Computers and Electronics in Agriculture 225 (2024) 109313

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Accurately estimate soybean growth stages from UAV imagery by accounting for spatial heterogeneity and climate factors across multiple environments

Yingpu Che [a,b,c,1], Yongzhe Gu [a,1], Dong Bai [a,1], Delin Li [a,1], Jindong Li [a,b], Chaosen Zhao [d], Qiang Wang [e], Hongmei Qiu [f], Wen Huang [g], Chunyan Yang [h], Qingsong Zhao [h], Like Liu [i], Xing Wang [j], Guangnan Xing [k], Guoyu Hu [l], Zhihui Shan [m], Ruizhen Wang [d], Ying-hui Li [a,b,*], Xiuliang Jin [a,b,*], Li-juan Qiu [a,b,*]

[a] *State Key Laboratory of Crop Gene Resources and Breeding/The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI)/ Key Laboratory of Grain Crop Genetic Resources Evaluation and Utilization (MARA), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[b] *National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, China*
[c] *HainanYazhou Bay Seed Lab, Sanya 572025, China*
[d] *Crops Research Institute of Jiangxi Academy of Agricultural Sciences/Nanchang Branch of National Centre of Oil Crops Improvement/Jiangxi Province Key Laboratory of Oil Crops Biology, Nanchang 330200, China*
[e] *Institute of Crop Resources, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China*
[f] *Soybean Research Institute, Jilin Academy of Agricultural Sciences/National Engineering Research Centre for Soybean, Changchun 130033, China*
[g] *Tong Hua Academy of Agricultural Sciences, Meihekou 135007, China*
[h] *Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang Branch Centre of National Centre for Soybean Improvement, The Key Laboratory of Crop Genetics and Breeding, Shijiazhuang 050031, China*
[i] *School of Life Sciences, Liaocheng University, Liaocheng 252059, China*
[j] *Xuzhou Institute of Agricultural Sciences in XuHuai Region of Jiangsu, Xuzhou 221131, China*
[k] *Soybean Research Institute/National Centre for Soybean Improvement/Key Laboratory for Biology and Genetic Improvement of Soybean (General), Ministry of Agriculture/National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China*
[l] *Crop Institute of Anhui Academy of Agricultural Sciences/Anhui Key Lab of Crops Quality Improving, Hefei 230031, China*
[m] *Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430062, China*

ARTICLE INFO

Keywords:
Soybean growth stages
Multi-environment trials
Photothermal accumulation area
Spatial heterogeneity
Unmanned aircraft vehicle

ABSTRACT

Multi-environment trials (METs) are widely used in soybean breeding to evaluate soybean cultivars' adaptability and performance in specific geographic regions. However, METs' reliability is affected by spatial and temporal variation in testing environments, requiring further knowledge to correct such changes. To improve METs' accuracy, the growth of 1303 soybean cultivars was accurately estimated by accounting for climatic effects and spatial heterogeneity using a linear mixed-effect model and a field spatial-correction model, respectively. The METs across 10 sites varied in climate and planting dates, spanning N16°41′52″ in latitude. A soybean growth and development monitoring algorithm was proposed based on the photothermal accumulation area (AUC_{pt}) rather than using calendar dates to reduce the impact of planting dates variability and climate factors. The AUC_{pt} correlates strongly with latitude of the above trial sites ($r > 0.77$). The proposed merit-based integrated filter decreases the influence of noise on photosynthetic vegetation (f_{PV}) and non-photosynthetic vegetation (f_{NPV}) more effectively than S-G filter and locally estimated scatterplot smoothing. The field spatial-correction model helped account for spatial heterogeneity with a better estimation accuracy ($R^2 \geq 0.62$, RMSE$\leq$0.17). Broad-sense heritability (H^2) with the field spatial-correction model outperformed the models without the model by an average of 52 % across the entire aerial surveys. Model transferability was evaluated across Sanya and Nanchang. Rescaled shape models in Sanya ($R^2 = 0.97$) were consistent with the growth curve in Nanchang ($R^2 = 0.89$). Finally, the methodology's precision estimations of crop genotypes' growth dynamics under differing

---

[*] Corresponding authors.
E-mail addresses: liyinghui@caas.cn (Y.-h. Li), jinxiuliang@caas.cn (X. Jin), qiulijuan@caas.cn (L.-j. Qiu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.compag.2024.109313
Received 28 September 2023; Received in revised form 29 June 2024; Accepted 5 August 2024
Available online 12 August 2024

environments displayed potential applications in precision agriculture and selecting high-yielding and stable soybean germplasm resources in METs.

1. Introduction

Soybean (*Glycine* max (L.) Merr.) is one of the most economically important leguminous seed crops (Shen et al., 2018; Liu et al., 2020; Lu et al., 2020). Multi-environment trials (METs) involve collecting germplasm resources that naturally occur in various geographical regions and cultivating them in one or more locations to compare their growth, morphology, and physiological characteristics (Rotili et al., 2020; Ortiz et al., 2023). These trials can aid in selecting the best-suited germplasm for specific regions in soybean production, as well as provide data and raw materials for genetic improvement by selecting high-yielding and stable germplasm resources (Jarquin et al., 2020). Nonetheless, the dependability of METs data could be influenced by fluctuations in testing environments over space and time, and there is presently inadequate comprehension regarding how to account for such variations. Temperature and photoperiod are the two dominant climatic factors that influence the phenotype traits of soybean growth dynamics (Zeng et al., 2016). Soybean growth and development was directly expressed on relevant thermal and light indices scale, rather than using calendar dates, reduce the impact of planting dates variability and climate effect, and facilitates effective evaluation of soybean growth across multiple environment test sites.

Precise, relevant, and cost-effective phenotypic data is a precondition for breeders to increase selection intensity and shorten the breeding cycle (Moreira et al., 2019). Recent advances in sensors and image-progressing algorithms have provided new opportunities to bridge the gap between genotypes and phenotypes (Zeng et al., 2020; Li et al., 2022a). Unmanned aircraft vehicle (UAV), a high-throughput phenotyping platform, can be used to quantitatively or qualitatively measure crop phenotyping traits, including plant height (Roth et al., 2022), yield traits (Cai et al., 2019; Wan et al., 2020), senescence rates (Hassan et al., 2021), crop aboveground biomass (Che et al., 2022), heading date (Lyu et al., 2023) and maturity date (Zhou et al., 2019) with a non-intrusive method (Koh et al., 2022; Li et al., 2022c). However, when monitoring crop growth dynamics over a prolonged period, the estimation accuracy of UAV-derived phenotypic traits may be negatively impacted by observation errors and noise caused primarily by atmospheric constituents, bi-directional reflectance distribution functions, and mixed-pixel effects. As a result, such factors may cause unstable fluctuations and anomalies in the time series of phenotypic traits, resulting in adversely affecting the long-term crop detection results.

Spatial heterogeneity is prevalent in a large number of genotypes breeding trials, thereby decreasing the repeatability and precision of the evaluated phenotypic traits (Yu et al., 2016; Moreira et al., 2019; Bernardeli et al., 2021). Spatial heterogeneity usually comes from agriculture management and variability in soil and field topography. Characterization of field random variability at the plot level and removal of systematic errors caused by sowing are critical when estimating phenotyping traits and genotypic effects (Robbins et al., 2012; Lee et al., 2013; Elias et al., 2018). Recently, Lado et al. (2013) and Rodríguez-Álvarez et al. (2018) introduced a field spatial-correction model that improves the adjustment in labor-intensive ground-measured amino nitrogen content and grain yield and increments the accuracy of measurements. The transposition of this field spatial-correction model to UAV observations would be very attractive, due to UAV' advantages of high throughput, flexibility, non-destructiveness, and non-invasiveness.

The main aim of breeding experiments is to estimate the breeding value that is stably inherited by the next generation for targeted genotypes from accurate phenotypic data. In standard genetic models, the observed phenotypic data p is the sum of an unobservable genotypic value g and a model residual ε; that is, $p = g + \varepsilon$. The best linear unbiased predictors (BLUP) (Henderson 1973) of breeding values are well-established methodologies in crop breeding (Lado et al., 2013), with the advantage of shrinking estimators towards the average performance of pure lines, reducing the variance, and increasing the predictive accuracy (Montesinos-López et al., 2016; Kaler et al., 2022). Broad-sense heritability (H^2) refers to the percentage of phenotypic total variance that is attributable to genetic variation due to substantial variability among genotypes. H^2 is associated with the coefficient of determination (R^2) obtained from standard genetic models and is also related to the squared correlation between predicted phenotypic values and genotypic values. In plant breeding, H^2 can be incorporated into the genetic gain (ΔG) equation to predict response to selection or used as a descriptive measure to determine the validity and accuracy of cultivar trial outcomes.

Based on the analysis above, our objectives in this study are (1) to analyze the relationship between accumulated photothermal areas (AUC_{pt}) of different soybean cultivars and latitude from emergence date (ED) to maturity date (MD); (2) to compare the ability to de-noise and rebuild cultivar shape models to obtain time-series phenotypic data with S-G filtering, LOESS, and the proposed merit-based integrated filter; (3) to investigate the adaptation of spatial heterogeneity correction using the field spatial correction model and UAV RGB images; and (4) to use BLUP to develop models of photosynthetic vegetation and non-photosynthetic vegetation and verify the environmental portability of soybean cultivar growth curves. Our goal in this study is to present a strategy for obtaining more accurate time-series growth curves of soybean cultivars by minimizing the effects of climate factors and field spatial variation.

2. Materials and methods

2.1. Study layout

Field experiments were carried out in Sanya city (N18°9′34″–18°37′30″, E108°56′30″–109°48′28″) in Hainan province and Nanchang city (N28°9′38″–29°7′30″, E115°26′6″–116°33′37″) in Jiangxi province, China. A panel of 350 soybean cultivars was selected from 2214 soybean accessions (Li et al., 2022a; Li et al., 2022b) from around the world presenting a broad genetic diversity (Fig. 1c) and planted at Sanya on December 3, 2020. The site experiment used a randomized block design with three plot replications. Each soybean accession was planted in a microplot of 1 m × 0.65 m with two rows and 10 cm spacing between the seedlings in each row. The 1303 soybean cultivars were sown and grown in Nanchang on July 15, 2020, of which 350 cultivars are the same as those planted in Sanya. Each soybean accession was planted in a 1.8 m × 0.8 m microplot with two rows and 10 cm spacing between seedlings in each row (Fig. 1b). The UAV-derived traits were obtained at 18 time points from December 23, 2020 to February 27, 2021 in Sanya and at 17 time points from August 1 to September 22, 2020 in Nanchang.

The emergence date (ED), flowering date (FD) and maturity date (MD) were manually measured in Nanchang, Harbin (N44°3′36″–46°40′21″, E125°40′47″–130°14′1″), Gongzhuling (N43°11′31″–44°16′34″, E124°02′08″–125°18′18″), Tonghua (N40°51′51″–43°1′4″, E125°15′24″–126°44′41″), Shijiazhuang (N37°26′12″–38°45′49″, E113°31′1″–115°28′39″), Liaocheng (N35°46′48″–37°1′54″, E115°16′36″–116°32′39″), Xuzhou (N33°43′5″–34°58′33″, E116°21′19″–118°40′21″), Nanjing (N31°13′50″–32°36′44″, E118°21′14″–119°14′12″), Hefei (N30°57′10″–32°32′28″, E116°40′58″–117°57′38″), and Wuhan (N29°58′19″–31°21′46″, E113°41′48″–115°4′37″) in 2018–2020. The same soybean cultivars as planted in

Nanchang were sown at nine other experimental sites, with the same planting strategy as in Nanchang. The ED corresponded to the emergence of seedling cotyledons at least 50 % of soybean in a plot. The FD corresponded to the flower of at least 50 % of soybean in a plot. The MD was when approximately 95 % of soybean pods achieved mature pod color.

2.2. *UAV image acquisition and preprocessing*

Aerial surveys were conducted under clear sky conditions from 10:00 to 14:00 local time using a Phantom 4 Multispectral Platform integrating an RGB sensor in Nanchang and a DJI Matrice 600 Hexacopter Platform integrating a Sony ILCE-7M2 RGB sensor in Sanya (Fig. 1a). The mission was planned at a flight height of 12 and 17 m above the ground in Nanchang and Sanya, respectively, with Ground Sampling Distances (GSD) of 0.58 cm and 0.47 cm. The trajectory overlap was 75 %, and the side overlaps were 60 % and 75 %, respectively, as described by Li et al. (2022a).

The image preprocessing was implemented by Agisoft metashape software Professional Edition (1.8.2, Formerly PhotoScan, Agisoft LLC, Russia) and Esri ArcGIS (10.7, ESRI, United States). The photogrammetric processes contained four steps: (1) the key points of UAV images were detected by applying the Scale Invariant Feature Transform (SIFT; Lowe, 2004) algorithm and the geometrically keypoint correspondences were filtered using the Random Sample Consensus (RANSAC; Fischler and Bolles, 1987) method; (2) point clouds were generated and georeferenced using the Structure From Motion- Multi View Stereo algorithm (SFM-MVS; Smith et al., 2016); (3) orthomosaic map generation required the creation of a surface mesh from the point cloud; (4) all orthomosaic maps were georeferenced using the extracted ground control points. Microplots were segmented from orthomosaic maps according to their ground control point position in the experimental design layout (Fig. 1b).

2.3. *Development of phenological models for soybean cultivars*

Fig. 2 describes the development scheme for soybean cultivar phenology models for UAV RGB images. This study uses as input variables to construct phenology models the ED and MD of soybean cultivars from ten environments in 2018–2020, weather information, and UAV datasets from Sanya and Nanchang. Five specific steps were involved: (1) calculate AUC_{pt} using recorded temperature, sunrise, and sunset during ED-MD; (2) vegetation can be classified as photosynthetic and non-photosynthetic using a vegetation classification model; (3) de-noise and rebuild cultivar shape models for time-series phenotypic data with S-G filtering, LOESS, and the proposed merit-based integrated filter; (4) investigate the adaptation of spatial heterogeneity correction using the field spatial correction model and UAV RGB images; (5) determine models for the BLUP of photosynthetic vegetation and non-photosynthetic vegetation and verify the environmental portability of soybean cultivar growth curves.

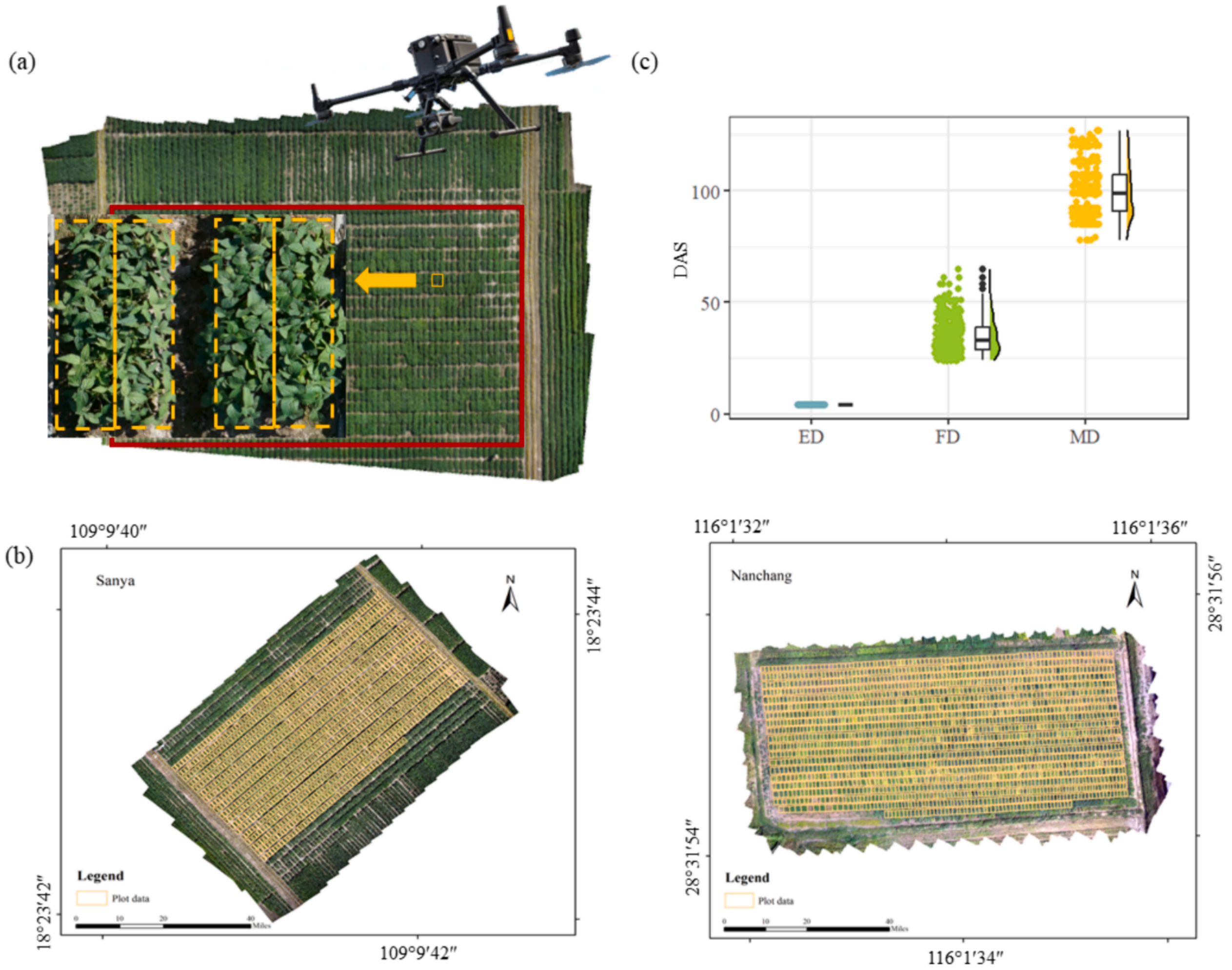

Fig. 1. Flowchart of the development and validation of phenological models for soybean cultivar time-series with AUC_{pt}. (a) UAV images acquisition. (b) Study layout in Nanchang and Sanya. (c) Phenotypic variation of flowering date and maturity date of soybean cultivars.

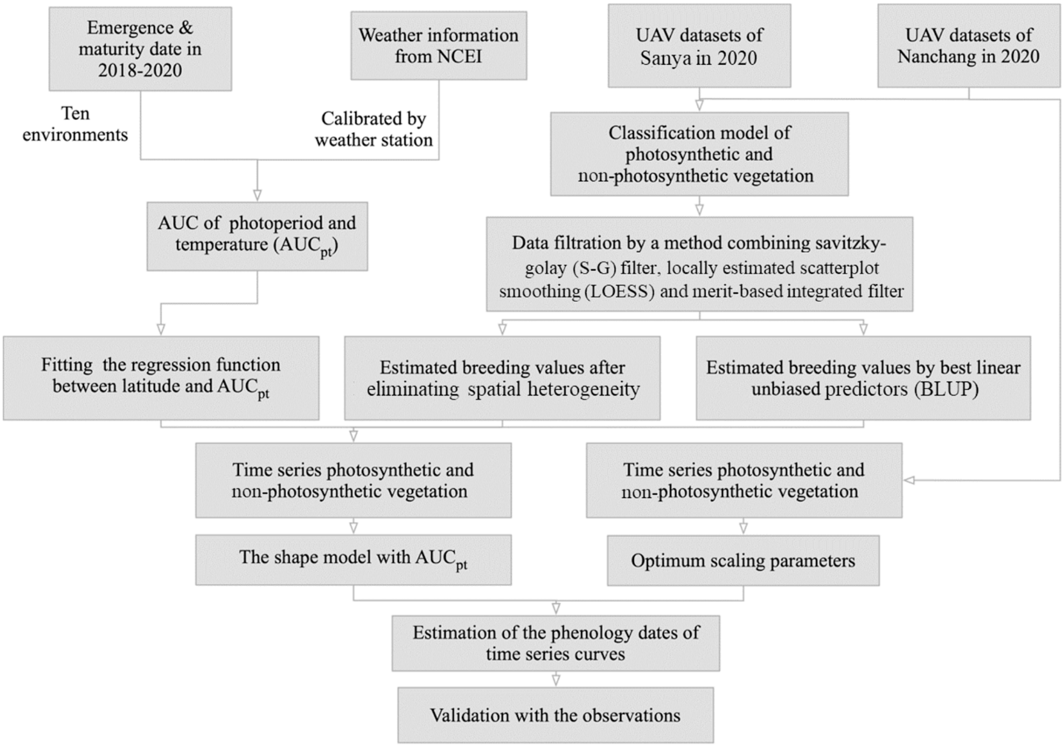

Fig. 2. Flowchart of the development and validation of phenological models for soybean cultivar time-series with AUC$_{pt}$.

2.3.1. *Collection of meteorological data*

As irrigation was implemented in both sites to alleviate water stress, the influence of precipitation was not taken into account in this study. The environmental factors considered included temperature and photoperiod. Air temperature was recorded every three hours by the ground weather station nearest to the 11 sites listed above. Data were downloaded from the NOAA's National Centers for Environmental Information (NCEI, https://www.ncei.noaa.gov/) and calibrated using the recorded values of the above four sites collected by field weather stations of the Chinese Academy of Agricultural Sciences (CAAS; $R^2 = 0.99$, Fig. S1).

Calculation of photothermal accumulation area (AUC$_{pt}$) consists of four steps: (1) fitting daily temperature curves to a double logistic function; (2) calculating sunrise and sunset of the above 11 sites using the NOAA Sunrise/Sunset and solar position calculators with equations from astronomical algorithms (Meeus 1998); (3) calculating daily photothermal values based on the area under daily temperature curves between sunrise and sunset; (4) obtaining AUC$_{pt}$ by summing the daily photothermal values. The relationship between latitude and AUC$_{pt}$ from ED to MD was analyzed by using monotonically increasing functions including linear ($y = ax + b$) and power [$y = a(x + b)^3$] functions.

2.3.2. *Photosynthetic and non-photosynthetic vegetation classification model*

Vegetation can be classified as photosynthetic (green leaves) and non-photosynthetic (i.e., senescence leaves, litter, and wood) by using a vegetation classification model constructed by the decision-tree-based segmentation model with two steps. The bounding boxes of photosynthetic vegetation (PV; binary value = 1), non-photosynthetic vegetation (NPV; binary value = 2), and bare soil (BS; binary value = 0) from 3500 selected images from different growth stage were first drawn by using the graphical image annotation tool LabelImg (https://github.com/heartexlabs/labelImg). The color features of annotated images including R, G, B, H, S, V, L, a, b from the color spaces of RGB, HSV, and CIELAB were divided randomly into the training and testing dataset according to the ratio 7:3. Calculating feature importance in the decision-tree-based segmentation model involves two steps: (1) calculate the importance for each node; and (2) calculate the importance for each feature by applying node importance splitting on that feature. The percentage of PV and NPV pixels in each microplot image was defined as the fraction f_{PV} and f_{NPV} of photosynthetic vegetation and non-photosynthetic vegetation, respectively.

2.3.3. *Merit-based integrated filter*

Many abnormally high values and noise artifacts in the f_{PV} and f_{NPV} time series caused by atmospheric variability, the bi-directional reflectance distribution function, and the mixed-pixel effects were first smoothed by using a S-G filter (Savitzky and Golay 1964), LOESS (Cleveland 1979), and the proposed merit-based integrated filter. Time series data (t, y_t) were smoothed based on S-G filtering by applying an unweighted linear least square fitting to a set of ($2m + 1$) adjacent data points with a K^{th} polynomial, which can increase the precision of the data without distorting the signal tendency. This work found that third-degree polynomials in the S-G filter produce better denoising than other low-degree polynomials. The LOESS employed locally weighted linear quadratic regressions to smooth data after computing the weight for each point within the span. The least squares criterion was modified by multiplication of a nonnegative weight factor w_i by the squared error at t_i within the span. Data points outside the span have zero weight. In this study, the proposed merit-based integrated filter consists of three steps (Fig. 3): (1) calculating the absolute difference between the observed time-series data and the value estimated by LOESS; (2) removing the outliers by comparing absolute differences with the triple median absolute deviation; (3) combining the averaged estimation from S-G and LOESS. By calculating the Sum of Squared Errors (SSE) between the fitted curve and the time series, it was found that the SSE was minimized when (2 m + 1), Kth, and span were set to 5, 3, and 0.3, respectively.

2.3.4. *Spatial heterogeneity correction model*

The spatial heterogeneity of f_{PV} and f_{NPV} caused by the different micro-environments between plots at each observation was corrected by using two-dimensional P-splines (Rodríguez-Álvarez et al., 2018). A

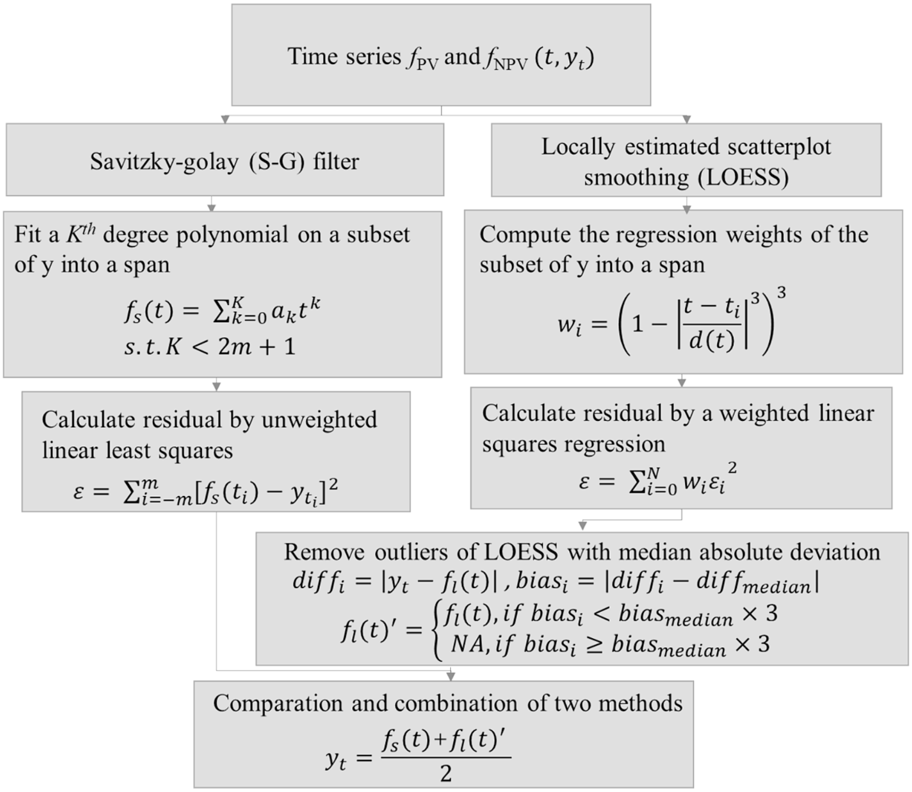

Fig. 3. Flowchart of proposed merit-based integrated filter. Note: f_{PV} and f_{NPV} are defined as the fractions of photosynthetic vegetation and non-photosynthetic vegetation. (t,y_t), $[t, f_s(t)]$, $[t, f_l(t)]$, respectively, and represent measured time series, filtered time series by S-G, and smoothed time series by LOESS. $f_l(t)$' is the time series after removing the smooth outliers of LOESS, where k, a, t_i, ε, w_i, $d(t)$ respectively represents a K^{th}-degree polynomial, a coefficient, the nearest neighbour of t, which is defined within the span, the residual, a nonnegative weight factor, and the maximum time between t and t_i.

smooth bivariate surface $f(r_i,c_i)$ was firstly modelled using a simplified model (Eqn.1). The P-spline approach is utilized to model the smooth bivariate function as a tensor product of B-spline bases (Dierckx, 1995). To avoid overfitting, difference penalties are added along the rows and columns, and the 2D P-spline model (Eqn.4) is estimated through the minimization of a penalized least squares objective function (Eilers and Marx, 2003).

$$y_i = f(r_i, c_i) + \varepsilon_i, \text{with}\varepsilon\ N(0, \sigma^2) \tag{1}$$

$$y = B\alpha + \varepsilon, \text{with}\varepsilon\ N(0, \sigma^2 I_n) \tag{2}$$

$$S^* = \|y - B\alpha\|^2 + \alpha^t P\alpha \tag{3}$$

where r_i and c_i denote the row and columns ($i = 1,2,\ldots,n$). B represent B-spline bases and α is a vector of unknown regression coefficients. P is the penalty matrix and ε is the error.

A more complex model was developed to account for genetic variation, row × column effects, sowing and duplication effects in the field experiment consisting of three duplications, 18 rows × 70 columns, and 350 genotypes. Our model included random factors for rows [$c_r \sim N(0,\sigma^2 I_{18})$], columns [$c_c \sim N(0,\sigma^2 I_{70})$], duplications [$c_d \sim N(0,\sigma^2 I_3)$], and soybean cultivars [$c_s \sim N(0,\sigma^2 I_{350})$]. For each trait ($f_{\mathrm{PV}}$ and f_{NPV}), the following model was applied to estimate fixed and random effect coefficients and variance components in mixed-effect models, by maximizing the REML log-likelihood function. The other linear mixed-effect model without P splines was constructed by using only duplications [$c_d \sim N(0,\sigma^2 I_3)$] and soybean cultivars [$c_s \sim N(0,\sigma^2 I_{350})$] as random factors based on the maximum likelihood or restricted maximum likelihood estimate methods (Bates et al., 2015).

$$y_p = f(r,c) + Z_s c_s + Z_r c_r + Z_c c_c + Z_d c_d + \varepsilon, \text{with}\varepsilon\ N\left(0, \sigma^2\right) \tag{4}$$

$$y_m = Z_s c_s + Z_d c_d + \varepsilon, \text{with}\varepsilon\ N\left(0, \sigma^2\right) \tag{5}$$

where Z_s, Z_r, Z_c, and Z_d represent column-partitioned matrices associated with soybean cultivars, rows, columns, and duplications, respectively. The smooth bivariate function f(r,c) was estimated by the tensor product of B-spline bases using the row and column as covariates (for details see Rodríguez-Álvarez et al., 2018). The quantity ε is the random error vector.

The transmissible or "additive" portion was due to the effects of individual alleles passed on by a parent to its offspring. The breeding values as the expected performance of offspring are predicted by using the BLUP (Henderson 1973), which is a well-established methodology in crop breeding, with the advantage that estimators shrink towards the average performance of pure lines, thereby reducing its variance and increasing its predictive accuracy (Montesinos-López et al., 2016; Kaler et al., 2022).

2.3.5. Shape model of soybean cultivars and space–time portability

The f_{PV} and f_{NPV} time series in Sanya combined with AUC_{pt} were fit by using the double logistic function (Eqn.6) and the S-curve (Eqn.7), respectively:

$$fd(t,p) = c_{min} + (c_{max} - c_{min}) \bullet \left[\left(\frac{1}{1 + p_2 \bullet e^{(t_g - t) \bullet k_g}} \right)^{p_1} - \left(\frac{1}{1 + p_4 \bullet e^{(t_s - t) \bullet k_s}} \right)^{p_3} \right] \tag{6}$$

$$fs(t,p) = c_{min} + \frac{c_{max} - c_{min}}{\left(1 + p_2 \bullet e^{(t_g - t) \bullet k_g}\right)^{p_1}} \tag{7}$$

The model parameters (Fig. 4a) include the minimum (c_{min}) and maximum (c_{max}), the inflection points k_g and k_s, the time reach of the

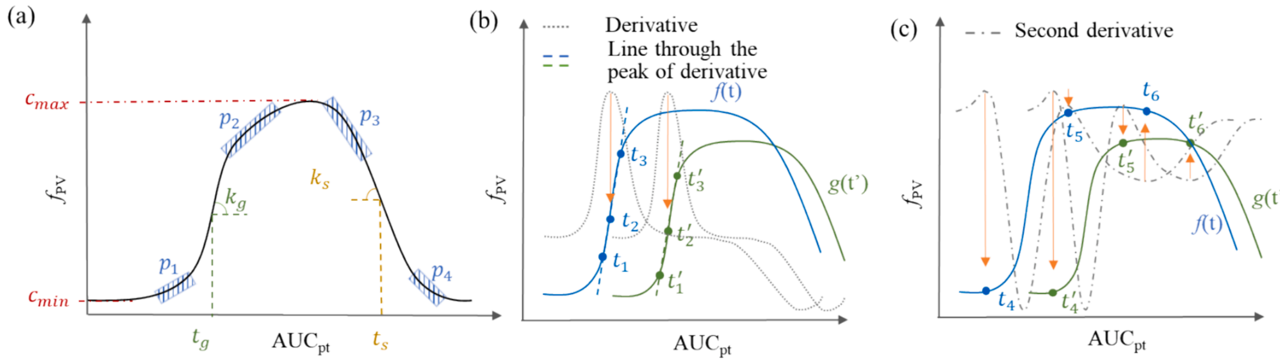

Fig. 4. Visualization of parameters for the double logistic function and scaled model. (a) Including minimum (c_{min}) and maximum (c_{max}), the inflection points (k_g and k_s), the time reach the inflection points (t_g and t_s), and the rates near the lower (p_1 and p_4) and upper (p_2 and p_3) asymptotes. Scaled model of *x* axis constructed by using six points derived from (b) first derivative and (c) second derivative.

inflection points (t_g and t_s), and the rates near the lower (p_1 and p_4) and upper (p_2 and p_3) asymptotes, which were used to fit the growth and development of soybean cultivars.

The model of the shape of the time-series growth curve *f*(t) for Sanya represents the growth curve for the target soybean cultivars, where *t* is AUC_{pt} and *f*(t) is a function of the time-series of f_{PV} and f_{NPV}. The growth curve *g*(*t*') in Nanchang is transformed from *f*(*t*) by stretching, compressing, and translating in the *t* and *y* directions by using xscale, yscale, and tshift, respectively, as follows:

$$g(t') = yscale \times f(xscale \times (t + tshift)) \tag{8}$$

Scaled models of the *x* axis were constructed by using six points (Fig. 4b–c), including the inflection point (t_2), intersections (t_1 and t_3) of the line through the inflection point, the time-series curves, and the peak (t_4 and t_5) and valley (t_6) positions of the second derivative; that is, $t' =$ xscale × (t + tshift). These six points were acquired by the first derivative, second derivative, and intersection points, which gave unique values that better describe the shape of the time-series curves. The yscale was calculated by using the lower and upper asymptotes; that is $g(t') =$ yscale × $f(t')$.

2.4. Statistical analysis and heritability estimates

The coefficient of determination R^2 and root mean square error (RMSE) were used to evaluate the degree of coincidence between the observed and estimated phenotyping BLUPs. The Pearson correlation coefficient, mean, maximum, minimum, standard deviation, and coefficient of variation were used to evaluate correlations between latitude and AUC_{pt}. The accuracy, precision, recall, F1-score, receiver operating characteristic (ROC) and area under ROC curve (AUC) served to evaluate the degree to which the measured value conformed to the estimated value and the generalizability of the vegetation classification model. The normal distribution statistics of residuals between observed and estimated values were analyzed to evaluate the repeatability and stability of the P-splines method.

The homogeneity of variance across the testing environments allowed for combined analyses of variance (ANOVA). The ANOVA's expected mean squares enabled estimation of variance components for both the testing germplasm and environments. Broad-sense heritability (H^2) based on plot means was determined as Eqn.(9).

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{GS}^2/S + \sigma_\varepsilon^2/r} \tag{9}$$

Where, σ_G^2, σ_{GS}^2 and σ_ε^2 were the genetic variance, the variance component for G by E and the residual variance, respectively. UAV data were collected for 350 soybean cultivars at two sites (S), Sanya and Nanchang, with three replications (r).

3. RESULTS

3.1. Relationships of FD and MD with temperature and latitude

In this study, the FD and MD of soybean cultivars was calculated from ten experimental sites with variability in climate and planting dates, spanning N16°41′52″ in latitude. Boxplots of FD and MD data from ten locations were arranged in parallel, clearly revealing the median, range, outliers, and distribution of each dataset (Fig. 5a–b). The FD and MD for Harbin, Gongzhuling, and Tongliao were higher than those in other sites, while FD and MD were shortest in Wuhan and Nanchang. The wide interquartile ranges of FD and MD indicated that the soybean materials present a broad genetic diversity. The FD and MD of soybean cultivars correlates highly with the temperature (r = -0.672 & −0.775, Fig. 5c). Their correlation with temperature was negative, indicating that the AUC_{pt} of soybean cultivars decreases with temperature.

The AUC_{pt} models were built from the relationships between latitude and AUC_{pt} from ED to MD using linear ($y = ax + b$) and power [$y = a(x + b)^3$] functions. The AUC_{pt} from ED to MD produces the most accurate estimates when using a power function with $R^2 = 0.77$–0.95, SD=0.044, CV=5.06 % (Fig. 5d), which is more accurate than the model constructed by the linear function (R^2 = 0.76–0.94, SD=0.045, CV=5.25 %).

3.2. Photosynthetic and non-photosynthetic vegetation classification

The classification accuracy was evaluated over the training and testing datasets. The testing dataset only slightly decreases as compared with the training dataset. The results indicate that the vegetation classification model is robust (Fig. 6a). The classification model is good for the PV and NPV of soybeans, with an accuracy of 0.984 and a ROC of 0.978. The diagonal elements of the confusion matrix list the correctly classified counts in image pixels of the training and testing datasets for NPV, PV, and BS, which are much greater than the confusion counts (Fig. 6b). The results show that the vegetation classification model is a well-functioning classifier (Fig. 6b).

The results demonstrate that the parameters *a*, *H*, and *b* from the HSV and CIELAB color models of annotated images are important color features for classifying PV and NPV of soybean (Fig. 6c). The parameters *a*, *H*, and *b* describe the color transformation process. The parameter *a* represents the color change from green to red, *b* represents the color change from blue to yellow, and *H* is the hue. These are useful not only because the scheme is more intuitive than raw RGB values but also because the conversions to and from RGB are extremely fast to compute.

3.3. Comparison of three curve-filtering methods

To denoise and rebuild the time series f_{PV} and f_{NPV}, we apply and

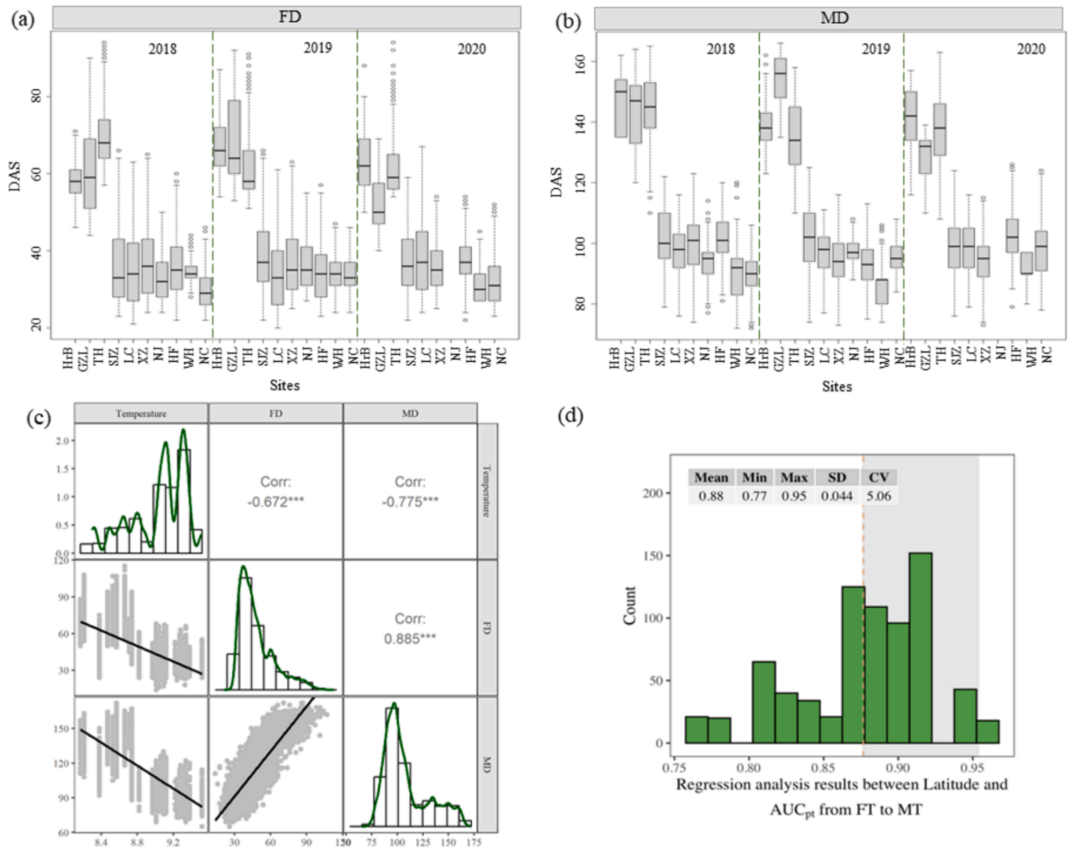

Fig. 5. Discrete distribution of flowering date (FD) and maturity date (MD) in soybean cultivars and their correlation with temperature and latitude. (a) and (b) The distribution of FD and MD from 2018 to 2020 at ten sites: HrB (Harbin), GZL (Gongzhuling), TH (Tonghua), SJZ (Shijiazhuang), LC (Liaocheng), XZ (Xuzhou), NJ (Nanjing), HF (Hefei), WH (Wuhan), and NC (Nanchang). (c) Represents the correlation coefficient among temperature, FD, and MD. (d) The regression analysis results between latitude and AUC$_{pt}$ from ED to MD of soybean cultivars.

compare the S-G filter and LOESS and propose a merit-based integrated filter (Fig. 7). In the time-series f_{PV}, LOESS proves to be a better method to reduce contamination compared with the other methods (Fig. S2). However, the S-G filter works better than LOESS to preserve the width and height of peaks in the waveform, and the performance of LOESS decreases for the time-series f_{NPV} (Fig. S3). The proposed merit-based integrated filter combines the virtues of the other two methods and resists their disadvantages, so it works better to minimize the influence of noise on time-series f_{PV} and f_{NPV}.

3.4. Correcting for spatial heterogeneity in plant breeding experiments with P splines

The two-dimensional smooth surfaces of each time-point f_{PV} and f_{NPV} datasets were modelled using P splines to estimate spatial trends across plots. Here, correcting for spatial heterogeneity in time-series f_{PV} is considered as an example in Fig. S4. The spatial trend of the raw data is recovered by using P splines (Fig. S4a–b). The raw data are consistent with the estimated results (Fig. S4a) with residuals showing a normal distribution with μ and σ values close to zero (Fig. S4c). There were highly significant variations among genotypes ($p \leq 0.001$) across UAV aerial surveys. When P-spline analysis was not performed, the average H^2 of time-series f_{PV} was 0.53, indicating that environmental factors had a significant impact. After P-spline analysis was performed, the average H^2 of time-series f_{PV} increased to 0.80, reducing the influence of environmental factors and increasing the prominence of genetic effects, which helps to explore the results caused by genetic effects.

Better predictions for the genotypic performance of field breeding experiments were obtained by correcting for the spatial effect. Fig. 5 demonstrates the good relationships between the measured and estimated f_{PV} and f_{NPV}. The estimation accuracy of the f_{PV} shape model (Fig. 8a) is as follows: with P splines ($R^2 \geq 0.62$, RMSE ≤ 0.17) and without P splines ($R^2 \geq 0.58$, RMSE ≤ 0.26). The shape model of f_{NPV} (Fig. 8b) with P splines ($R^2 \geq 0.67$ and RMSE$\leq$0.11) provided better accuracy than without P-splines ($R^2 \geq 0.60$, RMSE ≤ 0.19). The estimation accuracy of the shape model considering the spatial heterogeneity was improved when compared with results without the P-spline method. No significant systematic error appears in the time-series shape models for the three duplications. Time-series curves of five soybean cultivars selected from the best and worst estimation accuracy are shown and compared in the yellow and green boxes of Fig. 8a–b, respectively. The time-series f_{PV} and f_{NPV} shape model with P splines are an effective method for correcting spatial heterogeneity between different duplications of the same germplasm across a range of important growth stages.

3.5. Migrating time-series shape model from Sanya to Nanchang

The time-series shape model established in Sanya was rescaled by using xscale, yscale, and tshift to fit the curves of each soybean cultivar in Nanchang. The fitting accuracy of the target curve in Nanchang was as follows: $R^2 = 0.89$, RMSE = 0.11, and slope = 1 (Fig. 9a). The time-series shape model fits accurately with the observed target curve in Sanya with $R^2 = 0.97$, RMSE=0.039, and slope = 0.97 (Fig. 9b). 18 cultivars (Fig. 9c) were randomly taken as examples to illustrate the consistency of the two sites. Fig. 6b shows that the rescaled shape model in Sanya is consistent with the shape model in Nanchang. Upon considering the temperature and photoperiod, the AUC$_{pt}$ of each soybean cultivar at the vegetation and reproductive stages is almost the same at both sites and in both years (Fig. 9c). The shape model built with AUC$_{pt}$ minimizes the

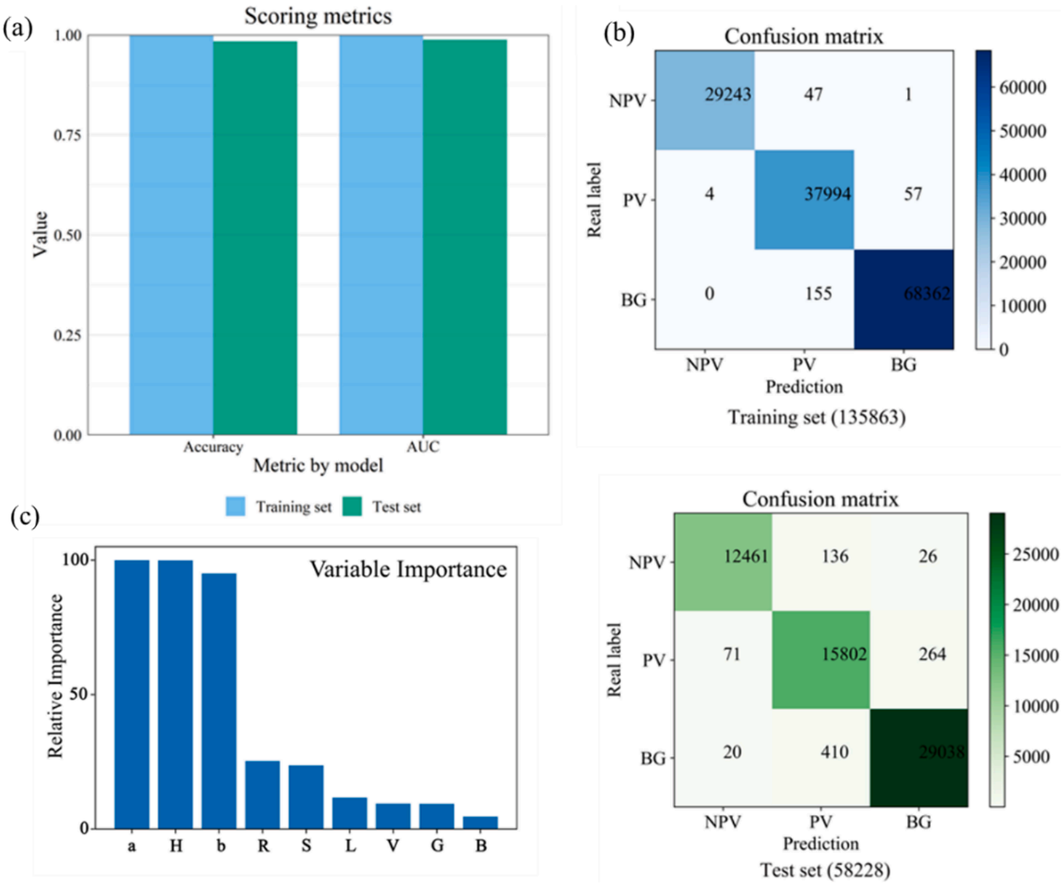

Fig. 6. Metrics of photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and bare soil (BS) classification model. (a) accuracy and AUC, (b) Confusion matrix, (c) Feature importance of f_{PV} and f_{NPV} classification model.

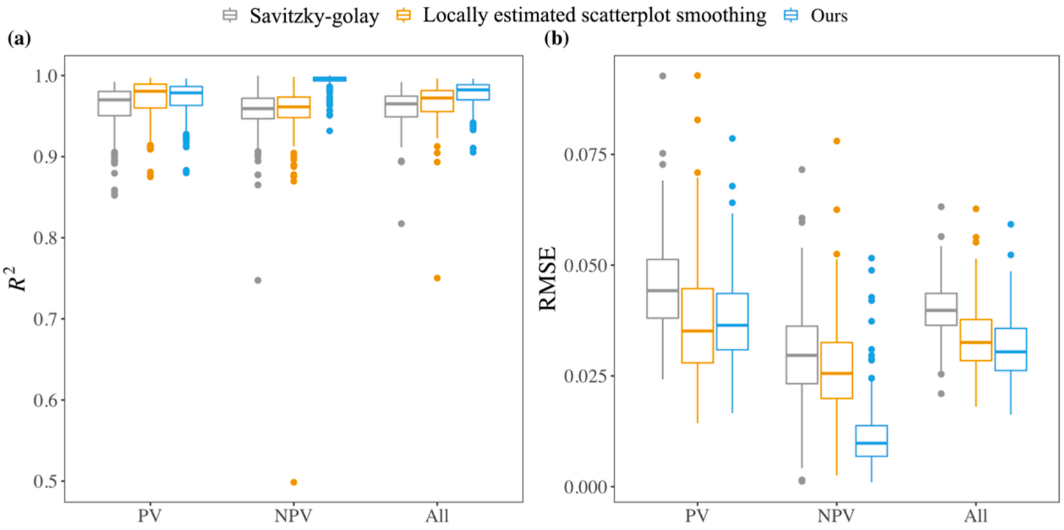

Fig. 7. Comparison of Savitzky-Golay filter (S-G filter), locally estimated scatterplot smoothing (LOESS), and proposed merit-based integrated filter.

estimation errors of phenological date given different sowing times and abnormal climatic conditions.

The measured maturity times of 36 randomly selected soybean cultivars were placed on the corresponding time-series curve (Fig. 10a). The results show that the maturity time on the time-series curves differ between soybean cultivars. Based on the degree of leaf drop at maturity, soybean cultivars can be categorized into three groups: no drop (all or most leaves remain on the stem), partial drop (few leaves remain on the stem), and complete drop (all leaves have fallen), thus determining the deciduousness of the germplasm. Accordingly, based on the maturity time, these cultivars may be divided into three maturation modes: near the upper (I) asymptote, inflection points (II), and near the lower (III) asymptote (Fig. 10b). The cultivars DaBaiPi, DaHuangDou1, and Qing6 came from Shandong, Shanxi, and Shandong, respectively, and belong to mode I. Energy conservation causes an earlier maturity and harvest of northern soybean cultivars grown in Sanya and Nanchang, and the

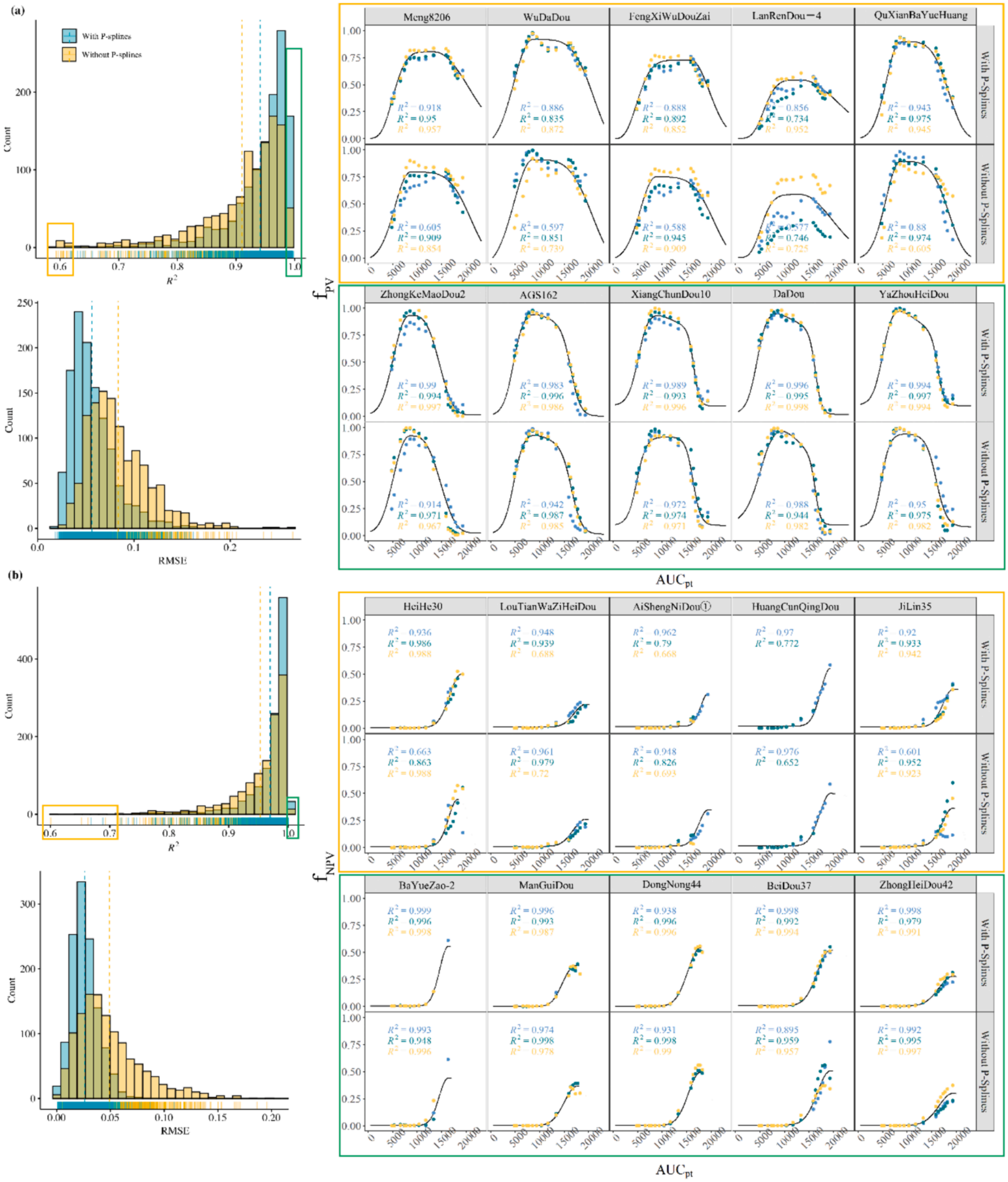

Fig. 8. Comparison of shape model with P-splines and shape model without P-splines for (a) f_{PV} and (b) f_{NPV}.

leaves remain green during senescence. The cultivars BaiMaoDou2, BenDiDaHuangDou, and HeiDouZai came from Guizhou, Hebei, and Guangdong, respectively, and belong to mode II. The cultivars Foster, LongChuanHuangNiDou, and WuYueHuang belong to mode II and came from Florida (United States), Guangdong, and Jiangxi, respectively. Soybean cultivars belonging to mode III were at a lower latitude than those belonging to mode II. These results indicated that soybean cultivars planted in Sanya and Nanchang mature later in southern districts and mature progressively earlier upon moving north (Fig. 10b).

4. Discussion

4.1. Soybean growth model with photothermal accumulation area provides a way to analyze the target growth curves of each soybean cultivar at different experimental sites

The MET of soybean varieties involved 10 sites with variability in climate and planting dates, spanning N16°41′52″ in latitude. The influence of climate factors can be minimized by introducing AUC_{pt}, as previously shown by Zeng et al. (2016) and Setiyono et al. (2007). The photothermal accumulation areas (AUC_{pt}) were calculated by

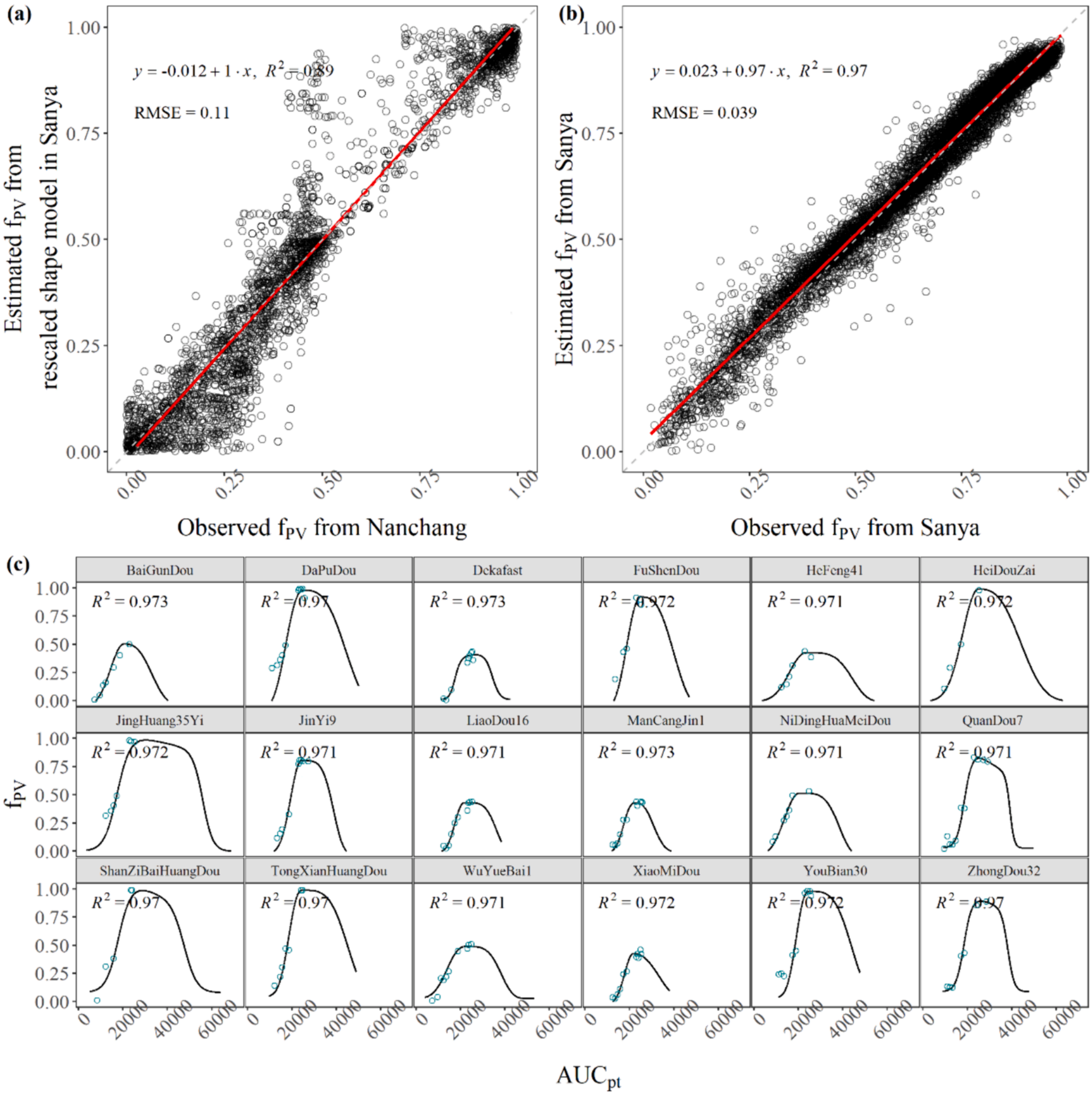

Fig. 9. Evaluation of the migration performance of time series shape models. (a) Relationship between estimated f_{PV} from rescaled shape model in Sanya and observed f_{PV} from Nanchang, (b) relationship between the estimated and observed f_{PV} from Sanya, and (c) visualization of time-series curves of 18 randomly selected soybean cultivars.

combining the daily temperature with local sunrise and sunset times in this study. The FD and MD of soybean cultivars correlates highly with the temperature ($r = -0.672$ & -0.775, Fig. 5c). Their correlation with temperature was negative, indicating that the AUC_{pt} of soybean cultivars decreases with temperature. The AUC_{pt} of soybean cultivars correlates strongly with the latitude ($r > 0.77$, Fig. 5d), and the relationship between latitude and AUC_{pt} from ED to MD follows a power function [$y = a\,(x + b)^3$; Fig. 5d] with $R^2 = 0.77$–0.95, SD = 0.044, and CV = 5.06 % (Fig. 5d). Based on the power function, the AUC_{pt} and its slope at northern China are greater than in Huang-Huai and the southern China soybean regions, which is consistent with the results of Yang et al. (2019) who reported a significant negative linear relationship between critical photoperiod and maturity group.

4.2. Performance comparison of denoising and rebuilding the crop time-series phenotypic data with S-G filtering, LOESS, and proposed merit-based integrated filter

The accurate extraction of the time-series f_{PV} and f_{NPV} depends on the deviations of the estimation, which in turn depend on wind and solar radiation. Most studies use the Savitzky-Golay (S-G) filter and locally estimated scatterplot smoothing (LOESS) to de-noise and rebuild crop time-series phenotypic data (Savitzky and Golay 1964; Cleveland 1979; Wang et al., 2014; Pan et al., 2015). S-G filtering, LOESS, and the proposed merit-based integrated filter (Fig. 7) were applied and compared to study how they affected the minimization of this remaining noise artifact in the time series f_{PV} and f_{NPV}. As shown in Figs. S2 and S3, the S-G filter tends to retain distribution features and preserve the width and

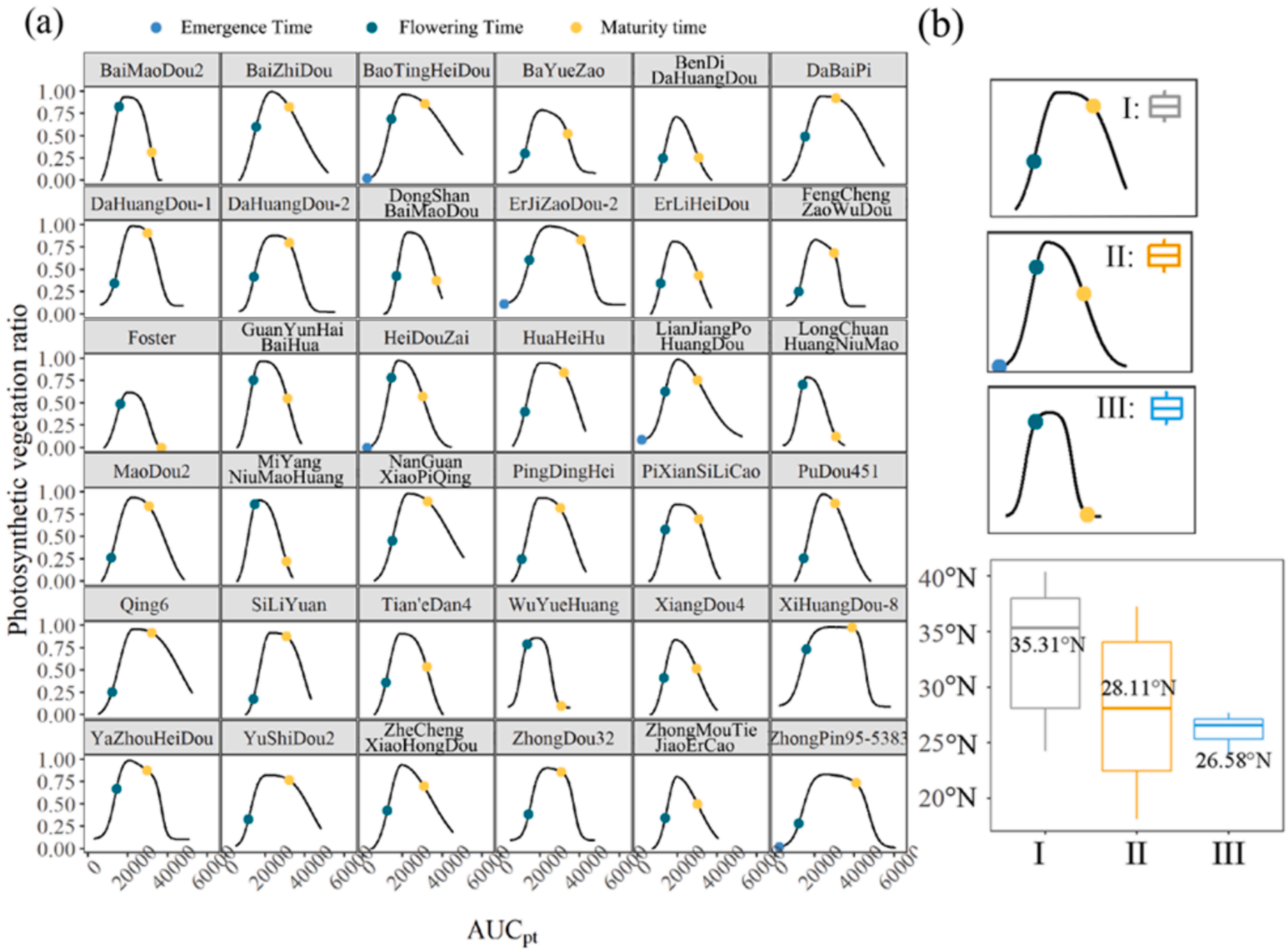

Fig. 10. Distinction and visualization of three maturation modes for different soybean varieties. (a) Patterns of emergence, flowering, and maturity date on f_{PV} time series curves for different soybean varieties. (b) The latitude of soybean cultivars selected from the three modes of maturation.

height of peaks in the spectral waveform. This remarkable advantage of S-G filtering has been mentioned by Azami *et al.* (2012) and Moosavi et al. (2018) and so was not unexpected. Compared with S-G filtering, LOESS smooths better and reduces contamination in the time-series data but flattens some of the sharp peaks that contain useful crop growth information (Fig. S3; Moosavi et al., 2018). The proposed merit-based integrated filter combines the merits of S-G filtering and LOESS to produce the best-performing filter (Figs. 3, S2 and S3).

4.3. Evaluation of omission of spatial heterogeneity of UAV phenotyping platform

The two-dimensional smooth surfaces of each time point in the f_{PV} and f_{NPV} datasets are fit by P-splines to assess the influence of spatial heterogeneity for the breeding experiment. The results indicate that the P-spline spatial heterogeneity model is promising for correcting spatial effects across a range of important growth stages. The P-spline spatial heterogeneity model not only recovers the spatial trend of the raw data but also is consistent with the raw data (Fig. S4). The genotypic performance of the field breeding experiments is better predicted by correcting the spatial effect with $R^2 \geq 0.62$, RMSE$\leq$0.17. In addition, no significant systematic error occurs with the time-series shape models for three duplications of the field. In short, the time-series f_{PV} and f_{NPV} shape model with P-splines is an effective method for correcting spatial heterogeneity between different duplications of the same germplasm. The effectiveness of the P-splines-Analysis of variance spatial field in breeding trails was tested on manual measurements or field phenotyping platforms (Lado et al., 2013; Rodríguez-Álvarez et al., 2018; Kronenberg et al., 2020). The present results prove the effectiveness of applying P-splines-ANOVA spatial field to breeding trials using UAV phenotyping platforms. The breeding values of f_{PV} and f_{NPV} are predicted by using the BLUPs after accounting for the effects of spatial heterogeneity and climate factors and building unique shape models of soybean cultivars with extensive genetic diversity.

4.4. Environmental portability of growth curve of soybean cultivars

The f_{PV} and f_{NPV} time series in Sanya combined with AUC$_{pt}$ were fit by using the double logistic function and S curve with $R^2 \geq 0.62$, RMSE$\leq$0.17. The soybean cultivars' growth curve in Nanchang was transformed from those in Sanya by stretching, compressing, and translating in the t and y directions with $R^2 = 0.89$, RMSE=0.11, and slope = 1. Considering the temperature and photoperiod, the AUC$_{pt}$ of each soybean cultivar at the vegetation and reproductive stages was almost the same at both sites and in both years (Fig. 9). The shape model built with AUC$_{pt}$ minimizes the estimation errors of phenological date when different sowing times and abnormal climatic conditions occur. Our results are entirely consistent with those recently presented by Zeng et al. (2016), Zhou et al. (2020), and Zhang et al. (2022) based on MODIS and UAV imagery. Our method further divides soybean cultivars according to their maturity time into three maturation modes: near the upper asymptote (I), inflection points (II), and near the lower asymptote (III) (Fig. 10b). Soybean cultivars planted in Sanya and Nanchang mature later than those from the southern district, and maturation times gradually advance as one moves north to the northern district. Energy conservation causes the leaves of northern soybean cultivars growing in Sanya and Nanchang to remain green during maturation (Fig. 10).

5. CONCLUSION

The multi-environment trials (METs) for soybean cultivars were conducted across 10 different sites, which exhibited variation in climate and planting dates, spanning N16°41′52″ in latitude. A proposed strategy was presented in this study to describe the target growth curve of soybean cultivars with extensive genetic diversity and phenotypic variation, and that accounts for the effects of spatial heterogeneity and climate factors in multi-environment trials (METs). The relationship between time and latitude AUC_{pt} from ED to MD is first explored to prepare for modelling subsequent AUC_{pt}-based growth across multiple regional test sites. The relationship between latitude and AUC_{pt} follows a power function with $R^2 = 0.77$–0.95, SD=0.044, and CV=5.06 %. Accurate shape models of soybean cultivars are established by using merit-based integrated filtering and a field spatial-correction model, which produces accurate results with $R^2 \geq 0.62$, RMSE≤0.17. The environmental portability of the growth curve of soybean cultivars is evaluated by using the time series f_{PV} in Sanya and Nanchang with a fitting accuracy of $R^2 = 0.89$, RMSE=0.11, and slope = 1. The resulted growth curve of soybean diversity panel offers a promising opportunity for genetic study and breeding program.

CRediT authorship contribution statement

Yingpu Che: Conceptualization, Methodology, Visualization, Writing – original draft. **Yongzhe Gu:** Data curation. **Dong Bai:** Investigation. **Delin Li:** Investigation. **Jindong Li:** Validation. **Chaosen Zhao:** Investigation. **Qiang Wang:** Investigation. **Hongmei Qiu:** Investigation. **Wen Huang:** Investigation. **Chunyan Yang:** Investigation. **Qingsong Zhao:** Investigation. **Like Liu:** Investigation. **Xing Wang:** Investigation. **Guangnan Xing:** Investigation. **Guoyu Hu:** Investigation. **Zhihui Shan:** Investigation. **Ruizhen Wang:** Investigation. **Ying-hui Li:** Funding acquisition, Resources, Writing – review & editing. **Xiuliang Jin:** Funding acquisition, Supervision, Writing – review & editing. **Li-juan Qiu:** Resources.

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 partially supported by the National Key R&D Program of China (2021YFD1201601), Hainan Yazhou Bay Seed Lab (B22C10204), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences, the project was supported by Nanfan special project, CAAS (YJTCY01, YBXM01), the earmarked fund for CARS (CARS-04-PS01).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2024.109313.

References

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Bernardeli, A., Rocha, J.R.A.S.deC., Borém, A., Lorenzoni, R., Aguiar, R., Silva, J.N.B., Bueno, R.D., Alves, R.S., Jarquin, D., Ribeiro, C., Costa, M.D.L., 2021. Modeling spatial trends and enhancing genetic selection: an approach to soybean seed composition breeding. Crop Sci. 61, 976–988. https://doi.org/10.1002/csc2.20364.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010.

Che, Y., Wang, Q., Zhou, L., Wang, X., Li, B., Ma, Y., 2022. The effect of growth stage and plant counting accuracy of maize inbred lines on LAI and biomass prediction. Precis. Agric. 23, 2159–2185. https://doi.org/10.1007/s11119-022-09915-1.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74, 829–836. https://doi.org/10.1080/01621459.1979.10481038.

Dierckx, P., 1995. Curve and Surface Fitting with Splines. Clarendon Press.

Eilers, P.H.C., Marx, B.D., 2003. Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. Chemom. Intel. Lab. Syst. 66, 159–174. https://doi.org/10.1016/S0169-7439(03)00029-7.

Elias, A.A., Rabbi, I., Kulakow, P., Jannink, J.L., 2018. Improving genomic prediction in cassava field experiments using spatial analysis. G3 Genes|Genomes|Genetics 8, 53–62. https://doi.org/10.1534/g3.117.300323.

Fischler, M.A., Bolles, R.C., 1987. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: Fischler, M.A., Firschein, O. (Eds.), Readings in Computer Vision. Morgan Kaufmann, San Francisco, pp. 726–740. https://doi.org/10.1016/B978-0-08-051581-6.50070-2.

Hassan, M.A., Yang, M., Rasheed, A., Tian, X., Reynolds, M., Xia, X., Xiao, Y., He, Z., 2021. Quantifying senescence in bread wheat using multispectral imaging from an unmanned aerial vehicle and QTL mapping. Plant Physiol. 187, 2623–2636. https://doi.org/10.1093/plphys/kiab431.

Henderson, C.R., 1973. Sire evaluation and genetic trends. J. Anim. Sci. 10–41. https://doi.org/10.1093/ansci/1973.Symposium.10.

Jarquin, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J.W.R., Covarrubias Pazaran, G., Burgueño, J., Pacheco, A., Grondona, M., Wimmer, V., Prasanna, B.M., 2020. Genomic prediction enhanced sparse testing for multi-environment trials. G3 (Bethesda) 10, 2725–2739. https://doi.org/10.1534/g3.120.401349.

Kaler, A.S., Purcell, L.C., Beissinger, T., Gillman, J.D., 2022. Genomic prediction models for traits differing in heritability for soybean, rice, and maize. BMC Plant Biol. 22, 87. https://doi.org/10.1186/s12870-022-03479-y.

Koh, J.C.O., Banerjee, B.P., Spangenberg, G., Kant, S., 2022. Automated hyperspectral vegetation index derivation using a hyperparameter optimisation framework for high-throughput plant phenotyping. New Phytol. 233, 2659–2670. https://doi.org/10.1111/nph.17947.

Kronenberg, L., Yates, S., Boer, M.P., Kirchgessner, N., Walter, A., Hund, A., 2020. Temperature response of wheat affects final height and the timing of stem elongation under field conditions. J. Exp. Bot. 72, 700–717. https://doi.org/10.1093/jxb/eraa471.

Lado, B., Matus, I., Rodríguez, A., Inostroza, L., Poland, J., Belzile, F., del Pozo, A., Quincke, M., Castro, M., von Zitzewitz, J., 2013. Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3 Genes|Genomes|Genetics 3, 2105–2114. https://doi.org/10.1534/g3.113.007807.

Lee, D.J., Durbán, M., Eilers, P., 2013. Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. Comput. Stat. Data Anal. 61, 22–37. https://doi.org/10.1016/j.csda.2012.11.013.

Li, D., Bai, D., Tian, Y., Li, Y., Zhao, C., Wang, Q., Guo, S., Gu, Y., Luan, X., Wang, R., Yang, J., Hawkesford, M.J., Schnable, J.C., Jin, X., Qiu, L., 2022a. Time series canopy phenotyping enables the identification of genetic variants controlling dynamic phenotypes in soybean. J. Integr. Plant Biol. 13 https://doi.org/10.1111/jipb.13380.

Li, Y., Qin, C., Wang, L., Jiao, C., Hong, H., Tian, Y., Li, Y., Xing, G., Wang, J., Gu, Y., Gao, X., Li, D., Li, H., Liu, Z., Jing, X., Feng, B., Zhao, T., Guan, R., Guo, Y., Liu, J., Yan, Z., Zhang, L., Ge, T., Li, X., Wang, X., Qiu, H., Zhang, W., Luan, X., Han, Y., Han, D., Chang, R., Guo, Y., Reif, J.C., Jackson, S.A., Liu, B., Tian, S., Qiu, L., 2022b. Genome-wide signatures of the geographic expansion and breeding of soybean. Sci. China Life Sci. 66, 350–365. https://doi.org/10.1007/s11427-022-2158-7.

Li, Z., Zhao, Y., Taylor, J., Gaulton, R., Jin, X., Song, X., Li, Z., Meng, Y., Chen, P., Feng, H., Wang, C., Guo, W., Xu, X., Chen, L., Yang, G., 2022c. Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. Remote Sens. Environ. 273, 112967 https://doi.org/10.1016/j.rse.2022.112967.

Liu, S., Zhang, M., Feng, F., Tian, Z., 2020. Toward a "green revolution" for soybean. Mol. Plant 13, 688–697. https://doi.org/10.1016/j.molp.2020.03.002.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., Chen, L., Su, T., Nan, H., Zhang, D., Zhang, L., Wang, Z., Yang, Y., Yu, D., Liu, X., Yang, Q., Lin, X., Tang, Y., Zhao, Xi, Yang, X., Tian, C., Xie, Q., Li, X., Yuan, X., Tian, Z., Liu, B., Kong, F., Weller, J.L., 2020. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. Nat. Gene. 52, 428–436. https://doi.org/10.1038/s41588-020-0604-7.

Lyu, M., Lu, X., Shen, Y., Tan, Y., Wan, L., Shu, Q., He, Yuhong, Yong, H., Cen, H., 2023. UAV time-series imagery with novel machine learning to estimate heading dates of rice accessions for breeding. Agric. Forest Meteorol. 341, 109646. https://doi.org/10.1016/j.agrformet.2023.109646.

Meeus, J., 1998. Astronomical Algorithms, seconnd ed. Willmann-Bell, Richmond, Va.

Montesinos-López, A., Montesinos-López, O.A., Crossa, J., Burgueño, J., Eskridge, K.M., Falconi-Castillo, E., He, X., Singh, P., Cichy, K., 2016. Genomic bayesian prediction model for count data with genotype × environment interaction. G3: Genes|Genomes|Genetics 6, 1165–1177. https://doi.org/10.1534/g3.116.028118.

Moosavi, S.R., Qajar, J., Riazi, M., 2018. A comparison of methods for denoising of well test pressure data. J. Pet. Explor. Prod. Technol. 8, 1519–1534. https://doi.org/10.1007/s13202-017-0427-y.

Moreira, F.F., Hearst, A.A., Cherkauer, K.A., Rainey, K.M., 2019. Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. Plant Methods 15, 139. https://doi.org/10.1186/s13007-019-0519-4.

Ortiz, R., Reslow, F., Huicho, J., Vetukuri, R.R., Crossa, J., 2023. Optimizing multi-environment testing in potato breeding: using heritability estimates to determine number of replications, sites, and years for field trials. Euphytica 219, 48. https://doi.org/10.1007/s10681-023-03174-5.

Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G.A., Yan, J., Liu, J., 2015. Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. Int. J. Appl. Earth Obs. Geoinf. 34, 188–197. https://doi.org/10.1016/j.jag.2014.08.011.

Robbins, K.R., Backlund, J.E., Schnelle, K.D., 2012. Spatial corrections of unreplicated trials using a two-dimensional spline. Crop. Sci. 52, 1138–1144. https://doi.org/10.2135/cropsci2011.08.0417.

Rodríguez-Álvarez, M.X., Boer, M.P., van Eeuwijk, F.A., Eilers, P.H.C., 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Stat. 23, 52–71. https://doi.org/10.1016/j.spasta.2017.10.003.

L. Roth, H.P. Piepho, A. Hund. Phenomics data processing: extracting dose–response curve parameters from high-resolution temperature courses and repeated field-based wheat height measurements. in silico Plants 1 (2022) diac007. 10.1093/insilicoplants/diac007.

Rotili, D.H., de Voil, P., Eyre, J., Serafin, L., Aisthorpe, D., Maddonni, G.Á., Rodríguez, D., 2020. Untangling genotype x management interactions in multi-environment on-farm experimentation. Field Crop. Res. 255, 107900 https://doi.org/10.1016/j.fcr.2020.107900.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047.

Setiyono, T.D., Weiss, A., Specht, J., Bastidas, A.M., Cassman, K.G., Dobermann, A., 2007. Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. Field Crop. Res. 100, 257–271. https://doi.org/10.1016/j.fcr.2006.07.011.

Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., Xing, S., Du, H., Ma, S., Tian, Z., 2018. De novo assembly of a Chinese soybean genome. Sci. China Life Sci. 61, 871–884. https://doi.org/10.1007/s11427-018-9360-0.

Smith, M.W., Carrivick, J.L., Quincey, D.J., 2016. Structure from motion photogrammetry in physical geography. Prog. Phys. Geogr.: Earth Environ. 40, 247–275. https://doi.org/10.1177/0309133315615805.

Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., Du, X., Zhai, L., Weng, H., Li, Y., Li, X., Bao, Y., Shou, J., He, Y., 2020. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer – a case study of small farmlands in the South of China. Agric. For. Meteorol. 291, 108096 https://doi.org/10.1016/j.agrformet.2020.108096.

Wang, L., Zhang, F., Jing, Y., Jiang, X., Yang, S., Han, X., 2014. Multi-temporal detection of rice phenological stages using canopy spectrum. Rice Sci. 21, 108–115. https://doi.org/10.1016/S1672-6308(13)60170-5.

Yang, W., Wu, T., Zhang, X., Song, W., Xu, C., Sun, S., Hou, W., Jiang, B., Han, T., Wu, C., 2019. Critical photoperiod measurement of soybean genotypes in different maturity groups. Crop. Sci. 59, 2055–2061. https://doi.org/10.1016/j.rse.2017.04.016.

Yu, N., Li, L., Schmitz, N., Tian, L.F., Greenberg, J.A., Diers, B.W., 2016. Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. Remote Sens. Environ. 187, 91–101. https://doi.org/10.1016/j.rse.2016.10.005.

Zeng, L., Wardlow, B.D., Wang, R., Shan, J., Tadesse, T., Hayes, M.J., Li, D., 2016. A hybrid approach for detecting corn and soybean phenology with time-series MODIS data. Remote Sens. Environ. 181, 237–250. https://doi.org/10.1016/j.rse.2016.03.039.

Zeng, L., Wardlow, B.D., Xiang, D., Hu, S., Li, D., 2020. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sens. Environ. 237, 111511 https://doi.org/10.1016/j.rse.2019.111511.

Zhang, C., Xie, Z., Shang, J., Liu, J., Dong, T., Tang, M., Feng, S., Cai, H., 2022. Detecting winter canola (*Brassica napus*) phenological stages using an improved shape-model method based on time-series UAV spectral data. Crop J. 10, 1353–1362. https://doi.org/10.1016/j.cj.2022.03.001.

Zhou, M., Ma, X., Wang, K., Cheng, T., Tian, Y., Wang, J., Zhu, Y., Hu, Y., Niu, Q., Gui, L., Yue, C., Yao, X., 2020. Detection of phenology using an improved shape model on time-series vegetation index in wheat. Comput. Electron. Agric. 173, 105398 https://doi.org/10.1016/j.compag.2020.105398.

Zhou, J., Yungbluth, D., Vong, C.N., Scaboo, A., Zhou, J., 2019. Estimation of the maturity date of soybean breeding lines using UAV-based multispectral imagery. Remote Sens. (Basel) 11, 2075. https://doi.org/10.3390/rs11182075.