packages within the statistical software R 3.6.2.

Before model construction, a correlation analysis was performed between each predictor variable and grassland variables to reduce input dimensionality and select appropriate input (Fig. s1 and Fig. s2 in the Supplementary Material). Variables that have insignificant correlation ($P > 0.01$) with each grassland variables were discarded to avoid bias and exclude the impractical variables, variables that were classified as unimportant by Boruta algorithm were also discarded and not used in the following analysis. The predictors used in each scenario are provided in the Table s1 of the Supplementary Material.

Table 1
Definition of the six scenarios set in this study.

Scenario	Scenario description
SC1	S3 data with all variables
SC2	S3 data with variables after optimization
SC3	Optimized S3 data without the red-edge bands and related VIs
SC4	MODIS data with all variables
SC5	MODIS data with variables after optimization
SC6	Optimized S3 data without SWIR bands and related VIs

2.6. Model evaluation

This study used a 5-fold cross-validation method to evaluate the predictive performance of the models. For the 5-fold cross-validation, the observed data sets were randomly divided into 5 groups. In each of the 5 folds, one group was selected as the test data set and the other 4 groups were used as the training set. Three validation criteria were calculated to evaluate the performance of the model, the root mean square error (RMSE) and the coefficient of determination (R^2).

3. Results

3.1. Descriptive statistics of the measured grassland variables

The mean values for measured LAI, FVC, and AGB were 0.82 m^2/m^2 (ranging from 0.25 to 1.85 m^2/m^2), 35.69 % (ranging from 3.58 % to 91.49 %), and 81.25 g/m^2 (ranging from 14.76 to 198.86 g/m^2), respectively (Table 2). Moreover, the three measured grassland variables all followed a positive skew distribution with the skewness value >0.50 (0.50, 0.53, and 0.71 for LAI, FVC, and AGB, respectively).

The omnidirectional variograms of the measured LAI, FVC, and AGB and their fitted models were determined by the criteria of the sum of squared errors reported by the "gstat" package of the R software (Fig. 2). In general, the three grassland variables all presented strong-medium spatial auto-correlations with varied range and nugget to sill ratio values of the fitted models. The fitted model of FVC has smaller range value (96 km) and nugget to sill ratio value (23.63 %) than LAI and AGB, which indicated a stronger spatial autocorrelation. LAI also showed a strong spatial autocorrelation with a range value of 228 km and nugget to sill ratio value of 31.58 %. AGB showed a moderate spatial autocorrelation with the largest nugget to sill ratio value of 45.97 %.

Table 2
Descriptive analysis of measured grassland LAI, FVC, and AGB.

	Mean	Min	Max	SD	Skewness
LAI (m^2/m^2)	0.82	0.25	1.85	0.35	0.50
FVC (%)	35.69	3.58	91.49	21.57	0.53
AGB (g/m^2)	81.25	14.76	198.86	48.70	0.71

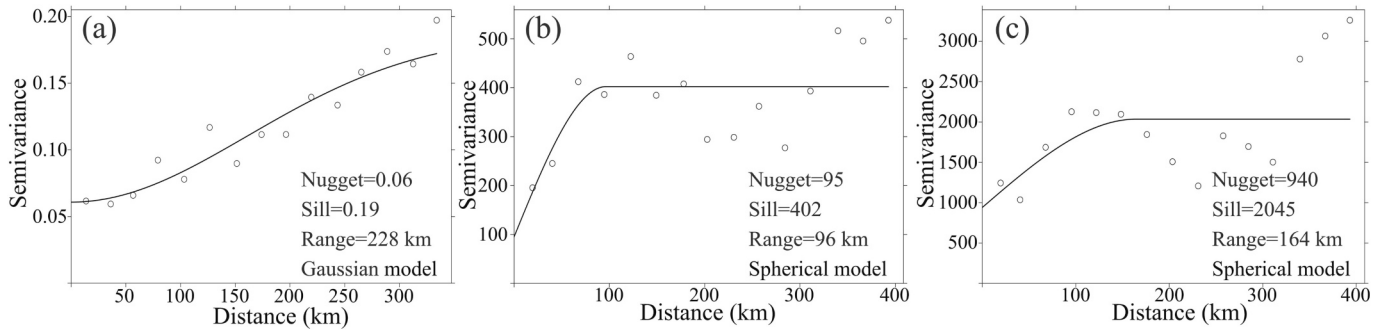


Fig. 2. Variograms, best fitted models, and model parameters for the measured LAI (a), FVC (b), and AGB (c).

3.2. Correlation analysis

The correlation coefficients between the measured grassland variables with predictors derived from S3 and MODIS data were illustrated in Fig. 3, Fig. s2, and Fig. s3. For S3 spectral bands, the three grassland variables are more correlated with the red (Oa8) and red edge bands (Oa9 and Oa10) than the other bands, the correlations between the three grassland variables and Oa4, Oa5, Oa6, and Oa7 are also significant (Fig. 3a). However, difference was observed for Oa12, Oa16, Oa17, Oa18, and Oa21, while FVC and AGB were significantly correlated with the five bands, the correlations between LAI and the five bands were insignificant. For the S3 derived VIs, EVI and $mNDVI_{705}$ presented the highest correlation with measured LAI, followed by $NDVI_{re}$, SAVI, and $CI_{rededge}$, MTCI was less correlated with LAI. FVC and AGB was significantly correlated with EVI, DVI, $mNDVI_{705}$, $NDVI_{re}$, NIRv, and SAVI, weaker correlation coefficients between MTCI with FVC and AGB were observed.

For the MODIS spectral bands, the two SWIR bands (B6 and B7) presented the highest correlation with the three grassland variables among all the seven bands, followed by B1, B3, and B2 (Fig. 3b). The lowest correlations were observed for green band (B4) and NIR band (B5). For the MODIS band derived VIs, the nine VIs all presented significant positive correlations with LAI, FVC, and AGB, among which, NDPI and SAVI presented higher correlation with the three grassland variables.

3.3. Prediction performance

INLA-SPDE method presented significant higher LAI prediction accuracy with larger R^2 and smaller RMSE values than the other two models, while RFK performed slightly better than RF (Fig. 4, Table s2). The R^2 values increased on average 0.11 and 0.10 for INLA-SPDE model when compared to RF and RFK, respectively, and the RMSE values for INLA-SPDE model were on the average 11.40 % and 9.83 % lower than the RF and RFK, respectively.

When all predictors (SC1 and SC4) and optimized predictors (SC2 and SC5) were included in the modeling, S3 derived predictors generated more accurate prediction results than MODIS. The R^2 for S3 predictions varied between 0.50 and 0.61 for SC1 and 0.52 and 0.63 for SC2, for MODIS the corresponding values were 0.49–0.58 and 0.45–0.59. The RMSE values for S3 models were 3.9–12.1 % lower than MODIS.

For different predictor combinations used in the INLA-SPDE model, the S3 optimized predictors (SC2) generated better prediction performance than SC1 and SC3. When the red-edge bands and VIs were removed from the S3 optimized predictors (SC3), the R^2 was 0.034 lower and RMSE was 4.6 % higher than that of SC2. For the MODIS product, optimized predictors without the SWIR bands and VIs (SC6) generated the best performance, while models with all predictors (SC4) performed the worst, the RMSE of SC6 was 4.2 % and 3.2 % lower than SC4 and SC5, respectively.

For FVC prediction, an overall better performance of INLA-SPDE model than RF and RFK model was observed (Fig. 5, Table s3). For

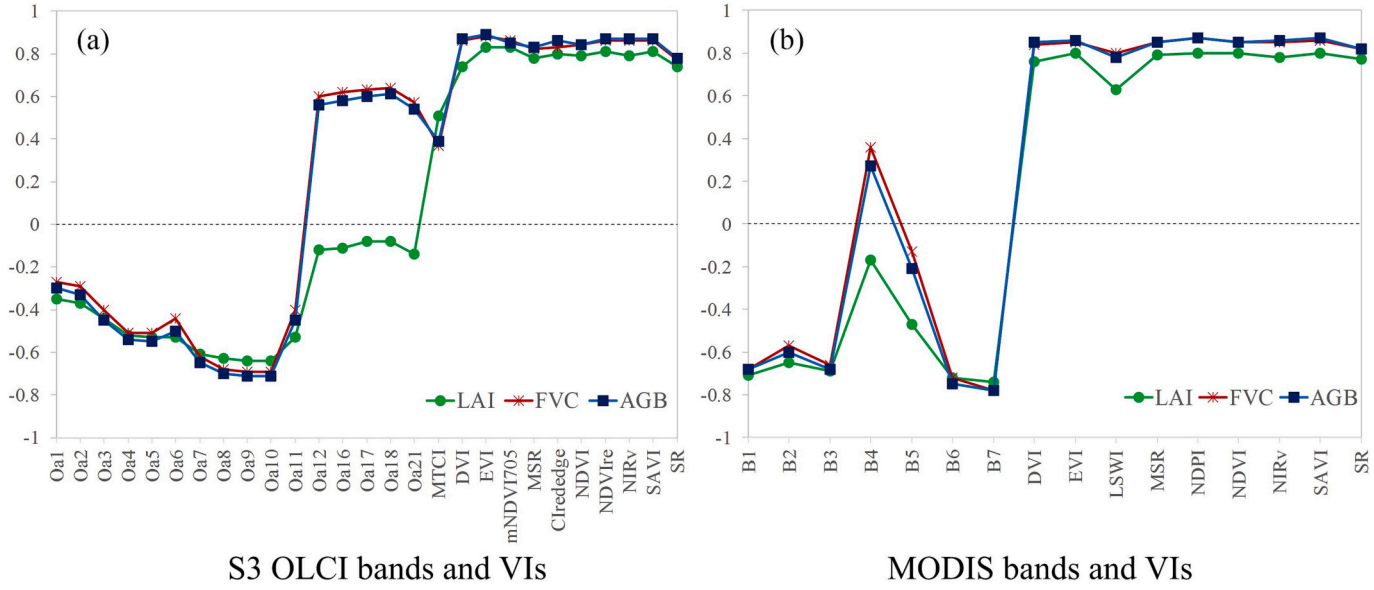


Fig. 3. Correlation coefficients between the three grassland variables and remote sensing predictors derived from S3 OLCI (a) and MODIS (b).

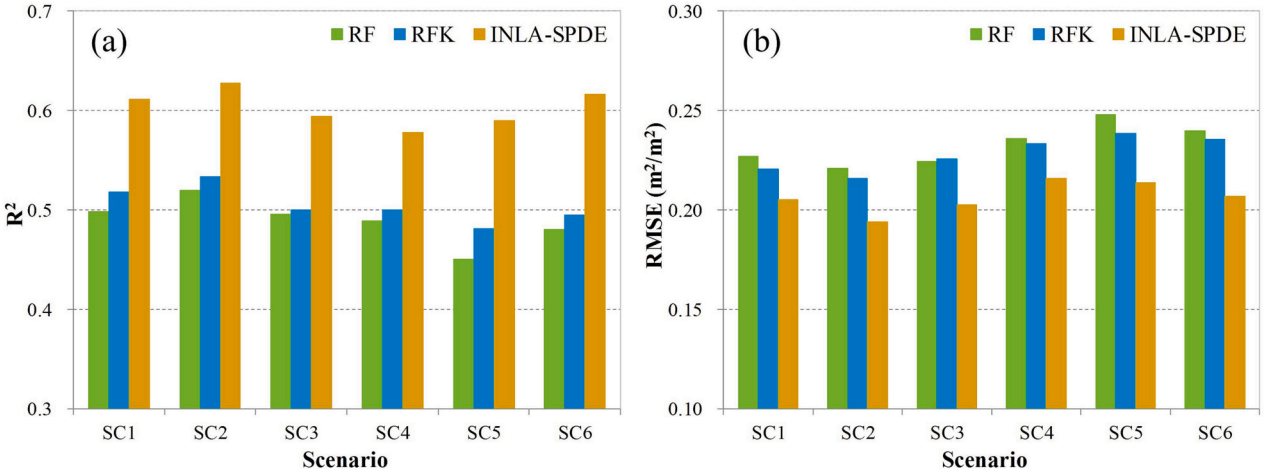


Fig. 4. Comparison of model performance for LAI prediction using RF, RFK, and INLA-SPDE models in 2016. The performance metrics ((a) for R^2 and (b) for RMSE) are calculated with ten-fold cross-validation method.

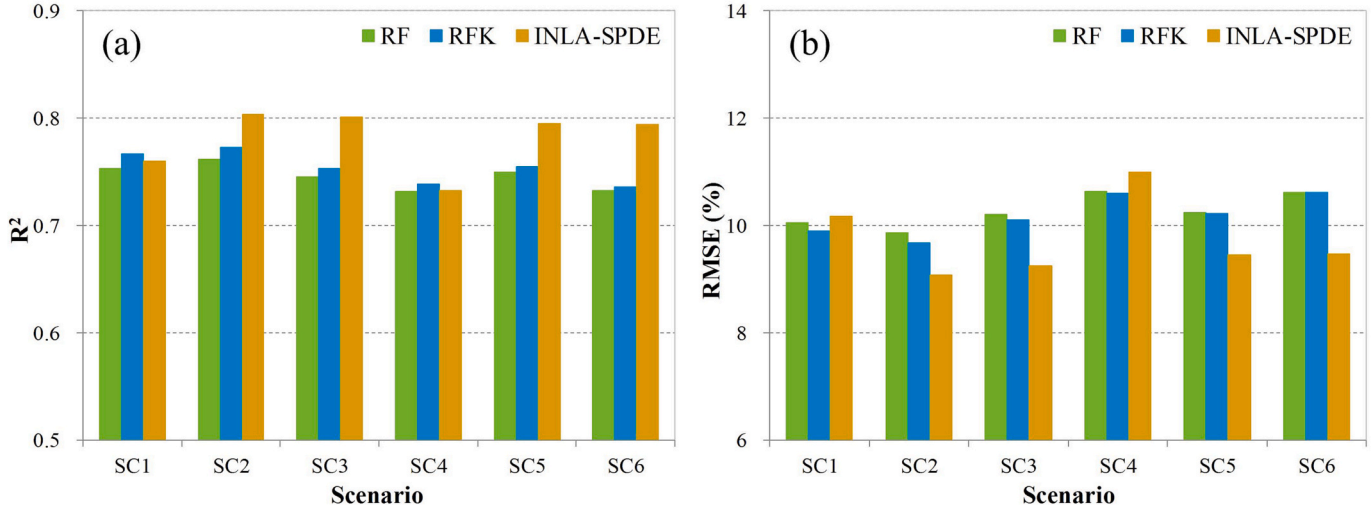


Fig. 5. Comparison of model performance for FVC prediction using RF, RFK, and INLA-SPDE models in 2017. The performance metrics ((a) for R^2 and (b) for RMSE) are calculated with ten-fold cross-validation method.

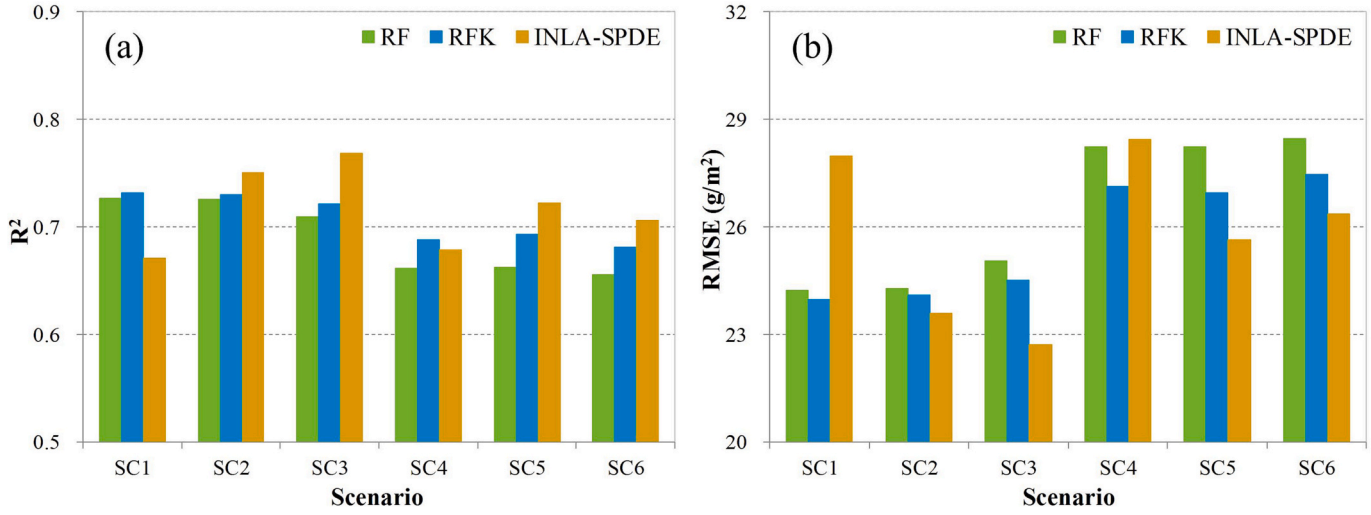


Fig. 6. Comparison of model performance for AGB prediction using RF, RFK, and INLA-SPDE models in 2017. The performance metrics ((a) for R^2 and (b) for RMSE) are calculated with ten-fold cross-validation method.

INLA-SPDE, the R^2 values increased on average 0.04 and 0.03 and the RMSE values decreased on average 5.22 % and 4.43 % when compared to RF and RFK, respectively.

S3 derived predictors generated more accurate prediction results than MODIS when all predictors or optimized predictors were included in the modeling. The R^2 values were 0.76 and 0.73 for INLA-SPDE predicted FVC when all predictors were included for S3 and MODIS, respectively, the corresponding RMSE values were 10.17 % and 11.00 %, respectively.

When using different predictor combinations, models with optimized predictors performed the best for RF and RFK, followed by the models with all the predictors, the models with predictors without red-edge or SWIR predictors performed the worst. For INLA-SPDE model, unlike LAI prediction, the removal of red-edge (SC3) or SWIR (SC6) predictors achieved comparative accuracy with S3 or MODIS optimized predictors (SC2 and SC5) and did not reduce the FVC prediction performance, the models with all predictors performed the worst.

Similar to LAI and FVC prediction, S3 derived predictors generated more accurate AGB prediction results than MODIS, the RMSE decreased by 9.14 % on average for S3 compared to MODIS (Fig. 6, Table s4). Considering different models, the INLA-SPDE performed better than the other two models when optimized predictors (SC2 and SC5) or predictors without red-edge or SWIR variables (SC3 and SC6) were included, followed by the RFK model, and RF model performed the worst. For these four scenarios, the R^2 values of INLA-SPDE predictions was 0.05 and 0.03 higher on average than predictions based on RF and RFK, respectively, the RMSE values of INLA-SPDE predictions were 7.19 % and 4.61 % lower on average than RF and RFK. However, when all the predictors were included in the models (SC1 and SC4), INLA-SPDE showed a worse performance than RFK.

With regards to different predictor combinations, models with optimized predictors derived from S3 or MODIS performed similarly with models with all predictors for RF and RFK predictions, removing the red-edge or SWIR variables from optimized predictors have resulted in worse model performance. For INLA-SPDE predictions, S3 optimized predictors without the red-edge variables (SC3) performed better than SC2 and SC1. MODIS optimized predictors (SC5) performed better than the other two MODIS scenarios, the removal of SWIR bands from predictors (SC6) resulted in a worse performance.

3.4. Comparison with global land products

INLA-SPDE predicted LAI using SC2 as the predict variables showed the lowest degree of uncertainty compared with measured LAI (R^2 = 0.6278, RMSE = 0.1939 m^2/m^2), followed by the PROBA-V LAI (R^2 = 0.6274, RMSE = 0.2974 m^2/m^2, the MODIS LAI has the largest uncertainty (R^2 = 0.4797, RMSE = 0.4281 m^2/m^2) (Fig. 7). Moreover, an overestimation was observed for the MODIS LAI product, which mainly occurred with LAI values >1.5 m^2/m^2. For PROBA-V LAI product, a slight underestimation was observed at LAI values smaller than 0.7 m^2/m^2. Comparatively, INLA-SPDE predicted LAI were closer to the observed values, and the scatter plots were located closer to the 1:1 line.

INLA-SPDE predicted FVC using SC2 as the predict variables were also observed to be closer to the measured FVC values (R^2 = 0.8039, RMSE = 9.0825 %), PROBA-V FVC product also showed a good agreement with measured FVC but with a larger bias (R^2 = 0.7612, RMSE = 12.3248 %) (Fig. 8). However, an underestimation was observed for the PROBA-V FVC product with FVC values smaller than 20 %.

3.5. Relative importance of predictor variables

The relative importance of all the S3 derived predictors in SC1 was ranked using the Boruta algorithm, the results are shown in Fig. 9. Generally, the VIs were identified more important than spectral bands with higher relative influence scores. For LAI prediction, EVI and three red-edge based VIs ($mNDVI_{705}$, $NDVI_{re}$, and $CI_{rededge}$) were ranked more important than the other predictors, DVI and MCTI were identified less important among all the VIs. The two red-edge bands (Oa09 and Oa10) and two red bands (Oa08 and Oa07) were identified more important than the other S3 bands, while Oa02, Oa04, Oa06, Oa12, and Oa16 were less important. For FVC prediction, EVI, NIRv, and $mNDVI_{705}$ were the most important predictors, among all the S3 bands, two red-edge bands (Oa09 and Oa10) and one red band (Oa08) were identified more important. Oa21, Oa11, Oa06 and MTCI contributed little to the FVC prediction. For AGB prediction, EVI and $mNDVI_{705}$ were the most important predictors, the contribution of Oa08, Oa09 and Oa10 are more significant than the other S3 bands.

The relative importance of all the MODIS derived predictors in SC4 was illustrated in Fig. 10. It was observed that the SWIR-based VI (NDPI) was recognized as the most important predictor for FVC and AGB prediction, which is also the top important predictor for LAI prediction (Fig. 10). The two SWIR bands (B6 and B7) were also ranked more important than the other bands. Similar with S3 predictors, the two NIR bands (B2 and B5) were identified as less important with small relative influence scores.

4. Discussion

4.1. Performance of grassland variable prediction models

Non-spatial ML models assume the ground observations as randomly distributed across space and time, spatial autocorrelation is rarely taken into account when predicting vegetation traits. In continuous fields under natural conditions, plant traits are often serially correlated in these two dimensions. Wang et al. (2014), Li et al. (2016) and Su et al. (2020) indicated strong spatial autocorrelations in the model residuals

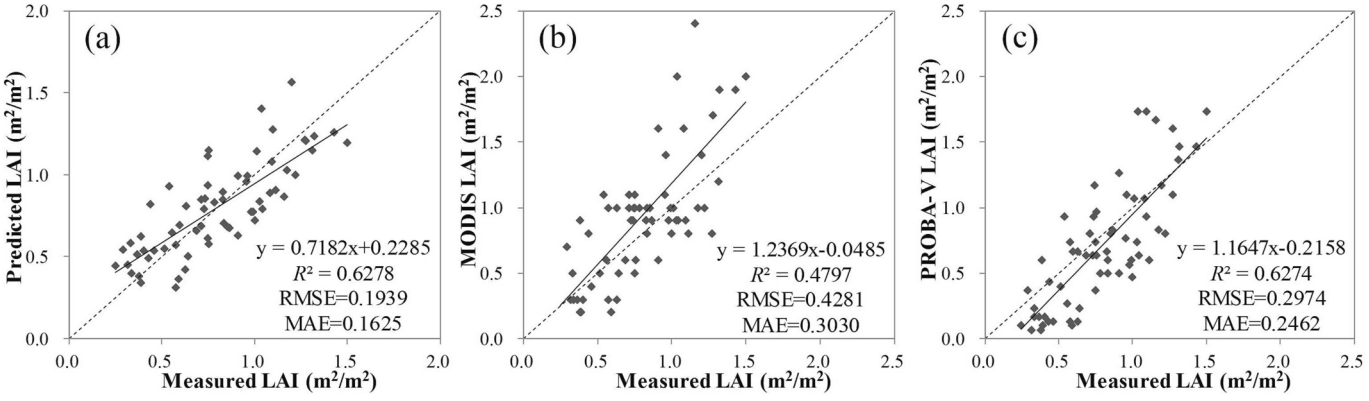


Fig. 7. Scatterplot between the ground measured LAI with INLA-SPDE modeled LAI (a), MODIS LAI (b), and PROBA-V LAI (c).

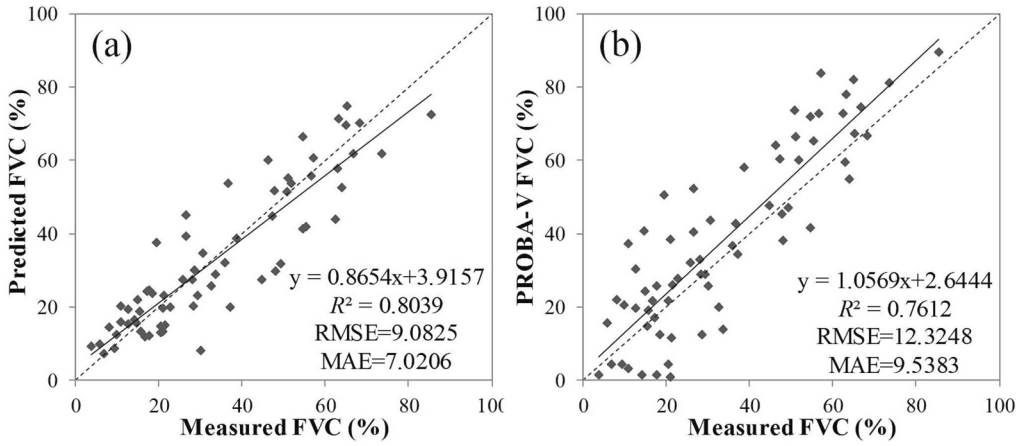


Fig. 8. Scatterplot between the ground measured FVC with INLA-SPDE modeled FVC (a) and PROBA-V FVC (b).

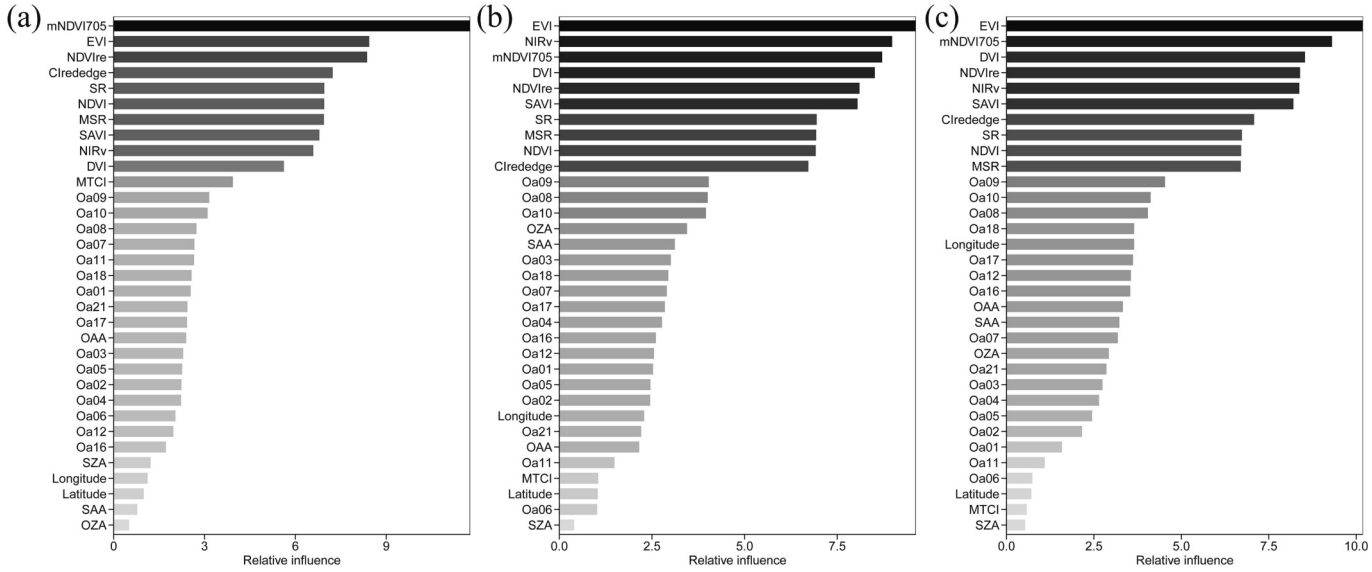


Fig. 9. Relative importance of S3 derived predictor variables ranked by Boruta algorithm for predicting LAI (a), FVC (b), and AGB (c). The depth of the color indicates the level of importance.

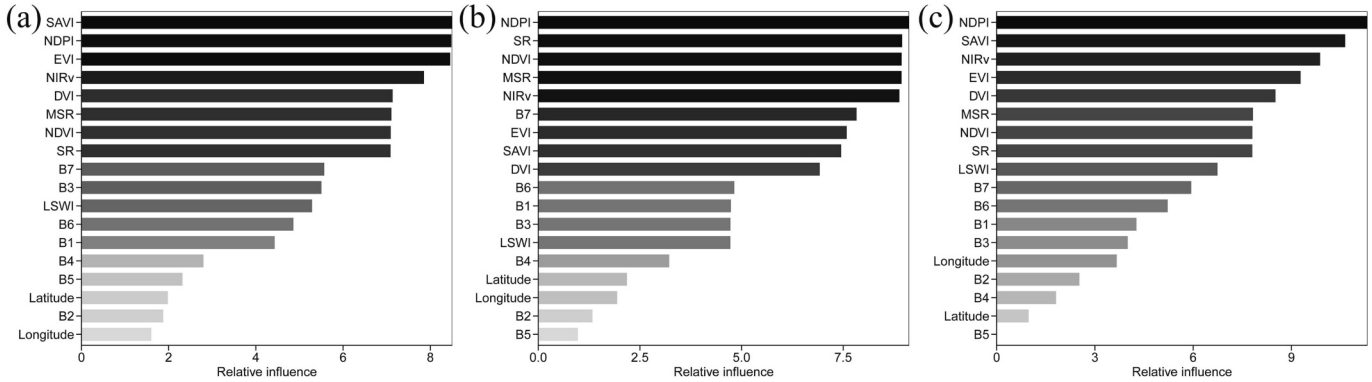


Fig. 10. Relative importance of MODIS derived predictor variables ranked by Boruta algorithm for predicting LAI (a), FVC (b), and AGB (c). The depth of the color indicates the level of importance.

of vegetation LAI and AGB. ML hybrid geostatistical models can compensate the gap to a certain extent by geostatistical modeling the residuals and supplying them to the regression models. Our results also presented an overall higher prediction accuracy of RFK compared to RF model. However, the improvement was limited, which was also observed in Su et al. (2020), the results can be partly attributed to the limited ground measurements used in such a large study region. The geostatistical analysis requires a certain number of ground

measurements with representative locations to include to ensure the model performance, the parameters to fit the semivariogram models of the residuals also have a significant influence to the model performance. Comparatively, INLA-SPDE presented more accurate predictions compared to RF and RFK model, as indicated by higher R^2 and lower RMSE values. INLA-SPDE accounts for the spatial dependency using a mesh to represent the Matérn function (Bakka et al., 2018; Rocha et al., 2019), the model had the potential to map the spatial distribution of environmental variables along with their posterior marginal (Huang et al., 2017). It was also found that stronger spatial autocorrelation of vegetation trait can lead to more accurate predictions for INLA-SPDE (Rocha et al., 2019). Moreover, Huang et al. (2017) indicated that INLA-SPDE is able to achieve robust model performance with sparse datasets, this suggests that INLA-SPDE could be employed when a detailed sampling campaign is not available.

A number of disadvantages of INLA-SPDE also need to be considered. Firstly, the spatial autocorrelation of the vegetation properties directly affects the prediction accuracy of INLA-SPDE, thus a certain number of ground measurements are needed to study the spatial autocorrelations and conduct a geostatistical analysis of the variables before model construction. Besides, we found INLA-SPDE was more susceptible to noise information of the predictor variables. As illustrated in Fig. 5, Fig. 6, and Fig. 7, the INLA-SPDE using all variables included (SC1 and SC4) presented lower prediction accuracy compared to optimized predictors (SC2 and SC5) with lower R^2 and higher RMSE values. In comparison, the RF and RFK algorithms are more robust under multicollinearity and that overfitting can be controlled at both scenarios. Therefore, a feature selection to remove the irrelevant predictors is necessary to decrease dimensionality before conducting INLA-SPDE.

4.2. Performance of S3 OLCI

Our results showed better predictive capability of the models using S3 OLCI data when compared to MODIS data. The better performance of S3 OLCI data can be attributed to the additional image channels in the visible-NIR spectral range, especially the red-edge band, and finer spectral resolution. Narrow and continuous reflectance bands in visible-NIR spectra can include more absorption features related to grassland traits, this allows for better results and higher accuracy of grassland traits prediction (Lee et al., 2004; Thenkabail et al., 2004). When using conventional bands or indices that could be calculated similarly for both sensors (SC3 and SC6), the S3 models were more accurate than MODIS models. In addition, the red-edge bands and VIs are more favorable to the estimation of grassland biophysical variables than the vegetative indices derived from visible and NIR wavelengths (Astola et al., 2019; Dong et al., 2020; Korhonen et al., 2017), the red-edge bands and VIs were consistently selected as the top important S3 predictors. The benefit of S3 with four red-edge bands for vegetation variable estimation is therefore highlighted in comparison with that of MODIS data.

The finer spatial resolution of S3 data (300 m) is more likely to reduce uncertainty in grassland traits estimates because structural and compositional variability over a landscape is better represented than in coarser resolution data such as MODIS imagery (500 m). Moreover, in contrast to coarse resolution data that record mixed spectral signatures within a MODIS pixel, the compatible sizes of S3 image pixels and ground sample plots also reduced the uncertainty of grassland traits estimation (Asam et al., 2013; Deo et al., 2018). The results of Deo et al. (2018) also demonstrated that the accuracy of AGB prediction models decreased with increasing pixel size of the predictors from 30 to 1000 m. Nonetheless, there are tradeoffs in the prediction accuracy versus operational efficiency of optical data as high spatial resolutions have to compromise with low temporal resolutions of sensors. S3-A OLCI allows global coverage to be provided in 2–3 days, while the value is 1–2 days for MODIS data. With the launch of S3—B satellite, the two satellites in polar orbit can provide a revisit time of less than two days.

4.3. The importance of red-edge and SWIR bands

Our results showed that the four S3 red-edge bands (Oa9, Oa10, Oa11, and Oa12) and related VIs ($mNDVI_{705}$, $NDVI_{re}$, and $CI_{rededge}$) all significantly correlated with LAI, FVC, and AGB, they were also ranked more important than the other predictors. The removal of red-edge bands and related VIs (SC3) from the optimized variables (SC2) has resulted in worse prediction performance with smaller R^2 and larger RMSE in most models. Compared to FVC and AGB, the performance of LAI prediction models was improved the most by including red-edge bands and VIs to the predictors. The shape of the red-edge region is strongly influenced by vegetation conditions, an increase in leaf chlorophyll concentration or biomass will result in the broadening of the absorption feature centered around 670 nm, causing the movement of the red-edge position to longer wavelengths (Dawson and Curran, 1998; Mutanga and Skidmore, 2007). At the canopy level, the shape of the red-edge region can also be strongly influenced by LAI because the canopy reflectance in the red-edge region mainly results from the multiple scattering between leaf layers (Delegido et al., 2008; Lee et al., 2004). Thus, the red-edge bands have been successfully used in studies to estimate chlorophyll concentration, biomass and LAI (Dong et al., 2019; Mutanga and Skidmore, 2007). Compared to the conventional red and NIR bands, the red-edge is less sensitive to soil background, leaf angle distribution, and atmospheric effects and thus can improve vegetation trait estimation (Clevers, 1999; Dong et al., 2019). The employment of the red-edge bands can alleviate the saturation effect that may occur for conventional VIs, this will improve the performance of vegetation variable estimation (Mutanga and Skidmore, 2004; Sun et al., 2020). However, red-edge bands cannot completely remove the saturation effect for large LAI and biomass. Using red-edge VIs, Sun et al. (2020) and Andreatta et al. (2022) found that saturation could be still observed in dense canopy for predicting LAI, FVC, and AGB, respectively. In this study, the grassland was mainly located in arid and semiarid regions of northern China, dominated by small and moderate grassland canopy ($LAI < 2.0\ m^2/m^2$, $AGB < 200\ g/m^2$). Accordingly, VI saturation can be overlooked when developing grassland variable estimation models.

Our correlation analysis and variable importance analysis also indicated the dominate role of the MODIS SWIR bands and VIs in grassland variable prediction. When removing these predictors from prediction models (SC6), worse performance for FVC and AGB prediction were observed compared to SC5. The SWIR region is related to the water absorption spectrum (Goward, 1985) and the bands present a faster response to leaf water status (Braga et al., 2021), the differences in leaf absorption properties of vegetation and non-vegetation and between vegetation species can be discriminated with SWIR measurements, which serves as a basis to effectively discriminate dry vegetation from green vegetation and the bare soil. Absorption in the SWIR is also associated with different components of dry vegetation tissues like cellulose, hemicellulose and lignin (Jacques et al., 2014; Kergoat et al., 2015). Thus, SWIR measurements contribute new information about FVC and AGB estimation that were not previously available in visible and near infrared measurements (Kergoat et al., 2015; Wu et al., 2021). Among all the MODIS variables, NDPI was identified the most important variable for FVC and AGB prediction, a comprehensive evaluation of NPDI performance compared to conventional VIs was conducted by Xu et al. (2021), the results found the superiority of NDPI for estimating grassland AGB by including the SWIR band to account for the canopy water content and reduce the effect of soil background. However, our results also found that adding SWIR bands and VIs to visible-NIR predictors did not improve LAI prediction accuracy, which indicated that no additional information beyond visible-NIR predictors was provided by SWIR predictors for LAI prediction.

5. Conclusion

We evaluated the potential of Sentinel-3 OLCI and MODIS data for

estimating three grassland biophysical variables using a Bayesian spatial model. Our results showed that the models based on S3 data outperformed MODIS models for predicting the three variables in terms of RMSE and R^2, which can be attributed to the higher spatial resolution and to the availability of the red-edge wavelength region. The selected INLA-SPDE spatial method that explicitly deal with spatial information resulted in low bias and higher prediction accuracy than RF and RFK in most of the models, the predicted LAI and FVC maps also showed a better agreement with ground observations than MODIS and PROBA-V land products. Our results also highlighted the unique value of red-edge bands in estimating grassland variables, the models utilizing red-edge bands yielded 3.5 %, 3.2 %, and 0.4 % lower RMSEs than conventional visible and NIR bands for LAI, FVC, and AGB prediction, respectively. SWIR bands also improved the prediction accuracy for FVC and AGB prediction, but no additional information beyond visible-NIR predictors was provided by SWIR predictors for LAI prediction. The use of Sentinel-3 OLCI imageries with the state-of-the-art INLA-SPDE spatial model is recommended for large scale vegetation parameters estimation and support grassland resource management.

CRediT authorship contribution statement

Zhenwang Li: Conceptualization, Methodology, Writing - Original draft preparation, Writing - Review and Editing; Lei Ding and Beibei Shen: Data curation, Writing - Original draft preparation; Dawei Xu and Xu Wang: Validation, Investigation; Jiquan Chen, Wei Fang, Alim Pulatov, Maira Kussainova, Amartuvshin Amarjargal, Erkin Isaev, Tao Liu, and Chengming Sun: Writing - Review and Editing; Xiaoping Xin: Conceptualization, Resources, Funding acquisition, Writing - Review and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2021YFD1300500, 2021YFF0703904); the National Natural Science Foundation of China (42301365); the Special Funding for Modern Agricultural Technology Systems from the Chinese Ministry of Agriculture (CARS-34); the Fundamental Research Funds Central Non-profit Scientific Institution (1610132021016); the Special Fund for Independent Innovation of Agricultural Science and Technology in Jiangsu, China (CX(21)3063); and a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Appendix A. The bands and acquisition date of Sentinel-3 and MODIS data used in this study

Sensors	Band central wavelength (unit: nm)	Acquisition date	Spatial resolution (m)
Sentinel-3 OLCI	Oa1: 400; Oa2: 412.5; Oa3: 442.5; Oa4: 490; Oa5: 510; Oa6: 560; Oa7: 620; Oa8: 665; Oa9: 673.75; Oa10: 681.25; Oa11: 708.75; Oa12: 753.75; Oa16: 778.75; Oa17: 865; Oa18: 885; Oa21: 1020	2016211; 2016212; 2016213; 2016214; 2016215; 2016216; 2016217; 2016218; 2017203; 2017205; 2017206; 2017208; 2017209; 2017210; 2017211	300 m
MODIS	B1: 469; B2: 555; B3: 645; B4: 858; B5: 1240; B6: 1640; B7: 2130	2016219 2017213	500 m

Appendix B. Definition of vegetation indices

Name	Abbreviation	Formula	Sensors	Reference
Enhanced vegetation index	EVI	$2.5 \times \frac{(NIR - Red)}{NIR + 6 \times Red - 7.5 \times Blue + 1}$	S3 OLCI MODIS	(Huete et al., 2002)
Simple ratio	SR	$\frac{NIR}{Red}$	S3 OLCI MODIS	(Tucker, 1979)
Normalized difference vegetation index	NDVI	$\frac{NIR - Red}{NIR + Red}$	S3 OLCI MODIS	(Tucker, 1979)
Difference vegetation index	DVI	$NIR - Red$	S3 OLCI MODIS	(Richardsons and Wiegand, 1977)
Modified simple ratio	MSR	$\frac{NIR/Red - 1}{\sqrt{NIR/Red + 1}}$	S3 OLCI MODIS	(Chen, 1996)
Soil-adjusted vegetation index	SAVI	$\frac{NIR - Red}{NIR + Red + L} \times (1 + L)$ where $L = 0.5$	S3 OLCI MODIS	(Huete, 1988)
Near-infrared reflectance of vegetation	NIRv	$NDVI \times NIR$	S3 OLCI MODIS	(Badgley et al., 2017)
Red-edge chlorophyll index	$CI_{rededge}$	$NIR/RE_{705} - 1$	S3 OLCI	(Haboudane et al., 2008)
Red-edge NDVI	$NDVI_{re}$	$\frac{NIR_{750} - RE_{705}}{NIR_{750} + RE_{705}}$	S3 OLCI	(Gitelson and Merzlyak, 1994)
Modified NDVI	$mNDVI_{705}$	$\frac{NIR_{750} - RE_{705}}{NIR_{750} + RE_{705}}$	S3 OLCI	(Datt, 1999)
MERIS total chlorophyll index	MTCI	$\frac{NIR_{750} - RE_{710}}{RE_{710} - RE_{680}}$	S3 OLCI	(Dash and Curran, 2007)

(continued on next page)

(*continued*)

Name	Abbreviation	Formula	Sensors	Reference
Normalized difference phenology index	NDPI	$\frac{NIR-(0.74\times Red+0.26\times SWIR)}{NIR+(0.74\times Red+0.26\times SWIR)}$	MODIS	(Wang et al., 2017)
Land surface water index	LSWI	$\frac{NIR-SWIR}{NIR+SWIR}$	MODIS	(Xiao et al., 2002)

Note: *Blue*, *Red*, *RE*, *NIR*, and *SWIR* represent the atmospherically-corrected surface reflectance in blue, red, red-edge, near-infrared (NIR), and short-wave infrared (SWIR) band, respectively. For S3, the Oa3, Oa8, and Oa17 were used as *Blue*, *Red*, and *NIR* bands, respectively; Oa10, Oa11, and Oa12 were used as RE_{680}, RE_{705}, and NIR_{750} band, respectively, to calculate $CI_{rededge}$, $NDVI_{re}$, $mNDVI_{705}$, and MTCI. For MODIS, the B1, B3, B4, and B6 bands were used as *Blue*, *Red*, *NIR,* and *SWIR* bands, respectively.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.168594.

References

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management. J. Plant Ecol. 9, 649–671.

Andreatta, D., Gianelle, D., Scotton, M., Dalponte, M., 2022. Estimating grassland vegetation cover with remote sensing: a comparison between Landsat-8, Sentinel-2 and PlanetScope imagery. Ecol. Indic. 141, 109102.

Asam, S., Pasolli, L., Notarnicola, C., Klein, D., 2013. Comparison of leaf area indices for grasslands within the Alpine upland based on multi-scale satellite data time series and radiation transfer modeling. In: MultiTemp 2013: 7th International Workshop on the Analysis of Multi-temporal Remote Sensing Images, pp. 1–4.

Astola, H., Häme, T., Sirro, L., Molinier, M., Kilpi, J., 2019. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sens. Environ. 223, 257–273.

Badgley, G., Field, C.B., Berry, J.A., 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. Sci. Adv. 3, e1602244.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., et al., 2018. Spatial modeling with R-INLA: a review. WIREs Comput. Stat. 10, e1443.

Banerjee, S., Fuentes, M., 2012. Bayesian modeling for large spatial datasets. WIREs Comput. Stat. 4, 59–66.

Baret, F., 2016. 2 - estimation of biophysical variables from satellite observations. In: Baghdadi, N., Zribi, M. (Eds.), Land Surface Remote Sensing in Agriculture and Forest. Elsevier, pp. 37–80.

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., et al., 2013. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: principles of development and production. Remote Sens. Environ. 137, 299–309.

Braga, P., Crusiol, L.G.T., Nanni, M.R., Caranhato, A.L.H., Fuhrmann, M.B., Nepomuceno, A.L., et al., 2021. Vegetation indices and NIR-SWIR spectral bands as a phenotyping tool for water status determination in soybean. Precis. Agric. 22, 249–266.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Catchpole, W.R., Wheeler, C.J., 1992. Estimating plant biomass: a review of techniques. Aust. J. Ecol. 17, 121–131.

Chen, J.M., 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. Can. J. Remote. Sens. 22, 229–242.

Clevers, J.G.P.W., 1999. The use of imaging spectrometry for agricultural applications. ISPRS J. Photogramm. Remote Sens. 54, 299–304.

Dash, J., Curran, P.J., 2007. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). Adv. Space Res. 39, 100–104.

Datt, B., 1999. A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using Eucalyptus leaves. J. Plant Physiol. 154, 30–36.

Dawson, T.P., Curran, P.J., 1998. A new technique for interpolating the reflectance red edge position. Int. J. Remote Sens. 19, 2133–2139.

De Grave, C., Verrelst, J., Morcillo-Pallarés, P., Pipia, L., Rivera-Caicedo, J.P., Amin, E., et al., 2020. Quantifying vegetation biophysical variables from the Sentinel-3/FLEX tandem mission: evaluation of the synergy of OLCI and FLORIS data sources. Remote Sens. Environ. 251, 112101.

De Keukelaere, L., Sterckx, S., Adriaensen, S., Knaeps, E., Reusen, I., Giardino, C., et al., 2018. Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: validation for coastal and inland waters. Eur. J. Remote Sens. 51, 525–542.

Delegido, J., Fernández, G., Gandía, S., Moreno, J., 2008. Retrieval of chlorophyll content and LAI of crops using hyperspectral techniques: application to PROBA/CHRIS data. Int. J. Remote Sens. 29, 7107–7127.

Deo, R.K., Domke, G.M., Russell, M.B., Woodall, C.W., Andersen, H.-E., 2018. Evaluating the influence of spatial resolution of Landsat predictors on the accuracy of biomass models for large-area estimation across the eastern USA. Environ. Res. Lett. 13, 055004.

Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J.M., et al., 2019. Assessment of red-edge vegetation indices for crop leaf area index estimation. Remote Sens. Environ. 222, 133–143.

Dong, T., Liu, J., Qian, B., He, L., Liu, J., Wang, R., et al., 2020. Estimating crop biomass using leaf area index derived from Landsat 8 and Sentinel-2 data. ISPRS J. Photogramm. Remote Sens. 168, 236–250.

van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., et al., 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. Environ. Sci. Technol. 50, 3762–3772.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.H., Féménias, P., Frerick, J., et al., 2012. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. Remote Sens. Environ. 120, 37–57.

Fang, W., 2005. Spatial analysis of an invasion front of Acer platanoides: dynamic inferences from static data. Ecography 28, 283–294.

Gilks, W.R., Richardson, S., Spiegelhalter, D., 1995. Markov Chain Monte Carlo in Practice. CRC press.

Gitelson, A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. spectral features and relation to chlorophyll estimation. J. Plant Physiol. 143, 286–292.

Goward, S.N., 1985. Shortwave infrared detection of vegetation. Adv. Space Res. 5, 41–50.

Guo, P.-T., Li, M.-F., Luo, W., Tang, Q.-F., Liu, Z.-W., Lin, Z.-M., 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: an application of random forest plus residuals kriging approach. Geoderma 237-238, 49–59.

Haboudane, D., Tremblay, N., Miller, J.R., Vigneault, P., 2008. Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. IEEE Trans. Geosci. Remote Sens. 46, 423–437.

Huang, J., Malone, B.P., Minasny, B., McBratney, A.B., Triantafilis, J., 2017. Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. Sci. Total Environ. 609, 621–632.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Jacques, D.C., Kergoat, L., Hiernaux, P., Mougin, E., Defourny, P., 2014. Monitoring dry vegetation masses in semi-arid areas with MODIS SWIR bands. Remote Sens. Environ. 153, 40–49.

Kang, L., Han, X., Zhang, Z., Sun, O.J., 2007. Grassland ecosystems in China: review of current knowledge and research advancement. Philos. Trans. R. Soc. B 362, 997–1008.

Kergoat, L., Hiernaux, P., Dardel, C., Pierre, C., Guichard, F., Kalilou, A., 2015. Dry-season vegetation mass and cover fraction from SWIR1.6 and SWIR2.1 band ratio: ground-radiometer and MODIS data in the Sahel. Int. J. Appl. Earth Obs. Geoinf. 39, 56–64.

Knyazikhin, Y., Martonchik, J.V., Myneni, R.B., Diner, D.J., Running, S.W., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. J. Geophys. Res. Atmos. 103, 32257–32275.

Korhonen, L., Hadi, Packalen P., Rautiainen, M., 2017. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. Remote Sens. Environ. 195, 259–274.

Kravitz, J., Matthews, M., Bernard, S., Griffith, D., 2020. Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: successes and challenges. Remote Sens. Environ. 237, 111562.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36, 1–13.

Lee, K.-S., Cohen, W.B., Kennedy, R.E., Maiersperger, T.K., Gower, S.T., 2004. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. Remote Sens. Environ. 91, 508–520.

Li, Z., Wang, J., Tang, H., Huang, C., Yang, F., Chen, B., et al., 2016. Predicting grassland leaf area index in the meadow steppes of northern China: a comparative study of regression approaches and hybrid geostatistical methods. Remote Sens. 8, 632.

Li, Z.-w., Xin, X.-p., Tang, H., Yang, F., Chen, B.-r., Zhang, B.-h., 2017. Estimating grassland LAI using the Random Forests approach and Landsat imagery in the meadow steppe of Hulunber, China. J. Integr. Agric. 16, 286–297.

Li, L., Chen, J., Han, X., Zhang, W., Shao, C., 2020. Overview of Chinese Grassland Ecosystems. Grassland Ecosystems of China: A Synthesis and Resume. Springer Singapore, Singapore, pp. 23–47.
Liang, S., Wang, J., 2020. Chapter 12 - fractional vegetation cover. In: Liang, S., Wang, J. (Eds.), Advanced Remote Sensing (Second Edition). Academic Press, pp. 477–510.
Liaw, A., Wiener, M.C., 2007. Classification and Regression by randomForest.
Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. J. Stat. Softw. 63, 1–25.
Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat Methodol. 73, 423–498.
Moraga, P., Dean, C., Inoue, J., Morawiecki, P., Noureen, S.R., Wang, F., 2021. Bayesian spatial modelling of geostatistical data using INLA and SPDE methods: a case study predicting malaria risk in Mozambique. Spat. Spatio-temp. Epidemiol. 39, 100440.
Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. Int. J. Remote Sens. 25, 3999–4014.
Mutanga, O., Skidmore, A.K., 2007. Red edge shift and biochemical content in grass canopies. ISPRS J. Photogramm. Remote Sens. 62, 34–42.
Poggio, L., Gimona, A., Spezia, L., Brewer, M.J., 2016. Bayesian spatial modelling of soil properties and their uncertainty: the example of soil organic matter in Scotland using R-INLA. Geoderma 277, 69–82.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.
Reyes-Muñoz, P., Pipia, L., Salinero-Delgado, M., Belda, S., Berger, K., Estévez, J., et al., 2022. Quantifying fundamental vegetation traits over Europe using the Sentinel-3 OLCI catalogue in Google Earth Engine. Remote Sens. 14, 1347.
Richardsons, A.J., Wiegand, A., 1977. Distinguishing vegetation from soil background information. Photogramm. Eng. Remote. Sens. 43, 1541–1552.
Rocha, A.D., Groen, T.A., Skidmore, A.K., Darvishzadeh, R., Willemen, L., 2018. Machine learning using hyperspectral data inaccurately predicts plant traits under spatial dependency. Remote Sens. 10, 1263.
Rocha, A.D., Groen, T.A., Skidmore, A.K., 2019. Spatially-explicit modelling with support of hyperspectral data can improve prediction of plant traits. Remote Sens. Environ. 231, 111200.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat Methodol. 71, 319–392.
Schaaf, C., Wang, Z., 2021. MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V061. In: DAAC NELP, editor.
Shen, M., Duan, H., Cao, Z., Xue, K., Qi, T., Ma, J., et al., 2020. Sentinel-3 OLCI observations of water clarity in large lakes in eastern China: implications for SDG 6.3.2 evaluation. Remote Sens. Environ. 247, 111950.
Strömberg, C.A.E., Staver, A.C., 2022. The history and challenge of grassy biomes. Science 377, 592–593.
Su, H., Shen, W., Wang, J., Ali, A., Li, M., 2020. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. For. Ecosyst. 7, 64.
Sun, Y., Qin, Q., Ren, H., Zhang, T., Chen, S., 2020. Red-edge band vegetation indices for leaf area index estimation from Sentinel-2/MSI imagery. IEEE Trans. Geosci. Remote Sens. 58, 826–840.
Tang, L., Tian .L., L, Steward, B., 2003. Classification of broadleaf and grass weeds using gabor wavelets and an artificial neural network. Trans. ASAE 46, 1247.
Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Van Der Meer, B., 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote Sens. Environ. 91, 354–376.
Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit Region. Econ. Geogr. 46, 234–240.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.
Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G. P.W., et al., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. ISPRS J. Photogramm. Remote Sens. 108, 273–290.
Wang, J., Brown, D.G., Bai, Y., 2014. Investigating the spectral and ecological characteristics of grassland communities across an ecological gradient of the Inner Mongolian grasslands with in situ hyperspectral data. Int. J. Remote Sens. 35, 7179–7198.
Wang, C., Chen, J., Wu, J., Tang, Y., Shi, P., Black, T.A., et al., 2017. A snow-free vegetation index for improved monitoring of vegetation spring green-up date in deciduous ecosystems. Remote Sens. Environ. 196, 1–12.
Wu, J., Kurosaki, Y., Gantsetseg, B., Ishizuka, M., Sekiyama, T.T., Buyantogtokh, B., et al., 2021. Estimation of dry vegetation cover and mass from MODIS data: verification by roughness length and sand saltation threshold. Int. J. Appl. Earth Obs. Geoinf. 102, 102417.
Xiao, X., Boles, S., Frolking, S., Salas, W., Moore, B., Li, C., et al., 2002. Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. Int. J. Remote Sens. 23, 3009–3022.
Xu, D., Wang, C., Chen, J., Shen, M., Shen, B., Yan, R., et al., 2021. The superiority of the normalized difference phenology index (NDPI) for estimating grassland aboveground fresh biomass. Remote Sens. Environ. 264, 112578.
Xue, K., Ma, R., Duan, H., Shen, M., Boss, E., Cao, Z., 2019. Inversion of inherent optical properties in optically complex waters using sentinel-3A/OLCI images: a case study using China's three largest freshwater lakes. Remote Sens. Environ. 225, 328–346.