



## OPEN ACCESS

## EDITED BY

Jianfeng Zhou,  
University of Missouri, United States

## REVIEWED BY

Wanneng Yang,  
Huazhong Agricultural University, China  
Jing Zhou,  
University of Wisconsin-Madison,  
United States  
Xu Wang,  
University of Florida, United States

## \*CORRESPONDENCE

Hiroyoshi Iwata  
✉ hiroiwata@g.ecc.u-tokyo.ac.jp

RECEIVED 07 April 2023

ACCEPTED 15 June 2023

PUBLISHED 05 July 2023

## CITATION

Sakurai K, Toda Y, Hamazaki K, Ohmori Y,  
Yamasaki Y, Takahashi H, Takanashi H,  
Tsuda M, Tsujimoto H, Kaga A,  
Nakazono M, Fujiwara T and Iwata H (2023)  
Random regression for modeling soybean  
plant response to irrigation changes using  
time-series multispectral data.  
*Front. Plant Sci.* 14:1201806.  
doi: 10.3389/fpls.2023.1201806

## COPYRIGHT

© 2023 Sakurai, Toda, Hamazaki, Ohmori,  
Yamasaki, Takahashi, Takanashi, Tsuda,  
Tsujimoto, Kaga, Nakazono, Fujiwara and  
Iwata. This is an open-access article  
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Random regression for modeling soybean plant response to irrigation changes using time-series multispectral data

Kengo Sakurai<sup>1</sup>, Yusuke Toda<sup>1</sup>, Kosuke Hamazaki<sup>1</sup>,  
Yoshihiro Ohmori<sup>1</sup>, Yuji Yamasaki<sup>2</sup>, Hirokazu Takahashi<sup>3</sup>,  
Hideki Takanashi<sup>1</sup>, Mai Tsuda<sup>4,5</sup>, Hisashi Tsujimoto<sup>2</sup>,  
Akito Kaga<sup>6</sup>, Mikio Nakazono<sup>3</sup>, Toru Fujiwara<sup>1</sup>  
and Hiroyoshi Iwata<sup>1\*</sup>

<sup>1</sup>Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo, Japan, <sup>2</sup>Arid Land Research Center, Tottori University, Tottori, Japan, <sup>3</sup>Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan, <sup>4</sup>Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan, <sup>5</sup>Tsukuba Plant Innovation Research Center, University of Tsukuba, Tsukuba, Japan, <sup>6</sup>Soybean and Field Crop Applied Genomics Research Unit, Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba, Japan

Plant response to drought is an important yield-related trait under abiotic stress, but the method for measuring and modeling plant responses in a time series has not been fully established. The objective of this study was to develop a method to measure and model plant response to irrigation changes using time-series multispectral (MS) data. We evaluated 178 soybean (*Glycine max* (L.) Merr.) accessions under three irrigation treatments at the Arid Land Research Center, Tottori University, Japan in 2019, 2020 and 2021. The irrigation treatments included W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, and D: no watering. To capture the plant responses to irrigation changes, time-series MS data were collected by unmanned aerial vehicle during the irrigation/non-irrigation switch of each irrigation treatment. We built a random regression model (RRM) for each of combination of treatment by year using the time-series MS data. To test the accuracy of the information captured by RRM, we evaluated the coefficient of variation (CV) of fresh shoot weight of all accessions under a total of nine different drought conditions as an indicator of plant's stability under drought stresses. We built a genomic prediction model (MT<sub>RRM</sub> model) using the genetic random regression coefficients of RRM as secondary traits and evaluated the accuracy of each model for predicting CV. In 2020 and 2021, the mean prediction accuracies of MT<sub>RRM</sub> models built in the changing irrigation treatments ( $r = 0.44$  and  $0.49$ , respectively) were higher than that in the continuous drought treatment ( $r = 0.34$  and  $0.44$ , respectively) in the same year. When the CV was predicted using the MT<sub>RRM</sub> model across 2020 and 2021 in the changing irrigation treatment, the mean prediction accuracy ( $r = 0.46$ ) was 42% higher than that of the simple genomic prediction model ( $r = 0.32$ ). The results suggest that this RRM method using the time-series MS data can effectively capture the genetic variation of plant response to drought.

## KEYWORDS

plant response, irrigation change, drought stress, single environmental trial, multispectral (MS), time-series, random regression model (RRM), *Glycine max* (L.) Merr.

## 1 Introduction

Soybean (*Glycine max* (L.) Merr.) exhibits a 40% reduction in yield due to drought (Specht et al., 1999), and genetic improvement of drought tolerance in soybeans is needed. Plant responses to drought stress are associated with drought tolerance (Pathan et al., 2014; Ye et al., 2020), and phenotypic data can be collected non-destructively from plants using high-throughput phenotyping (HTP). The relationship between spectral reflectance collected by hyperspectral (HS) and multispectral (MS) cameras and drought stress has been reported in several studies (Winterhalter et al., 2011; Bi et al., 2018; Wijewardana et al., 2019). Relationships between the normalized difference vegetation index (NDVI), which is calculated from the reflectance of near-infrared and red spectra, and the level of wilting have been reported in soybeans (Zhou et al., 2020). In soybeans, genotypes with slow-wilting traits exhibit high yield under drought conditions (Du et al., 2009; Pathan et al., 2014). Among several vegetation indices (VIs), normalized difference red-edge (NDRE), calculated from the reflectance of red-edge and red spectra, is reported to be the best vegetation index for detecting drought stress (Zygielbaum et al., 2009). These plant responses to irrigation changes are continuous, and time-series data can be collected using HTP (Yang et al., 2017). New insights can be obtained by capturing and analyzing time-series plant responses to irrigation changes, which are collected using HTP (Moreira et al., 2020). Chen et al. collected time-series changes in the digital volume of barley calculated from plant images collected using HTP during drought and recovery periods. The speed of recovery differs among genotypes; the faster the speed of recovery, the larger the final biomass (Chen et al., 2014).

Because the shape of time-series changes of biomass, leaf area index, and plant height was an S-shaped curve, they were commonly modeled using a logistic function (Setiyono et al., 2008; Sun and Frelich, 2011; Paine et al., 2012). When the shape of the curve or the function describing the curve is not obvious, smooth functions, such as Legendre polynomials or spline functions, can be used to model time series changes (Yang et al., 2006; van Eeuwijk et al., 2019). The random regression model (RRM) has been widely used for genetic analysis of time-series data (Campbell et al., 2018; Oliveira et al., 2019; Freitas Moreira et al., 2021). In the RRM, the covariance between each time point in a multivariate mixed model is modeled using Legendre polynomials and spline functions based on the assumption that time-series data are changing continuously (Henderson, 1984). The RRM makes it possible to describe time-series random genetic effects using a small number of parameters, i.e., regression coefficients (Huisman et al., 2002; Schaeffer, 2004). Estimated random regression coefficients can be used in a genome-wide association study (GWAS) to search for new genes (Campbell et al., 2019) or as secondary traits in a multi-trait model (MTM) to increase the prediction accuracy of the target trait (Sun et al., 2017).

With respect to drought tolerance, one of the important breeding targets is “stability under drought stress”, which is a phenotype that is stable under different drought levels (i.e., severe,

moderate, and mild) caused by different rainfall and water availability in different years and locations (Pidgeon et al., 2006; Kumar et al., 2012; Kumar et al., 2018; Torres and Henry, 2018; Ayed et al., 2021). Coefficient of variation (CV) of phenotypes in a target trait calculated from multi-environmental trials has long been used as an indicator for the stability over environments (Rao and Willey, 1980; Mohammadi and Amri, 2008; Das et al., 2010; Küchenmeister et al., 2012; Ray et al., 2015; Di Matteo et al., 2016). CV is calculated by dividing the standard deviation of phenotypes of the target trait by its mean, with smaller values indicating greater stability (Francis and Kannenberg, 1978). To evaluate the under-drought-stress stability with CV of the trait, field trials must be conducted at various drought levels. Conducting multi-environmental trials and evaluating the CVs for new genotypes is, however, time-consuming and costly. Predicting CV based on plant responses to irrigation changes may greatly reduce the time and cost of the CV evaluation. Although the relationship between the plant responses to irrigation changes and drought tolerance has been reported (Chen et al., 2014; Marchetti et al., 2019; Dodig et al., 2021), no studies have modeled plant response to irrigation changes and, based on the model, evaluated its relationship with plant’s stability under drought stress.

The objective of this study was to develop a method to measure and model plant responses to changes in irrigation. The developed method was applied to a single environmental trial with multiple irrigation patterns, and its effectiveness was evaluated based on its relationship to plant’s stability under drought stress. In this study, we evaluated 178 soybean accessions under different irrigation treatments in a 3-year trial. CV was calculated using fresh shoot weights observed in nine combinations of treatments by years and was used as an indicator of the plant’s stability under drought stress. Time-series MS data for each year were modeled using RRM, and the calculated genetic random regression coefficients were used as secondary traits for the genomic prediction of CV. If genomic prediction models using the calculated genetic random regression coefficients as secondary traits show higher prediction accuracy than those of the simple genomic prediction model without secondary traits, this suggests that time-series MS data are useful in predicting CV. If the genomic prediction models using the time-series MS data collected in the changing irrigation treatments show higher prediction accuracy than those in the continuous drought treatment, this suggests that time-series changes in MS data caused by irrigation change are useful in evaluating plant’s stability under drought stress. Additionally, we built three different prediction models: (1) within each combination of treatments by years, (2) using a small dataset of secondary traits, and (3) across years for the same type of treatment. Prediction model 1 could be used to predict the CV of novel genotypes within the same year. Prediction model 2 aimed to reduce the amount of data with secondary traits in the training set and the cost of collecting secondary traits. Prediction model 3 was intended to predict the CV of novel genotypes over the years using a previously prepared prediction model. Based on these three cases, we evaluated the effectiveness of the method for modeling time-series MS data using RRM.

## 2 Materials and methods

### 2.1 Experimental data

In this study, the accessions and experimental fields were the same as in the previous study (Sakurai et al., 2022). The diverse panel of 178 soybean accessions used were obtained from the gene bank of the National Institute of Agrobiological Sciences, Tsukuba, Japan (Table S1). These 178 accessions mainly consisted of Japanese and global soybean minicore collections (Kaga et al., 2011; Kajiya-Kanegae et al., 2021). A total of 178 soybean accessions were grown in three years 2019, 2020, and 2021 in the same experimental field at the Arid Land Research Center, Tottori University, Japan (35°32' N lat, 134°12' E long, 14 m above sea level). The experimental field soil was sandy and retained high water permeability. We used four types of drought treatments: no watering (drought, Treatment D),

watering for 5 d followed by no watering for 5 d (Treatment W5), watering for 10 d followed by no watering for 10 d (Treatment W10), and no watering for 10 d followed by watering for 10 d (Treatment D10). Treatments D, W5, and D10 were implemented in 2019. Treatments D, W10, and D10 were implemented in 2020 and 2021 (Figure 1). Each treatment consisted of two ridges. Each ridge consisted of two rows and one irrigation tube, and microplots were placed parallel to each other on either side of the tubes. The ridge height and width were 30 and 136 cm, respectively. For each treatment, 178 accessions were randomly assigned to microplots per year. As each accession was assigned one microplot per treatment, there were no replicates per treatment. The reason for not taking replications but prioritizing the number of genotypes is that previous studies have shown that genomic predictive modeling and QTL analysis are better with a higher number of genotypes, even if the number of replications is set to one

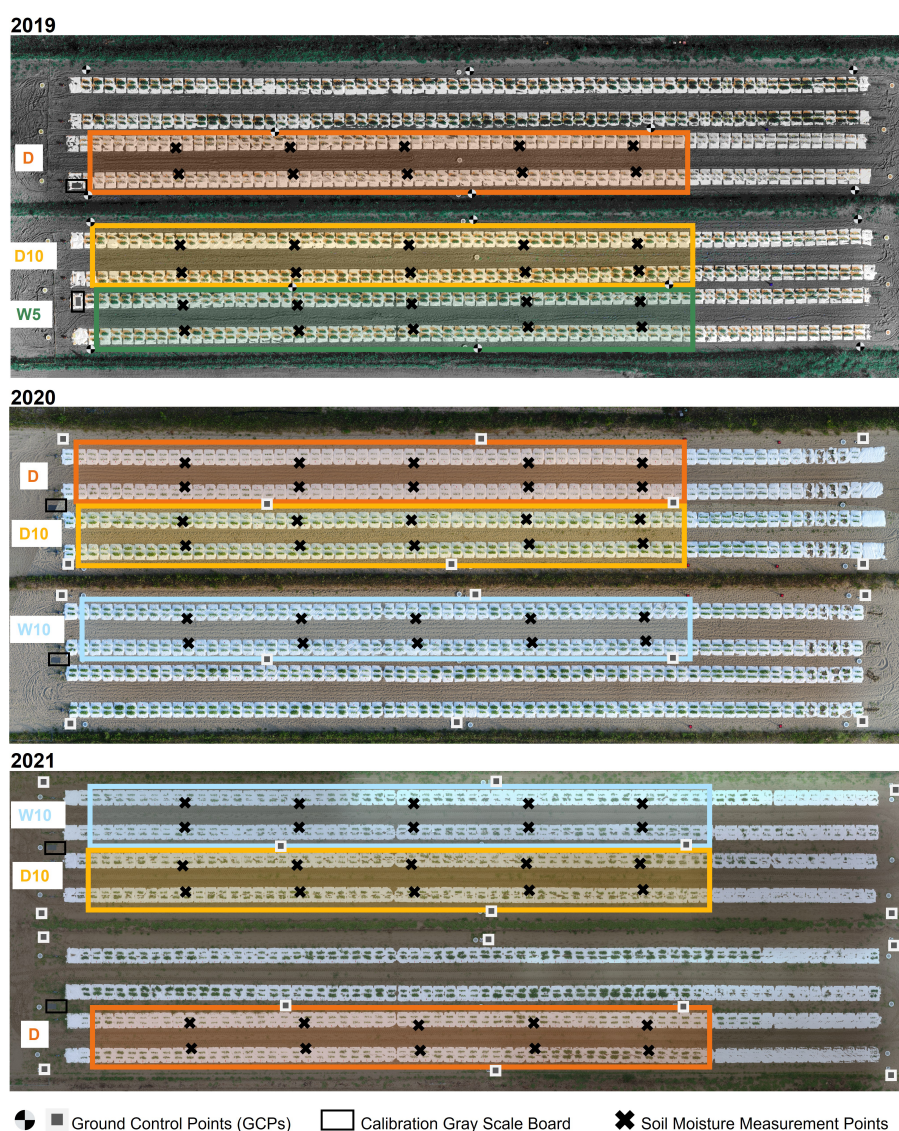


FIGURE 1

Experimental field overview and treatments in each year. W5: watering for 5 d followed by no watering for 5 d, W10: watering for 10 d followed by no watering for 10 d, D10: no watering for 10 d followed by watering for 10 d, D: no watering treatment.

(Knapp and Bridges, 1990; Lorenz, 2013). Four plants of each accession were grown in each microplot. The distances between the rows, microplots, and plants were 50, 80, and 20 cm, respectively (Figure 2). Fertilizers (13, 6.0, 20, 11, 7.0 g/m<sup>2</sup> 115 of N, P, K, Mg, and Ca, respectively) were applied to the field prior to sowing. Sowing was performed on 10 July 2019, 8 July 2020, and 6 July 2021. Two to three seeds were sown at each position, after which the germinated seedlings were thinned to one per position 2 weeks after sowing.

White mulch sheets (Tyvec; DuPont, Wilmington, USA) were laid over the ridges to prevent rainwater infiltration into the soil and control soil drought levels (Figure 1). A watering tube was installed under the mulch sheets at the center of each row. The watering tube (JKC Agro, Kumamoto, Japan) irrigated at a flow rate of 1.1 L/h m. Watering was done for over 5 h daily (7:00–9:00, 12:00–14:00, and 16:00–17:00), starting the day after seedling thinning for Treatments W5, W10, and D10. The irrigation cycle for each treatment is shown in Figure S1. Soil moisture was measured using a soil moisture meter (TDR-341F; Fujiwara Seisakusho, Tokyo, Japan) at 10 sites for each treatment over 30 d. Except for rainy days, no more than 3 d were allowed between soil moisture measurements. Soil moisture in each treatment was 127 calculated as the average of the treatments in the year (Figure S1). Days to

flowering (DTF), which was defined as the date that 50% of plants flowered in each microplot (Figure S2), were determined. To record DTF, we visited each microplot every other day during the flowering period. In each microplot of the four plants, two centrally located plants were collected 62 d after sowing, and the fresh aboveground weight of each plant was measured. The phenotypic value of the fresh weight of each microplot was calculated as the average of the two plants measured for each microplot. The number of rainy days was 19, 12, and 22 during the field experiment in 2019, 2020, and 2021, respectively.

## 2.2 Multispectral data collection and processing

The MS image collection and image analysis referred to the method employed by Sakurai et al. (Sakurai et al., 2022). In each treatment of a year, MS images were collected using unmanned aerial vehicles (UAVs). In 2019 and 2020, MS images (1.0 cm/pixel) were collected using a four-eye MS camera (Xacti, Osaka, Japan) mounted on a quadcopter UAV (DJI Matrice 100; DJI, Shenzhen, China). The MS camera has four independent lenses and sensors attached to different filters (MidOpt, Palatine, USA), including a

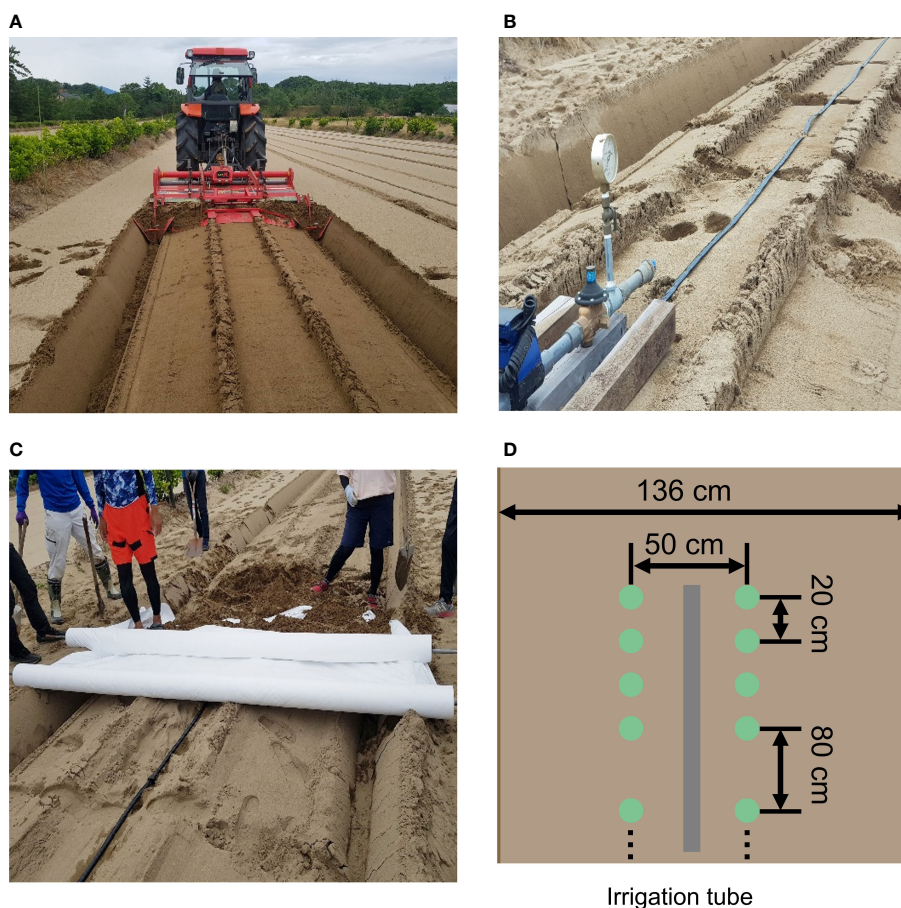


FIGURE 2

The setup of the experimental field. (A) Ridge for each treatment. (B) Irrigation tube on each ridge. (C) White multi-sheet. (D) Planting pattern in each ridge. Pictures (A–C) were used in Bui et al. (Bui et al., 2022). Permission to use these pictures was granted by the author.

triple-bandpass filter (TB550/660/850) and a red-edge bandpass. The TB550/660/850 can collect spectral intensities at 550 nm (green), 660 nm (red), and 850 nm (near-infrared). Bi725 can collect the spectral intensity at 725 nm (red-edge). The MS camera was set for continuous data capture at two frames per second per lens for a total of eight frames per second. The overlap and sidelap rates were set to 90% and the flights were set to an altitude of 20 m. In 2021, MS images (0.74 cm/pixel) were collected using DJI Phantom 4 Multispectral (P4M; DJI, Shenzhen, China). P4M has one RGB sensor and five spectral-band sensors at 450 nm (blue), 560 nm (green), 650 nm (red), 730 nm (red-edge), and 840 nm (near-infrared). P4M continuously collected MS images every two seconds during each flight. The overlap and sidelap rates were set to 75% and the flights were set to an altitude of 15 m.

Each spectral reflectance was calculated as the ratio of each spectral intensity from a grey scale panel set in the experimental field (Figure 1). All the flights were scheduled for 11:00-13:00 under clear sky conditions. Images were collected four, six, and seven times in 2019, 2020, and 2021, respectively (Figure 3). UAVs measurements were scheduled before and after the irrigation treatment was switched on W10 and D10 to capture the response of plants to changes in irrigation. Orthomosaic images of each spectral reflectance were obtained using a Pix4Dmapper (Pix4D, Prilly, Switzerland). Sixteen ground control points (GCPs) were set up in the field annually. The positions of the GCPs were measured using Aeropoints (Propeller, Sydney, Australia). Using geolocation information from the GCPs, 792 microplots, each with a maximum of four plants, were segmented from the orthomosaic images. Plants were segmented from an image

of each microplot using the NDVI-based segmentation method to extract and analyze the MS image data from only the plants. In this experimental field, it was reported that the NDVI differed significantly from the soil surface, white mulch sheets, and plants in 2019 (Sakurai et al., 2022); thus, plants were segmented with the NDVI threshold set to 0.15, as applied by Sakurai et al. (Sakurai et al., 2022). These analyses (extraction of spectral reflectance, calculation of NDVI, and NDVI-based segmentation) were performed using the OpenCV v3.3.1, library in Python v3.6.8.

Based on the spectral reflectance values of each plant pixel segmented in each microplot, we calculated two types of VIs: NDVI (Price and Bausch, 1995; Zarate-Valdez et al., 2015) and NDRE (Jorge et al., 2019). The equations for these VIs are listed in Supplementary Table S2. As there were four plants in each microplot, MS data were collected as a community of four plants within each microplot without considering the overlap between plants. Although we attempted to segment only the plant pixels, background noise was not completely removed. In each microplot, the average VI value may have been heavily influenced by the background noise. To reduce this effect, the median of all the segmented plant pixels was used as the representative value of each microplot. The VIs were calculated using R v4.1.2.

## 2.3 Genotyping

The genome dataset was the same as that used by Sakurai et al. (Sakurai et al., 2022). Whole-genome sequence data were obtained

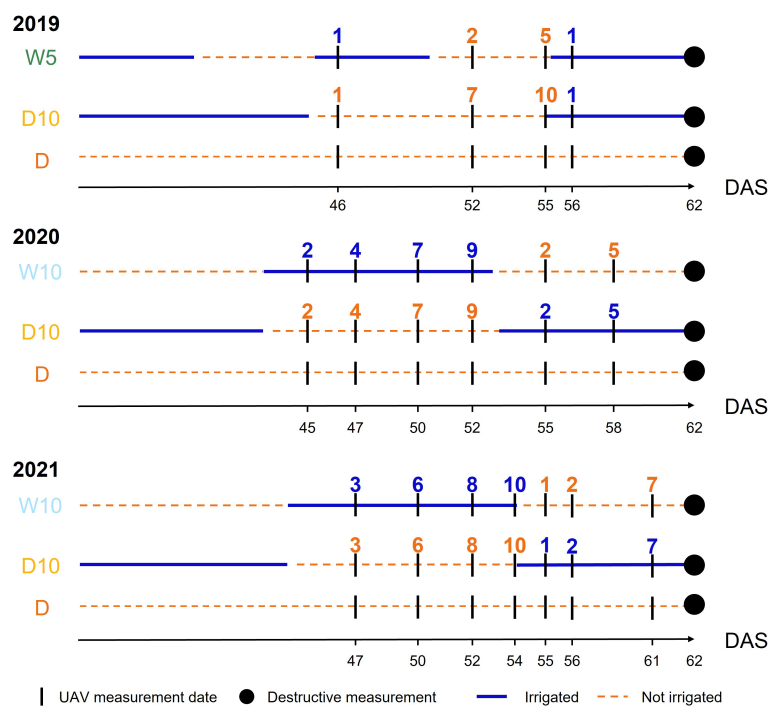


FIGURE 3

The timings of unmanned aerial vehicle measurements in each year. Blue (respectively Orange) numbers indicate the number of days elapsed since the start of irrigation (respectively drought) treatment. DAS: days after sowing, W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, D: no watering treatment.

for all accessions (Kajiya-Kanegae et al., 2021). All accessions were genotyped using an Illumina HiSeq X Ten or HiSeq 4000 (Illumina, San Diego, USA), and 4,776,813 single-nucleotide polymorphisms (SNPs) were identified. Among these SNPs, those that were heterozygous or those in which >95% of the individuals had missing data were excluded. Markers were also filtered for a minor allele frequency <0.025, and missing data were imputed based on the mean, after which they were filtered again for a minor allele frequency <0.05. Finally, linkage disequilibrium was computed only for SNP pairs for which the distance was <25,000 base pairs, and SNPs with linkage disequilibrium below 0.8 were selected, resulting in 173,583 SNP markers. Using these SNP markers, the additive numerator relationship matrix  $G$  was estimated using the ‘calcGRM’ function in the ‘RAINBOWR’ package in R v0.1.25 (Hamazaki and Iwata, 2020).

## 2.4 Random regression model

The MS data were collected at four, six, and seven time points in 2019, 2020, and 2021, respectively. The ratio of flowering at the timings of UAV measurements in each combination of treatments was calculated using DTF data (Table S3). A time series of VI values was modeled using an RRM (Mrode, 2014) for each treatment of each year separately. The RRM takes the following form:

$$y_{it} = b_t + \sum_{k=0}^{nr} \phi_k(t) u_{ik} + e_{it} \quad (1)$$

where  $y_{it}$  is the phenotypic value of each VI (NDVI or NDRE) at the time point  $t$  ( $t = 1, \dots, 4$  in 2019,  $t = 1, \dots, 6$  in 2020,  $t = 1, \dots, 7$  in 2021) for genotype  $i$  ( $i = 1, \dots, 178$ ),  $b_t$  is the fixed effect of each time point,  $nr$  is the order of Legendre polynomial for the genetic effects,  $\phi_k(t)$  is the  $k$ th ( $k = 0, \dots, nr$ ) Legendre polynomials for time point  $t$ ,  $u_{ik}$  is the genetic effect for the  $k$ th coefficients of Legendre polynomials, and  $e_{it}$  is the random residual effect. Vector  $\mathbf{u}_k = (u_{1k}, \dots, u_{178k})^T$ ,  $\mathbf{u} = (\mathbf{u}_0^T, \dots, \mathbf{u}_{nr}^T)^T$  follows the multivariate normal (MVN) distribution,  $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{Q} \otimes \mathbf{I}_{178})$  where  $\mathbf{Q}$  is  $(nr + 1) \times (nr + 1)$  (co)variance matrix for the Legendre polynomials and  $\mathbf{I}_{178}$  is  $178 \times 178$  identical matrix. Vector  $\mathbf{e}_k = (e_{1k}, \dots, e_{178k})^T$ ,  $\mathbf{e} = (\mathbf{e}_0^T, \dots, \mathbf{e}_{nr}^T)^T$  follows the MVN distribution  $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_{178(nr+1)} \sigma_e^2)$ , where  $\mathbf{I}_{178(nr+1)}$  is the identical matrix and  $\sigma_e^2$  is the residual variance. To determine the order of  $nr$ , we built RRMs with three values of  $nr$  ( $nr = 0, 1, 2$ ) using the data collected in 2021, because the number of time points was the largest in 2021 among other years. The goodness of the model fit was assessed by computing Akaike’s information criterion (AIC) (Akaike, 1974). The best value of  $nr$  is selected based on the lowest AIC value. This RRM model was built using the “ASREML” R package v4.1.0.154 (Gilmour et al., 2015).

## 2.5 Coefficient of variation

We used CV (Francis and Kannenberg, 1978) of fresh weight as an indicator of plant’s stability under drought stresses. CV was

calculated as follows:

$$CV_i = \frac{\sigma_i}{\mu_i} \times 100 \quad (2)$$

where  $\mu_i$  is the mean fresh weight in all nine combinations of treatments by years for genotype  $i$  ( $i = 1, \dots, 178$ ),  $\sigma_i$  is the standard deviation of the fresh weight in the combinations. A low CV value indicates high environmental stability. Before calculating  $\mu_i$  and  $\sigma_i$ , fresh weight was scaled by min-max normalization, ranging from 0 to 1 for each combination of treatments by years. This is because  $CV_i$  is significantly affected by specific combinations of treatments by years, which have large mean values.

## 2.6 Genomic heritability and genomic prediction

Simple genomic prediction models were built to calculate the genomic heritability (Litchfield et al., 2015) for three traits; fresh weight in each combination of treatments by years, the genetic random regression coefficient of RRMs, and CV. Also, a simple genomic prediction model was built to predict the genetic value of the CV. The simple genomic prediction model has the following form:

$$y_i = \mu + u_i + e_i \quad (3)$$

where  $y_i$  is the phenotypic value for each trait of genotype  $i$  ( $i = 1, \dots, 178$ ),  $\mu$  is the overall mean,  $u_i$  is the genetic random effect, and  $e_i$  is the residual random effect. The vector  $\mathbf{u} = (u_1, \dots, u_{178})^T$  follows the MVN distribution,  $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \mathbf{G} \sigma_g^2)$  where  $\mathbf{G}$  is the additive numerator relationship matrix, and  $\sigma_g^2$  is the additive genetic variance. Vector  $\mathbf{e} = (e_1, \dots, e_{178})^T$  follows the MVN distribution  $\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_{178} \sigma_e^2)$ , where  $\mathbf{I}_{178}$  is the identical matrix and  $\sigma_e^2$  is the residual variance. Based on the estimated parameters of genetic and residual variances, genomic heritability was calculated as  $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$ . The model was implemented using the ‘EMM.cpp’ function in the ‘RAINBOWR’ package in R v0.1.25 (Hamazaki and Iwata, 2020).

## 2.7 Multitrait model

For each combination of treatments by years, the MTM was built as a Bayesian multivariate Gaussian model (Montesinos-López et al., 2016; Lado et al., 2018) to predict the CV of a genotype. An MTM using  $M$  secondary traits takes the following form:

$$y_{im} = \mu_m + u_{im} + e_{im} \quad (4)$$

where  $y_{i0}$  is the phenotypic value of the CV for genotype  $i$  ( $i = 1, \dots, 178$ );  $y_{im}$  ( $m = 1, \dots, M$ ) is the phenotypic value of the secondary traits;  $\mu_m$  is the overall mean for trait  $m$  ( $m = 0, \dots, M$ );  $u_{im}$  is the genetic random effect ( $m = 0, \dots, M$ ); and  $e_{im}$  is the residual random effect ( $m = 0, \dots, M$ ). Vector  $\mathbf{u}_m = (u_{1m}, \dots, u_{178m})^T$ ,  $\mathbf{u} = (\mathbf{u}_0^T, \dots, \mathbf{u}_M^T)^T$  follows the MVN distribution,  $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \Sigma \otimes \mathbf{G})$  where  $\Sigma$  is  $(M + 1) \times (M + 1)$  genetic (co)variance matrix across

traits and  $\mathbf{G}$  is the additive numerator relationship matrix. Vector  $\mathbf{e}_m = (e_{1m}, \dots, e_{178m})^T$ ,  $\mathbf{e} = (\mathbf{e}_0^T, \dots, \mathbf{e}_M^T)^T$  follows the MVN distribution  $\mathbf{e} \sim \text{MVN}(0, \mathbf{R} \otimes \mathbf{I}_{178})$ , where  $\mathbf{R}$  is  $(M + 1) \times (M + 1)$  residual (co)variance matrix across traits and  $\mathbf{I}_{178}$  is the identical matrix.

We built two different types of MTMs: (1) an MTM directly using the VI value at each time point as a secondary trait ( $\text{MT}_{\text{All}}$  model), and (2) an MTM using the genetic random regression coefficients of RRM (Equation 1) as a secondary trait ( $\text{MT}_{\text{RRM}}$  model).

## 2.8 Cross-validation cases

We assumed three cases using MTM (Figure 4) and compared the prediction accuracies of  $\text{MT}_{\text{All}}$  model and  $\text{MT}_{\text{RRM}}$  model. The prediction accuracy was evaluated using a 10-fold cross-validation with 10 replicates. Pearson's correlations were calculated between the observed and predicted CV values in each replicate, and the average of these correlations was used as the prediction accuracy.

### 2.8.1 Case 1: within each combination of treatments by years

In the first case (Case1), the whole data set in each combination of treatments by years was divided into training and test sets. Predicting the CV of the test set using secondary traits and genome-wide marker data. Because an MTM was built in each combination of treatments by years in Case1, the number of secondary traits in an MTM need not be equal among years. In  $\text{MT}_{\text{All}}$  model,  $M$  (the number of secondary traits) was four, six, and seven in 2019, 2020, and 2021, respectively. In  $\text{MT}_{\text{RRM}}$  model,  $M$  was equal to  $nr + 1$  (the order in Equation 1), which was the same for all years.

### 2.8.2 Case 2: using a small dataset of secondary traits

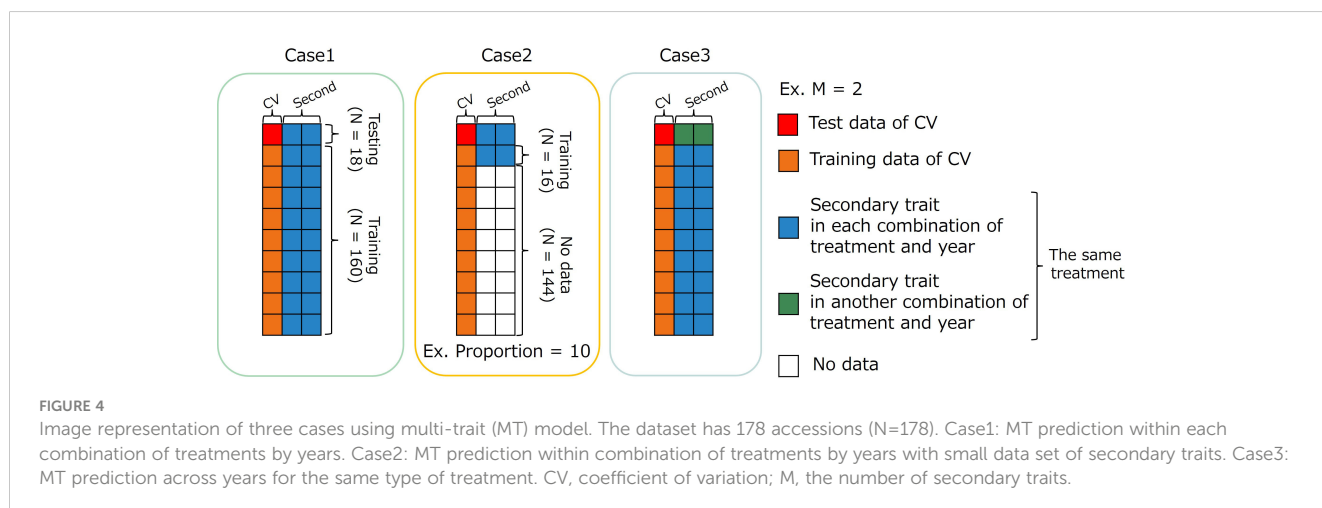
In the second case (Case2), we assumed a situation that CV had already been calculated from past data. In MTM, secondary trait data are usually collected for all genotypes in the training and test sets. We have to conduct the field experiment and collect time-

series MS data to build an MTM in our study. The small number of genotypes measured for secondary traits in the training set resulted in cost reduction. The proportions of data with secondary traits were set to 10%, 25%, and 50%. First, the whole data set for each combination of treatments by years was divided into training and test sets. Second, genotypes with secondary trait data in the training data were randomly selected five times for each proportion to reduce the effect of specific datasets with secondary traits. Therefore, the prediction accuracy was evaluated via a 10-fold cross-validation with 50 replicates. Because an MTM was built in each combination of treatments by years in Case2, the number of secondary traits in an MTM need not be equal among years. In  $\text{MT}_{\text{All}}$  model,  $M$  was four, six, and seven in 2019, 2020, and 2021, respectively. In  $\text{MT}_{\text{RRM}}$  model,  $M$  was equal to  $nr + 1$  for all the years.

### 2.8.3 Case 3: across years for the same type of treatment

The third case (Case3) was intended to create a prediction model using specific year data and predict CV using data from another year. In this case, we predicted the CV of novel genotypes over the years with their secondary trait data and genome-wide marker data using a previously prepared prediction model built using another year's data. The prediction accuracy was calculated via cross-year cross-validation using the same treatment. Because Treatment W5 did not have yearly replications, this validation was performed only for treatments W10, D10, and D.

In  $\text{MT}_{\text{All}}$  model,  $M$  was only three because the measurement timings differ among the years. We set the start of irrigation (and drought) in the irrigation changing treatments as a reference point and considered the time difference within a day as the same measurement timing. Therefore, MS data on the dates after sowing (DAS) of 52, 55, and 56 in 2019; 50, 52, and 55 in 2020; and 52, 54, and 55 in 2021 were used in  $\text{MT}_{\text{All}}$  model (Figure 3). However, in  $\text{MT}_{\text{RRM}}$  model, all time point data were available for modeling because time-series MS data were modeled with  $nr + 1$  orders, and the genetic random regression coefficients were used as secondary trait data. In  $\text{MT}_{\text{RRM}}$  model,  $M$  can be fixed at  $nr + 1$  for all years.



In all three cases, to compare the prediction accuracy between  $MT_{All}$  and  $MT_{RRM}$  models, we calculated the proportion of improvement. The proportion of improvement is defined as follows:

$$POV(\%) = \left( \frac{PA_{RRM} - PA_{All}}{PA_{All}} \right) \times 100 \quad (5)$$

where  $POV$  denotes the proportion of improvement,  $PA_{All}$  denotes the prediction accuracy of  $MT_{All}$  model, and  $PA_{RRM}$  denotes the prediction accuracy of  $MT_{RRM}$  model.

## 3 Results

### 3.1 Relationship between coefficient of variation and fresh weight

Fresh weight varied among the nine combinations of treatments by years (Figure 5). As the fresh weight in Treatment D was smaller than that in the other treatments each year, Treatment D was the treatment with the most severe drought stress. For Treatment D10, the fresh weight in 2019 was higher than that in the other two years. Genomic heritability of fresh weight ranged from 0.18 to 0.64 and that of CV was 0.35 (Table 1). Next, we calculated the phenotypic correlation between CV and fresh weight for each combination of treatments by years. No positive phenotypic correlation was observed between the CV and fresh weight. In addition, the phenotypic correlation between CV and fresh weight was negative

in seven out of nine combinations of treatments by years. This result indicates a positive relationship between fresh weight and the stability under drought stress as assessed by CV and in the seven combinations of treatments and years.

### 3.2 Model selection for random regression model

We evaluated the goodness-of-fit of the RRM using NDVI and NDRE values for each treatment in 2021. Under all treatments, the best model was based on NDVI values using linear Legendre polynomials, that is, the order of  $nr$  was 1 (Table 2). The model with  $nr = 0$  exhibited the best NDRE values under all treatments (Table S4). This result indicates that only the intercept varied among genotypes in the RRM of NDRE values. Therefore, we employed the RRM of NDVI using  $nr = 1$  in later analyses. We built each RRM (Equation 1) using the time-series NDVI data for 2019, 2020, and 2021. We then calculated the genetic effect for the intercept and 1st coefficients of the Legendre polynomials (L0 and L1) for each year.

### 3.3 Genetic correlation and genomic heritability of secondary traits

We estimated the genetic correlations between CV and L0 and CV and L1, and estimated the genomic heritability of L0, L1, and

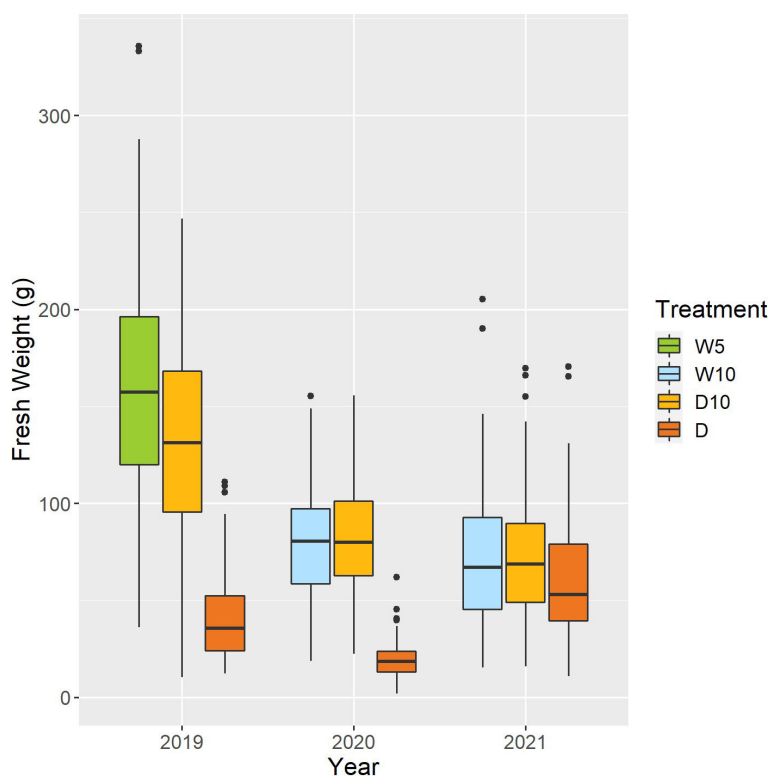


FIGURE 5

Boxplot of observed fresh weights of 178 soybean accessions in each combination of treatments by years. W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, D: no watering treatment.



**TABLE 1** Genomic heritability of fresh weight in each combination of treatments by years, coefficient of variation (CV) calculated over the nine combinations, and phenotypic correlation ( $r$ ) between fresh weight in each combination of treatments by years and the CV.

Year	Treatment/Index	$h^2$	$r$
2019	W5	0.41	-0.12
	D10	0.23	-0.12
	D	0.23	-0.35***
2020	W10	0.64	-0.40***
	D10	0.33	-0.41***
	D	0.24	-0.42***
2021	W10	0.18	-0.44***
	D10	0.40	-0.45***
	D	0.43	-0.43***
	CV	0.35	

$h^2$ : genomic heritability,  $r$ : phenotypic correlation between fresh weight and coefficient of variation (CV), W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, D: no watering treatment, \*\*\*: significant at  $p < 0.001$ .

CV. In seven out of nine combinations of treatments by years, negative genetic correlations between CV and L0 or CV and L1 were observed (Table 3). L0 and L1 values for each genotype were associated with the intercepts and slopes of the time-series NDVI values. Therefore, these negative genetic correlations indicate that a low CV is associated with large intercepts and slopes of the time-series NDVI values. In 2019, there were no obvious patterns in genetic correlations or genomic heritability among the treatments (Table 3). However, in 2020, treatments W10 and D10 showed higher genetic correlations and genomic heritability than Treatment D for all the traits. In 2021, the genetic correlation of L0 was -0.2 ( $p < 0.01$ ) in Treatment D, whereas W10 and D10 showed higher genetic correlations of -0.4 ( $p < 0.001$ ) and -0.51 ( $p < 0.001$ ), respectively.

### 3.4 Case 1: within each combination of treatments by years

In Case1, we used CV, which was calculated from the fresh weight collected in nine combinations of treatments by years, as the same target trait and built  $MT_{RRM}$  model and  $MT_{All}$  model in each combination of treatments by years. The prediction accuracy of the simple genomic prediction model was 0.32. The prediction accuracy of  $MT_{RRM}$  model was higher than that of the simple genomic prediction model for all year treatments (Figure 6A). Three out of nine  $MT_{All}$  models evaluating plant's stability under drought. We focused on the  $MT_{RRM}$  model. In 2019, there was no difference in the prediction accuracy among the treatments in  $MT_{RRM}$  model. However, the prediction accuracies of treatments W10 and D10 in 2020 were 25% and 30% higher, respectively than those of Treatment D in 2020. In addition, the prediction accuracies of treatments W10 and D10 in 2021 were higher by 14% and 11%, respectively than those of Treatment D in 2021. These results indicate that the time-series MS data collected during the treatment, which changed the irrigation pattern, were more useful than those in the continuous drought treatment for predicting the CV in 2020 and 2021. To evaluate the importance of using RRM, we compared  $MT_{RRM}$  model and  $MT_{All}$  model. In the seven combinations of treatments by years, the prediction accuracies of  $MT_{RRM}$  model were higher than those of  $MT_{All}$  model (Figure 6C). In particular, the proportion of improvement was 14% for Treatment W10 in 2021.

### 3.5 Case 2: using a small dataset of secondary traits

To reduce the cost and labor required to create the prediction model, we changed the proportion of genotypes collected as secondary trait data in the training set. In 2020 and 2021, except

**TABLE 2** The goodness-of-fit of random regression models (RRMs) with the normalized difference vegetation index (NDVI) values in 2021.

Treatment	$nr$	Loglik	AIC	$p$
W10	0	527.2572655	-1038.514531	8
	<b>1</b>	<b>739.1732966</b>	<b>-1458.346593</b>	<b>10</b>
	2	459.0197275	-892.0394551	13
D10	0	409.6136183	-803.2272366	8
	<b>1</b>	<b>553.0469996</b>	<b>-1086.093999</b>	<b>10</b>
	2	256.6404682	-487.2809363	13
D	0	704.2429918	-1392.485984	8
	<b>1</b>	<b>725.6443146</b>	<b>-1431.288629</b>	<b>10</b>
	2	458.2977954	-890.5955909	13

$nr$ , the order of Legendre polynomial for the genetic effect; Loglik, log likelihood; AIC, Akaike's information criterion;  $p$ , the number of parameters; W10, watering for 10 d followed by no watering 10 d; D10, no watering for 10 d followed by watering 10 d; D, no watering treatment. The best model in each treatment is bolded based on AIC.

**TABLE 3** Genetic correlation between each parameter calculated in random regression models (RRMs) and coefficient of variation, and genomic heritability of genetic random regression coefficient of RRMs.

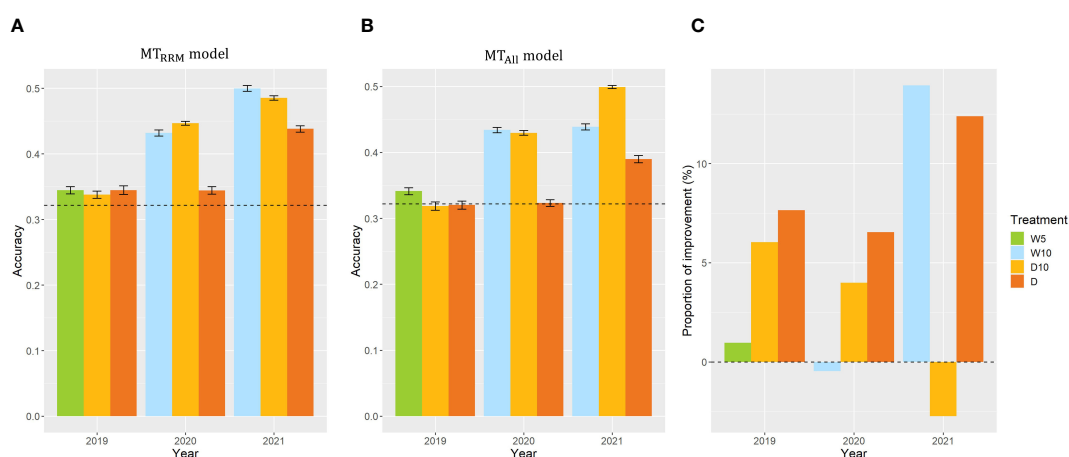
Year	Trait	Index	W5	W10	D10	D
2019	L0	$r_g$	-0.26***		-0.03	-0.16*
		$h^2$	0.59		0.33	0.1
	L1	$r_g$	0.08		-0.03	-0.17*
		$h^2$	0.79		0.54	0.31
2020	L0	$r_g$		-0.32***	-0.4***	-0.15
		$h^2$		0.41	0.29	0.21
	L1	$r_g$		-0.22**	-0.33***	-0.03
		$h^2$		0.79	0.43	0.1
2021	L0	$r_g$		-0.4***	-0.51***	-0.2**
		$h^2$		0.11	0.39	0.38
	L1	$r_g$		-0.35***	-0.31***	-0.34***
		$h^2$		0.7	0.67	0.29

L0: genetic effect for the intercept of Legendre polynomials in RRM, L1: the genetic effect for the 1st coefficient of Legendre polynomials in RRM,  $r_g$ : genetic correlation,  $h^2$ : genomic heritability, W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, D: no watering treatment, \*, \*\*, \*\*\*: significant at  $p < 0.05$ , 0.01, and, 0.001, respectively.

for Treatment D in 2020, the prediction accuracy for all combinations of treatments by years increased as the proportion of genotypes with secondary trait data increased (Figure 7). In Treatment D10 of 2020 and 2021, even when the proportions of data with secondary traits were 10%, the prediction accuracies of  $MT_{RRM}$  models were 0.40 and 0.42. These prediction accuracies were 23% and 30% higher than those of the simple genomic prediction model, respectively. Moreover, there was no obvious pattern in the proportion of improvement (Figure S3).

### 3.6 Case 3: across years for the same type of treatment

In Case3, we evaluated the across-year predictions for each treatment. In all three treatments,  $MT_{RRM}$  models and  $MT_{All}$  models outperformed the simple genomic prediction model (Figures 8A, B). In cross-validations with the 2020 and 2021 datasets, the prediction accuracy of  $MT_{RRM}$  models for Treatments W10 and D10 were on average 42% and 42% higher



**FIGURE 6**

Prediction accuracy of  $MT_{RRM}$  model and  $MT_{All}$  model, and the comparison of  $MT_{RRM}$  model and  $MT_{All}$  model in Case1. (A, B) Prediction accuracy of  $MT_{RRM}$  model and  $MT_{All}$  model within each combination of treatments by years. Error bars represent standard error over 10 replicate cross-validations. A dashed line represents the prediction accuracy of the simple genomic prediction model. (C) The proportion of improvement calculated using the prediction accuracy of  $MT_{RRM}$  model and  $MT_{All}$  model. W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, D: no watering treatment.

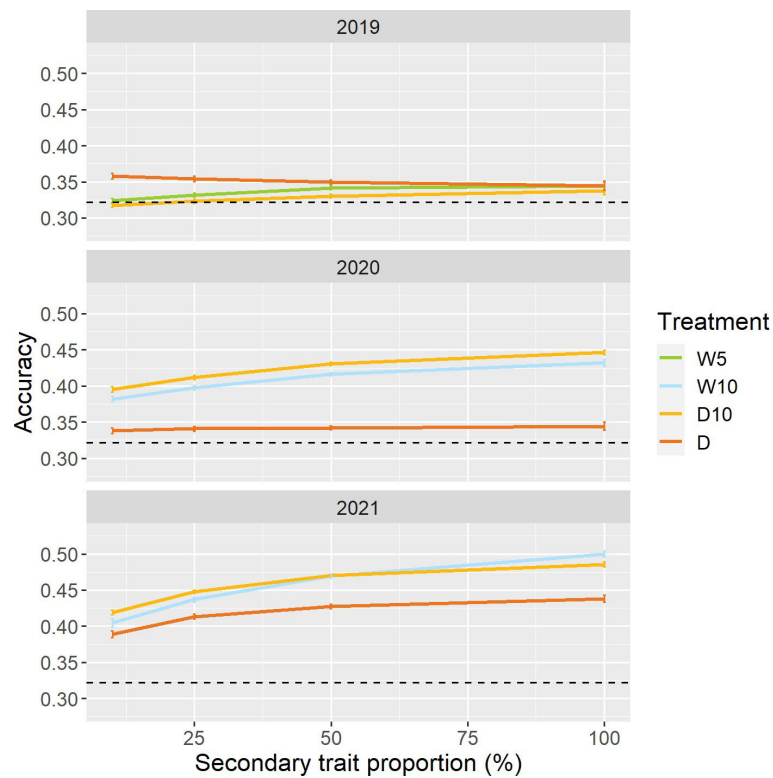


FIGURE 7

Prediction accuracy of  $MT_{RRM}$  model when a specific proportion of genotypes in the training set did not have secondary trait data. Dashed line represents the prediction accuracy of simple genomic prediction model. Error bars represent standard error over 50 replicate cross-validations. W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, D: no watering treatment.

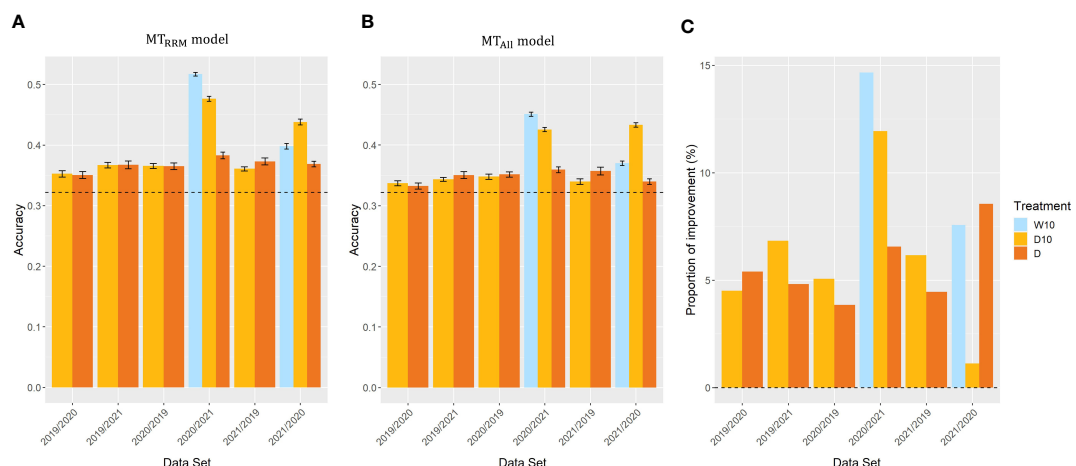


FIGURE 8

Prediction accuracy of  $MT_{RRM}$  model and  $MT_{All}$  model, and the proportion of improvement in Case3. Item names, such as 2019/2020, represent the training year and test year. (A, B) Prediction accuracy of  $MT_{RRM}$  model and  $MT_{All}$  model among years in the same treatment. Dashed line represents the prediction accuracy of simple genomic prediction model. Error bars represent standard error over 10 replicate cross-validations. (C) The proportion of improvement in each combination. W5: watering for 5 d followed by no watering 5 d, W10: watering for 10 d followed by no watering 10 d, D10: no watering for 10 d followed by watering 10 d, D: no watering treatment.

than that of the simple genomic prediction model, respectively. The proportion of improvement in all  $MT_{RRM}$  models was greater than zero (Figure 8C). The proportion of improvement was, on average, 11, 6, and 6% for treatments W10, D10, and D, respectively. This result indicates that the time-series MS data should be modeled using RRM for across-year predictions.

## 4 Discussion

### 4.1 Usefulness of MS data for predicting plant's stability under drought stress

In this study, we compared the prediction accuracy of a simple genomic prediction model and a genomic prediction model using time-series NDVI values as secondary traits to predict the CV (Figures 6–8). In soybeans, a relationship between NDVI and drought tolerance has been reported (Zhou et al., 2020). Except for Case2 in 2019, all genomic prediction models using time-series NDVI values as a secondary trait showed higher prediction accuracy than a simple genomic prediction model (Figures 6–8). These results suggest that time-series MS data are useful for predicting the CV.

In 2020 and 2021, the prediction accuracy of treatments W10 and D10, which changed irrigation treatments, was higher than that of Treatment D, which was the no watering treatment, in the same year (Figures 6, 7). A relationship between plant responses to changes in irrigation and drought tolerance has been reported. In soybeans, slow wilting resulted in drought tolerance (Fletcher et al., 2007; Pathan et al., 2014; Valliyodan et al., 2017; Ye et al., 2020) and a relationship between slow wilting and NDVI values under drought is observed (Zhou et al., 2020). Additionally, the speed of recovery from drought stress is associated with drought tolerance in several crop species (Hayano-Kanashiro et al., 2009; Kränzlein et al., 2021), including soybean (Hossain et al., 2014). However, no study has evaluated the plant's stability under drought stress using time-series NDVI changes caused by changes in irrigation. This result indicates that plant responses to irrigation changes are useful for evaluating plant's stability under drought stress.

In this study, the overall prediction accuracy of CV was lower in 2019 than that in the other years. There are two possible reasons for this result. One possible reason is that NDVI was measured in 2019 (four times) less frequently than in 2020 (six times) and 2021 (seven times). The small number of measurements may not capture the time-series changes in NDVI well. In addition, we collected NDVI values on the first day after the irrigation change in Treatment D10 in 2019 (Figure 3). The second possible reason is the weather conditions. In 2020 and 2021, the soil moisture content increased during the irrigation period and decreased during the non-irrigation period (Figure S1). In contrast, in 2019, treatments W5 and D10 showed the opposite trend of increasing soil moisture content during the last non-irrigation period. In 2019, it rained for 6 out of 10 days from 25 August to 3 September during the period of no irrigation in Treatment D10. The lack of drought stress during this period might have resulted

in a failure to capture the plant response to irrigation change and low prediction accuracy in 2019.

### 4.2 The advantage of using random regression

In Case1, MTRRM model, which modeled time-series NDVI data and used L0 and L1 (genetic random regression coefficients) as secondary traits in the MTM, generally showed a higher prediction accuracy than MTAll model, which treated time-series NDVI data as independent traits and used each day's NDVI value as a secondary trait in the MTM (Figure 6). It was reported that the RRM is superior to the MTM in modeling time-series data (Kranis et al., 2007; Mota et al., 2013). In addition, there are problems with over-parameterization and high computational requirements when treating time-series data as independent traits and constructing MTMs (Misztal et al., 2000; Oh and See, 2008; Moreira et al., 2020). In terms of calculation efficiency, RRM is useful for modeling time-series data.

Campbell et al. (Campbell et al., 2018) built RRM using time-series projected shoot area (PSA) data for 179 rice lines obtained in their own study. Using the built-in RRM (Campbell et al., 2018), Campbell predicted time-series changes in PSA for new 178 rice lines collected in a different year (Campbell et al., 2017). This result suggests that RRM is useful for predictions across years. In our study, MTRRM model is superior to MTAll model in terms of prediction across years in Case3 (Figure 8). MTRRM model can build a model using all-day data of NDVI value in each year, whereas MTAll model can only use dates that are common across all years. This difference in the amount of data may be the reason for the difference in prediction accuracy. When using MTAll model for predictions across years, it is necessary to match the data measurement dates. It was difficult to collect data on a desired date because of weather conditions or equipment problems. At this point, the RRM can be built without considering the measurement date and number of days of data measurement.

### 4.3 Utility for breeding

Environmental stability, also, the stability of phenotypes over environments, is an important target in breeding (Tollenaar and Lee, 2002; Sabaghnia et al., 2008; Mickelbart et al., 2015; Blum, 2018). The CV of phenotypic values over environments can be a good indicator for selecting stable genotypes. To evaluate genotypes' stability across environments, it is necessary to conduct multi-environmental experiments which require high costs and labor (Fan et al., 2007; Karimizadeh et al., 2013; Akter et al., 2015; Lal et al., 2022). In this study, we proposed a method to measure the response to changes in irrigation by remote sensing and model it with RRM at a single location in multiple years. We found that genetic variation captured by the RRM was associated with plant's stability under drought stress, even with only one location in three years. The results of Case1 validation suggest that

genetic variation captured by measuring and modeling plant responses to irrigation in a single environment changes can predict the stability over nine environments (Figure 6). The results of Case2 validation indicate that even when 10% of genotypes in a training set had secondary trait data, MTRRM model was more accurate than a simple genomic prediction model in treatments W10 and D10 in 2020 and 2021 (Figure 7). Therefore, it is possible to reduce the number of genotypes measured for secondary traits in the training set. The results of Case3 validation also suggest that we can predict the plant's stability across years (Figure 8). Once a prediction model is constructed, the stability of unknown genotypes can be predicted based on time-series NDVI data. The results of Case2 and Case3 validations also demonstrate the flexibility of MTRRM model. When the stability of unknown genotypes under drought is evaluated in a single environment, it greatly reduces the time and cost, and thus streamlines breeding schemes. When calculating CVs, experimental designs with repetitions per treatment and year combination are usually used (Mohammadi and Amri, 2008; Das et al., 2010; Küchenmeister et al., 2012; Di Matteo et al., 2016). In this study, priority was given to increasing the number of genotypes over replications. This is because it is known that maximizing the number of genotypes (even with one replication) is the most effective when the number of plots is limited for QTL analysis and genomic prediction (Knapp and Bridges, 1990; Lorenz, 2013). Given the lack of replications and the small size of microplots, the fresh weight is probably influenced by non-genetic factors. However, the prediction accuracy of CV was improved by using information obtained from the methods for measuring and modeling changes in time-series NDVI data. Therefore, the effectiveness of the developed method was confirmed. In order to apply these measuring and modeling methods in breeding selection, it is necessary to associate the information obtained from the developed methods with drought tolerance-related traits and even yield-related traits in expanded field trials.

In this study, we developed a method to measure and model plant responses to irrigation changes and confirmed their association with plant's stability under drought stress. Drone-based HTP allows us to capture time-series plant responses to environmental and irrigation changes (Chen et al., 2014; Dodig et al., 2021), and measuring and modeling plant responses to these environmental changes will provide important insights that have not yet been obtained previously (Arnold et al., 2019; Moreira et al., 2020). This developed method will contribute to the study of abiotic stress and genetic improvement of soybean.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

KS: Methodology, software, formal analysis, data curation, writing—original draft preparation and visualization. YT: investigation, data curation and project administration. KM: methodology and software. HI: conceptualization, investigation, data curation, writing—review and editing, supervision, project administration and funding acquisition. AK: Conceptualization, resources, data curation, and funding acquisition. YO, YY, HirT, HidT, MT, HTs, MN, and TF: Conceptualization, data curation, and funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by JST CREST (<https://www.jst.go.jp/kisoken/crest/en/index.html>) Grant Number JPMJCR16O2, Japan. The funder played no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

## Acknowledgments

We are grateful to the technical staff of Arid Land Research Center, Tottori University, and Izumi Higashida. The content of the manuscript has previously appeared online in a preprint.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1201806/full#supplementary-material>

## References

- Akaike, H. (1974). *A new look at the statistical model identification* (New York: Springer Series in Statistics). doi: 10.1007/978-1-4612-1694-016
- Akter, A., Hasan, M. J., Kulsum, U., Rahman, M., Khatun, M., and Islam, M. (2015). Gge biplot analysis for yield stability in multi-environment trials of promising hybrid rice (*oryza sativa* L.). *Bangladesh Rice J.* 19, 1–8. doi: 10.3329/brj.v19i1.25213
- Arnold, P. A., Kruuk, L. E., and Nicotra, A. B. (2019). How to analyse plant phenotypic plasticity in response to a changing climate. *New Phytologist* 222, 1235–1241. doi: 10.1111/nph.15656
- Ayed, S., Othmani, A., Bouhaouel, I., and da Silva, J. A. (2021). Multi-environment screening of durum wheat genotypes for drought tolerance in changing climatic events. *Agronomy* 11, 1–15. doi: 10.3390/agronomy11050875
- Bi, Y., Kong, W., and Huang, W. (2018). Hyperspectral diagnosis of nitrogen status in arbuscular mycorrhizal inoculated soybean leaves under three drought conditions. *Int. J. Agric. Biol. Engineering* 11, 126–131. doi: 10.25165/ij.ijabe.20181106.4019
- Blum, A. (2018). *Plant breeding for stress environments* (Boca Raton: CRC press).
- Bui, K. T., Naruse, T., Yoshida, H., Toda, Y., Omori, Y., Tsuda, M., et al. (2022). Effects of irrigation on root growth and development of soybean: a 3-year sandy field experiment. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1047563
- Campbell, M. T., Du, Q., Liu, K., Brien, C. J., Berger, B., Zhang, C., et al. (2017). A comprehensive image-based phenomic analysis reveals the complex genetic architecture of shoot growth dynamics in rice (*oryza sativa*). *Plant Genome* 10, plantgenome2016.07.0064. doi: 10.3835/plantgenome2016.07.0064
- Campbell, M., Momen, M., Walia, H., and Morota, G. (2019). Leveraging breeding values obtained from random regression models for genetic inference of longitudinal traits. *Plant Genome* 12, 180075. doi: 10.3835/plantgenome2018.10.0075
- Campbell, M., Walia, H., and Morota, G. (2018). Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. *Plant Direct* 2, 1–11. doi: 10.1002/pld3.80
- Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., et al. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* 26, 4636–4655. doi: 10.1105/tpc.114.129601
- Das, S., Misra, R. C., Patnaik, M. C., and Das, S. R. (2010). Gxe interaction, adaptability and yield stability of mid-early rice genotypes. *Indian J. Agric. Res.* 44, 104–111.
- Di Matteo, J. A., Ferreyra, J. M., Cerrudo, A. A., Echarte, L., and Andrade, F. H. (2016). Yield potential and yield stability of Argentine maize hybrids over 45 years of breeding. *Field Crops Res.* 197, 107–116. doi: 10.1016/j.fcr.2016.07.023
- Dodig, D., Božinović, S., Nikolić, A., Zorić, M., Vančetović, J., Ignjatović-Mičić, D., et al. (2021). Dynamics of maize vegetative growth and drought adaptability using image-based phenotyping under controlled conditions. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.652116
- Du, W., Yu, D., and Fu, S. (2009). Detection of quantitative trait loci for yield and drought tolerance traits in soybean using a recombinant inbred line population. *J. Integr. Plant Biol.* 51, 868–878. doi: 10.1111/j.1744-7909.2009.00855.x
- Fan, X. M., Kang, M. S., Chen, H., Zhang, Y., Tan, J., and Xu, C. (2007). Yield stability of maize hybrids evaluated in multi-environment trials in Yunnan, China. *Agron. J.* 99, 220–228. doi: 10.2134/agronj2006.0144
- Fletcher, A. L., Sinclair, T. R., and Allen, L. H. (2007). Transpiration responses to vapor pressure deficit in well watered 'slow-wilting' and commercial soybean. *Environ. Exp. Botany* 61, 145–151. doi: 10.1016/j.envexpbot.2007.05.004
- Francis, T. R., and Kannenberg, L. W. (1978). Yield stability studies in short-season maize. I. a descriptive method for grouping genotypes. *Can. J. Plant Sci.* 58, 1029–1034. doi: 10.4141/cjps78-157
- Freitas Moreira, F., Rojas de Oliveira, H., Lopez, M. A., Abughali, B. J., Gomes, G., Cherkaoui, K. A., et al. (2021). High-throughput phenotyping and random regression models reveal temporal genetic control of soybean biomass production. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.715983
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2015). *ASReml user guide release 4.1 structural specification* (Hemel Hempstead: VSN international Ltd).
- Hamazaki, K., and Iwata, H. (2020). Rainbow: haplotype-based genome-wide association study using a novel snp-set method. *PLoS Comput. Biol.* 16, 1–17. doi: 10.1371/journal.pcbi.1007663
- Hayano-Kanashiro, C., Calderón-Vásquez, C., Ibarra-Laclette, E., Herrera-Estrella, L., and Simpson, J. (2009). Analysis of gene expression and physiological responses in three Mexican maize landraces under drought stress and recovery irrigation. *PLoS One* 4, e7531. doi: 10.1371/journal.pone.0007531
- Henderson, C. R. (1984). *Applications of linear models in animal breeding* (Guelph: University of Guelph).
- Hossain, M. M., Liu, X., Qi, X., Lam, H. M., and Zhang, J. (2014). Differences between soybean genotypes in physiological response to sequential soil drying and rewetting. *Crop J.* 2, 366–380. doi: 10.1016/j.cj.2014.08.001
- Huisman, A. E., Veerkamp, R. F., and Van Arendonk, J. A. (2002). Genetic parameters for various random regression models to describe the weight data of pigs. *J. Anim. Sci.* 80, 575–582. doi: 10.2527/2002.803575x
- Jorge, J., Vallbè, M., and Soler, J. A. (2019). Detection of irrigation inhomogeneities in an olive grove using the ndre vegetation index obtained from UAV images. *Eur. J. Remote Sensing* 52, 169–177. doi: 10.1080/22797254.2019.1572459
- Kaga, A., Shimizu, T., Watanabe, S., Tsubokura, Y., Katayose, Y., Harada, K., et al. (2011). Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breed. Sci.* 61, 566–592. doi: 10.1270/jsbbs.61.566
- Kajiya-Kanegae, H., Nagasaki, H., Kaga, A., Hirano, K., Ogiso-Tanaka, E., Matsuoka, M., et al. (2021). Whole-genome sequence diversity and association analysis of 198 soybean accessions in mini-core collections. *DNA Res.* 28, 1–13. doi: 10.1093/dnares/dsaa032
- Karimizadeh, R., Mohammadi, M., Sabaghni, N., Mahmoodi, A. A., Roustami, B., Seyyedi, F., et al. (2013). Gge biplot analysis of yield stability in multi-environment trials of lentil genotypes under rainfed condition. *Notulae Scientia Biologica* 5, 256–262. doi: 10.15835/nsb529067
- Knapp, S. J., and Bridges, W. C. (1990). Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* 126, 769–777. doi: 10.1093/genetics/126.3.769
- Kranis, A., Su, G., Sorensen, D., and Woolliams, J. A. (2007). The application of random regression models in the genetic analysis of monthly egg production in turkeys and a comparison with alternative longitudinal models. *Poultry Sci.* 86, 470–475. doi: 10.1093/ps/86.3.470
- Kränzlein, M., Geilfus, C. M., Franzisky, B. L., Zhang, X., Wimmer, M. A., and Zörb, C. (2021). Physiological responses of contrasting maize (*zea mays* L.) hybrids to repeated drought. *J. Plant Growth Regulation* 41, 2708–2718. doi: 10.1007/s00344-021-10468-2
- Küchenmeister, F., Küchenmeister, K., Wrage, N., Kayser, M., and Isselstein, J. (2012). Yield and yield stability in mixtures of productive grassland species: does species number or functional group composition matter? *Grassland Sci.* 58, 94–100. doi: 10.1111/j.1744-697X.2012.00242.x
- Kumar, S., Kumari, J., Bansal, R., Kuri, B. R., Upadhyay, D., Srivastava, A., et al. (2018). Multi-environmental evaluation of wheat genotypes for drought tolerance. *Indian J. Genet. Plant Breeding* 78, 26–35. doi: 10.5958/0975-6906.2018.00004.4
- Kumar, A., Verulkar, S. B., Mandal, N. P., Variar, M., Shukla, V. D., Dwivedi, J. L., et al. (2012). High-yielding, drought-tolerant, stable rice genotypes for the shallow rainfed lowland drought-prone ecosystem. *Field Crops Res.* 133, 37–47. doi: 10.1016/j.fcr.2012.03.007
- Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., and Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (*triticum aestivum* L.) baking quality. *Theor. Appl. Genet.* 131, 2719–2731. doi: 10.1007/s00122-018-3186-3
- Lal, R. K., Chanotiya, C. S., Gupta, P., and Mishra, A. (2022). Influences of traits associations for essential oil yield stability in multi-environment trials of vetiver (*chrysopogon zizanioides* L. roberly). *Biochem. Systematics Ecol.* 103, 104448. doi: 10.1016/j.bse.2022.104448
- Litchfield, K., Thomsen, H., Mitchell, J. S., Sundquist, J., Houlston, R. S., Hemminki, K., et al. (2015). Quantifying the heritability of testicular germ cell tumour using both population-based and genomic approaches. *Sci. Rep.* 5, 1–7. doi: 10.1038/srep13889
- Lorenz, A. J. (2013). Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. *G3: Genes Genomes Genet.* 3, 481–491. doi: 10.1534/g3.112.004911
- Marchetti, C. F., Ugena, L., Humplik, J. F., Polák, M., Čavar Zeljković, S., Podlešáková, K., et al. (2019). A novel image-based screening method to study water-deficit response and recovery of barley populations using canopy dynamics phenotyping and simple metabolite profiling. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01252
- Mickelbart, M. V., Hasegawa, P. M., and Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nat. Rev. Genet.* 16, 237–251. doi: 10.1038/nrg3901
- Misztal, I., Strabel, T., Jamrozik, J., Mäntysaari, E. A., and Meuwissen, T. H. (2000). Strategies for estimating the parameters needed for different test-day models. *J. Dairy Sci.* 83, 1125–1134. doi: 10.3168/jds.S0022-0302(00)74978-2
- Mohammadi, R., and Amri, A. (2008). Comparison of parametric and non-parametric methods for selecting stable and adapted durum wheat genotypes in variable environments. *Euphytica* 159, 419–432. doi: 10.1007/s10681-007-9600-6
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., et al. (2016). A genomic Bayesian multi-trait and multi-environment model. *G3: Genes Genomes Genet.* 6, 2725–2774. doi: 10.1534/g3.116.032359
- Moreira, F. F., Oliveira, H. R., Volenec, J. J., Rainey, K. M., and Brito, L. F. (2020). Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00681

- Mota, R. R., Marques, L. F., Lopes, P. S., da Silva, L. P., Neto, F. R., de Resende, M. D., et al. (2013). Genetic evaluation using multi-trait and random regression models in simmental beef cattle. *Genet. Mol. Res.* 12, 2465–2480. doi: 10.4238/2013.July.24.2
- Mrode, R. A. (2014). *Linear models for the prediction of animal breeding values* (Cambridge: Cabi).
- Oh, S. H., and See, M. T. (2008). Comparison of genetic parameter estimates of total sperm cells of boars between random regression and multiple trait animal models. *Asian-Australasian J. Anim. Sci.* 21, 923–927. doi: 10.5713/ajas.2008.70383
- Oliveira, H. R., Brito, L. F., Lourenco, D. A., Silva, F. F., Jamrozik, J., Schaeffer, L. R., et al. (2019). Invited review: advances and applications of random regression models: from quantitative genetics to genomics. *J. Dairy Sci.* 102, 7664–7683. doi: 10.3168/jds.2019-16265
- Paine, C. E., Marthews, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A., et al. (2012). How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods Ecol. Evol.* 3, 245–256. doi: 10.1111/j.2041-210X.2011.00155.x
- Pathan, S. M., Lee, J. D., Slepner, D. A., Fritschi, F. B., Sharp, R. E., Carter, T. E., et al. (2014). Two soybean plant introductions display slow leaf wilting and reduced yield loss under drought. *J. Agron. Crop Sci.* 200, 231–236. doi: 10.1111/jac.12053
- Pidgeon, J. D., Ober, E. S., Qi, A., Clark, C. J., Royal, A., and Jaggard, K. W. (2006). Using multi-environment sugar beet variety trials to screen for drought tolerance. *Field Crops Res.* 95, 268–279. doi: 10.1016/j.fcr.2005.04.010
- Price, J. C., and Bausch, W. C. (1995). Leaf area index estimation from visible and near-infrared reflectance data. *Remote Sens. Environment* 52, 55–65. doi: 10.1016/0034-4257(94)00111-Y
- Rao, M. R., and Willey, R. W. (1980). Evaluation of yield stability in intercropping: studies on sorghum/pigeonpea. *Exp. Agriculture* 16, 105–116. doi: 10.1017/S0014479700010796
- Ray, D. K., Gerber, J. S., Macdonald, G. K., and West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nat. Commun.* 6, 1–9. doi: 10.1038/ncomms6989
- Sabaghnia, N., Sabaghpour, S. H., and Dehghani, H. (2008). The use of an ammi model and its parameters to analyse yield stability in multi-environment trials. *J. Agric. Sci.* 146, 571–581. doi: 10.1017/S0021859608007831
- Sakurai, K., Toda, Y., Kajiya-Kanegae, H., Ohmori, Y., Yamasaki, Y., Takahashi, H., et al. (2022). Time-series multispectral imaging in soybean for improving biomass and genomic prediction accuracy. *Plant Genome* 15, e20244. doi: 10.1002/tpg2.20244
- Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Sci.* 86, 35–45. doi: 10.1016/S0301-6226(03)00151-9
- Setiyono, T. D., Weiss, A., Specht, J. E., Cassman, K. G., and Dobermann, A. (2008). Leaf area index simulation in soybean grown under near-optimal conditions. *Field Crops Res.* 108, 82–92. doi: 10.1016/j.fcr.2008.03.005
- Specht, J. E., Hume, D. J., and Kumudini, S. V. (1999). Soybean yield potential - a genetic and physiological perspective. *Crop Sci.* 39, 1560–1570. doi: 10.2135/cropsci1999.3961560x
- Sun, S., and Frelich, L. E. (2011). Flowering phenology and height growth pattern are associated with maximum plant height, relative growth rate and stem tissue mass density in herbaceous grassland species. *J. Ecol.* 99, 991–1000. doi: 10.1111/j.1365-2745.2011.01830.x
- Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J.-L., and Sorrells, M. E. (2017). Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome* 10, plantgenome2016.11.0111. doi: 10.3835/plantgenome2016.11.0111
- Tollenaar, M., and Lee, E. A. (2002). Yield potential, yield stability and stress tolerance in maize. *Field Crops Res.* 75, 161–169. doi: 10.1016/S0378-4290(02)00024-2
- Torres, R. O., and Henry, A. (2018). Yield stability of selected rice breeding lines and donors across conditions of mild to moderately severe drought stress. *Field Crops Res.* 220, 37–45. doi: 10.1016/j.fcr.2016.09.011
- Valliyodan, B., Ye, H., Song, L., Murphy, M., Grover Shannon, J., and Nguyen, H. T. (2017). Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans. *J. Exp. Botany* 68, 1835–1849. doi: 10.1093/jxb/erw433
- van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., et al. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci.* 282, 23–39. doi: 10.1016/j.plantsci.2018.06.018
- Wijewardana, C., Alsajri, F. A., Irby, J. T., Krutz, L. J., Golden, B., Henry, W. B., et al. (2019). Physiological assessment of water deficit in soybean using midday leaf water potential and spectral features. *J. Plant Interactions* 14, 533–543. doi: 10.1080/17429145.2019.1662499
- Winterhalter, L., Mistele, B., Jampatong, S., and Schmidhalter, U. (2011). High throughput phenotyping of canopy water mass and canopy temperature in well-watered and drought stressed tropical maize hybrids in the vegetative stage. *Eur. J. Agronomy* 35, 22–32. doi: 10.1016/j.eja.2011.03.004
- Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01111
- Yang, R., Tian, Q., and Xu, S. (2006). Mapping quantitative trait loci for longitudinal traits in line crosses. *Genetics* 173, 2339–2356. doi: 10.1534/genetics.105.054775
- Ye, H., Song, L., Schapaugh, W. T., Ali, M. L., Sinclair, T. R., Riar, M. K., et al. (2020). The importance of slow canopy wilting in drought tolerance in soybean. *J. Exp. Botany* 71, 642–652. doi: 10.1093/jxb/erz150
- Zarate-Valdez, J. L., Metcalf, S., Stewart, W., Ustin, S. L., and Lampinen, B. (2015). Potentials and limits of vegetation indices for lai and apar assessment. *Precis. Agriculture* 16, 161–173. doi: 10.2136/sssaj1977.03615995004100040037x
- Zhou, J., Zhou, J., Ye, H., Ali, M. L., Nguyen, H. T., and Chen, P. (2020). Classification of soybean leaf wilting due to drought stress using uav-based imagery. *Comput. Electron. Agriculture* 175, 105576. doi: 10.1016/j.compag.2020.105576
- Zygielbaum, A. I., Gitelson, A. A., Arkebauer, T. J., and Rundquist, D. C. (2009). Non-destructive detection of water stress and estimation of relative water content in maize. *Geophys. Res. Letters* 36, 2–5. doi: 10.1029/2009GL038906