Phenomic prediction in drought-stressed faba bean across spectral, structural, and fused canopy predictors

Lennart Scheer

lennart.scheer@agrar.uni-giessen.de

Department of Plant Breeding, Justus Liebig University Giessen https://orcid.org/0009-0002-5463-4407

Lennard Roscher-Ehrig

Department of Plant Breeding, Justus Liebig University Giessen

Benjamin Wittkop

Department of Plant Breeding, Justus Liebig University Giessen

Jason Wenzig

Department of Plant Breeding, Justus Liebig University Giessen

Andreas Stahl

Institute for Resistance Research and Stress Tolerance, Julius Kühn-Institute

Olaf Sass

NPZ Hans-Georg Lembke KG

Gregor Welna

NPZ Hans-Georg Lembke KG

Hanna Tietgen

NPZ Innovation GmbH (NPZi)

Rod J. Snowdon

Department of Plant Breeding, Justus Liebig University Giessen

Research Article

Keywords: high-throughput, multispectral, 3-dimensional, phenotyping, faba bean, phenomic selection

Posted Date: May 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9714994/v1

Additional Declarations: The authors declare no competing interests.

Phenomic prediction in drought-stressed faba bean across spectral, structural, and fused canopy predictors

Lennart Scheer[1], Lennard Roscher-Ehrig[1], Benjamin Wittkop[1], Jason Wenzig[1], Andreas Stahl[2], Olaf Sass[3], Gregor Welna[3], Hanna Tietgen[4] and Rod J. Snowdon[1]

[1] Department of Plant Breeding, Justus Liebig University Giessen, Germany
[2] Institute for Resistance Research and Stress Tolerance, Julius Kühn-Institute (JKI), Quedlinburg, Germany
[3] NPZ Hans-Georg Lembke KG, Holtsee, Germany
[4] NPZ Innovation GmbH (NPZi), Holtsee, Germany

E-mail of corresponding author: Lennart.scheer@agrar.uni-giessen.de

Abstract

Background Faba bean is an important grain legume in temperate cropping systems because it provides protein-rich seed and contributes biological nitrogen fixation. However, its productivity is highly sensitive to drought, and breeding for improved drought performance is constrained by complex genotype by environment interactions and the difficulty of measuring relevant traits at scale. This study evaluated whether scanner-derived vegetation indices (VI), 3D canopy traits, and their combination can predict key agronomic and physiological traits in drought-stressed faba bean, and how predictive ability changes when information is used from single dates or cumulatively across the season.

Results Predictive performance was strongly trait dependent and varied with predictor set and temporal strategy. Combined VI + 3D predictors generally produced the highest and most consistent predictive ability for major traits. Total grain yield reached 0.75 under cumulative VI + 3D prediction at 93 days after sowing (DAS 93), cumulative water uptake peaked at 0.80 at DAS 97, and total straw biomass reached 0.66 at DAS 104. In contrast, some component traits were predicted equally well or better by 3D information alone, including grain number with 0.70 and pod number with 0.55 under cumulative 3D prediction. Useful prediction windows also differed among traits, with broad late-season windows for major agronomic traits but narrower, more stage-specific windows for productive tillers, thousand kernel weight, and water-use efficiency.

Conclusion Phenomic prediction under drought in faba bean was strongly shaped by trait type, predictor composition, and temporal design. Combined VI + 3D predictors were most effective for integrative traits, whereas several component traits were predicted equally well or better by 3D information alone. These findings highlight the potential of scanner-based multisensor phenotyping to support drought-related selection in faba bean breeding.

Keywords: high-throughput, multispectral, 3-dimensional, phenotyping, faba bean, phenomic selection

Introduction

Faba bean (*Vicia faba* L.) is a major cool-season grain legume in temperate and Mediterranean cropping systems. It contributes substantially to protein supply for food and feed and provides important ecosystem services through biological nitrogen fixation and beneficial effects on soil properties (Jensen et al. 2010; Karkanis et al. 2018). Despite these advantages, the cropped area of faba bean remains limited compared with cereals, partly because of its high yield variability and its sensitivity to abiotic stresses, particularly drought (Jensen et al. 2010; Khan et al. 2010; Karkanis et al. 2018). Water deficits during flowering and early pod filling can markedly reduce pod set, seed number, and seed weight, whereas stress at earlier developmental stages restricts canopy expansion and radiation interception (Mwanamwenge et al. 1999). As heat and drought events are expected to become more frequent under ongoing climate change, improving performance and stability under water-limited conditions has become a major objective in faba bean breeding (Khan et al. 2010; Papastylianou et al. 2021; Heino et al. 2023).

Breeding for drought adaptation in faba bean is complicated by the physiological and developmental complexity of the trait. Plant responses to soil water deficit involve changes in transpiration, leaf area development, canopy structure, and pigment composition, which interact with phenology and plant architecture (Khan et al. 2010; Muktadir et al. 2020). Conventional field selection based on final yield under stress and non-stress conditions captures the integrated outcome of these processes, but provides limited insight into the dynamics underlying stress adaptation. More process-related traits, including transpiration efficiency, stay-green behavior, canopy temperature, or chlorophyll-related parameters, may offer a more mechanistic perspective, yet their measurement is often laborious and difficult to implement at the scale required in breeding populations (Khan et al. 2010; Muktadir et al. 2020; Ninomiya 2022).

High-throughput phenotyping offers a promising route to address this limitation. Ground-based scanning systems that combine multispectral sensing with three-dimensional canopy reconstruction enable repeated, non-destructive characterization of crop development over time (Ninomiya 2022; Xu and Li 2022). Vegetation indices such as normalized difference vegetation index (NDVI), green leaf index (GLI), normalized pigment chlorophyll index (NPCI), or plant senescence reflectance index (PSRI) capture aspects of canopy greenness, pigment composition, and biomass-related variation (Tayade et al. 2022), while color-space metrics such as hue or lightness describe temporal changes in canopy appearance (Tripodi et al. 2024). In parallel, 3D structural traits provide information on canopy architecture, growth, and spatial development (Ninomiya 2022; Tripodi et al. 2024). Under drought, repeated measurements of these features can reveal genotype-specific differences in growth trajectories, senescence dynamics, and canopy responses to water limitation (Tripodi et al. 2024; Adak et al. 2023b).

Phenomic prediction extends this idea by using high-dimensional phenotypic data to predict complex target traits. The concept is closely related to phenomic selection, which was formalized by (Rincent et al. 2018) using near-infrared spectroscopy (NIRS) as a low-cost alternative to marker-based prediction in wheat and poplar. Since then, phenomic prediction has been explored in different types of predictors, including NIRS-based approaches and image-derived time-series data in multiple major crops like wheat, maize, and rapeseed as well as legume crops like soy bean (Rincent et al. 2018; Weiß et al. 2022; Roth et al. 2023; Adak et al. 2023b; Roscher-Ehrig et al. 2024; Braun et al. 2025; Parmley et al. 2019; Zhu et al. 2021). While most studies used NIRS due to its rapid and inexpensive acquisition, more recent work has increasingly incorporated RGB, multispectral, and hyperspectral traits, vegetation indices, and structural descriptors derived from digital surface models or point clouds (Ninomiya 2022; Roth et al. 2021; Adak et al.

2023b). These predictor classes capture partly distinct biological information: spectral and color-related traits are more closely linked to canopy greenness, pigment status, and senescence, whereas 3D traits reflect canopy height, volume, architecture, and growth dynamics. Their combination may therefore be particularly informative under drought, where physiological and structural responses can diverge over time.

In faba bean, UAV-based studies have already demonstrated the potential of RGB, multispectral, thermal, and fused sensor data for predicting traits such as plant height, biomass, chlorophyll status, harvest index, and yield (Ji et al. 2022; Cui et al. 2023; Ji et al. 2024; Mohammadi et al. 2024). However, compared with cereals and with the broader phenomic selection literature, explicit phenomic prediction in faba bean based on repeated canopy time series remains limited, especially under drought stress and in studies directly comparing spectral predictors, 3D structural predictors, and their combination for end-of-season trait prediction.

A key question is how these phenotypic time series can be translated into robust prediction of agronomic and physiological target traits. Beyond the choice of prediction model, three aspects are particularly relevant. First, it remains unclear whether multispectral and RGB-derived canopy traits, 3D structural traits, or their combination provide the strongest predictive signal. Second, it is uncertain whether prediction should rely on information from individual measurement dates or benefit from cumulative integration across the season. Third, the extent to which dense temporal sampling is required, as opposed to a smaller number of strategically selected dates, remains insufficiently understood. Addressing these questions is essential not only for improving phenomic prediction itself, but also for identifying efficient phenotyping strategies that balance predictive performance with practical feasibility (Adak et al. 2023b; Chatterjee et al. 2023).

The present study aimed to provide a basis for more efficient scanner-based phenotyping and drought-related selection in faba bean. The objectives were to quantify the relative predictive value of spectral and structural canopy information, assess the benefit of combining both data domains, determine the value of cumulative temporal integration, and identify the earliest measurement windows at which reliable prediction of harvest traits becomes possible. For this purpose, the present study evaluated a diverse set of faba bean genotypes grown in containers under controlled drought conditions and monitored throughout the season with a multispectral 3D scanning platform. From repeated scans, canopy traits derived from RGB and multispectral information, including color-space metrics and spectral indices, as well as 3D structural traits, were extracted and used to predict agronomic traits measured at harvest. For each day after sowing included in the analysis, six prediction scenarios were compared, defined by three predictor sets and two temporal strategies: RGB and spectral traits only, 3D structural traits only, and combined spectral plus structural traits, each analyzed either as single-day predictors or as cumulative predictor sets including all scans up to the respective date. Prediction performance was then compared across linear mixed models, kernel-based approaches, and machine-learning algorithms.

Material and Methods

Experimental Setup

The high-throughput phenotyping platform Droughtspotter XXL (**Figure 1**; Phenospex, Heerlen, the Netherlands) at the Rauischholzhausen field station of Justus Liebig University (50°45'40.05"N, 8°52'45.96"E) comprises 240 large containers arranged in six rows with 40 containers each. The experiment was conducted in 120 L containers designed to provide a rootable soil profile of 90 cm. To create conditions resembling those of an agronomic field setting, each container was filled with approximately 168 kg of a clay and sand substrate. Throughout the experimental period, every container was positioned on a separate high precision gravimetric balance, enabling continuous monitoring of container weight at 5 min intervals with a measurement resolution of ±20 g. Water supply was regulated individually for each container by an automated irrigation system, which operated on the basis of the continuously recorded weight data. The resulting gravimetric data were used to assess container specific water status and to establish predefined drought stress treatments. Irrigation was carried out once per day at midnight, thereby avoiding the main period of transpiration.

Automated canopy phenotyping was performed using two PhenoSpex PlantEye F600 sensors operated in a dual-scan configuration. This setup was implemented to reduce occlusion and shading effects caused by overlapping plant material and thereby improve the completeness of canopy reconstruction. The two sensors were positioned at a fixed angle of 145° relative to each other and mounted on a vertical lifting device. Scanning was fully automated, with each container measured once per day at a constant time. Identically positioned barcodes served as spatial reference markers, enabling both the alignment and merging of the two scan perspectives into a single 3D point cloud and the definition of a standardized virtual bounding box for each container. This virtual box defined the spatial area considered for scan evaluation, and only points located within this fixed 3D volume were retained for further analysis. Each scan generated a high-resolution point cloud in which every point was assigned x, y, and z coordinates together with multispectral reflectance information in the red, green, blue, and near-infrared channels. The maximum measurable canopy height was 210 cm above the container.

Fig. 1 High-throughput phenotyping platform “Droughtspotter XXL” comprising 240 containers, each with a rootable volume of 120 liters, housed under a foil greenhouse providing semi-controlled environmental conditions (**A**). Automated phenotyping with a multispectral 3D scanner enabled high-resolution, non-destructive monitoring of plant growth and canopy development throughout the experiment (**B**).

Genotype Set

The experiment used a panel of 80 genetically diverse faba bean (*Vicia faba* L.) genotypes. This plant material comprised both registered cultivars and advanced breeding lines, which were provided by Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, Hohenlieth, Germany.

Experimental Design

Each genotype was evaluated in three replications within a randomized complete block design (RCBD). Per container, 18 seeds were sown and subsequently thinned to nine plants, which were arranged in a 3 × 3 grid covering 0.16 m^2, corresponding to a field equivalent planting density of 56 plants m^{-2}. Before sowing, all genotypes were inoculated with *Rhizobium fabae* using the commercial inoculant Rhizofix RF-20 (Feldsaaten Freudenberger GmbH & Co. KG, Krefeld, Germany). Soil nutrient status was determined prior to the start of the experiment and fertilization followed the recommendations of the Hessian state advisory service. Nutrient supply was split into two applications: the first, applied 19 days after sowing, contained ammonium molybdate, manganese sulfate, zinc sulfate, copper sulfate, boric acid, potassium dihydrogen sulfate and magnesium sulfate, while the second, applied 35 days after sowing, provided potassium dihydrogen sulfate and potassium chloride. No mineral nitrogen fertilizer was applied. Aphid infestations were controlled when necessary, using the commercial pyrethroid insecticide Karate Zeon (Syngenta Agro GmbH, Frankfurt am Main, Germany) at the recommended label rate. To minimize lodging, all plots were stabilized using a twine support system.

Phenotypic data

Phenotypic data comprised end season agronomic traits and time resolved canopy traits derived from vegetation indices, color metrics, and 3D reconstructions. Agronomic traits were measured destructively at the end of the experiment and included grain yield, straw biomass, grain number, pod number, mean internode length, water use efficiency, cumulative water uptake, number of productive tillers, and thousand kernel weight. The canopy predictor set comprised vegetation indices and color traits, namely green leaf index (GLI), normalized difference vegetation index (NDVI), normalized pigment chlorophyll index (NPCI), plant senescence reflectance index (PSRI), hue, lightness, and saturation, as well as 3D-derived structural traits, including averaged plant height, maximum plant height, digital biomass, projected leaf area, 3D leaf area, voxel volume, canopy light penetration depth, surface angle, convex hull area coverage, convex hull area, convex hull circumference, convex hull maximum width, and convex hull aspect ratio **(Supplementary Table S1)**. For each agronomic trait, adjusted genotype means were obtained from a linear mixed model of the form

$$y_{ijklm} = \mu + G_i + r_j + c_k + b_l + w_m + e_{ijklm}, \qquad \text{(Equation 1)}$$

where y_{ijklm} is the observed plot mean of genotype $i$in row j, column k, block l, and replicate m, μ is the overall intercept, $g_i$is the genotypic effect, r_j, c_k, b_l, and w_m are the effects of row, column, block, and replicate, and e_{ijklm} is the residual error. Genotype was fitted as a fixed effect while all other effects were fitted as random effects to obtain adjusted genotype means as best linear unbiased estimates (BLUEs) (Piepho et al. 2012).

Because the phenomic prediction framework relied on time specific information, vegetation indices and 3D traits were additionally adjusted separately for each imaging date. For this purpose, the data was split by date, and, for each trait by date combination, the mixed model

$$y_{ijklm}^{(d)} = \mu^{(d)} + G_i^{(d)} + r_j^{(d)} + c_k^{(d)} + b_l^{(d)} + w_m^{(d)} + e_{ijklm}^{(d)} \qquad \text{(Equation 2)}$$

was fitted to the subset of observations recorded on date d. The model structure was identical to that used for the agronomic trait model. These date specific adjusted genotype means were then used as phenomic predictors in the downstream prediction analyses (Piepho et al. 2012; Roth et al. 2021).

Model training and evaluation

Prediction models were fitted separately for each target trait, each type of predictor (i.e. vegetation indices and color traits (VI), 3D-derived structural traits (3D), and their combination), each day after sowing (DAS), and each model type. Two temporal scenarios were considered. In the single day scenario, only predictors from the focal DAS were used. In the cumulative scenario, predictors from all DAS from the first included day up to the focal DAS were combined. For a given trait and DAS, predictive ability was assessed using repeated random training test splits at the genotype level. In each repetition, 80% of genotypes were randomly assigned to the training set and the remaining 20% were held out as an independent test set (Adak et al. 2023b; Rincent et al. 2018; Arlot and Celisse 2010).

For mixed models and Bayesian regression models implemented in Bayesian Generalized Linear Regression, phenotypic values of test set genotypes were set to missing during model fitting so that parameter estimation was based exclusively on the training genotypes. For random forest, extreme gradient boosting, and support vector regression, only the training genotypes were used for model fitting. Predictions were then generated for the genotypes in the test set (Pérez and los Campos 2014; Covarrubias-Pazaran 2016).

Predictive performance was quantified as the Pearson correlation between observed adjusted means and predicted values in the test set,

$$r = cor(y_{test}, \hat{y}_{test}) \qquad \text{(Equation 3)}$$

This measure was chosen because it is scale invariant and directly comparable across traits, scenarios, and model classes. To obtain stable estimates of predictive performance, the random training test split and model fitting procedure were repeated 200 times for each combination of trait, DAS, predictor and model. In each repetition, a new random partition of genotypes into training and test sets was drawn. For each combination, the resulting 200 correlation values were summarized by their median and standard deviation. The median correlation reflects typical predictive ability across random splits, whereas the standard deviation characterizes the variability among splits (Krause et al. 2019; Rincent et al. 2018; Arlot and Celisse 2010).

Statistical and machine learning models

General prediction framework

All prediction analyses were based on scanner derived vegetation indices, color metrics, and/or 3D canopy traits as predictors, and on adjusted genotypic means of the target traits as response variables. For each focal DAS, a wide genotype by feature matrix was constructed in which rows represented genotypes and columns represented scanner derived predictors. In the cumulative scenario, predictors from all included DAS up to the time point of interest were combined, with each predictor retaining its DAS specific identity. In the single day scenario, only predictors from the day of interest were included. Before model fitting, all predictor variables were centered and scaled to zero mean and unit variance within each prediction scenario.

For a given trait and prediction scenario, let n denote the number of genotypes and p the number of phenomic predictors. Let $y \in \mathbb{R}^n$ denote for the vector of adjusted genotypic means of the target trait and $W \in \mathbb{R}^{n \times p}$ for the corresponding matrix of standardized phenomic predictors. All prediction models followed the general form

$$y = 1_n\mu + f(W) + \varepsilon \quad \text{(Equation 4)}$$

with

$$\varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$$

where 1_n is a vector of ones of length n, μ is the overall intercept, $f(W)$ is a model specific function of the phenomic predictors, and ε is the residual term. Depending on the model class, $f(W)$ was represented by a phenomic relationship kernel, a shrinkage regression on individual predictors, or a nonlinear machine learning function (VanRaden 2008; Pérez and los Campos 2014; Gianola and van Kaam 2008).

VIBLUP

The VIBLUP model is conceptually analogous to GBLUP but uses a phenomic relationship matrix derived from standardized predictor values instead of a genomic relationship matrix. Let W denote the $n \times p$ matrix of standardized phenomic predictors for the current trait, DAS, and scenario. The phenomic relationship matrix was defined as

$$G_P = \frac{1}{p} W W^\top, \quad \text{(Equation 5)}$$

The prediction model was

$$y = \mathbf{1}_n\mu + u + \varepsilon, \qquad \text{(Equation 6)}$$

with random genotypic effects

$$u \sim N(0, \sigma_u^2 G_{\text{VI}}), \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n),$$

where u is the vector of random genotype effects structured by the phenomic relationship matrix and σ_u^2 and σ_ε^2 are the corresponding variance components. The model was fitted by restricted maximum likelihood using sommer, supplying G_{VI} as covariance structure for the random genotype effect (VanRaden 2008; Rincent et al. 2018; Krause et al. 2019; Covarrubias-Pazaran 2016).

RKHS regression

To allow for nonlinear relationships between phenomic predictors and target traits, a reproducing kernel Hilbert space model was implemented. Let w_i denote the standardized predictor vector of genotype i. Pairwise squared Euclidean distances between genotypes were computed and scaled by the number of predictors p

$$D_{ij} = \frac{\|wi - wj\|^2}{p} \qquad \text{(Equation 7)}$$

Based on this scaled distance matrix, three Gaussian kernels were defined as

$$K_m(i,j) = \exp(-h_m D_{ij}) \qquad \text{(Equation 8)}$$

with

$$h_m \in \{0.5, 1.0, 2.5\}$$

We used a small set of fixed scaling factors h_m to capture short, medium and long-range similarities in phenomic space and combined the corresponding kernels in an additive multi kernel model. The RKHS model was

$$y = \mathbf{1}_n\mu + \sum_{m=1}^{M} g_m + \varepsilon, \qquad \text{(Equation 9)}$$

with random kernel specific genetic effects

$$g_m \sim N\left(0, \sigma_{g_m}^2 K_m\right), \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n),$$

where $\sigma_{g_m}^2$ denotes the variance component associated with kernel K_m. Model fitting and prediction were carried out in BGLR using a Bayesian Gibbs sampling scheme with 5000 iterations and a burn in of 1000 iterations. Predictions for test genotypes were obtained from posterior mean fitted values (Gianola and van Kaam 2008; los Campos et al. 2010; Pérez and los Campos 2014; Krause et al. 2019).

Bayesian LASSO

In the Bayesian LASSO (BL) model we modeled the effect of each individual phenomic predictor explicitly. In this case,

$$f(W) = W\beta, \quad \text{(Equation 10)}$$

and the model can be written as

$$y = \mathbf{1}_n\mu + W\beta + \varepsilon, \qquad \text{(Equation 11)}$$

with $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$

and regression coefficients $\beta = (\beta_1, \ldots, \beta_p)^\top$. (Equation 12)

The Bayesian LASSO prior imposes a double exponential (Laplace) distribution on each coefficient,

$$\beta_j \sim \mathrm{DE}(0, \lambda), j = 1, \ldots, p,$$

which induces strong shrinkage of small effects while allowing a subset of predictors to retain larger effects. The global shrinkage parameter was estimated from the data. The model was implemented in BGLR using the same Gibbs sampling settings as for RKHS, and predictions were obtained as posterior mean fitted values for the test genotypes (Park and Casella 2008; Pérez and los Campos 2014).

Random forest

Random forest regression was used as a non-parametric tree based method to capture nonlinear effects and interactions among predictors. For each trait, DAS, and scenario, a random forest was

fitted to the training genotypes. An ensemble of B regression trees was grown on bootstrap samples of the training set, with random subsets of predictors considered at each split. Let $\hat{y}^{(b)}(i)$ denote the prediction for genotype i from tree b. The random forest prediction is given by

$$\hat{y}(i) = \frac{1}{B}\sum_{b=1}^{B}\hat{y}^{(b)}(i), \qquad \text{(Equation 13)}$$

with $$B = 500.$$

The main hyperparameters, namely the number of predictors sampled at each split ($mtry$) and the minimum node size, were tuned by Bayesian optimization on the training set. The search space for $mtry$ ranged from $\max(1, \lfloor p/10 \rfloor)$ to $\lfloor p/3 \rfloor$, and the minimum node size ranged from 3 to 50. Bayesian optimization was based on a Gaussian process surrogate model with expected improvement as acquisition criterion, using an initial Latin hypercube design followed by iterative infill steps. The best hyperparameter combination according to out of bag root mean squared error in the training set was then used to fit the final model and generate predictions for the test genotypes (Breiman 2001; Wright and Ziegler 2017; Bischl et al. 2017; Jones et al. 1998; McKay et al. 1979).

Extreme gradient boosting

Extreme gradient boosting (XGB) was used as a second tree-based method that constructs an additive ensemble of shallow regression trees by sequentially minimizing a regularized loss function. For genotype i with predictor vector w_i, the XGB model can be written as

$$\hat{y}(i) = \sum_{k=1}^{K} f_k(w_i), \quad \text{(Equation 14)}$$

where each f_k denotes a regression tree and K is the number of boosting rounds. Model fitting aimed to minimize the regularized objective function

$$L = \sum_{i=1}^{n}(y_i - \hat{y}(i))^2 + \sum_{k=1}^{K}\Omega(f_k), \text{ (Equation 15)}$$

where $\Omega(f_k)$ penalizes tree complexity. For most traits, key hyperparameters were tuned by Bayesian optimization on the training set, including the learning rate η with a search range from 0.01 to 0.9, maximum tree depth from 2 to 20, minimum child weight from 1 up to approximately one third of the training sample size, subsampling rate from 0.1 to 1, column subsampling rate from 0.1 to 1, and the tree specific regularization parameter γ from 0.01 to 5. For each candidate setting, up to 7000 boosting rounds were fitted with early stopping if cross validated error did not improve for 25 rounds. For water use efficiency, a moderately flexible fixed configuration was used with $\eta = 0.1$, maximum depth $= 4$, minimum child weight $= 1$, subsampling $= 0.8$, column subsampling $= 0.8$, $\gamma = 0$, and 500 boosting rounds. All XGB models used squared error loss and two computational threads. The final model was then used to predict the test genotypes (Chen et al. 2016; Jones et al. 1998; Bischl et al. 2023).

Support vector regression

Support vector regression (SVR) with a radial basis function kernel was implemented as an additional nonlinear prediction model. For genotype i with predictor vector w_i, the prediction can be expressed as

$$\hat{y}(i) = \sum_{j=1}^{n_{\text{train}}} \alpha_j K(w_i, w_j) + b, \qquad \text{(Equation 16)}$$

where the sum runs over training genotypes, $K(\cdot,\cdot)$ is the radial basis function kernel,

$$K(w_i, w_j) = \exp\left(-\gamma \parallel w_i - w_j \parallel^2\right), \quad \text{(Equation 17)}$$

α_j are the coefficients associated with support vectors, and b is the bias term. The regularization parameter C and the epsilon insensitive loss parameter ε were tuned by Bayesian optimization on the training set, with search ranges $C \in [10^{-3}, 2^{10}]$ and $\varepsilon \in [0,0.5]$. The best configuration according to internal cross validation error was used to fit the final SVR model and generate predictions for the test genotypes (Drucker et al. 1997; Smola and Schölkopf 2004; Karatzoglou et al. 2004; Jones et al. 1998; Bischl et al. 2023).

Data Analysis

Data analysis was performed in R (version 4.1.2). Data cleaning, filtering, and reshaping of time series data were conducted using dplyr (Wickham et al. 2022), tidyr (Wickham et al. 2019), stringr (Wickham 2019), and purrr (Henry and Wickham 2022). Mixed models used to derive adjusted means and phenomic relationship-based predictions were fitted with sommer (Covarrubias-Pazaran 2016). Bayesian regression models, including RKHS regression and Bayesian LASSO, were implemented with BGLR (Pérez and los Campos 2014). Random forest regression was carried out with ranger (Wright and Ziegler 2017), gradient boosting with xgboost (Chen et al. 2022), and support vector regression with kernlab (Karatzoglou et al. 2004). Hyperparameter optimization for machine learning models was based on Bayesian optimization using mlrMBO (Bischl et al. 2017) together with smoof (Bossek 2017), ParamHelpers (Bischl et al. 2022), lhs (Carnell 2022), DiceKriging (Roustant et al. 2012), and mlr (Bischl et al. 2016). Parallel computation across traits was implemented with foreach (Microsoft Coorperation and Weston 2022), doParallel (Microsoft Coorperation and Weston 2022), and parallel (R Core Team 2021), while the number of computational threads per model was explicitly restricted to control computational load.

Results

Temporal dynamics of prediction accuracy across predictor sets

Prediction accuracy varied strongly among traits, predictor sets, and time points (**Figure 2**). Across the major traits, the highest values were generally achieved when spectral and structural information were combined, particularly under cumulative prediction. Total grain yield reached a maximum median prediction accuracy of 0.75 for cumulative VI + 3D at DAS 93, while cumulative water uptake showed the strongest overall performance in the dataset, peaking at 0.80 for cumulative VI + 3D at DAS 97. A similar tendency was observed for total straw biomass, for which cumulative VI + 3D yielded the highest value at 0.66 at DAS 104. These patterns indicate that the joint use of VI and 3D information was especially advantageous for complex biomass- and yield-related traits and that cumulative predictor integration often increased predictive performance relative to single-day prediction.

For several component traits, however, the strongest performance was obtained with 3D predictors alone. Grain number reached its highest median accuracy under cumulative 3D prediction at 0.70 at DAS 108, whereas pod number was best predicted by cumulative 3D at 0.55 at DAS 63. Mean internode length showed comparably high values for both 3D and combined predictors, with the overall maximum observed for single day VI + 3D at 0.67 at DAS 82. In contrast, productive tillers, thousand kernel weight, and WUE were predicted with only moderate accuracy overall. Productive tillers peaked at 0.38 for single-day VI + 3D at DAS 109, thousand kernel weight at 0.43 for single-day VI + 3D at DAS 91, and WUE at 0.44 for single-day 3D at DAS 67. Taken together, these results show that both the optimal prediction window and the most informative predictor set were clearly trait specific, although combined VI + 3D predictors provided the most consistent advantage for the main agronomic target traits.

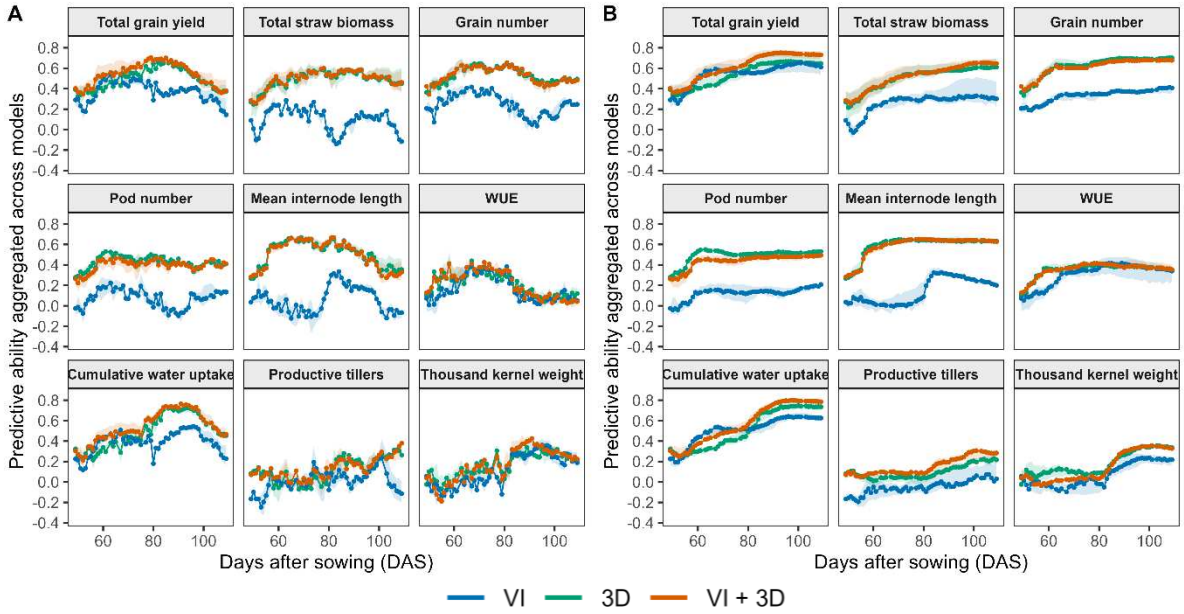


Figure 2 Temporal dynamics of prediction accuracy across days after sowing for nine agronomic traits under single-day (A) and cumulative (B) prediction. Lines represent the median prediction accuracy aggregated across the six prediction models VIBLUP, RKHS, BL, RF, XGB, and SVM for each day after sowing. Predictor sets are shown for vegetation indices (VI), three-dimensional traits (3D), and their combination (VI + 3D).

The timing of near-peak prediction performance, defined as the first time point reaching 95% of peak prediction accuracy, differed markedly among traits and predictor sets. (**Figure 3**). For the major traits, cumulative prediction not only achieved the highest peak accuracies but also provided comparatively broad temporal windows in which prediction performance remained close to the maximum. Total grain yield under cumulative VI + 3D reached 95% of its peak already at DAS 84 and attained its maximum at DAS 93, while the corresponding window extended until DAS 109 and covered 26 days. A similarly broad and late prediction window was observed for cumulative water uptake, where cumulative VI + 3D reached 95% of peak performance at DAS 90, peaked at DAS 97, and retained high performance until DAS 109, corresponding to a plateau width of 20 days. Total straw biomass followed the same general pattern, although the useful prediction window began somewhat later, with cumulative VI + 3D reaching 95% of peak accuracy at DAS 96 and peaking at DAS 104. These results indicate that for the main biomass- and yield-related traits, cumulative prediction often enabled robust forecasting across an extended late-season period

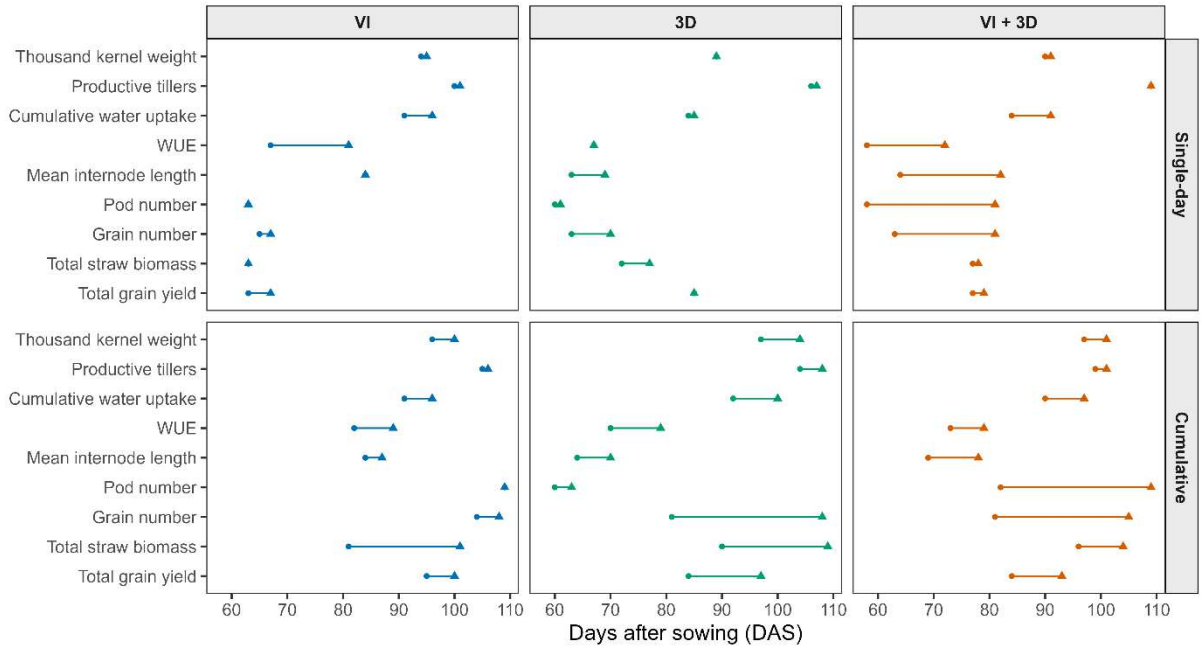


Figure 3 Timing of earliest near-peak prediction and peak prediction across traits and predictor sets. Circles indicate the earliest day after sowing (DAS) at which predictive ability reached at least 95% of the peak, and triangles indicate the DAS of maximum predictive ability. Results are shown for single-day and cumulative prediction using VI, 3D, and VI + 3D predictors. Horizontal segments connect the earliest useful prediction day to the peak day.

rather than at a single isolated time point.

For several component traits, the temporal pattern was more strongly predictor-set specific. Grain number showed a particularly broad window under cumulative 3D prediction, with 95% of peak accuracy already reached at DAS 81 and the maximum observed only at DAS 108, resulting in a 29-day plateau. Pod number displayed the most extended window in the entire dataset, with cumulative 3D reaching 95% of peak performance at DAS 60 and maintaining this level until DAS 109, corresponding to a plateau width of 50 days, even though the peak itself occurred much earlier at DAS 63. Mean internode length likewise showed a long high-performance phase, especially under cumulative 3D and cumulative VI + 3D, where prediction remained near peak across 46 and 41 days, respectively. In contrast, productive tillers, thousand kernel weight, and WUE were characterized by distinctly narrower useful windows. Productive tillers under single-day VI + 3D peaked only at DAS 109 and showed a plateau width of just 1 day, while WUE under single-day 3D also reached both 95% of peak and its absolute maximum at DAS 67, again with a

plateau width of 1 day. Thousand kernel weight was similarly constrained, with the best-performing setup, single-day VI + 3D, reaching 95% of peak accuracy at DAS 90 and peaking at DAS 91, resulting in a plateau width of only 2 days. Taken together, these results show that some traits can be predicted reliably across broad developmental windows, whereas others require a much narrower and more precisely targeted prediction period.

Prediction performance under single-day and cumulative approaches

Within the combined VI + 3D predictor set, prediction performance differed among models, although the magnitude of these differences depended strongly on the trait and prediction mode (**Figure 4 A, B, VI and 3D can be found in Supplementary Figure S1 and Figure S2**). For the major agronomic target traits, Bayesian LASSO and VIBLUP generally showed the strongest and most stable performance across the season. Under cumulative prediction, total grain yield reached the highest across-season median accuracy with both, VIBLUP and Bayesian LASSO at 0.75. A similar pattern was observed for total straw biomass, where Bayesian LASSO achieved a median of 0.66 and VIBLUP 0.65, and for cumulative water uptake, where both models performed best with a median of 0.64. Grain number and mean internode length followed the same general trend, with Bayesian LASSO showing the highest median accuracies under both cumulative and single-day prediction. These results indicate that for the principal integrative and biomass-related traits, the strongest predictive performance within the fused VI + 3D setup was consistently obtained by the linear shrinkage-based and relationship-based models.

For several secondary traits, the model ranking was less uniform and overall accuracies remained substantially lower. Pod number was again best predicted by Bayesian LASSO, with median accuracies of 0.51 under cumulative and 0.45 under single-day prediction. In contrast, productive tillers and WUE showed their highest cumulative medians under RKHS, reaching 0.12 and 0.43, respectively, while thousand kernel weight under cumulative prediction was best captured by SVM with a median of 0.12. Under single-day prediction, however, Bayesian LASSO was the best-performing model for all three of these traits, although absolute accuracies remained modest, particularly for productive tillers and WUE. Across traits, XGBoost most often showed the weakest performance and was especially inferior for single-day prediction, where its median accuracy dropped to 0.45 for total grain yield, 0.36 for total straw biomass, 0.27 for pod number, and only 0.08 for thousand kernel weight. Taken together, these results show that model choice mattered substantially within the VI + 3D predictor set, with Bayesian LASSO and VIBLUP providing the most robust overall performance for the main target traits, whereas nonlinear approaches only occasionally offered an advantage for more difficult component traits.

The effect of cumulative prediction relative to single-day prediction was strongly trait dependent (**Figure 5**). Clear gains were observed for several major agronomic traits, especially total grain yield, cumulative water uptake, total straw biomass, and grain number. The strongest improvement was found for total grain yield with VI predictors, where cumulative prediction increased peak predictive ability by 0.11, followed by cumulative water uptake with VI at 0.10. Positive but smaller gains were also observed for total straw biomass across all predictor sets and for grain number, particularly under 3D prediction. **Table 1** further shows that these differences were often accompanied by marked shifts in prediction timing. For grain yield, cumulative prediction delayed the DAS of peak predictive ability by 33 days and VI, 12 days under 3D und 14 days und VI + 3D. In addition, the earliest DAS at which at least 95% of peak predictive ability was reached was often shifted, indicating that cumulative prediction changed not only the magnitude of prediction accuracy but also the temporal position of useful prediction windows. These results indicate that cumulative integration often improved prediction of broad biomass- and yield-related traits. In contrast, cumulative prediction was not consistently advantageous for

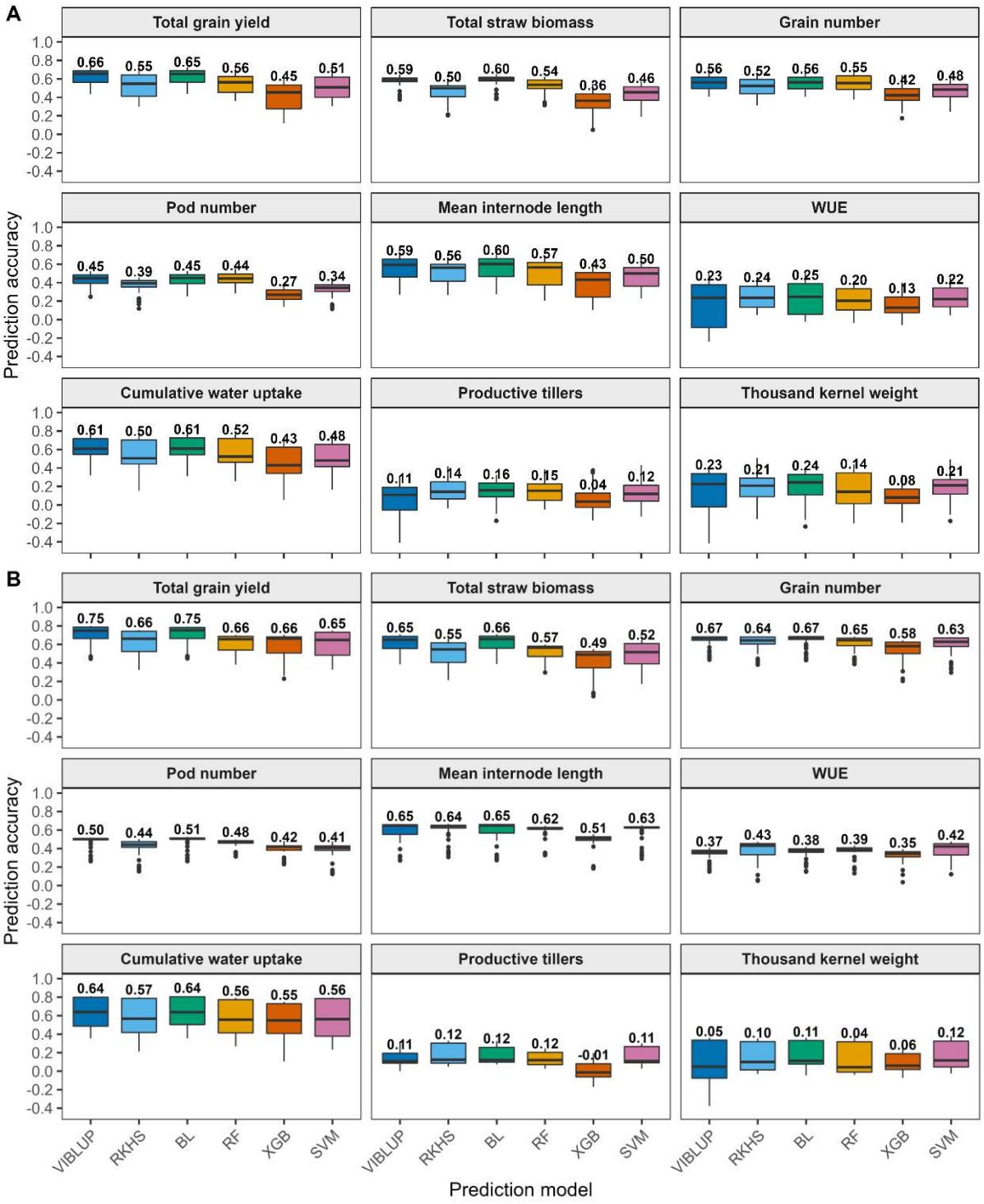


Figure 4 Model-specific prediction performance within the combined VI + 3D predictor set. Boxplots summarize the distribution of DAS-specific predictive abilities across the season for each prediction model and trait, shown separately for single-day predictions (A) and cumulative predictions (B). Number over the quartile represents the median prediction accuracy.

all traits and frequently shifted high prediction windows toward later developmental stages. Productive tillers showed the strongest decline under cumulative prediction with reductions in peak predictive ability of -0.16, -0.09, and -0.07 for VI, 3D, and VI + 3D, respectively (**Table 1**). Thousand kernel weight and mean internode length also tended to perform less well when information across time points was combined. WUE and pod number showed mixed responses depending on the predictor set. Table 1 further indicates that cumulative prediction altered not only prediction strength but also the temporal position and width of useful prediction windows,

with trait- and predictor-set-specific shifts in peak DAS and in the onset and end of the 95% performance interval. Overall, cumulative prediction was most beneficial for integrative target traits, whereas several component traits were predicted equally well or better from single-day information.

Table 1 Differences between cumulative and single-day prediction with respect to prediction strength and prediction timing.MIL = Mean internode length, WUE = water use efficiency, CWU = cumulative water uptake, TKW = thousand kernel weight, VI = vegetation indices only; 3D = 3D structural traits only, Cum = cumulative prediction; Sd = single-day prediction; DAS = days after sowing. Δ↑ denote differences in peak predictive ability. Δx̄ denote differences in mean predictive ability across all DAS. Δ↑ DAS denote differences in the DAS at which peak predictive ability occurred. ΔDAS95 denote differences in the earliest DAS at which at least 95% of the scenario-specific peak predictive ability was reached. Δ late DAS95 denote differences in the latest DAS at which at least 95% of the scenario-specific peak predictive ability was reached. Positive values indicate larger values for cumulative prediction than for single-day prediction; negative values indicate smaller values.

Trait	Predictor set	Δ↑ Cum / Sd	Δx̄ Cum / Sd	Δ↑ DAS Cum / Sd	Δ DAS95 Cum / Sd	Δ late DAS95 Cum / Sd
Grain yield	VI	0.114	0.176	33	32	41
Grain yield	3D	-0.034	0.048	12	-1	22
Grain yield	VI + 3D	0.048	0.077	14	7	22
Straw Biomass	VI	0.045	0.163	38	18	42
Straw Biomass	3D	0.023	0.015	32	18	16
Straw Biomass	VI + 3D	0.056	0.024	26	19	16
Grain number	VI	-0.002	0.091	41	39	34
Grain number	3D	0.057	0.095	38	18	26
Grain number	VI + 3D	0.025	0.083	24	18	26
Pod number	VI	-0.016	0.052	46	46	46
Pod number	3D	0.020	0.064	2	0	44
Pod number	VI + 3D	0.026	0.042	28	24	5
MIL	VI	-0.009	0.084	3	0	5
MIL	3D	-0.020	0.076	1	1	23
MIL	VI + 3D	-0.015	0.082	-4	5	27
WUE	VI	0.035	0.145	8	15	11
WUE	3D	-0.036	0.147	12	3	31
WUE	VI + 3D	-0.010	0.142	7	15	12
CWU	VI	0.099	0.125	0	0	11
CWU	3D	0.007	0.019	15	8	14
CWU	VI + 3D	0.033	0.031	6	6	12
Nr. of tillers	VI	-0.161	-0.060	5	5	5
Nr. of tillers	3D	-0.086	-0.017	1	-2	0
Nr. of tillers	VI + 3D	-0.071	0.024	-8	-10	-5
TKW	VI	-0.127	-0.045	5	2	9
TKW	3D	0.017	0.017	15	8	19
TKW	VI + 3D	-0.068	-0.026	10	7	14

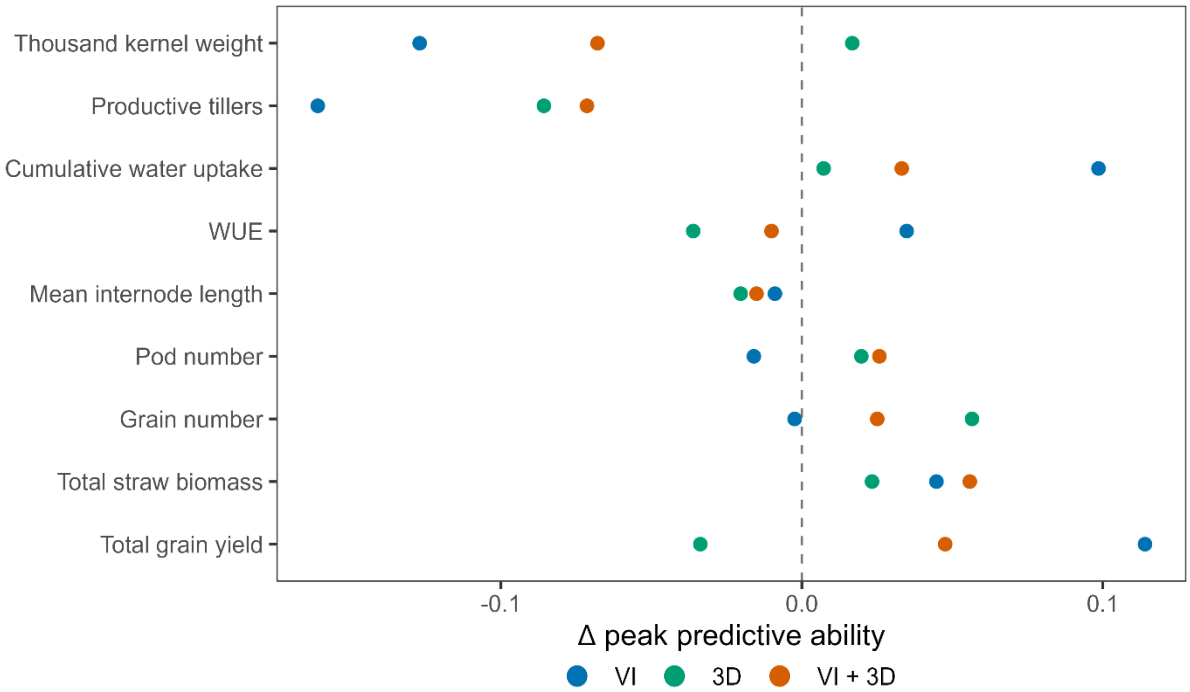


Figure 5 Difference in peak predictive ability between cumulative and single-day prediction across traits and predictor sets. Positive values indicate an advantage of cumulative prediction, whereas negative values indicate higher peak performance for single-day prediction.

Prediction performance across different predictor sets

The comparison of predictor sets across the full range of DAS-specific prediction results confirmed that predictive performance was strongly trait specific, but also revealed consistent differences among VI, 3D, and combined VI + 3D predictors (**Figure 6 A, B, Table 2, Figure 7**). Detailed model-specific comparisons across predictor sets for both the single-day and cumulative approaches are provided in **Supplementary Figures S4** and **S5**. Under single-day prediction, the combined predictor set generally provided the strongest overall performance for the major target traits. Total grain yield showed the highest across-season median predictive ability for single-day VI + 3D at 0.56, compared with 0.52 for 3D and 0.38 for VI alone. A very similar pattern was observed for total straw biomass, where VI + 3D reached a median of 0.52, narrowly exceeding 3D at 0.51 and clearly outperforming VI at 0.12. Cumulative water uptake likewise benefited from predictor fusion, with single-day VI + 3D showing the highest median value at 0.50, followed by 3D at 0.47 and VI at 0.41. The same tendency was observed for grain number and mean internode length, for which the combined predictor set again yielded the highest median performance across the season. These results indicate that, when only one timepoint was used, integrating spectral and structural information generally improved the robustness of prediction across developmental stages. **Table 2** supports this pattern by showing positive differences in peak predictive ability of VI + 3D relative to VI for all single-day traits, ranging from 0.04 for WUE to 0.33 for mean internode length, while the corresponding advantage of VI + 3D over 3D was generally smaller and in several cases close to zero.

Some traits, however, showed weaker or more variable gains from predictor fusion. Pod number was more accurately predicted by 3D alone, with a median predictive ability of 0.43 under single-day prediction, compared with 0.41 for VI + 3D and only 0.08 for VI. Productive tillers and

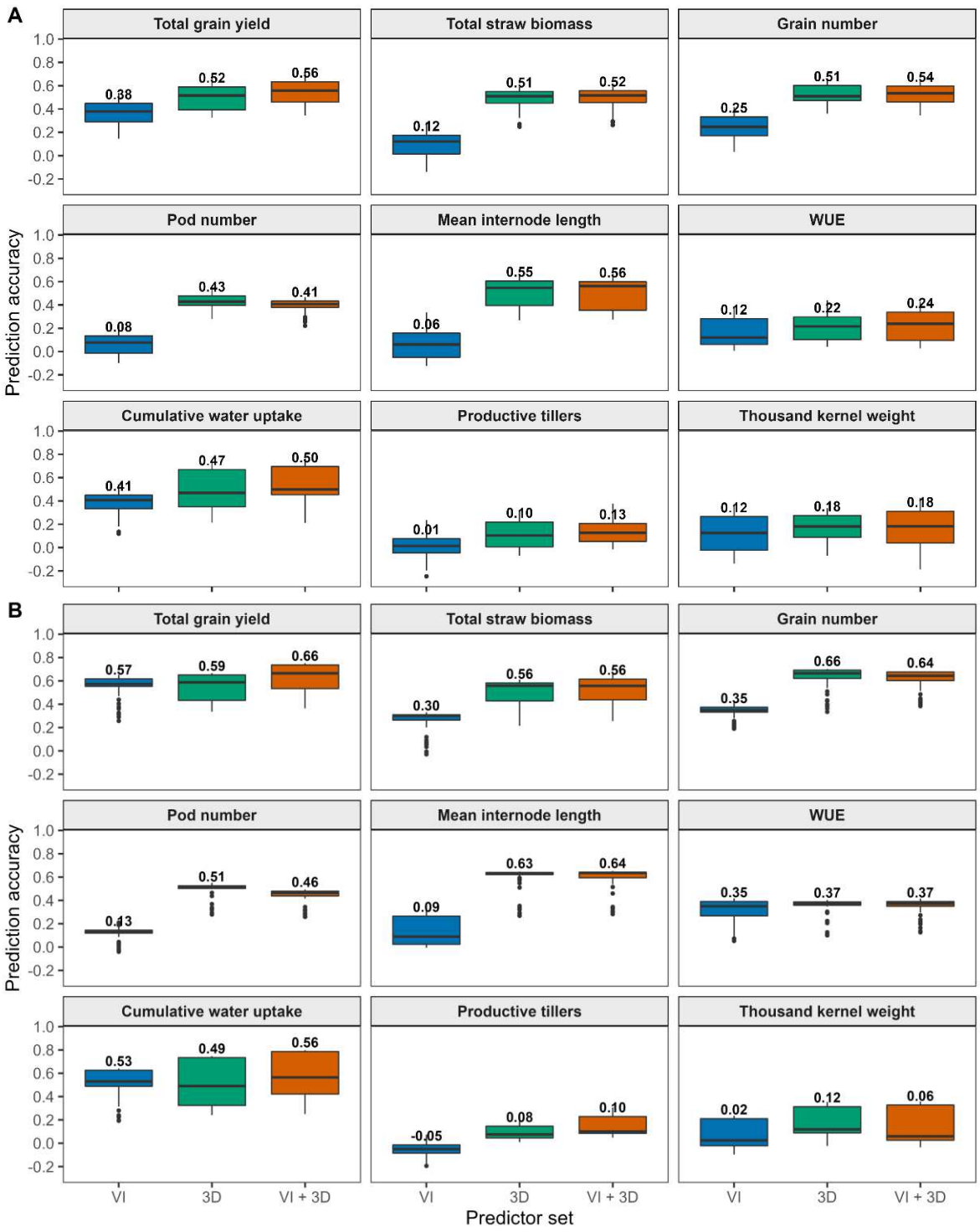


Figure 6 Comparison of predictor sets across traits based on DAS-specific predictive abilities aggregated across models. Boxplots show the distribution of model-aggregated predictive abilities across all evaluated DAS for each trait and predictor set, presented separately for single-day predictions (A) and cumulative predictions (B). Number over the quartile represents the median prediction accuracy.

thousand kernel weight showed only modest overall prediction levels, although VI + 3D still provided the highest median performance for both traits. WUE also remained moderate overall, but single-day VI + 3D again slightly outperformed the individual predictor sets. Under cumulative prediction, the relative ranking of predictor sets became more differentiated. VI + 3D remained superior for total grain yield, cumulative water uptake, mean internode length, and productive tillers, with median values of 0.66, 0.56, 0.64, and 0.10, respectively. In contrast, 3D alone

provided the strongest cumulative predictions for grain number, pod number, thousand kernel weight, and WUE, and it performed nearly identically to VI + 3D for total straw biomass. VI alone was consistently the weakest predictor set for most traits and was particularly inferior for total straw biomass, grain number, pod number, and mean internode length. This shift is also reflected in **Table 2**, where cumulative VI + 3D retained a clear advantage over VI for grain yield, straw biomass, cumulative water uptake, and productive tillers, but not consistently over 3D. For grain number and pod number, for example, the cumulative difference in peak predictive ability between VI + 3D and 3D was negative (-0.02 and -0.06, respectively), indicating that 3D alone slightly outperformed the fused predictor set for these traits. **Figure 7** further highlights these contrast patterns by directly comparing the peak predictive ability of different predictor-set pairs within single-day and cumulative prediction. Across traits, the largest positive differences were generally observed when VI + 3D was compared with VI, whereas contrasts between VI + 3D and 3D were often much smaller and in some cases negative. This was especially evident for cumulative grain number and pod number, where 3D exceeded VI3D, and for single-day thousand kernel weight, where VI + 3D still showed an advantage over both VI and 3D.Taken together, these results show that predictor fusion most consistently improved performance for the main integrative agronomic traits, whereas 3D predictors alone were often sufficient, and in some cases superior, for more structural component traits.

Table 2 Differences between the combined predictor set (VI3D) and the individual predictor sets VI and 3D with respect to prediction strength and prediction timing. MIL = Mean internode length, WUE = water use efficiency, CWU = cumulative water uptake, TKW = thousand kernel weight, VI3D = combined vegetation indices and 3D structural traits; VI = vegetation indices only; 3D = 3D structural traits only; DAS = days after sowing. Δ↑ denote differences in peak predictive ability. Δx̄ denote differences in mean predictive ability across all DAS. Δ↑ DAS denote differences in the DAS at which peak predictive ability occurred. ΔDAS95 denote differences in the earliest DAS at which at least 95% of the scenario-specific peak predictive ability was reached. Positive values indicate larger values for VI3D than for the respective comparator; negative values indicate smaller values.

Trait	Temporal strategie	Δ↑ VI3D/ VI	Δ↑ VI3D/ 3D	Δx̄ VI3D/ 3D	Δx̄ VI3D / VI	Δ↑ DAS VI3D/ VI	Δ↑ DAS VI3D / 3D	ΔDAS95 VI3D / VI	ΔDAS95 VI3D / 3D
Grain yield	Sd	0.166	0.004	0.175	0.046	12	-6	14	-8
Grain Yield	Cum	0.100	0.085	0.076	0.076	-7	-4	-11	0
Straw Biomass	Sd	0.314	0.011	0.403	0.007	15	1	14	5
Straw Biomass	Cum	0.325	0.043	0.264	0.016	3	-5	15	6
Grain number	Sd	0.239	0.010	0.286	0.000	14	11	-2	0
Grain number	Cum	0.267	-0.022	0.278	-0.012	-3	-3	-23	0
Pod number	Sd	0.243	-0.063	0.327	-0.033	18	20	-5	-2
Pod number	Cum	0.285	-0.057	0.317	-0.054	0	46	-27	22
MIL	Sd	0.332	0.000	0.438	-0.010	-2	13	-20	1
MIL	Cum	0.326	0.006	0.436	-0.004	-9	8	-15	5
WUE	Sd	0.043	-0.013	0.050	0.009	-9	5	-9	-9
WUE	Cum	-0.002	0.013	0.046	0.004	-10	0	-9	3
CWU	Sd	0.222	0.025	0.159	0.047	-5	6	-7	0
CWU	Cum	0.156	0.051	0.065	0.058	1	-3	-1	-2
Nr. of tillers	Sd	0.144	0.063	0.122	0.016	8	2	9	3
Nr. of tillers	Cum	0.233	0.077	0.206	0.057	-5	-7	-6	-5
TKW	Sd	0.063	0.088	0.058	0.009	-4	2	-4	1

TKW	Cum	0.122	0.003	0.076	-0.034	1	-3	1	0

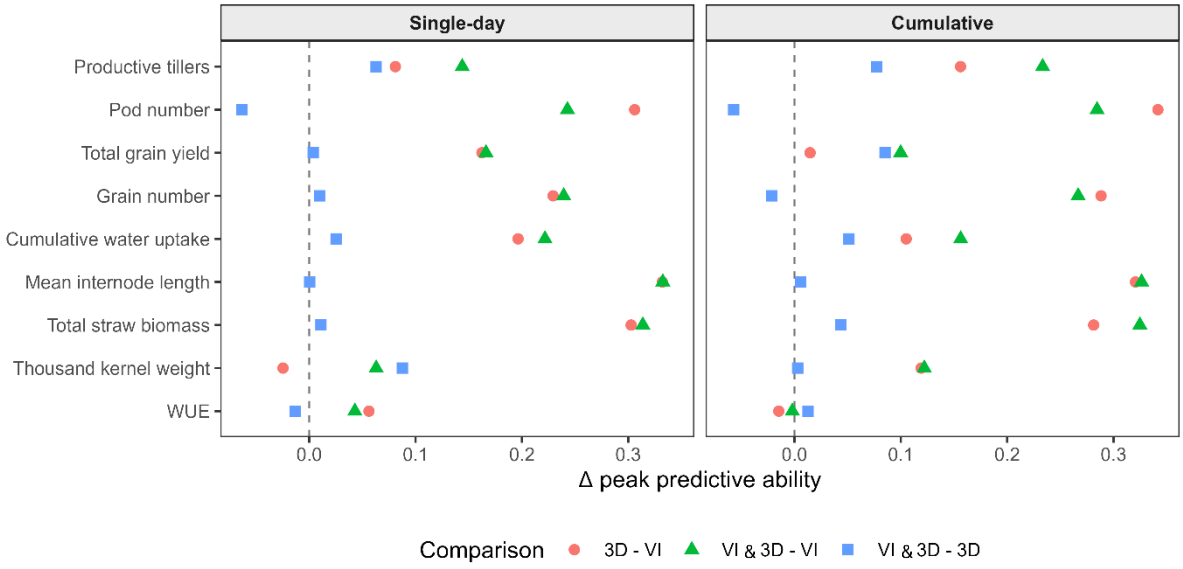


Figure 7 Difference in peak predictive ability between different predictor sets across traits and time periods. Positive values indicate an advantage of the respective predictor set.

Observation versus predicted values

The observed versus predicted scatterplots confirmed the strong predictive performance obtained for the main target traits (**Figure 8; Supplementary Table S2**). The best correlation between observed and predicted values was found for cumulative water uptake, with a correlation of 0.79 for cumulative VI + 3D prediction using VIBLUP at DAS 94. Total grain yield showed similarly strong performance, reaching a correlation of 0.77 for cumulative VI + 3D with VIBLUP at DAS 93. Grain number was also predicted reliably, with the best setup given by cumulative 3D prediction with VIBLUP at DAS 63, resulting in a correlation of 0.69. Total straw biomass showed the weakest, but still moderate, correlation among the four displayed traits, with a value of 0.65 for cumulative VI + 3D prediction using Bayesian LASSO at DAS 105. The scatterplots of the remaining traits are provided in the **Supplementary Figure S3**.

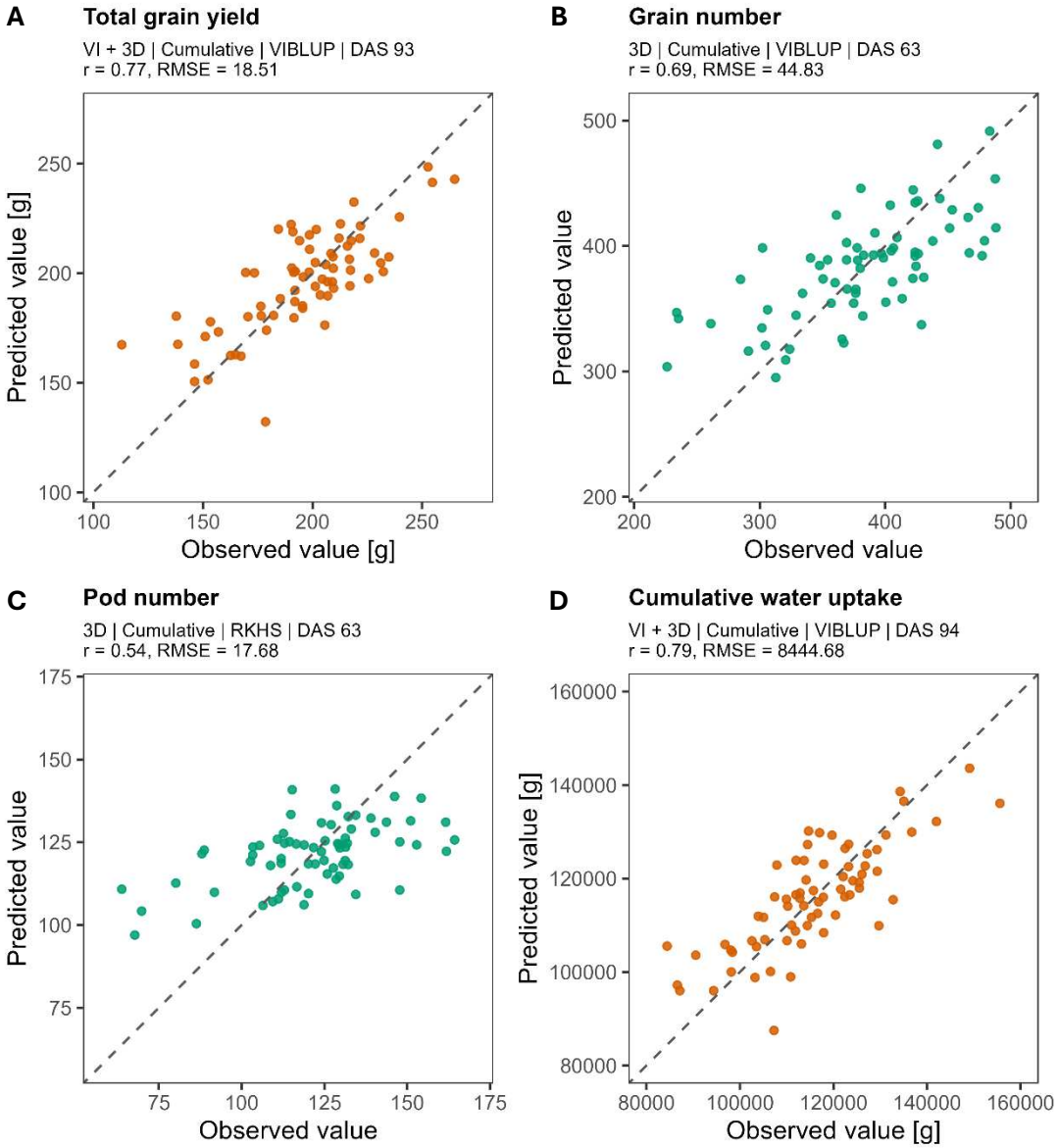


Figure 8 Observed versus predicted values for the selected target traits under the best-performing prediction setup for each trait. Predictions are shown for total grain yield (A), grain number (B), pod number (C), and cumulative water uptake (D), using the predictor set, model, and day after sowing (DAS) that yielded the highest performance for each trait. Each point represents one genotype, and the predicted value corresponds to the genotype-wise median across all repeated cross-validation runs. The dashed line indicates the 1:1 relationship between observed and predicted values.

Discussion

Trait-specific differences between different predictor sets

A central finding of the present study is that the predictive value of phenomic data was strongly trait dependent and that the relative contribution of vegetation indices, 3D-derived traits, and their combination differed systematically among trait classes. The combined VI + 3D predictor set was most advantageous for the major integrative target traits, particularly grain yield, straw biomass, and cumulative water uptake, whereas several component traits, including grain number and pod number, were predicted equally well or better by 3D information alone. This pattern is biologically plausible because vegetation indices and 3D descriptors capture partly distinct, but complementary, dimensions of plant performance. Vegetation indices primarily reflect canopy-level variation associated with chlorophyll content, greenness, canopy cover, and stress-related changes in plant status, whereas 3D phenotyping describes structural properties such as canopy height, volume, surface configuration, and architectural uniformity (Tayade et al. 2022; Liu et al. 2019). Under drought, these dimensions jointly contribute to final plant performance, but they do not necessarily vary in parallel across development (Mansoor and Chung 2024). Predictor fusion is therefore likely to be most effective for traits that integrate both physiological condition and structural development over time. This interpretation is consistent with the broader drought-phenotyping literature, which shows that water limitation simultaneously affects canopy function, pigment dynamics, biomass formation, and plant architecture, and that these responses are only partly recoverable from any single sensor domain (Mansoor and Chung 2024; Tayade et al. 2022; Liu et al. 2019).

The same reasoning also explains why 3D predictors were particularly competitive for several structural or component traits. Traits such as pod number, grain number, and internode-related morphology are often more directly linked to plant architecture than to canopy spectral behavior alone. In cereals and legumes, studies combining structural descriptors with multispectral information have repeatedly shown that canopy geometry contributes substantially to the prediction of biomass-related and architecture-related traits, and in some cases already captures most of the usable signal (Ji et al. 2024; Panigrahi et al. 2025; Sankaran et al. 2018). Earlier work in faba bean and related crops has likewise shown that plant height and other structure-derived descriptors can be strongly associated with growth stage and yield formation (Ji et al. 2022; Mohammadi et al. 2024), while the addition of structural information to spectral predictors does not invariably increase predictive ability for every target trait. The present results therefore support a biologically differentiated interpretation in which predictor fusion is particularly valuable for highly integrative agronomic traits, whereas 3D descriptors alone may already be close to sufficient for more architecture-driven components. More broadly, these findings indicate that the optimal sensing strategy depends less on maximizing data volume than on matching predictor composition to the functional nature of the target trait.

Direct evidence for the relative value of fused spectral and structural predictors in faba bean remains limited, which makes the present results particularly relevant in the context of legume phenomics (Ji et al. 2022; Mohammadi et al. 2024; Ji et al. 2024).

Cumulative versus single-day prediction

The comparison between cumulative and single-day prediction represents a second key result of the present study. Cumulative prediction improved predictive ability for traits such as grain yield, straw biomass, cumulative water uptake, and partly grain number, but this advantage was not universal across all traits. This pattern is biologically meaningful because these traits are not

instantaneous properties, but seasonally integrated outcomes of canopy establishment, stress adaptation, biomass accumulation, and source-sink regulation (Adak et al. 2024; Chatterjee et al. 2023). In such cases, cumulative predictors do not merely summarize a larger number of observations but rather encode the developmental trajectory itself. Repeated phenomic measurements can therefore retain information on how a genotype reached a given phenotypic state, which may be more informative than the state observed at any single timepoint alone (Adak et al. 2023b; Adak et al. 2023a). This interpretation is consistent with the broader literature on longitudinal crop phenotyping, where repeated measurements have frequently improved the prediction of growth-related and stress-related traits when trait expression reflects cumulative developmental processes rather than isolated snapshots (Adak et al. 2023b; Chatterjee et al. 2023).

At the same time, the present results clearly indicate that cumulative prediction should not be regarded as inherently superior. Traits such as productive tillers, thousand kernel weight, and, depending on the predictor set, WUE and mean internode length did not consistently benefit from temporal accumulation. This is likewise biologically plausible. Some traits are governed by relatively narrow developmental phases, and the most informative signal may therefore be concentrated at a specific stage rather than distributed across the season. When information from multiple dates is combined, such stage-specific signals may be diluted by less informative observations from earlier or later developmental phases. This interpretation is supported by crop-specific studies in faba bean and cereals showing that the predictive relationship between image-derived traits and agronomic performance is often strongly stage dependent and that the most informative time points are not uniformly distributed across the season (Mohammadi et al. 2024; Ji et al. 2022; Adak et al. 2023a). The present findings therefore suggest that cumulative prediction should be understood as a trait-specific strategy rather than a generally preferable one. For traits with strongly integrated seasonal biology, cumulative information can substantially improve predictive performance, whereas for developmentally narrower or more discrete component traits, carefully targeted single-day models may remain both more efficient and more biologically appropriate (Mohammadi et al. 2024). The present results also suggest that dense temporal sampling is not equally necessary for all traits. For traits with broad high-performance windows, a reduced number of strategically placed measurements may capture much of the relevant predictive signal, whereas narrow-window traits are more likely to require finer temporal resolution (Chatterjee et al. 2023; Mohammadi et al. 2024).

A closely related aspect is the width of the useful prediction window. The analysis of earliest useful prediction and peak prediction showed that some traits were associated with broad windows of near-maximal performance, whereas others exhibited distinctly narrow high-performance phases. Broad plateaus may indicate that the relevant phenotypic information remains sufficiently stable across an extended developmental period, which is likely to be the case for strongly integrative traits such as grain yield, biomass-related traits, and cumulative water uptake. In contrast, narrow windows suggest that the relevant signal is more tightly linked to a restricted developmental stage or is only transiently expressed at the canopy level (Adak et al. 2024; Mohammadi et al. 2024). This distinction is not only statistically relevant, but also biologically informative, because it reflects how directly and how persistently a given trait is represented in canopy-level phenomic measurements. It also has clear methodological implications, since the value of repeated phenotyping depends not only on peak accuracy, but also on whether informative phenotypic differences remain detectable across a broad or narrow temporal interval (Chatterjee et al. 2023; Adak et al. 2023b).

Added value of multispectral information beyond 3D phenotyping

An additional practical question arising from the present results is whether multispectral information is always needed when high-quality 3D canopy data is already available. For several major integrative target traits, especially grain yield, straw biomass, and cumulative water uptake, the combined VI + 3D predictor set slightly outperformed 3D alone, indicating that spectral information captured biologically relevant variation that was not fully represented by canopy structure. By contrast, for several component traits, 3D predictors were similarly effective or even superior, suggesting that structural information already contained most of the usable predictive signal.

This distinction is biologically plausible. Three-dimensional phenotyping primarily captures architectural properties such as canopy height, volume, and spatial organization (Liu et al. 2019; Tripodi et al. 2024), whereas vegetation indices are more closely related to canopy greenness, pigment status, moisture-related variation, and general plant health or stress condition (Tayade et al. 2022; Tripodi et al. 2024). In other words, 3D data can describe how the canopy is built, but not to the same extent how physiologically functional or stressed the canopy is at a given time (Liu et al. 2019; Ninomiya 2022). Reviews of spectral phenotyping emphasize that vegetation indices are frequently used to predict chlorophyll status, health, moisture, and production-related variation, whereas digital plant phenotyping platforms are particularly powerful for retrieving architectural traits.

From a practical breeding perspective, the key question is therefore not whether multispectral sensing is always superior, but whether its added value is large enough to justify the additional investment in sensors, calibration, data processing, and workflow complexity (Young 2019; Araus et al. 2018; Ninomiya 2022). Recent faba bean and legume studies support this differentiated view. In faba bean, multimodal fusion improved estimation accuracy for several biomass- and yield-related targets (Ji et al. 2024), while other studies showed that RGB or structural descriptors could already provide strong prediction for some traits or stages (Ji et al. 2022; Mohammadi et al. 2024; Panigrahi et al. 2025). Thus, multispectral information appears most worthwhile when the target trait integrates both structural development and physiological status, whereas for more architecture-driven traits a simpler 3D-focused workflow may already be sufficient.

Interpretation of model-specific performance

Although predictor set and temporal strategy had the strongest overall influence on prediction performance, model choice also contributed meaningfully to trait-specific differences in accuracy. In particular, the consistently strong performance of Bayesian LASSO and VIBLUP for several major target traits suggests that a substantial proportion of the predictive signal was captured by distributed, partly correlated effects across phenomic predictors rather than by highly irregular local nonlinearities. This interpretation is compatible with previous work in phenomic and relationship-based prediction, where shrinkage-based and kernel- or relationship-based approaches often provided robust performance for complex agronomic traits (Pérez and los Campos 2014; Rincent et al. 2018; Krause et al. 2019; Roscher-Ehrig et al. 2024). At the same time, the occasional advantages of more flexible nonlinear models for specific component traits indicate that trait architecture may differ in the extent to which nonlinear predictor interactions contribute to predictive performance. Model choice should therefore be interpreted as trait dependent rather than universally hierarchical.

Practical implications for phenotyping and breeding

The present findings have direct implications for the design of phenotyping strategies under drought stress. Most importantly, they indicate that sensor selection and temporal design should be tailored to the biology of the target trait rather than maximizing sensor complexity by default (Young 2019; Araus et al. 2018; Ninomiya 2022). For highly integrative target traits such as grain yield and cumulative water uptake, the combined use of VI and 3D information appears justified because it provided the strongest and most stable predictions. In these cases, the added complexity of multisensor acquisition may be justified by improved predictive performance and more reliable genotype ranking under stress (Ji et al. 2024; Panigrahi et al. 2025). By contrast, for traits in which 3D predictors alone performed similarly to or better than fused predictors, a more architecture-focused workflow may be sufficient. This is particularly relevant from the perspective of throughput, acquisition costs, data-processing demands, and operational scalability (Young 2019; Ninomiya 2022). Field-based high-throughput phenotyping is constrained not only by sensor capability, but also by the need to balance biological relevance, logistical feasibility, and breeding utility (Araus et al. 2018; Young 2019).

The temporal results are equally important for practical deployment. Broad high-performance windows imply that measurements can be scheduled with a certain degree of flexibility without major loss of predictive ability, which is advantageous under field conditions where weather, labor availability, and platform access often constrain the timing of flights or scans (Chatterjee et al. 2023; Hein et al. 2021). Narrow windows, in contrast, indicate that successful phenotyping requires more precise developmental targeting. From a breeding perspective, this distinction is highly relevant because it helps define where flexible routine monitoring is sufficient and where high-resolution or stage-specific phenotyping is necessary. More broadly, the results support the view that phenomic prediction can serve as an efficient indirect selection tool when it is aligned with the temporal biology of the target trait (Rincent et al. 2018; Rebetzke et al. 2019). This is especially relevant for breeding targets that are difficult, labor-intensive, or destructive to phenotype directly at scale (Rebetzke et al. 2019; Hein et al. 2021). The present findings therefore do not merely compare alternative predictor sets, but also provide a practical framework for prioritizing sensing modalities and measurement schedules according to trait class and breeding objective (Chawade et al. 2019; Araus et al. 2018). In practice, the value of additional sensor domains will depend not only on predictive gain but also on whether these gains justify the added complexity of data acquisition, processing, and deployment in breeding pipelines (Young 2019; Araus et al. 2018).

Limitations and outlook

Several limitations should be considered when interpreting the present findings. First, the observed ranking of predictor sets, models, and temporal strategies should not be regarded as universally transferable without further validation across years, environments, and drought scenarios. Canopy reflectance and structural development are both strongly shaped by environmental conditions, and the optimal combination of sensor modality and timing may shift depending on stress intensity, stress timing, radiation regime, or differences in plant architecture across germplasm (Roth et al. 2023; Marsh et al. 2021; Ji et al. 2024). This limitation is widely recognized in image-based phenotyping studies, where predictive performance often proves context dependent unless explicitly evaluated across multiple environments (Roth et al. 2023; Kaushal et al. 2024). The present results should therefore be interpreted as robust within the

experimental setting examined here, while broader generalization will require independent validation under additional conditions.

An additional limitation relates to the experimental platform. The study was conducted in the DroughtSpotter XXL, a semi-controlled container-based phenotyping system designed to approximate field-like growth conditions. While this setup enables high-resolution control of water supply and canopy development, it also reduces some aspects of field heterogeneity and constrains rooting volume relative to open-field conditions. Predictor rankings and optimal prediction windows identified under such conditions may therefore not fully transfer to field environments without independent validation (Moritz et al. 2024; Scheer et al. 2026).

Second, the lower predictive ability observed for traits such as WUE, thousand kernel weight, and productive tillers should not be interpreted simply as a methodological shortcoming. These traits are not directly observable canopy properties, but rather downstream, composite, or ratio-type outcomes that integrate complex interactions among multiple traits and may be only indirectly reflected in image-derived phenotypes (Cai et al. 2021; Rebetzke et al. 2019). Weaker predictions for such traits may therefore indicate a biological mismatch between the target variable and what canopy-level spectral or structural descriptors can realistically encode. This point is important because it cautions against assessing phenomic prediction solely in terms of model performance without considering trait observability. Some agronomic traits are intrinsically more difficult to infer from canopy-level imagery than others, even under otherwise well-designed modeling frameworks (Rebetzke et al. 2019; Cai et al. 2021).

A further issue concerns prediction timing. For several major agronomic traits, the strongest prediction windows were located relatively late in development. This is scientifically informative because it indicates strong end-of-season inferential power, but it is less directly aligned with breeding scenarios in which early selection is particularly valuable. One of the most relevant next steps will therefore be to determine whether similar predictive robustness can be achieved earlier in the season through improved temporal feature engineering, stage-aware modeling, or alternative predictor construction (Mohammadi et al. 2024; Ji et al. 2022; Adak et al. 2024). In addition, the moderate performance of image-based prediction for some traits suggests that integrated phenomic and genomic approaches may be especially promising where canopy imagery alone does not capture sufficient information (Marsh et al. 2021; Volk et al. 2021; Kaushal et al. 2024). Rather than indicating that one sensor domain or one temporal design is generally optimal, the present results emphasize that successful drought phenotyping depends on matching predictor composition, temporal strategy, and model choice to the biological properties of the target trait, and on validating these choices under environments relevant to breeding practice (Marsh et al. 2021; Roth et al. 2023).

Conclusion

Phenomic prediction under drought was strongly trait dependent and influenced by both predictor composition and temporal design. Combined VI + 3D predictors provided the clearest advantage for major integrative traits such as grain yield, straw biomass, and cumulative water uptake, whereas several component traits were predicted equally well or better by 3D information alone. Cumulative prediction improved performance for traits with a strong seasonal integrative basis, but this advantage was not universal across all targets.

These findings indicate that effective phenomic prediction depends on matching sensor domain, temporal strategy, and modeling approach to the biological nature of the target trait. From a

breeding perspective, multisensor prediction appears particularly promising for complex agronomic targets, while simpler trait-specific phenotyping strategies may be sufficient for more architecture-related components.

Acknowledgments

The authors thank Maximilian Henning, Bernhard Schick and Andreas Eckert for valuable technical support during the DroughtSpotter XXL experiment.

Authors' contributions

RJS and AS conceived the study, LS and LRE developed the method, LS performed the experiment, data curation. LS, LRE and JW performed the data analysis and interpreted the results, LS wrote the manuscript with contribution from RJS & BW. General guidance was provided by BW. Guidance regarding faba bean cultivation was provided by OS and GW. Project coordination on the NPZ side was carried out by HT. All authors revised the manuscript and approved the final manuscript.

Corresponding author

Correspondence to Lennart Scheer.

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author LS.

Funding

RJS was supported by grants 469336000 from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and grant 28AD105B21 from the German Federal Ministry of Nutrition and Agriculture (BMEL). The phenotyping facility was partially funded by infrastructure grant 461868281 to RJS from the DFG.

Competing Interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

Clinical trial number

Not applicable.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the corresponding author used ChatGPT in order to improve writing style, grammar and spelling. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Publication bibliography

Adak, Alper; DeSalvio, Aaron J.; Arik, Mustafa A.; Murray, Seth C. (2024): Field-based high-throughput phenotyping enhances phenomic and genomic predictions for grain yield and plant height across years in maize. In *G3 (Bethesda, Md.)* 14 (7). DOI: 10.1093/g3journal/jkae092.

Adak, Alper; Kang, Myeongjong; Anderson, Steven L.; Murray, Seth C.; Jarquin, Diego; Wong, Raymond K. W.; Katzfuß, Matthias (2023a): Phenomic data-driven biological prediction of maize through field-based high-throughput phenotyping integration with genomic data. In *Journal of experimental botany* 74 (17), pp. 5307–5326. DOI: 10.1093/jxb/erad216.

Adak, Alper; Murray, Seth C.; Anderson, Steven L. (2023b): Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. In *G3 (Bethesda, Md.)* 13 (1). DOI: 10.1093/g3journal/jkac294.

Araus, José Luis; Kefauver, Shawn C.; Zaman-Allah, Mainassara; Olsen, Mike S.; Cairns, Jill E. (2018): Translating High-Throughput Phenotyping into Genetic Gain. In *Trends in plant science* 23 (5), pp. 451–466. DOI: 10.1016/j.tplants.2018.02.001.

Arlot, Sylvain; Celisse, Alain (2010): A survey of cross-validation procedures for model selection. In *Statist. Surv.* 4 (none). DOI: 10.1214/09-SS054.

Bischl, B; Lang, M; Marwedel, P; Weihs, C; Rahnenfuehrer, J; Bisch, B (2016): mlr: Machine Learning in R. Edited by Journal of Machine Learning Research (170). Available online at https://jmlr.org/papers/v17/15-066.html.

Bischl, Bernd; Binder, Martin; Lang, Michel; Pielok, Tobias; Richter, Jakob; Coors, Stefan et al. (2023): Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. In *WIREs Data Min & Knowl* 13 (2), Article e1484. DOI: 10.1002/widm.1484.

Bischl, Bernd; Lang, Michel; Richter, Jakob; Horn, Daniel; Kerschke, Pascal (2022): ParamHelpers: Helpers for Parameters in Black-Box Optimization, Tuning and Machine Learning. R package version 1.14.1. Available online at https://CRAN.R-project.org/package=ParamHelpers.

Bischl, Bernd; Richter, Jakob; Bossek, Jakob; Horn, Daniel; Thomas, Janek; Lang, Michel (2017): mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions.

Bossek, Jakob (2017): smoof: Single- and Multi-Objective Optimization Test Functions. The R Journal. Available online at https://journal.r-project.org/archive/2017/RJ-2017-004/index.html.

Braun, Vincent; Zhu, Xintian; Meyenberg, Carina; Hahn, Volker; Maurer, Hans Peter; Würschum, Tobias; Thorwarth, Patrick (2025): Phenomic Prediction Can Be Improved by Optimization of NIRS Preprocessing. In *Plant Breeding* 144 (4), pp. 454–469. DOI: 10.1111/pbr.13265.

Breiman, Leo (2001): Random Forests. In *Machine Learning* 45 (1), pp. 5–32. DOI: 10.1023/A:1010933404324.

Cai, Wanyuan; Ullah, Sana; Yan, Lei; Lin, Yi (2021): Remote Sensing of Ecosystem Water Use Efficiency: A Review of Direct and Indirect Estimation Methods. In *Remote Sensing* 13 (12), p. 2393. DOI: 10.3390/rs13122393.

Carnell, Rob (2022): lhs: Latin Hypercube Samples. R package version 1.1.6. Available online at https://CRAN.R-project.org/package=lhs.

Chatterjee, Sumantra; Adak, Alper; Wilde, Scott; Nakasagga, Shakirah; Murray, Seth C. (2023): Cumulative temporal vegetation indices from unoccupied aerial systems allow maize (Zea mays L.) hybrid yield to be estimated across environments with fewer flights. In *PloS one* 18 (1), e0277804. DOI: 10.1371/journal.pone.0277804.

Chawade, Aakash; van Ham, Joost; Blomquist, Hanna; Bagge, Oscar; Alexandersson, Erik; Ortiz, Rodomiro (2019): High-Throughput Field-Phenotyping Tools for Plant Breeding and Precision Agriculture. In *Agronomy* 9 (5), p. 258. DOI: 10.3390/agronomy9050258.

Chen, Tianqi; He, Tong; Benesty, Michael; Khotilovich, Vadim; Tang, Yuan; Cho, Hyunsu et al. (2016): xgboost: Extreme Gradient Boosting. In Balaji Krishnapuram, Mohak Shah, Alex Smola, Charu Aggarwal, Dou Shen, Rajeev Rastogi (Eds.): Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA, 13 08 2016 17 08 2016. New York, NY, USA: ACM, pp. 785–794.

Chen, Tianqi; He, Tong; Benesty, Michael; Khotilovich, Vadim; Tang, Yuan; Cho, Hyunsu et al. (2022): xgboost: Extreme Gradient Boosting. R package version 1.6.0.1. Available online at https://CRAN.R-project.org/package=xgboost.

Covarrubias-Pazaran, Giovanny (2016): Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. In *PloS one* 11 (6), e0156744. DOI: 10.1371/journal.pone.0156744.

Cui, Yuxing; Ji, Yishan; Liu, Rong; Li, Weiyu; Liu, Yujiao; Liu, Zehao et al. (2023): Faba Bean (Vicia faba L.) Yield Estimation Based on Dual-Sensor Data. In *Drones* 7 (6), p. 378. DOI: 10.3390/drones7060378.

Drucker, Harris; Burges, Chris; Kaufman, Linda; Smola, Alex; Vapnik, Vladimir (1997): Support vector regression machines 28 (7), pp. 779–784.

Gianola, Daniel; van Kaam, Johannes B. C. H. M. (2008): Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. In *Genetics* 178 (4), pp. 2289–2303. DOI: 10.1534/genetics.107.084285.

Hein, Nathan T.; Ciampitti, Ignacio A.; Jagadish, S. V. Krishna (2021): Bottlenecks and opportunities in field-based high-throughput phenotyping for heat and drought stress. In *Journal of experimental botany* 72 (14), pp. 5102–5116. DOI: 10.1093/jxb/erab021.

Heino, Matias; Kinnunen, Pekka; Anderson, Weston; Ray, Deepak K.; Puma, Michael J.; Varis, Olli et al. (2023): Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. In *Scientific reports* 13 (1), p. 3583. DOI: 10.1038/s41598-023-29378-2.

Henry, Lionel; Wickham, Hadley (2022): purrr: Functional Programming Tools. R package version 0.3.4. Available online at https://CRAN.R-project.org/package=purrr.

Jensen, Erik S.; Peoples, Mark B.; Hauggaard-Nielsen, Henrik (2010): Faba bean in cropping systems. In *Field Crops Research* 115 (3), pp. 203–216. DOI: 10.1016/j.fcr.2009.10.008.

Ji, Yishan; Chen, Zhen; Cheng, Qian; Liu, Rong; Li, Mengwei; Yan, Xin et al. (2022): Estimation of plant height and yield based on UAV imagery in faba bean (Vicia faba L.). In *Plant methods* 18 (1), p. 26. DOI: 10.1186/s13007-022-00861-7.

Ji, Yishan; Liu, Zehao; Liu, Rong; Wang, Zhirui; Zong, Xuxiao; Yang, Tao (2024): High-throughput phenotypic traits estimation of faba bean based on machine learning and drone-based

multimodal data. In *Computers and Electronics in Agriculture* 227, p. 109584. DOI: 10.1016/j.compag.2024.109584.

Jones, Donald R.; Schonlau, Matthias; Welch, William J. (1998): Efficient Global Optimization of Expensive Black-Box Functions. In *Journal of Global Optimization* 13 (4), pp. 455–492. DOI: 10.1023/A:1008306431147.

Karatzoglou, Alexandros; Smola, Alex; Hornik, Kurt; Zeileis, Achim (2004): kernlab - An S4 Package for Kernel Methods in R. In *J. Stat. Soft.* 11 (9). DOI: 10.18637/jss.v011.i09.

Karkanis, Anestis; Ntatsi, Georgia; Lepse, Liga; Fernández, Juan A.; Vågen, Ingunn M.; Rewald, Boris et al. (2018): Faba Bean Cultivation - Revealing Novel Managing Practices for More Sustainable and Competitive European Cropping Systems. In *Frontiers in plant science* 9, p. 1115. DOI: 10.3389/fpls.2018.01115.

Kaushal, Swas; Gill, Harsimardeep S.; Billah, Mohammad Maruf; Khan, Shahid Nawaz; Halder, Jyotirmoy; Bernardo, Amy et al. (2024): Enhancing the potential of phenomic and genomic prediction in winter wheat breeding using high-throughput phenotyping and deep learning. In *Frontiers in plant science* 15, p. 1410249. DOI: 10.3389/fpls.2024.1410249.

Khan, H. R.; Paull, J. G.; Siddique, K.H.M.; Stoddard, F. L. (2010): Faba bean breeding for drought-affected environments: A physiological and agronomic perspective. In *Field Crops Research* 115 (3), pp. 279–286. DOI: 10.1016/j.fcr.2009.09.003.

Krause, Margaret R.; González-Pérez, Lorena; Crossa, José; Pérez-Rodríguez, Paulino; Montesinos-López, Osval; Singh, Ravi P. et al. (2019): Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. In *G3 (Bethesda, Md.)* 9 (4), pp. 1231–1247. DOI: 10.1534/g3.118.200856.

Liu, Shouyang; Martre, Pierre; Buis, Samuel; Abichou, Mariem; Andrieu, Bruno; Baret, Frédéric (2019): Estimation of Plant and Canopy Architectural Traits Using the Digital Plant Phenotyping Platform. In *Plant physiology* 181 (3), pp. 881–890. DOI: 10.1104/pp.19.00554.

los Campos, Gustavo de; Gianola, Daniel; Rosa, Guilherme J. M.; Weigel, Kent A.; Crossa, José (2010): Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. In *Genetics research* 92 (4), pp. 295–308. DOI: 10.1017/S0016672310000285.

Mansoor, Sheikh; Chung, Yong Suk (2024): Functional phenotyping: Understanding the dynamic response of plants to drought stress. In *Current Plant Biology* 38, p. 100331. DOI: 10.1016/j.cpb.2024.100331.

Marsh, Jacob I.; Hu, Haifei; Gill, Mitchell; Batley, Jacqueline; Edwards, David (2021): Crop breeding for a changing climate: integrating phenomics and genomics with bioinformatics. In *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* 134 (6), pp. 1677–1690. DOI: 10.1007/s00122-021-03820-3.

McKay, M. D.; Beckman, R. J.; Conover, W. J. (1979): Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. In *Technometrics* 21 (2), pp. 239–245. DOI: 10.1080/00401706.1979.10489755.

Microsoft Coorperation; Weston, Steve (2022): doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.17. Available online at https://CRAN.R-project.org/package=doParallel.

Mohammadi, Shirin; Uhlen, Anne Kjersti; Lillemo, Morten; Ergon, Åshild; Shafiee, Sahameh (2024): Enhancing phenotyping efficiency in faba bean breeding: integrating UAV imaging and machine learning. In *Precision Agric* 25 (3), pp. 1502–1528. DOI: 10.1007/s11119-024-10121-4.

Moritz, Anna; Eckert, Andreas; Vukasovic, Stjepan; Snowdon, Rod; Stahl, Andreas (2024): Physiological phenotyping of transpiration response to vapour pressure deficit in wheat. In *BMC plant biology* 24 (1), p. 1032. DOI: 10.1186/s12870-024-05692-3.

Muktadir, Md Abdul; Adhikari, Kedar Nath; Merchant, Andrew; Belachew, Kiflemariam Y.; Vandenberg, Albert; Stoddard, Frederick L.; Khazaei, Hamid (2020): Physiological and Biochemical Basis of Faba Bean Breeding for Drought Adaptation—A Review. In *Agronomy* 10 (9), p. 1345. DOI: 10.3390/agronomy10091345.

Mwanamwenge, J.; Loss, S.P; Siddique, K.H.M; Cocks, P.S (1999): Effect of water stress during floral initiation, flowering and podding on the growth and yield of faba bean (Vicia faba L.). In *European Journal of Agronomy* 11 (1), pp. 1–11. DOI: 10.1016/S1161-0301(99)00003-9.

Ninomiya, Seishi (2022): High-throughput field crop phenotyping: current status and challenges. In *Breeding science* 72 (1), pp. 3–18. DOI: 10.1270/jsbbs.21069.

Panigrahi, Shubham Subrot; Singh, Keshav D.; Balasubramanian, Parthiba; Wang, Hongquan; Natarajan, Manoj; Ravichandran, Prabahar (2025): UAV-Based LiDAR and Multispectral Imaging for Estimating Dry Bean Plant Height, Lodging and Seed Yield. In *Sensors (Basel, Switzerland)* 25 (11). DOI: 10.3390/s25113535.

Papastylianou, Panayiota; Vlachostergios, Dimitrios N.; Dordas, Christos; Tigka, Evangelia; Papakaloudis, Paschalis; Kargiotidou, Anastasia et al. (2021): Genotype X Environment Interaction Analysis of Faba Bean (Vicia faba L.) for Biomass and Seed Yield across Different Environments. In *Sustainability* 13 (5), p. 2586. DOI: 10.3390/su13052586.

Park, Trevor; Casella, George (2008): The Bayesian Lasso. In *Journal of the American Statistical Association* 103 (482), pp. 681–686. DOI: 10.1198/016214508000000337.

Parmley, Kyle; Nagasubramanian, Koushik; Sarkar, Soumik; Ganapathysubramanian, Baskar; Singh, Asheesh K. (2019): Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. In *Plant phenomics (Washington, D.C.)* 2019, p. 5809404. DOI: 10.34133/2019/5809404.

Pérez, Paulino; los Campos, Gustavo de (2014): Genome-wide regression and prediction with the BGLR statistical package. In *Genetics* 198 (2), pp. 483–495. DOI: 10.1534/genetics.114.164442.

Piepho, Hans-Peter; Möhring, Jens; Schulz-Streeck, Torben; Ogutu, Joseph O. (2012): A stage-wise approach for the analysis of multi-environment trials. In *Biometrical journal. Biometrische Zeitschrift* 54 (6), pp. 844–860. DOI: 10.1002/bimj.201100219.

R Core Team (2021): R: A Language and Environment for Statistical Computing. Edited by R Foundation for Statistical Computing. Available online at https://www.R-project.org/.

Rebetzke, G. J.; Jimenez-Berni, J.; Fischer, R. A.; Deery, D. M.; Smith, D. J. (2019): Review: High-throughput phenotyping to enhance the use of crop genetic resources. In *Plant science : an international journal of experimental plant biology* 282, pp. 40–48. DOI: 10.1016/j.plantsci.2018.06.017.

Rincent, Renaud; Charpentier, Jean-Paul; Faivre-Rampant, Patricia; Paux, Etienne; Le Gouis, Jacques; Bastien, Catherine; Segura, Vincent (2018): Phenomic Selection Is a Low-Cost and

High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar. In *G3 (Bethesda, Md.)* 8 (12), pp. 3961–3972. DOI: 10.1534/g3.118.200760.

Roscher-Ehrig, Lennard; Weber, Sven E.; Abbadi, Amine; Malenica, Milka; Abel, Stefan; Hemker, Reinhard et al. (2024): Phenomic Selection for Hybrid Rapeseed Breeding. In *Plant phenomics (Washington, D.C.)* 6, p. 215. DOI: 10.34133/plantphenomics.0215.

Roth, Lukas; Fossati, Dario; Krähenbühl, Patrick; Walter, Achim; Hund, Andreas (2023): Image-based phenomic prediction can provide valuable decision support in wheat breeding. In *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* 136 (7), p. 162. DOI: 10.1007/s00122-023-04395-x.

Roth, Lukas; Rodríguez-Álvarez, María Xosé; van Eeuwijk, Fred; Piepho, Hans-Peter; Hund, Andreas (2021): Phenomics data processing: A plot-level model for repeated measurements to extract the timing of key stages and quantities at defined time points. In *Field Crops Research* 274, p. 108314. DOI: 10.1016/j.fcr.2021.108314.

Roustant, Oliver; Ginsbourger, David; Deville, Yves (2012): DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. Edited by Journal of Statistical Software (1). Available online at https://www.jstatsoft.org/v51/i01/.

Sankaran, Sindhuja; Zhou, Jianfeng; Khot, Lav R.; Trapp, Jennifer J.; Mndolwa, Eninka; Miklas, Phillip N. (2018): High-throughput field phenotyping in dry bean using small unmanned aerial vehicle based multispectral imagery. In *Computers and Electronics in Agriculture* 151, pp. 84–92. DOI: 10.1016/j.compag.2018.05.034.

Scheer, Lennart; Moritz, Anna; Wittkop, Benjamin; Stahl, Andreas; Sass, Olaf; Tietgen, Hanna et al. (2026): Precise Evaluation of Transpiration Patterns in Relation to Grain Yield Under Drought Stress in Faba Bean. In *Plant direct* 10 (2), e70141. DOI: 10.1002/pld3.70141.

Smola, Alex J.; Schölkopf, Bernhard (2004): A tutorial on support vector regression. In *Statistics and Computing* 14 (3), pp. 199–222. DOI: 10.1023/B:STCO.0000035301.49549.88.

Tayade, Rupesh; Yoon, Jungbeom; Lay, Liny; Khan, Abdul Latif; Yoon, Youngnam; Kim, Yoonha (2022): Utilization of Spectral Indices for High-Throughput Phenotyping. In *Plants (Basel, Switzerland)* 11 (13). DOI: 10.3390/plants11131712.

Tripodi, Pasquale; Vincenzo, Cono; Venezia, Accursio; Cocozza, Annalisa; Pane, Catello (2024): Precision Phenotyping of Wild Rocket (Diplotaxis tenuifolia) to Determine Morpho-Physiological Responses under Increasing Drought Stress Levels Using the PlantEye Multispectral 3D System. In *Horticulturae* 10 (5), p. 496. DOI: 10.3390/horticulturae10050496.

VanRaden, P. M. (2008): Efficient methods to compute genomic predictions. In *Journal of dairy science* 91 (11), pp. 4414–4423. DOI: 10.3168/jds.2007-0980.

Volk, Gayle M.; Byrne, Patrick F.; Coyne, Clarice J.; Flint-Garcia, Sherry; Reeves, Patrick A.; Richards, Chris (2021): Integrating Genomic and Phenomic Approaches to Support Plant Genetic Resources Conservation and Use. In *Plants (Basel, Switzerland)* 10 (11). DOI: 10.3390/plants10112260.

Weiß, Thea Mi; Zhu, Xintian; Leiser, Willmar L.; Li, Dongdong; Liu, Wenxin; Schipprack, Wolfgang et al. (2022): Unraveling the potential of phenomic selection within and among diverse breeding material of maize (Zea mays L.). In *G3 (Bethesda, Md.)* 12 (3). DOI: 10.1093/g3journal/jkab445.

Wickham, Hadley (2019): stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. Available online at https://CRAN.R-project.org/package=stringr.

Wickham, Hadley; Averick, Mara; Bryan, Jennifer; Chang, Winston; McGowan, Lucy; François, Romain et al. (2019): Welcome to the Tidyverse. In *JOSS* 4 (43), p. 1686. DOI: 10.21105/joss.01686.

Wickham, Hadley; Francois, Romain; Henry, Lionel; Müller, Kirill (2022): dplyr: A Grammar of Data Manipulation. R package version 1.0.10. Available online at https://CRAN.R-project.org/package=dplyr.

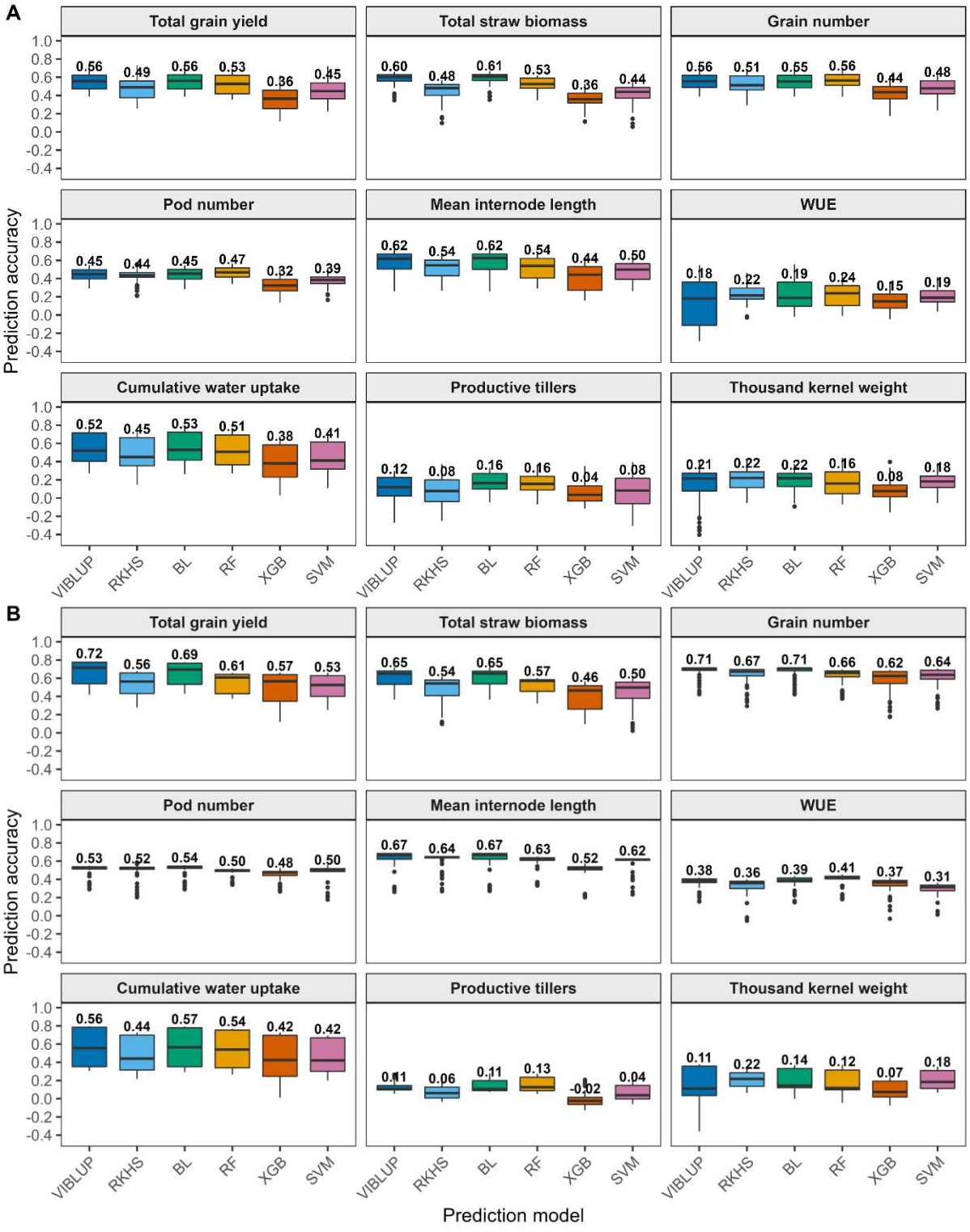
Wright, Marvin N.; Ziegler, Andreas (2017): ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. In *J. Stat. Soft.* 77 (1). DOI: 10.18637/jss.v077.i01.

Xu, Rui; Li, Changying (2022): A Review of High-Throughput Field Phenotyping Systems: Focusing on Ground Robots. In *Plant phenomics (Washington, D.C.)* 2022, p. 9760269. DOI: 10.34133/2022/9760269.

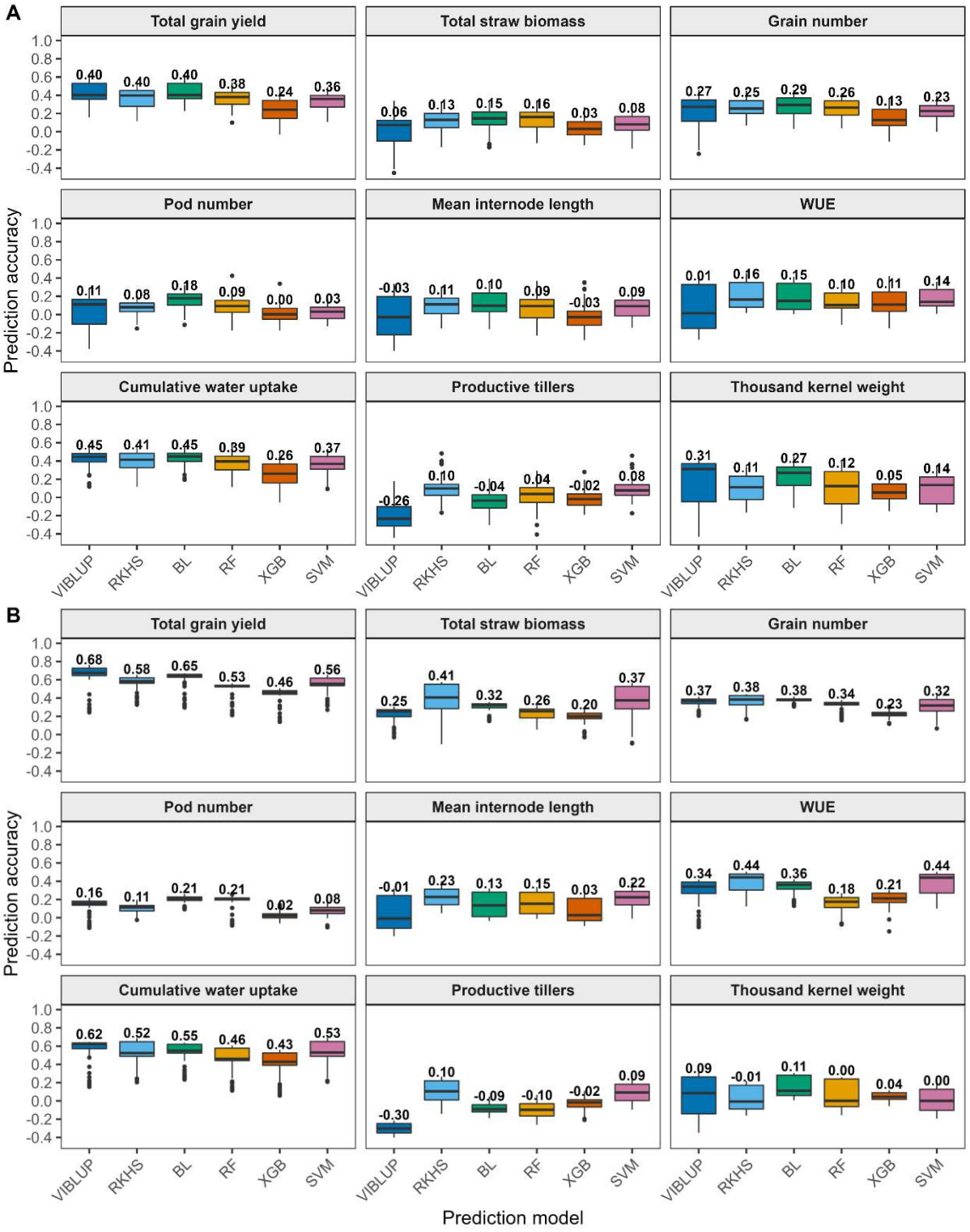
Young, Sierra N. (2019): A Framework for Evaluating Field-Based, High-Throughput Phenotyping Systems: A Meta-Analysis. In *Sensors (Basel, Switzerland)* 19 (16). DOI: 10.3390/s19163582.

Zhu, Xintian; Leiser, Willmar L.; Hahn, Volker; Würschum, Tobias (2021): Phenomic selection is competitive with genomic selection for breeding of complex traits. In *The Plant Phenome Journal* 4 (1), Article e20027. DOI: 10.1002/ppj2.20027.

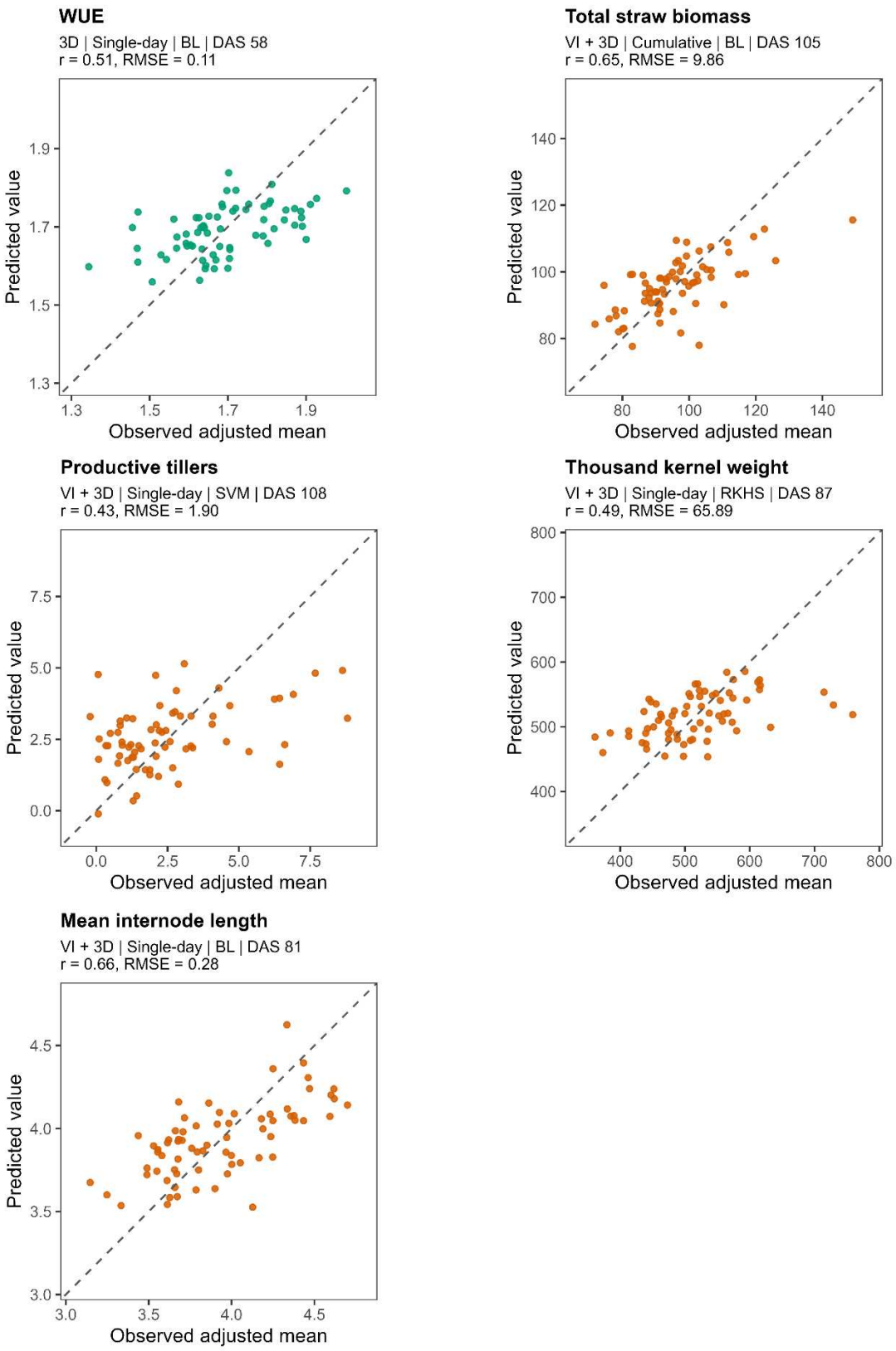
Supplementary Figures



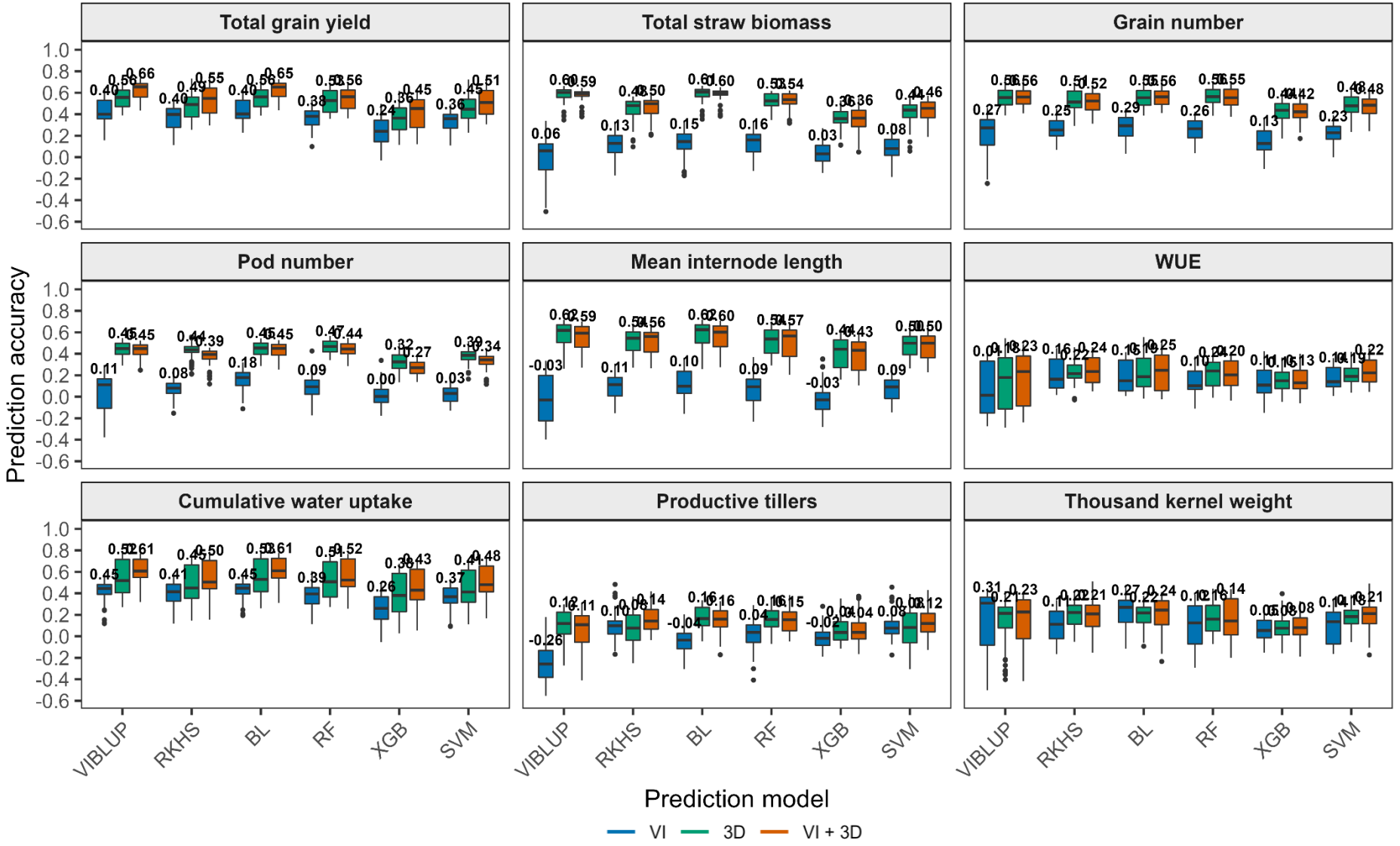
Supplementary Figure S1 Model-specific prediction performance within the combined 3D predictor set. Boxplots summarize the distribution of DAS-specific predictive abilities across the season for each prediction model and trait, shown separately for single-day predictions (A) and cumulative predictions (B). Number over the quartile represents the median prediction accuracy.



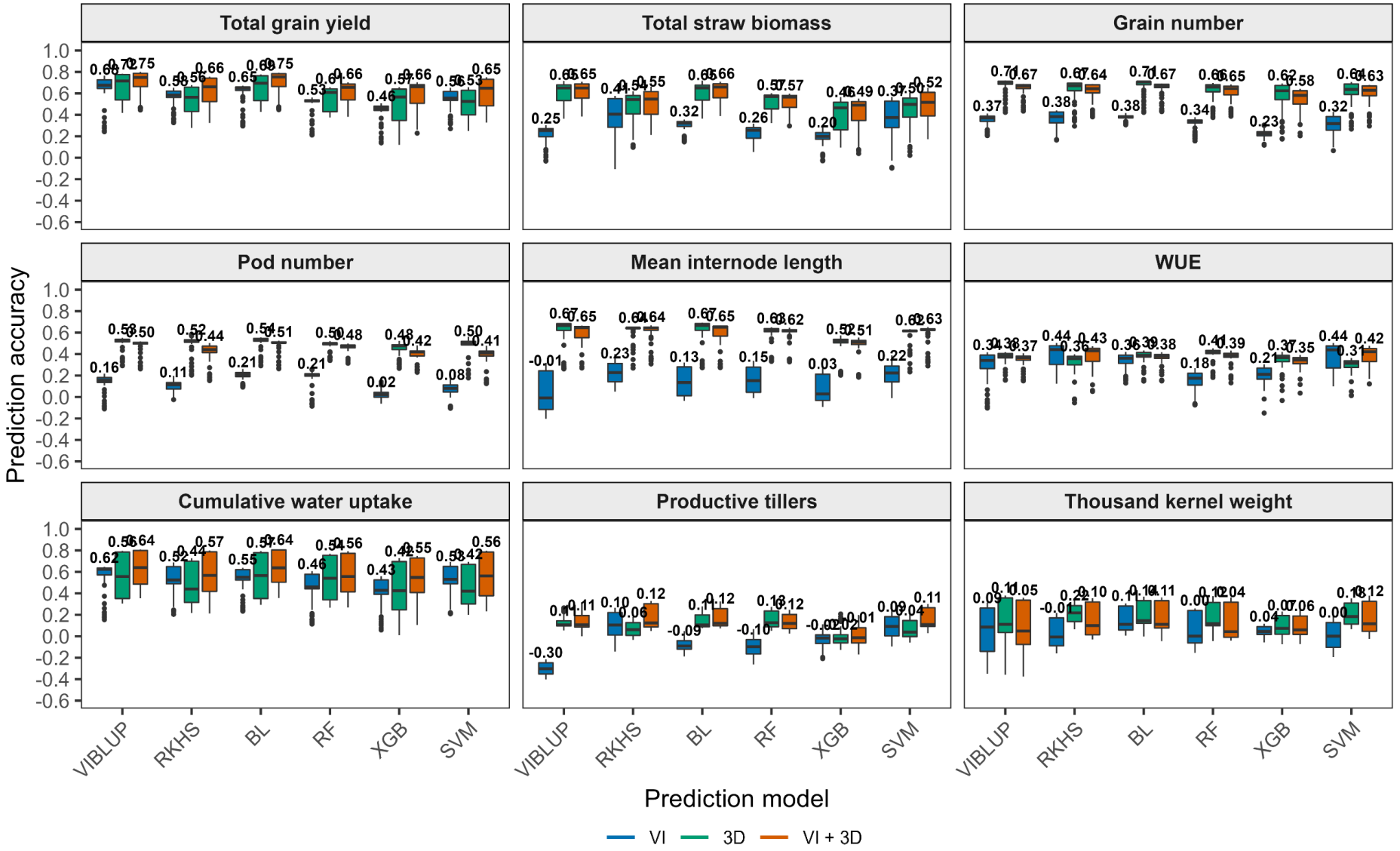
Supplementary Figure S2 Model-specific prediction performance within the combined VI predictor set. Boxplots summarize the distribution of DAS-specific predictive abilities across the season for each prediction model and trait, shown separately for single-day predictions (A) and cumulative predictions (B). Number over the quartile represents the median prediction accuracy.



Supplementary Figure S3 Observed versus predicted values for the selected target traits under the best-performing prediction setup for each trait. Predictions are shown for total grain yield, grain number, cumulative water uptake, and total straw biomass, using the predictor set, model, and day after sowing (DAS) that yielded the highest performance for each trait. Each point represents one genotype, and the predicted value corresponds to the genotype-wise median across all repeated cross-validation runs. The dashed line indicates the 1:1 relationship between observed and predicted values.



Supplementary Figure S4 Grouped boxplots of predictive ability for the Single-day setting across all traits. Within each trait, prediction accuracies are shown for six models and three predictor sets (VI, 3D, and VI + 3D) across the evaluated days after sowing.



Supplementary Figure S5 Grouped boxplots of predictive ability for the cumulative setting across all traits. Within each trait, prediction accuracies are shown for six models and three predictor sets (VI, 3D, and VI + 3D) across the evaluated days after sowing.

Variable	Unit	Definition	Biological meaning
Grain yield	g	Total grain mass per experimental unit.	Final grain production.
Straw biomass	g	Total above-ground straw mass per experimental unit after harvest.	Above-ground non-grain biomass.
Grain number	count	Total number of grains per experimental unit.	Yield component related to sink size.
Pod number	count	Total number of pods per experimental unit.	Yield component reflecting reproductive success.
Mean internode length	cm	Arithmetic mean of internode lengths measured per experimental unit at day of harvest.	Describes stem elongation and canopy architecture.
Water use efficiency	$g\ g^{-1}$	Calculated as grain yield divided by cumulative water uptake.	Grain production per unit of water consumed.
Cumulative water uptake	g	Total amount of water used by the experimental unit.	Plant water consumption over time.
Number of productive tillers	count	Number of tillers bearing harvestable reproductive structures.	Describes effective tillering contributing to yield.
Thousand kernel weight	g	Grain weight expressed as the mass of 1000 kernels.	Reflects average grain size and grain filling.
Green Leaf Index	dimensionless	GLI = (2G - R - B) / (2G + R + B) where R, G, and B denote red, green, and blue reflectance or digital intensity values.	Emphasizes relative canopy greenness.
Normalized Difference Vegetation Index	dimensionless	NDVI = (NIR - R) / (NIR + R), where NIR is near-infrared reflectance and R is red reflectance.	Proxy for canopy vigor and photosynthetically active vegetation.
Normalized Pigment Chlorophyll Index	dimensionless	NPCI = (R - B) / (R + B).	Sensitive to pigment composition and often interpreted in relation to chlorophyll status.
Plant Senescence Reflectance Index	dimensionless	PSRI = (R - B) / NIR.	Indicates canopy senescence and changes in pigment composition.
Hue	software-derived, color trait	Color descriptor derived from RGB information after	Describes the dominant canopy color tone.

		transformation into a color space.	
Lightness	software-derived, color trait	Brightness descriptor derived from transformed RGB information.	Reflects the overall brightness of the canopy surface.
Saturation	software-derived, color trait	Color intensity descriptor derived from transformed RGB information.	Describes how vivid or dull the canopy color appears.
Averaged plant height	cm	Mean canopy height derived from the 3D point cloud.	Represents average vertical development of the canopy.
Maximum plant height	cm	Maximum canopy height detected in the 3D reconstruction.	Captures the tallest plant structures within the canopy.
Digital biomass	software-derived, structural trait	Proxy for canopy biomass estimated from 3D structural information.	Describes total structural canopy mass in a non-destructive digital way.
Projected leaf area	cm^2	Two-dimensional projected area of the canopy as seen from the sensor perspective.	Represents canopy cover and horizontal leaf display.
3D leaf area	cm^2	Surface area estimate derived from the 3D canopy reconstruction.	Reflects total leaf surface represented in the point cloud.
Voxel volume	software-derived, cm^3	Canopy volume estimated by dividing 3D space into voxels and summing occupied voxels.	Describes volumetric canopy size.
Canopy light penetration depth	cm	Depth to which light penetrates into the canopy structure, derived from the 3D reconstruction.	Indicates canopy openness and the internal light environment.
Surface angle	degree	Mean orientation angle of canopy surfaces derived from the 3D reconstruction.	Describes canopy surface inclination and architectural arrangement.
Convex hull area coverage	software-derived, dimensionless	Fraction of the convex hull area occupied by the reconstructed canopy.	Describes how densely the canopy fills its outer boundary.
Convex hull area	cm^2	Area enclosed by the convex hull around the canopy structure or projection.	Represents the outer spatial extent of the canopy.
Convex hull circumference	cm^2	Perimeter of the convex hull enclosing the canopy.	Describes canopy boundary length and horizontal spread.
Convex hull maximum width	cm^2	Maximum horizontal width of the convex hull.	Captures the widest canopy extension.

Convex hull aspect ratio	software-derived, dimensionless	Ratio between major and minor hull dimensions.	Describes canopy shape, especially elongation versus compactness.

Supplementary Table S1 Definition, units, and biological interpretation of response traits and canopy-derived predictor variables used in the prediction analyses. The table summarizes all target traits and sensor-derived predictor variables, including their units, mathematical definitions where applicable, and their biological or structural relevance for characterizing plant performance, canopy color, and three-dimensional canopy architecture.

Trait	Time mode	Scenario	peak_r	peak DAS agg	peak model r	DAS	Model	cor obs pred	rmse	mae	bias	intercept pred on obs	slope pred on obs	mean pred sd	median pred sd
Cumulative water uptake	Cumulative	VI_3D	0.8	97	0.81	94	VIBLUP	0.79	8444.68	6814.1	187.24	41340.94	0.64	2041.12	1967.14
Grain number	Cumulative	D3	0.7	108	0.72	63	VIBLUP	0.69	44.83	35.36	0.06	208.65	0.46	7.44	7.07
Mean internode length	Single-day	VI_3D	0.67	82	0.71	81	BL	0.66	0.28	0.24	0	2.38	0.4	0.04	0.04
Pod number	Cumulative	D3	0.55	63	0.59	63	RKHS	0.54	17.68	13.74	0.05	91.47	0.25	2.66	2.48
Productive tillers	Single-day	VI_3D	0.38	109	0.43	108	SVM	0.43	1.9	1.47	0.08	2	0.23	0.51	0.47
Thousand kernel weight	Single-day	VI_3D	0.43	91	0.51	87	RKHS	0.49	65.89	49.75	-0.22	402.9	0.22	10.29	9.99
Total grain yield	Cumulative	VI_3D	0.75	93	0.8	93	VIBLUP	0.77	18.51	14.16	0.66	79.26	0.6	4.25	4.1
Total straw biomass	Cumulative	VI_3D	0.66	104	0.71	105	BL	0.65	9.86	7.53	-0.05	56.85	0.41	1.67	1.62
WUE	Single-day	D3	0.44	67	0.56	58	BL	0.51	0.11	0.09	0	1.27	0.25	0.01	0.01

Table S2. Final prediction setup and performance metrics for each trait. For each trait, the table reports the selected time mode (Time mode), predictor set (Scenario), peak predictive ability at the aggregated level (peak r) and the corresponding day after sowing (peak DAS agg), followed by the peak predictive ability of the best individual model (peak model r), the selected day after sowing for that model (DAS), and the final selected prediction model (Model). Final predictive performance is further summarized by the correlation between observed and predicted values (cor obs pred), root mean squared error (rmse), mean absolute error (mae), prediction bias (bias), regression intercept (intercept pred on obs) and slope (slope pred on obs) of predicted on observed values, as well as the mean and median standard deviation of genotype wise predictions across seeds (mean pred sd, median pred sd).