



# Early prediction of biomass in hybrid rye based on hyperspectral data surpasses genomic predictability in less-related breeding material

Rodrigo José Galán<sup>1</sup> · Angela-Maria Bernal-Vasquez<sup>2</sup> · Christian Jebsen<sup>2</sup> · Hans-Peter Piepho<sup>3</sup> · Patrick Thorwarth<sup>1,2</sup> · Philipp Steffan<sup>4</sup> · Andres Gordillo<sup>4</sup> · Thomas Miedaner<sup>1</sup>

Received: 21 August 2020 / Accepted: 19 January 2021 / Published online: 17 February 2021  
© The Author(s) 2021

## Abstract

**Key message** Hyperspectral data is a promising complement to genomic data to predict biomass under scenarios of low genetic relatedness. Sufficient environmental connectivity between data used for model training and validation is required.

**Abstract** The demand for sustainable sources of biomass is increasing worldwide. The early prediction of biomass via indirect selection of dry matter yield (DMY) based on hyperspectral and/or genomic prediction is crucial to affordably untap the potential of winter rye (*Secale cereale* L.) as a dual-purpose crop. However, this estimation involves multiple genetic backgrounds and genetic relatedness is a crucial factor in genomic selection (GS). To assess the prospect of prediction using reflectance data as a suitable complement to GS for biomass breeding, the influence of trait heritability ( $H^2$ ) and genetic relatedness were compared. Models were based on genomic (GBLUP) and hyperspectral reflectance-derived (HBLUP) relationship matrices to predict DMY and other biomass-related traits such as dry matter content (DMC) and fresh matter yield (FMY). For this, 270 elite rye lines from nine interconnected bi-parental families were genotyped using a 10 k-SNP array and phenotyped as testcrosses at four locations in two years (eight environments). From 400 discrete narrow bands (410 nm–993 nm) collected by an uncrewed aerial vehicle (UAV) on two dates in each environment, 32 hyperspectral bands previously selected by Lasso were incorporated into a prediction model. HBLUP showed higher prediction abilities (0.41 – 0.61) than GBLUP (0.14 – 0.28) under a decreased genetic relationship, especially for mid-heritable traits (FMY and DMY), suggesting that HBLUP is much less affected by relatedness and  $H^2$ . However, the predictive power of both models was largely affected by environmental variances. Prediction abilities for DMY were further enhanced (up to 20%) by integrating both matrices and plant height into a bivariate model. Thus, data derived from high-throughput phenotyping emerges as a suitable strategy to efficiently leverage selection gains in biomass rye breeding; however, sufficient environmental connectivity is needed.

**Keywords** Biomass · Genetic relatedness · High-throughput phenotyping · Genomic prediction · Prediction ability · Rye

---

Communicated by Thomas Lubberstedt.

✉ Thomas Miedaner  
thomas.miedaner@uni-hohenheim.de

<sup>1</sup> State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany

<sup>2</sup> KWS SAAT SE, Grimsehlstraße 31, 37574 Einbeck, Germany

<sup>3</sup> Biostatistics Unit, Institute of Crop Science, University of Hohenheim, 70593 Stuttgart, Germany

<sup>4</sup> KWS LOCHOW GMBH, Ferdinand-von-Lochow Straße 5, 29303 Bergen, Germany

## Introduction

Worldwide, the consumption of energy obtained from renewable origins, especially bio-based sources, is rising (World Bioenergy Association 2019). In the European Union (EU), for instance, the share of renewable energy is expected to be between 55 and 75% of the total energy consumption in 2050, increasing in proportion the needs for biomass (European Commission 2011). New policy directives have established sustainability guidelines for bioenergy production (European Union 2010). For example, in Germany, the principal European biogas producer, the permitted share of maize (*Zea mays* L.) as the most

common fermentation substrate has been limited to 44% by 2021 (Renewable Energy Sources Act “EEG”, EEG 2017). Thus, suitable alternatives are welcome to diversify maize-based biomass production.

Among the small-grain cereals, winter rye (*Secale cereale* L.) stands out for its vigorous growth and enhanced tolerance to abiotic and biotic stress factors. Europe is the largest rye grower worldwide covering about 81% of the global area with Russia, Poland, and Germany being the main producers (FAO 2019). In a previous study, rye demonstrated its high dry matter yield (DMY) potential even on sandy soils and under drought stress (Galán et al. 2020a). Under these conditions, rye yielded 8.4 t dry matter ha<sup>-1</sup>, and under better environmental conditions, yields were up to 14.7 t dry matter ha<sup>-1</sup>. Rye can, therefore, represent a suitable alternative for biomass production in a variety of agroecological conditions, including areas where the cultivation of other cereal crops would not be competitive (Geiger and Miedaner 2009). Considering that three quarters of the rye harvest is used for non-food purposes, rye appears as a sustainable alternative source of biomass (Geiger and Miedaner 2009; Miedaner et al. 2012).

In Germany, only 4 rye varieties are currently registered for whole plant silage (Bundessortenamt 2019). Rye is, however, mainly bred for grain yield (GY; Haffke et al. 2014) which is, in the breeding scheme proposed here, already assessed in the first year of general combining ability testing (GCA-1), generally sharing only less-related genotypes over the years (Suppl. Fig. 1). Then, within each selection cycle, a selected fraction of GCA-1 is re-evaluated for GY and additionally for DMY by destructive methods in duplicated GCA-2 experiments the following year, mainly due to the high costs of assessing DMY in a large GCA-1 population. At these first selection stages, the enhancement of DMY is, therefore, heavily dependent on the adequate exploitation of indirect selection (Falconer and Mackay 1996).

Higher selection gains have been reported when plant height (PH) was used as a secondary trait instead of GY to indirectly estimate DMY in hybrid rye (Haffke et al. 2014; Galán et al. 2020a). Recently, multi-kernel models jointly using reflectance and genomic data as alternative sources of information and bivariate models including also the routinely assessed PH were suggested as superior strategies to leverage rye as a dual-purpose crop in an affordable manner for the breeder (Galán et al. 2020b). By this, the available genetic variation present in the GCA-1 population may be better exploited without the need to duplicate these large-scale trials and, therefore, the selection gain for DMY could be further enhanced. In consequence, fewer and superior DMY-genotypes being tested in GCA-1 trials could be identified, reducing the amount of capital, time, and labor needed to conduct the destructive sampling of DMY in GCA-2

trials. In this context, the non-destructive assessment of DMY at earlier stages arises as a crucial prerequisite.

Imaging-based phenotyping quantitatively measures the interaction (e.g., absorbance, reflectance, or transmittance of photons) between the incident light and plant tissues, which at specific regions of the electromagnetic spectrum is linked to a wide range of morphological and physio-chemical canopy properties (Li et al. 2014). As observed by Rincent et al. (2018), this interaction is mainly mediated by the chemical composition of the tissues, which is itself determined by endophenotypes, intermediate molecular phenotypes associated with a quantitative trait (Mackay et al. 2009), and genetics. Thus, based on reflectance data, high-throughput phenotyping (HTP) can acquire a considerable amount of detailed phenotypic information of key traits from a large number of genotypes, emerging as a valuable breeding tool (Montes et al. 2007; Cabrera-Bosquet et al. 2012; Würschum 2019).

Examples of the application of HTP in plant breeding are among others, the estimation of above-ground biomass (Babar et al. 2006; Montes et al. 2011; Busemeyer et al. 2013; Fu et al. 2014; Barmeier and Schmidhalter 2017; Yue et al. 2017, 2018) as well as GY, plant responses to biotic and abiotic stress, nitrogen use efficiency, nutrient status, early plant vigor, seeds quality traits, leaf physiology and biochemistry, vegetation cover fraction, and leaf area index (reviewed by Fahlgren et al. 2015; Yang et al. 2017; Würschum 2019). Therefore, it has been proposed to remotely phenotype large breeding populations in a reliable and cost-effective manner (Furbank and Tester 2011; White et al. 2012). HTP platforms, including uncrewed aerial vehicles (UAVs) such as drones mounted with hyperspectral cameras, can simultaneously collect hundreds of high-resolution images, screening the electromagnetic spectrum (from 400 up to 2500 nm) in a continuous mode (Araus and Cairns 2014). Consequently, this noninvasive technology represents a valuable tool for the improvement of complex traits (Finkel 2009; Fiorani and Schurr 2013).

Genome-wide molecular markers integrated into genomic selection (Meuwissen et al. 2001) have been successfully applied in several study cases in hybrid rye breeding for relevant traits, e.g., GY and GY components (Auinger et al. 2016; Bernal-Vasquez et al. 2017; Miedaner et al. 2019). Moreover, in previous studies, reflectance fingerprints recorded by HTP platforms represented a valuable tool to improve the prediction ability of DMY in hybrid rye of models based on agronomic (Galán et al. 2020a) and genomic information (Galán et al. 2020b). These studies have shown the benefits of integrating hyperspectral and molecular information for predicting DMY of unphenotyped candidates within single or closely related populations. The proposed models were cross-validated, where rye lines derived from the same cross were randomly allocated to the training (TRN) or validation (VAL) sets. Considering

the breeding scheme at hand, where DMY is tested at later stages, predictions of candidates of subsequent selection cycles, where TRN and VAL correspond to different, largely independent genetic backgrounds, would be of utmost interest. This “across-cycles” prediction would allow, for instance, estimating the DMY performance of GCA-1 candidates (being tested only for GY at this stage) by training the model with GCA-2 phenotypic data from one or several previous selection cycles (Suppl. Fig. 1). It is under these scenarios where the largest contribution of predictive breeding towards an affordable dual-purpose rye breeding program is expected. If the data available consist of multiple connected cycles, breeders could consider to combine them to improve the predictive power of models (Auinger et al. 2016).

However, the predictive power of GS critically depends on a close relationship between TRN and VAL (Habier et al. 2007; Miedaner et al. 2019). Reduced or even negative prediction accuracies were reported for GS among less related bi- and multiparental families in several crops, including wheat (Herter et al. 2019), maize (Riedelsheimer et al. 2013; Lehermeier et al. 2014), sugar beet (Würschum et al. 2013), and barley (Thorwarth et al. 2017). Similarly, genomic prediction models showed modest prediction ability for complex traits in rye (e.g., GY) when applied between bi-parental families even though they were connected by a common parental line (Wang et al. 2014). Here, the question of whether alternative or complementary approaches to GS for leveraging prediction accuracies across less connected datasets emerges as highly relevant for biomass breeding in rye.

The aim of our study was, therefore, to answer this question by evaluating and comparing genomic- and hyperspectral-enabled predictions for three biomass-related traits (DMY, FMY, and DMC) in rye under a varying degree of relatedness between TRN and VAL. Additionally, the advantages of combining different sources of information in multi-kernel and bivariate models to leverage the prediction of DMY were evaluated. We employed 270 winter rye lines from nine interconnected bi-parental families, including their parental components tested as testcrosses in 8 environments (= location-year combination). While keeping the TRN size constant, our specific objectives were to perform (1) prediction of progenies from half-sib and unrelated parents, (2) prediction using only progenies from unrelated parents, and (3) prediction of new progenies in a new environment.

## Materials and methods

### Plant materials, field experiments, hyperspectral and molecular data

The plant materials, field experiments, molecular and hyperspectral data analyzed in the present study have been

described before in detail by Galán et al. (2020b). In short, ten diverse parental lines of the Petkus (seed parent) gene pool were crossed following a single-round robin design (Verhoeven et al. 2006). F1 progenies were derived from each of the chain crosses, i.e., line 1 × line 2, line 2 × line 3, ..., line 10 × line 1. After self-fertilization of single F1 plants for four consecutive years ( $S_4$  generation), 264 recombinant inbred lines (RILs) were obtained. The ten bi-parental families ranged from 4 up to 32 RILs (Suppl. Fig. S2) and were clearly distinct in a principal component analysis (PCA) based on molecular data with little overlap between unrelated crosses in the first two dimensions (Suppl. Fig. S3). A total of 274 three-way hybrids [(A • B) × C] were produced from the cross of these 264 RILs and their ten parental components with a single-cross tester from the opposite (pollinator) gene pool. They were evaluated in two adjacent trials laid out as a resolvable incomplete block design ( $\alpha$ -lattice design) with two replicates in 2017 and 2018 at each of four ecologically different locations (Bernburg, Petkus, Wohlde and Prislich) in Northern Germany (i.e., eight location-year combinations hereafter referred as “environments”). All 274 testcrosses were used for estimating means, variance components, and heritabilities (Table 2), whereas 4 genotypes were not considered for prediction modeling as described in later sections. Plots were harvested at the late milk stage (Meier 1997) to get the respective fresh biomass yield (FMY, dt ha<sup>-1</sup>) per plot. During harvest, representative samples of about 1000 g were weighed from each plot and oven-dried to a constant weight at 110 °C. Dry matter content (DMC, %) was determined by weight differences. Then, DMY (dt ha<sup>-1</sup>) per plot was estimated as  $DMY = FMY \times DMC / 100$ . Also, PH (cm) was recorded at each plot.

During the grain-filling stage, an UAV (Camflight FX8HL, Sandnes, Norway) fitted with a hyperspectral camera (HySpex Mjolnir V-1240, Skedsmokorset, Norway) collected reflectance fingerprints consisting of 400 bands (410 nm – 993 nm) for all genotypes in all environments. The UAV flew at about 25 m above plots, around solar noon-time two times per environment (except in Bernburg 2017 where only one flight took place). Then each plot was identified on the obtained images by a polygon. Raw data were radiometrically calibrated (HySpex PostProcessor Version 1.2) and normalized based on the incident sunlight as well as orthorectified and georeferenced via the PARGE Software (ReSe Applications LLC, Wil, Switzerland). Lastly, all data points within each wavelength and polygon were averaged, resulting in one spectrum per plot. Then, these data were transferred to a tabular data frame, including the computed reflectance values of all bands for all genotypes for further analysis.

The 264 RILs and their ten parental components were also genotyped with an Illumina INFINIUM chip with 9,963 single nucleotide polymorphisms (SNPs) assays (KWS SAAT

SE & Co. KG, Einbeck, Germany). Data quality analysis consisted of the exclusion of SNPs showing more than 10% missing values or a minor allele frequency (MAF) < 0.05. Missing values in the remaining data were then imputed by the software Linkimpute (Money et al. 2015). Then, data were again screened for MAF < 0.05. After this procedure, 6,420 markers remained for further analyses.

## Phenotypic data analysis

The analyses were based on adjusted entry means (best linear unbiased estimators, BLUEs) for all agronomic traits estimated within and across environments for subsequent incorporation into prediction models. The combined analysis across environments as well as the data adjustment within single environments were conducted following model (1) and model (2) from Galán et al. (2020b), respectively. The full model can also be found in the Supplementary File 1. For the analysis across environments within the same year, the year main effect and corresponding interactions with genotypes were dropped from the mixed model. Phenotypic data were filtered for outliers at the trial level using the Bonferroni-Holm test (Bernal-Vasquez et al. 2016). Bands were deleted from plots identified as an outlier for DMY.

## Stage-wise procedure for biomass traits prediction

The incorporation of genomic and hyperspectral data for predicting DMY, FMY, and DMC was conducted by a three-stage procedure (Piepho et al. 2012). This analysis, together with the corresponding linear mixed and prediction models employed at each stage, was previously described in detail in Galán et al. (2020b). All statistical analyses were performed within the R-environment v. 3.4.4 (R Core Team 2018).

In the first stage of this analysis, bands were adjusted across flight dates per environment. Then, the obtained adjusted entry means (BLUEs) were used in the second stage for the estimation of BLUEs per genotypes across environments. At this second stage, heritability ( $H^2$ ) was estimated for all agronomic traits and each band across environments as

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\bar{v}}{2}}$$

where  $\bar{v}$  is the mean variance of a difference of two adjusted genotype means (BLUEs) estimated for phenotypic and hyperspectral data (Piepho and Möhring 2007). BLUEs of genotypes were calculated with the software package ASReml-R v. 3.0 (Gilmour et al. 2009).

In the third stage, the phenotypic and hyperspectral BLUEs were used for fitting prediction models to estimate

best linear unbiased predictions (BLUP) of genotypic effects for each agronomic trait based on genetic and hyperspectral data. Two single-kernel prediction models were fitted with genetic (genomic BLUP, GBLUP) or hyperspectral (hyperspectral BLUP, HBLUP) data with  $n = 270$  individuals, based on the  $m = 6,420$  conserved SNP markers or  $b = 32$  bands, respectively.

For GBLUP, the random genetic values (effects) were estimated based on genetic data incorporated into  $\mathbf{G}$ , a genomic additive relationship matrix (Habier et al. 2013).  $\mathbf{G}$  was calculated with the *synbreed* package (Wimmer et al. 2012) in R according to the “method I” of VanRaden (VanRaden 2008) as  $G = \frac{ZZ'}{2 \sum p_i(1-p_i)}$ , where  $Z = M - P$ ,  $M$  is the  $n \times m$  marker matrix reflecting the SNP genotype of  $n$ th individual at the  $m$ th SNP position the of alleles coded as 0, 1, and 2 for  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$ , respectively,  $P$  contains a  $n \times m$  matrix of allele frequencies multiplied by 2,  $p_i$  is the allele frequency of the  $i$ th allele. For the prediction scenario S2 (described below), the GBLUP model was adapted from the model (7) in Bernal-Vasquez et al. (2017) as

$$y = Xb + Z_g u_g + Z_{ge} u_{ge} + e \quad (1)$$

where  $\mathbf{y}$  is the vector of BLUEs of genotype trait values obtained from within-environments,  $\mathbf{X}$  is the design matrix of the environments,  $\boldsymbol{\beta}$  is the vector of environments effects,  $\mathbf{Z}_g$  is the marker matrix for genotypes, and  $\mathbf{u}_g$  the vector of marker effects. The genotype-by-environment effects is modelled by  $\mathbf{w} = \mathbf{Z}_{ge} \mathbf{u}_{ge}$ , with  $\mathbf{Z}_{ge}$  standing for the marker matrix for genotypes-by-environment effects and  $\mathbf{u}_{ge}$  the vector of marker-by-environment effects with variance  $\text{var}(\mathbf{u}_{ge}) = \mathbf{I} \sigma_{ge}^2$ , thus  $\text{var}(\mathbf{w}) = \mathbf{Z}_{ge} \mathbf{Z}_{ge}^T \sigma_{ge}^2$ .  $\mathbf{Z}_{ge}$  is a block-diagonal matrix with blocks given by the marker coefficient matrices of genotypes in a given environment ( $\mathbf{Z}_{ge_r}$ ) and for the eight environments considered in the present study, it can

be defined as  $\begin{pmatrix} \mathbf{Z}_{ge_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{Z}_{ge_8} \end{pmatrix}$ . The variance of  $\mathbf{w}$  stands for

the linear structure of the genotype-by-environment variance-covariance matrix with the covariance of two genotypes within the same environment depending on the similarity in their marker profiles (Piepho 2009). Since the covariance among different environments is zero, any covariance between environments is captured by  $\mathbf{Z}_g$ .

As a measure of the genetic similarity among all  $n$  candidates, the Pearson’s coefficients of correlation among rows of  $M$  were calculated. Based on their SNP alleles, this genomic correlation ( $r_{GC}$ ) reflects the correlation pattern among individuals (Riedelsheimer et al. 2013). In contrast, for HBLUP the estimation of the random genetic values was based on reflectance data integrated into the hyperspectral reflectance-based relationship matrix  $\mathbf{H}$  defined as  $\mathbf{H} = \mathbf{D}\mathbf{D}'$ , where  $\mathbf{D}$  is a  $n \times b$  hyperspectral matrix of the standardized

BLUES of the bands, with  $b=32$ . These 32 bands belong to the visible spectrum (VS) and the infrared radiation (IR), and they were selected in a previous study (Galán et al. 2020b) using the least absolute shrinkage and selection operator (Lasso; Tibshirani 1996) for reducing the multicollinearity observed among continuous bands and increasing, therefore, the predictive power of reflectance-based models. Following the same procedure as described before for  $r_{GC}$ , a second correlation ( $r_{HC}$ ) among tested genotypes was developed based on hyperspectral data incorporated into  $\mathbf{H}'$ , which was derived from  $b=400$  available bands. By this, the correlation pattern among lines was estimated based on their unique reflectance fingerprints along the whole spectrum.

For the prediction scenario S1B (described below), the advantages of integrating different information sources to improve the predictive ability of DMY were assessed following the procedures described in Galán et al. (2020b). For this, genetic and hyperspectral data were combined in a multi-kernel prediction model (G+H), which was further extended to a bivariate model (Bivariate\_G+H) by incorporating PH as predictor.

All third-stage prediction models were fitted using the *sommer* package in R (Covarrubias-Pazarán 2016), except model (1), which was fitted within the R package *ASReml-R* v. 3.0 (Gilmour et al. 2009).

## Prediction schemes

To address the objectives of the present study, nine bi-parental families with a size of 24 to 32 individuals (Suppl. Fig. S2) and their parental components were divided into TRN and VAL following different schemes. The family 4×5 was not considered due to its reduced size ( $n=4$ ). The TRN composition varied in a controlled manner for testing the effect of the relatedness between this set and VAL on a genotypic level (S1) and both genotypic and environmental levels simultaneously (S2). An overview of the different prediction schemes is given in Table 1.

In S1, three different scenarios were analyzed, namely S1CV, S1A, and S1B, which have a decreasing genotypic relationship between TRN and VAL. Scenario S1CV

consisted in ninefold cross-validation (CV) of the whole data set (the nine bi-parental families and their parental components), with eight folds were used for model training and the remaining fold for validation purposes. In contrast, in S1A and S1B, a leave-one-out (LOO) family validation scheme was followed. Here, TRN ranged from six to eight bi-parental families, and VAL consisted of single families with variable size. Whereas in S1A half-sibs (HS) and unrelated lines (UR) were sampled in TRN, in S1B, it included only UR. Parental lines of VAL were available for model training only in S1CV. In contrast, under S1A and S1B, the parents of the validation family were excluded from TRN. The remaining eight parents were considered as UR and could be incorporated accordingly. Genotypes were classified as “unrelated” to distinguish this cross type from FS and HS and, therefore, this term does not have the same meaning as in a population genetics.

To avoid the influence of the TRN size on the prediction ability, for all three scenarios, the TRN size was fixed to 174, which was the largest possible common size among scenarios. If TRN was initially larger than 174, a random sampling without replacement was conducted among possible candidates in order to achieve the targeted size of 174. This procedure was repeated 9,000 times, each repetition consisting of a random composition of TRN to assess model error. The phenotypic and hyperspectral data included in S1 validation scenarios were adjusted across the same four (within-year analysis) or eight (combined years analysis) environments.

In S2, the predictive power of models was assessed by a LOO family validation scheme, as described above for S1. An important difference between S1 and S2 scenarios, is that for S2, data not connected to TRN, either by environments nor by genotypes, was used as VAL. For this, data for model training was collected on UR from six to seven bi-parental families at three or seven environments, while validation data came from single families of variable size evaluated at a fourth or eighth disconnected environment, for the within-year or combined years predictions, respectively.

**Table 1** Overview over the validation scenarios (TRN, training set; VAL, validation set; UR, Unrelated; HS, Half sibs; FS, Full sibs; P: Parental lines)

Name	TRN <sup>a</sup>	VAL	Relationship	No. environments sampled <sup>b</sup>	
				TRN	VAL
S1CV	8 random folds	1 Random fold	UR + HS + FS + P	8	8
S1A	8 families	1 Family	UR + HS	8	8
S1B	6 or 7 families	1 Family	UR	8	8
S2	6 or 7 families	1 Family	UR	7	1

<sup>a</sup>The TRN size remained constant across all S1-scenarios ( $n=174$ )

<sup>b</sup>Corresponds to combined years predictions

For all prediction schemes, prediction ability was assessed as the Pearson’s coefficients of correlation  $r$  between predicted breeding values and observed BLUEs derived from the combined analysis across environments

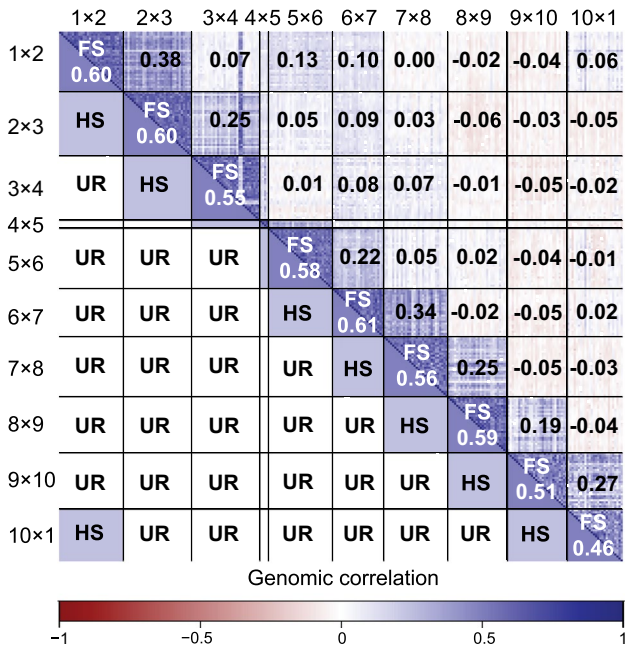
for S1 and data adjustment within single environments for S2.

### Results

#### Population structure, phenotypic and hyperspectral data analysis

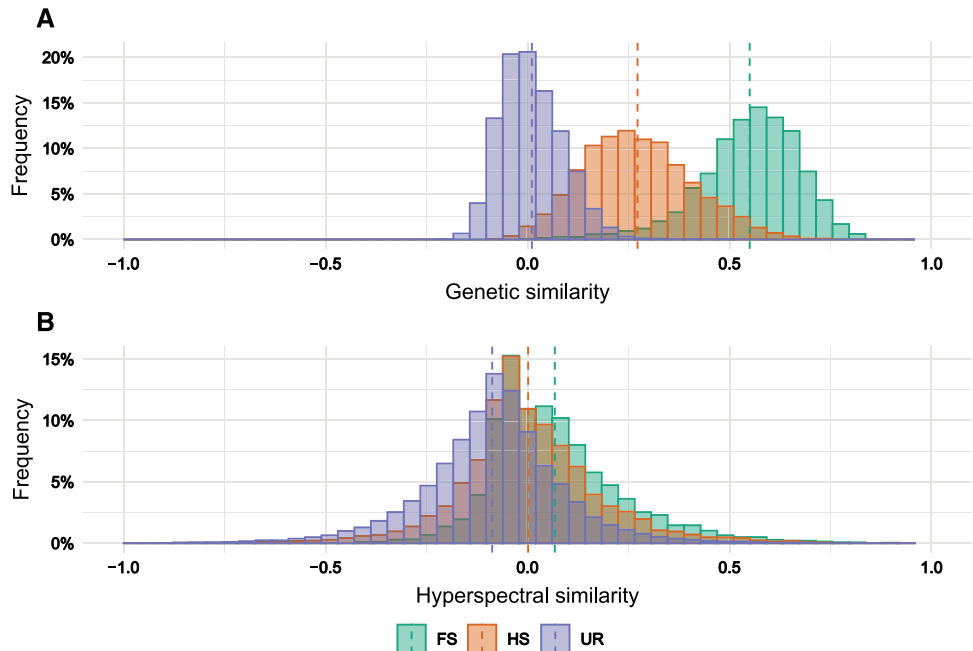
The population showed a genomic correlation pattern (Fig. 1), which clearly reproduces the SRR mating design used in the present study (Suppl. Fig. S2). Thus, the mean genomic relationship among full sibs (FS), half sibs (HS), and unrelated lines (UR) followed the expected decay based on prior pedigree information (Fig. 2a). Nevertheless, a substantial overlap between the  $r_{GC}$  values from HS with FS and UR was observed. The mean  $r_{GC}$  among the nine FS families was 0.55, with a range from 0.61 to 0.46. For HS, the average  $r_{GC}$  was 0.27, almost the mean between FS and UR (0.01). The highest  $r_{GC}$  among HS was 0.38, while the smallest correlation coefficient was 0.06. Among UR,  $r_{GC}$  ranged from 0.04 to  $-0.04$ . Interestingly, no clear distinction among lines could be drawn based on reflectance data (Fig. 2b). The  $r_{HC}$  for FS, HS, and UR was close to zero, with mean estimates equal to 0.07, zero, and  $-0.09$ , respectively.

In 2017, FMY and DMY had higher mean estimates than in 2018 (Table 2). In the first year, these values were 355.96 dt ha<sup>-1</sup> and 124.18 dt ha<sup>-1</sup>, respectively, whereas in the second year, they dropped correspondingly to 304.82 dt ha<sup>-1</sup> and 114.68 dt ha<sup>-1</sup>. The contrary was observed for DMC, which showed a higher mean in 2018 (38.84%) than



**Fig. 1** Heatmap showing the relatedness based on prior pedigree information (below diagonal) and the genomic correlation (above diagonal) among 264 rye lines distributed among ten bi-parental families. The numbers in the blocks refer to average genomic correlations between all pairs of individuals. FS, full sibs; HS, half sibs; UR, unrelated (color figure online)

**Fig. 2** Histograms of (A) genetic similarity and (B) hyperspectral similarity for full sibs (FS), half sibs (HS), and unrelated (color figure online)



**Table 2** Means, ranges, estimates of variance components (genotypic,  $\sigma_g^2$ ; genotype-by-location interaction,  $\sigma_{gl}^2$ ; genotype-by-year-by-location interaction,  $\sigma_{gyl}^2$ ; and residual error  $\sigma_\epsilon^2$ ), heritabilities  $H^2$  determined from 274 winter rye hybrids assessed in two years, which were individually or combined analyzed

Trait <sup>a</sup>	Means and ranges			Variance components				$H^2$
	Mean	Min	Max	$\sigma_g^2$	$\sigma_{gl}^2$	$\sigma_{gyl}^2$	$\sigma_\epsilon^2$	
2017								
FMY (dt ha <sup>-1</sup> )	355.96	332.85	386.17	41.61***	43.76***	–	190.02	0.56
DMY (dt ha <sup>-1</sup> )	124.18	116.63	131.74	4.97***	6.04***	–	16.62	0.53
DMC (%)	35.24	34.02	37.06	0.23***	0.04***	–	0.36	0.80
2018								
FMY (dt ha <sup>-1</sup> )	304.82	284.92	323.87	25.80***	27.15***	–	213.14	0.46
DMY (dt ha <sup>-1</sup> )	114.68	105.85	122.31	5.85***	6.79***	–	26.22	0.54
DMC (%)	38.84	37.21	40.62	0.27***	0.13***	–	1.51	0.70
Combined								
FMY (dt ha <sup>-1</sup> )	330.68	312.29	351.91	21.31***	15.15***	19.04***	203.13	0.47
DMY (dt ha <sup>-1</sup> )	119.48	113.31	126.33	3.41***	2.54***	3.64***	21.49	0.50
DMC (%)	37.02	35.74	38.45	0.23***	0.02	0.07***	0.94	0.81

<sup>a</sup>Traits are fresh matter yield (FMY), dry matter yield (DMY), and dry matter content (DMC)

\*\*\*Significant at the 0.001 probability level

in 2017 (35.24%). The estimated genotypic variance ( $\sigma_g^2$ ) was significantly greater than zero ( $p < 0.001$ ) for all traits. With one minor exception, the same holds for the genotype-by-location interaction ( $\sigma_{gl}^2$ ) and genotype-by-year-by-year interaction ( $\sigma_{gyl}^2$ ) variances. The estimates of  $H^2$  were in general higher in 2017 than in 2018. DMC displayed the higher  $H^2$  estimates, which ranged from 0.70 to 0.81, whereas  $H^2$  for FMY and DMY varied from 0.46 to 0.56. Across the analyzed hyperspectral spectrum,  $H^2$  was highly heterogeneous. The mean value across the 32 selected hyperspectral bands (Suppl. F4) was higher when both years were analyzed together ( $H^2 = 0.63$ ), followed by 2018 ( $H^2 = 0.54$ ) and 2017 ( $H^2 = 0.43$ ). Mean correlations with agronomic traits (considering absolute values) were rather low for all traits ( $r \leq 0.16$ ) with relatively broad ranges (up to  $r \leq 0.41$ , Suppl. Fig. S4). FMY mostly displayed the highest correlation estimates, followed by DMY and DMC.

### Prediction abilities under declining genotypic relationships (S1)

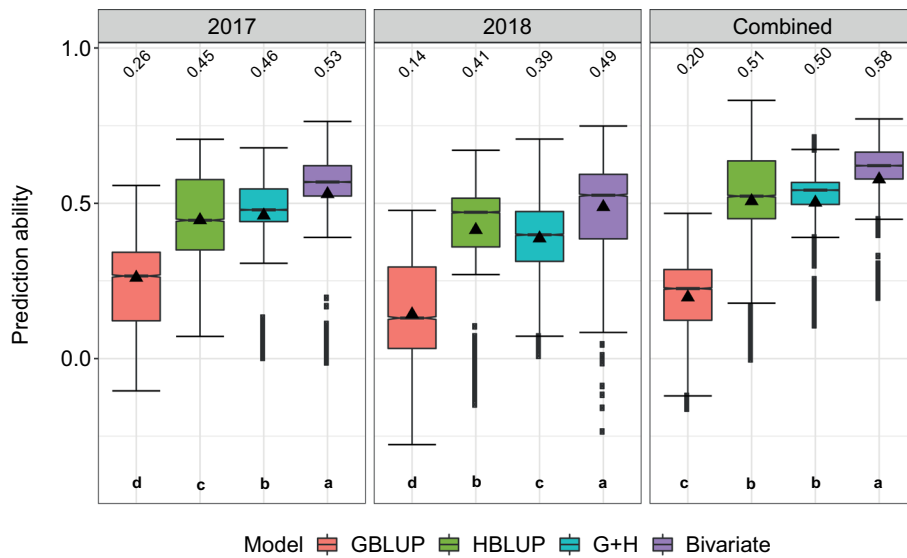
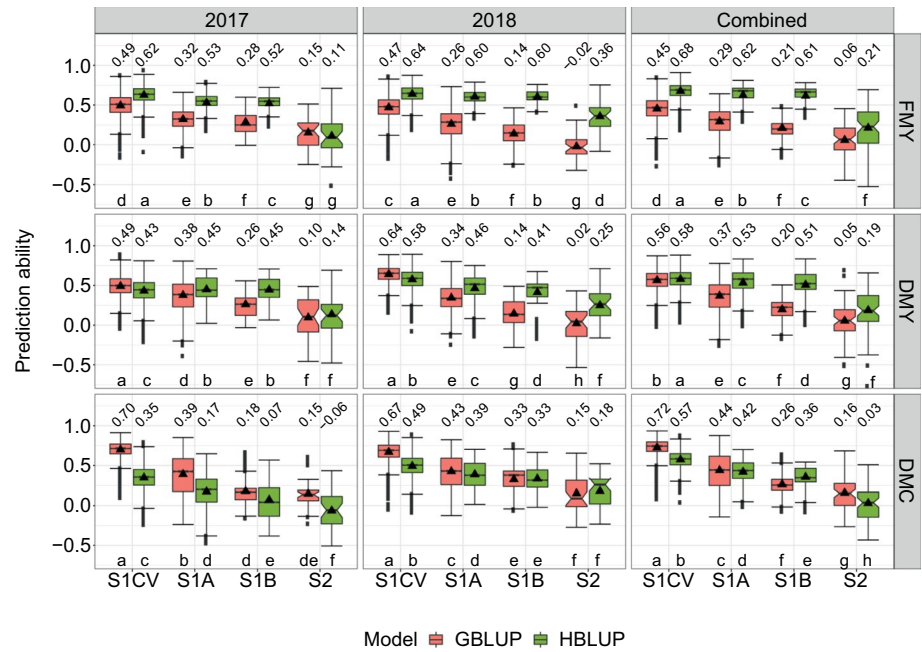
Overall, HBLUP was significantly more accurate than GBLUP for FMY and DMY, while the opposite was observed for DMC (Fig. 3). Combining the data across years was beneficial for HBLUP for all traits, and in the case of GBLUP only for DMC, while for FMY and DMY, GBLUP was mostly more accurate within single-year analysis. The highest prediction abilities for all models and traits were observed under validation scenario S1CV, which has the closest relationship between genotypes used for model training and validation, followed by scenarios S1A and S1B (Fig. 3). However, the impact of a reduced degree of genetic

relatedness between TRN and VAL on the prediction ability was unequal between models. Interestingly, the reduction of the predictive power of GBLUP was considerably higher than for HBLUP when validated on less related sets. For instance, under S1CV for DMY adjusted across years (Fig. 3), GBLUP showed a prediction ability of 0.56, while it dropped to 0.37 and 0.20 under scenarios S1A and S1B, respectively. This represents a decay of about one third and two thirds, respectively. In contrast, the prediction abilities of HBLUP were more stable since they ranged between 0.58 (S1CV) and 0.51 (S1B). Thus, HBLUP retained about 90% of the predicted power shown under S1CV when predicting UR genotypes (S1B). A similar trend was observed for all other traits within as well as across years. The prediction of DMY in validation scenario S1B could be further enhanced by a bivariate model combining hyperspectral and genomic data as well as PH (up to 0.58, Fig. 4). In contrast, the predictive power of the multi-kernel model was very similar to the achieved by HBLUP, although a slight reduction in the variability of the predictions was observed in 2017 and combined-years analysis.

### Predicting environmentally and genetically unconnected candidates (S2)

In the present study, the prediction models were additionally trained on unrelated data, either at a genotypic or environmental level, with the data used for model validation (scenario S2). Under S2, the prediction abilities of all models for all traits was significantly lower and displayed broader ranges when compared to their performance under S1 (Fig. 3). Under S2, HBLUP was also mostly more accurate than GBLUP for FMY and DMY, while the opposite was

**Fig. 3** Prediction abilities for fresh matter yield (FMY), dry matter yield (DMY), and dry matter content (DMC) of genomic (GBLUP) and hyperspectral (HBLUP) best linear unbiased predictions under four different validation schemes assessed across two years (2017 and 2018), which were individually and combined analyzed. Mean values are shown above each box plot and by black triangles and are significantly different, within each subplot, when no letter in common is shared (Tukey’s honestly significant difference test;  $\alpha = 0.01$ ) (color figure online)



**Fig. 4** Prediction abilities for dry matter yield of single-kernel (Genomic best linear unbiased predictor, GBLUP and Hyperspectral best linear unbiased predictor, HBLUP), multi-kernel (G+H), and bivariate (Bivariate\_G+H) models assessed across two years (2017 and 2018), which were individually and combined analyzed. Models

were tested under validation scenario S1B. Mean values are shown above each box plot and by black triangles and are significantly different, within each subplot, when no letter in common is shared (Tukey’s honestly significant difference test;  $\alpha = 0.01$ ) (color figure online)

observed for DMC. Ranges of mean prediction ability of HBLUP for FMY, DMY, and DMC were from 0.11 to 0.36, from 0.14 to 0.25, and from -0.06 to 0.18, respectively, while for GBLUP ranges lay between -0.02 to 0.15, 0.02 to 0.10, and 0.15 to 0.16, respectively. In contrast to S1, no clear benefits of combining the data collected across years were observed under the S2 validation scheme.

### Discussion

The accurate prediction of biomass at early stages via indirect selection for DMY based on GY trials is a fundamental requirement for the implementation of a resource-efficient dual-purpose breeding program in rye. In this way, the entire genetic variance could be exploited, leveraging



the expected selection gain. In our breeding program, each year represents a new selection cycle, where genotypes with different genetic backgrounds are evaluated in new GCA trials. The prediction of subsequent selection cycles implies an additional challenge since the data used for model training and validation are highly unconnected. Nevertheless, it is mainly under this scenario that breeding programs can benefit the most because the biomass improvement can be conducted at the first stage of test-cross evaluation without an increase of the number of field plots. The objective of this study was, therefore, to assess and compare the prediction ability of genetic and hyperspectral data under varying genetic relationships between the training and validation sets.

### Influence of the genetic composition of the TRN and traits characteristics

The degree of relatedness between individuals used for model training (TRN) and validation (VAL) directly influenced the prediction ability of all models; however, this impact was remarkably lower for HBLUP than for GBLUP (Fig. 3). The prediction abilities observed under scenario SICV can be considered as an upper limit, where model training is performed across FS, HS, UR, and parental lines of genotypes used for model validation. Then, a systematic reduction in the predictive performance of all models accompanied the exclusion of genotypes genetically closest to VAL. The exclusion of FS and parental lines from TRN (S1A) represented, averaged across single and combined years analyzes and traits, a reduction of about 40% on the performance of GBLUP, while the further removal of HS signified an additional penalization of around 20%. The larger drop in the prediction abilities observed for S1A compared to those of S1B can be explained by the asymmetrical relevance of using closest relatives for genomic model training (Albrecht et al. 2011; Technow et al. 2014; Juliana et al. 2019). In contrast to GBLUP, the penalization observed for HBLUP in S1A was, on average, only nearly 15% and an additional 6% in S1B, allowing this model to show the highest prediction abilities between the single-kernel models in these scenarios. Model performance was also dissimilar across traits. GBLUP showed mostly the higher abilities for DMC, whereas HBLUP performed better for FMY and DMY. The differences in predictive abilities are most likely a consequence of both trait  $H^2$  and the different information sources used by GS and reflectance-based models.

To adequately predict the performance of untested candidates, genomic models exploit the genetic relationships between them and individuals whose genotypic and phenotypic information is available, as previously shown in

many empirical and simulation-based studies in animal and plant breeding (Habier et al. 2007 2010; Roos et al. 2009; Pszczola et al. 2012; Riedelsheimer et al. 2013; Würschum et al. 2013; Crossa et al. 2014; Lehermeier et al. 2014; Technow et al. 2014; Wang et al. 2014; Thorwarth et al. 2017; Herter et al. 2019). In line with these observations, our results also showed that the predictive power of GS dropped substantially when predictions are made among lowly related populations. For predictions across subsequent cycles in rye, GS could represent a suitable strategy when TRN is represented by aggregated multi-year data from several cycles (Auinger et al. 2016; Bernal-Vasquez et al. 2017). Nevertheless, the authors of these papers concluded that GS still relies heavily on a sufficient relationship between predicted candidates and those used for model training. Selection cycles need, for instance, to be connected by a sufficient number of common ancestors. This prerequisite may not be easily fulfilled in practical rye breeding since subsequent breeding cycles usually are largely unconnected. In addition, the success of GS depends, among others, on trait related factors, such as heritability (Jia and Jannink 2012; Marulanda et al. 2015). Thus, the better and less variable GBLUP performance observed for DMC (Fig. 3) is likely explained by the larger  $H^2$  estimated for this trait in comparison to FMY and DMY (Table 2).

The reflectance fingerprints of the genotypes were more similar than their allelic status across relationship groups (Fig. 2), suggesting that the information imprinted among the spectrum is less sensitive to genetic distinctiveness among individuals than molecular data. These observations can explain why reflectance data allowed higher prediction abilities than marker data under decreased genetic relationships between TRN and VAL. In contrast to GBLUP, more highly heritable traits were not better predicted by HBLUP. In turn, for HBLUP to perform well, plant canopies should display specific absorption patterns related to some extent to the trait of interest as shown, for instance, by the correlations between the analyzed traits and bands. The most effectively predicted traits (FMY and DMY) showed higher correlations than the lowest predicted trait (DMC, Suppl. Fig. S4). Thus, the higher performance of HBLUP for FMY and DMY might be explained by the higher informativeness of the collected reflectance data for those traits than for DMC. Since the absorption of water and DMC is almost constant across the visible spectrum and the absorbance of these two features starts around 950 nm (Jacquemoud et al. 2000), where our spectrum was from 410 nm to 993 nm, further research could investigate the prospects of HBLUP based on reflectance data beyond 1000 nm to better predict DMC.

Several strategies have been investigated for taking advantage of reflectance data in predictive breeding.

Summarizing the reflectance characteristics of plants into simple vegetation indices (VIs) has been proposed to assess vegetation characteristics of interest like grain and biomass yields under different environmental conditions (Xue and Su 2017). However, prediction models benefited the most by the exploitation of whole-spectrum data (Aguate et al. 2017; Montesinos-López et al. 2017b; Krause et al. 2019; Galán et al. 2020b). Recently, highly heritable VIs genetically correlated with the trait of interest such as the Normalized Difference Vegetation Index (NDVI; Rouse et al. 1974; Tucker 1979) and the green NDVI (GNDVI; Gitelson et al. 1996), have been incorporated as secondary traits into multivariate pedigree and genomic prediction models to increase accuracy within the same wheat population and selection cycle (Rutkoski et al. 2016; Sun et al. 2017) as well as across selection cycles composed by closely related populations (Sun et al. 2019). Juliana et al. (2019) found that similar multivariate equations were superior to univariate genomic prediction models when predicting across populations and years. Still, the relationship between TRN and VAL was found crucial also for multivariate models, although the populations used for model training and validation were genetically related to some extent, and predictions were made within the same stressed environments. The results of the present study also showed that combining hyperspectral and genomic data in a multi-kernel model yielded only limited advantages over HBLUP for DMY prediction of less related progenies (Fig. 4). In this context, the prediction ability for DMY could be further increased up to 20% by a bivariate model including also PH. Nevertheless, the performance of G + H and the bivariate model in the present study were lower than when used for DMY prediction of highly related rye progenies, as reported in a previous research (Galán et al. 2020b). These findings reveal, on the one hand, the advantages of incorporating HTP data into prediction routines, and, on the other hand, the limits of GS in the context of across cycle predictions.

## Prediction of new genotypes in untested environments

In validation scenario S1B, UR genotypes were assessed across the same environments (Table 1). In contrast, in S2, unrelated individuals were tested under new environmental conditions, allowing the simultaneous assessment of the genotypic and environmental sampling on the predictive power of marker- and hyperspectral-based models. Predictive abilities in S2 were significantly lower than in S1B, suggesting that predicting the performance of genetically and environmentally highly unconnected individuals is challenging. This is consistent with studies showing that the prediction of new

candidates is less accurate when model training is performed without borrowing information of environments correlated to the one used for validation (Crossa et al. 2014; Krause et al. 2019). These poor predictions obtained in S2 might be explained by the substantial genotype-by-environment interactions ( $G \times E$ ) estimated for the predicted traits (Table 2) as well as by the high variability observed for hyperspectral bands among environments, resulting mainly from the extremely different conditions observed between growing seasons as reported in a previous study (Galán et al. 2020b). It seems, therefore, plausible that heterogeneous marker-to-trait and band-to-trait (Montesinos-López et al. 2017a) signals among environments adversely affected the prediction abilities from GBLUP and HBLUP. Therefore, to adequately predict untested genotypes under new environmental conditions, prediction equations need to be extended by environmental and genetic covariates for proper  $G \times E$  modeling (Piepho 2009; Burgueño et al. 2012; Resende et al. 2020).

A forward-validation approach aims to predict the performance of new genotypes by exploiting the data from previous years (Bernal-Vasquez et al. 2017). Considering our breeding scheme (Suppl. Fig. 1), data for model training could be obtained from split GCA-2 trials with biomass and grain yield plots, whereas model validation could be performed on GCA-1 data from a subsequent selection cycle. It should be kept in mind that we need large-drilled plots for biomass model training because this trait cannot be reliably measured on smaller observation plots. As different selection cycles involve new individuals from multiple genetic backgrounds, and usually hardly any common progeny is shared across cycles, the genetic relationship between the data used for model training and validation across cycles is expected to be substantially lower than in within-cycle predictions.

However, data used in across-cycles predictions is environmentally connected because GCA-2 genotypes are tested more intensively in a larger number of locations, within which the same environments as in GCA-1 are typically found. In practical plant breeding, large testing locations within the targeted environment are common, since they are more efficient in terms of logistics, trained personnel requirements, as well as field evaluation and management. In these testing sites, yield trials from different stages are planted next to each other, being reliable, large-scaled training data readily available for model calibration. Thus, scenario S1B mimics this practical situation much better than S2. Our results showed that, in this context, models incorporating hyperspectral data emerge as a promising strategy to achieve superior improvements in DMY in hybrid rye. Still, the relevance of S2 outcomes lies in a consistently unbiased estimation of the prediction abilities of the models (Utz et al. 2000), revealing the high impact of  $G \times E$  not only on GBLUP but also on HBLUP.

## Conclusions for biomass breeding in hybrid rye

Traditionally, biomass is estimated destructively at an earlier growth stage, preventing grain yield from being evaluated in those same plots. The effective indirect assessment of biomass at the early stages of the breeding program is crucial to entirely untap the potential of rye as a dual-purpose crop affordably. In this sense, prediction models accurately estimating the biomass yield of genotypes of diverse genetic backgrounds across selection cycles represent a valuable tool. In the present study, GBLUP achieved acceptable prediction abilities only for highly heritable traits across closely related individuals. In contrast, HBLUP was substantially less affected by genetic relatedness and trait heritability emerging as a suitable approach for predicting complex traits across highly distinct populations.

Considering that in modern plant breeding genomic information is usually already available before the candidate lines are evaluated as testcrosses in the expensive GCA trials, breeders usually perceive marker and HTP data as a complement, rather than an alternative. Here, HTP offers the possibility of screening large-scale field trials with reduced capital and time expenditures, than conventional methods (e.g. destructive sampling and visual scores). Moreover, combining hyperspectral, genomic, and PH in bivariate models allows more effective DMY predictions of genotypes showing low genetic connectivity to ones used for model training. The bivariate model here presented is flexible and allows the incorporation of GY and other correlated traits to DMY aiming superior predictive power. Nonetheless, by including several predictors, the complexity of the models increases in proportion.

Our results also show that not only GBLUP but also HBLUP was largely affected by  $G \times E$  interactions, resulting in poor to negligible predictive power when the environments used for model training and validation were different. To fully exploit the advantages of hyperspectral-based models, it is, therefore, highly recommended to incorporate reflectance fingerprints of genotypes collected in the respective environment. Our study demonstrates the capability of hyperspectral-enabled predictions to leverage selection gains to meet the increasing demand for sustainable biomass sources worldwide. Lastly, the prospects of HTP as an economical alternative to traditional biomass sampling are expected to increase in proportion to future improvements in terms of image data acquisition and management.

**Supplementary Information** The online version contains supplementary material available at (<https://doi.org/10.1007/s00122-021-03779-1>).

**Acknowledgements** This study was funded by the German Federal Ministry of Food and Agriculture (BMEL) through the German Agency

for Renewable Resources (FNR), grant number FKZ 22019716 to TM. We gratefully acknowledge the excellent support of the technical staff at each experimental station. We are particularly grateful to Hans-Otto Wegener, Jörn-Claus Gudehus, Karsten Sell, KWS LOCHOW GmbH, Bergen, Germany, for seed production and conducting field trials. We also thank Dr. Peer Wilde, KWS LOCHOW GmbH, Bergen, for his valuable contributions to this project.

**Author contributions statement** RG analyzed the data and wrote the manuscript, AMBV, HPP, and PT supported with statistical advice, CJ conducted hyperspectral phenotyping at all environments, PS supervised data collection at Wohlde, Prislisch, and Bernburg and provided scientific advice, TM and AG designed the research project and edited the manuscript. All the authors read and approved the manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical standards** The authors declare that the experiments comply with the current laws of Germany.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aguate FM, Trachsel S, Pérez LG, Burgueño J, Crossa J, Balzarini M, Gouache D, Bogard M, Gdl C (2017) Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield. *Crop Sci* 57(5):2517–2524. <https://doi.org/10.2135/crosci2017.01.0007>
- Albrecht T, Wimmer V, Auinger H-J, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön C-C (2011) Genome-based prediction of testcross values in maize. *Theor Appl Genet* 123(2):339. <https://doi.org/10.1007/s00122-011-1587-7>
- Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19(1):52–61. <https://doi.org/10.1016/j.tplants.2013.09.008>
- Auinger H-J, Schönleben M, Lehermeier C, Schmidt M, Korzun V, Geiger HH, Piepho H-P, Gordillo A, Wilde P, Bauer E (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *Theor Appl Genet*. 129(11):2043–2053. <https://doi.org/10.1007/s00122-016-2756-5>
- Babar MA, Reynolds MP, van Ginkel M, Klatt AR, Raun WR, Stone ML (2006) Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in

- wheat. *Crop Sci* 46(3):1046–1057. <https://doi.org/10.2135/cropsci2005.0211>
- Barmeier G, Schmidhalter U (2017) High-throughput field phenotyping of leaves, leaf sheaths, culms and ears of spring barley cultivars at anthesis and dough ripeness. *Front in Plant Sci* 8:1920. <https://doi.org/10.3389/fpls.2017.01920>
- Bernal-Vasquez A-M, Utz H-F, Piepho H-P (2016) Outlier detection methods for generalized lattices: a case study on the transition from ANOVA to REML. *Theor Appl Genet* 129(4):787–804. <https://doi.org/10.1007/s00122-016-2666-6>
- Bernal-Vasquez A-M, Gordillo A, Schmidt M, Piepho H-P (2017) Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. *BMC Genet* 18(1):51. <https://doi.org/10.1186/s12863-017-0512-8>
- Bundessortenamt (2019) Beschreibende Sortenliste Getreide, Mais, Öl-und Faserpflanzen, Leguminosen, Rüben. Zwischen-früchte, Hannover, Bundessortenamt
- Burgueño J, de Los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype  $\times$  environment interaction using pedigree and dense molecular markers. *Crop Sci* 52(2):707–719. <https://doi.org/10.2135/cropsci2011.06.0299>
- Busemeyer L, Ruckelshausen A, Möller K, Melchinger AE, Alheit KV, Maurer HP, Hahn V, Weissmann EA, Reif JC, Würschum T (2013) Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. *Sci Rep* 3:2442
- Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Luis Araus J (2012) High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J Integr Plant Biol* 54(5):312–320. <https://doi.org/10.1111/j.1744-7909.2012.01116.x>
- Covarrubias-Pazarán G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0156744>
- Crossa J, Pérez P, Hickey J, Burgueño J, Ornela L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112(1):48–60. <https://doi.org/10.1038/hdy.2013.16>
- de Roos APW, Hayes BJ, Goddard ME (2009) Reliability of genomic predictions across multiple populations. *Genetics* 183(4):1545–1553. <https://doi.org/10.1534/genetics.109.104935>
- EEG (2017) Gesetz für den ausbau erneuerbarer energien (Erneuerbare-Energien-Gesetz - EEG). [http://www.gesetze-im-internet.de/eeeg\\_2014/EEG\\_2017.pdf](http://www.gesetze-im-internet.de/eeeg_2014/EEG_2017.pdf). Accessed 02 Nov 2019
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Essex
- Fahlgren N, Gehan MA, Baxter I (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr Opin Plant Biol* 24:93–99
- FAO (2019) FAOSTAT database. Food and agriculture organization of the united nations. <http://www.fao.org/faostat/en/#data/QC>. Accessed 05 Nov 2019
- European Commission (2011) Energy roadmap 2050. communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions, com (2011) 885 final. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0885:FIN:EN:PDF>. Accessed 02 Nov 2019
- Finkel E (2009) With ‘phenomics’, plant scientists hope to shift breeding into overdrive. *Science* 325(5939):380–381. [https://doi.org/10.1126/science.325\\_380](https://doi.org/10.1126/science.325_380)
- Fiorani F, Schurr U (2013) Future scenarios for plant phenotyping. *Annu Rev Plant Biol* 64:267–291. <https://doi.org/10.1146/annurev-arplant-050312-120137>
- Fu Y, Yang G, Wang J, Song X, Feng H (2014) Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. *Comput Electron Agric* 100:51–59. <https://doi.org/10.1016/j.compag.2013.10.010>
- Furbank RT, Tester M (2011) Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* 16(12):635–644. <https://doi.org/10.1016/j.tplants.2011.09.005>
- Galán RJ, Bernal-Vasquez A-M, Jebsen C, Piepho H-P, Thorwarth P, Steffan P, Gordillo A, Miedaner T (2020a) Hyperspectral reflectance data and agronomic traits can predict biomass yield in winter rye hybrids. *BioEnergy Res* 13(1):168–182. <https://doi.org/10.1007/s12155-019-10080-z>
- Galán RJ, Bernal-Vasquez A-M, Jebsen C, Piepho H-P, Thorwarth P, Steffan P, Gordillo A, Miedaner T (2020b) Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-020-03651-8>
- Geiger HH, Miedaner T (2009) Rye breeding. In: Carena MJ (ed) *Cereals*, vol 3. Springer, New York, pp 157–181
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R, Butler D (2009) *ASReml user guide release 3.0*. VSN International Ltd, UK
- Gitelson AA, Kaufman YJ, Merzlyak MN (1996) Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens Environ* 58(3):289–298
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177(4):2389–2397. <https://doi.org/10.1534/genetics.107.081190>
- Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet Sel Evol* 42(1):5. <https://doi.org/10.1186/1297-9686-42-5>
- Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194(3):597–607. <https://doi.org/10.1534/genetics.113.152207>
- Haffke S, Kusterer B, Fromme FJ, Roux S, Hackauf B, Miedaner T (2014) Analysis of covariation of grain yield and dry matter yield for breeding dual use hybrid rye. *BioEnergy Res* 7(1):424–429. <https://doi.org/10.1007/s12155-013-9383-7>
- Herter CP, Ebmeyer E, Kollers S, Korzun V, Würschum T, Miedaner T (2019) Accuracy of within-and among-family genomic prediction for Fusarium head blight and Septoria tritici blotch in winter wheat. *Theor Appl Genet* 132(4):1121–1135. <https://doi.org/10.1007/s00122-018-3264-6>
- Jacquemoud S, Bacour C, Poilve H, Frangi J-P (2000) Comparison of four radiative transfer models to simulate plant canopies reflectance: direct and inverse mode. *Remote Sens Environ* 74(3):471–481. [https://doi.org/10.1016/S0034-4257\(00\)00139-5](https://doi.org/10.1016/S0034-4257(00)00139-5)
- Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192(4):1513–1522. <https://doi.org/10.1534/genetics.112.144246>
- Juliana P, Montesinos-López OA, Crossa J, Mondal S, González Pérez L, Poland J, Huerta-Espino J, Crespo-Herrera L, Govindan V, Dreisigacker S, Shrestha S, Pérez-Rodríguez P, Pinto Espinosa F, Singh RP (2019) Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theor Appl Genet* 132(1):177–194. <https://doi.org/10.1007/s00122-018-3206-3>
- Krause MR, González-Pérez L, Crossa J, Pérez-Rodríguez P, Montesinos-López O, Singh RP, Dreisigacker S, Poland J, Rutkoski J, Sorrells M, Gore MA, Mondal S (2019) Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3 Genes Genomes Genet* 9(4):1231–1247. <https://doi.org/10.1534/g3.118.200856>
- Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger AE, Menz M, Meyer N, Moreau L, Moreno-González J, Ouzunova M, Pausch H, Ranc N, Schipprack

- W, Schönleben M, Walter H, Charcosset A, Schön C-C (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics* 198(1):3–16. <https://doi.org/10.1534/genetics.114.161943>
- Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. *Sensors* 14(11):20078–20111. <https://doi.org/10.3390/s141120078>
- Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10(8):565–577. <https://doi.org/10.1038/nrg2612>
- Marulanda JJ, Melchinger AE, Würschum T (2015) Genomic selection in biparental populations: assessment of parameters for optimum estimation set design. *Plant Breed* 134(6):623–630. <https://doi.org/10.1111/pbr.12317>
- Meier U (1997) Growth stages of mono- and dicotyledonous plants. Blackwell Wissenschafts-Verlag, Berlin
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829
- Miedaner T, Koch S, Seggl A, Schmiedchen B, Wilde P (2012) Quantitative genetic parameters for selection of biomass yield in hybrid rye. *Plant Breed* 131(1):100–103. <https://doi.org/10.1111/j.1439-0523.2011.01909.x>
- Miedaner T, Korzun V, Bauer E (2019) Genomics-based hybrid rye breeding. In: Miedaner T, Korzun V (eds) Applications of genetic and genomic research in cereals. Elsevier, Amsterdam, Netherlands, pp 329–348
- Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong G-Y, Myles S (2015) LinkImpute fast and accurate genotype imputation for nonmodel organisms. *G3 Genes Genomes Genet* 5(11):2383–2390. <https://doi.org/10.1534/g3.115.021667>
- Montes JM, Melchinger AE, Reif JC (2007) Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci* 12(10):433–436. <https://doi.org/10.1016/j.tplants.2007.08.006>
- Montes JM, Technow F, Dhillon BS, Mauch F, Melchinger AE (2011) High-throughput non-destructive biomass determination during early plant development in maize under field conditions. *Field Crops Res* 121(2):268–273
- Montesinos-López A, Montesinos-López OA, Cuevas J, Mata-López WA, Burgueño J, Mondal S, Huerta J, Singh R, Autrique E, González-Pérez L, Crossa J (2017a) Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods* 13:62. <https://doi.org/10.1186/s13007-017-0212-4>
- Montesinos-López OA, Montesinos-López A, Crossa J, de Los Campos G, Alvarado G, Suchismita M, Rutkoski J, González-Pérez L, Burgueño J (2017b) Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13(1):4. <https://doi.org/10.1186/s13007-016-0154-2>
- Piepho H-P (2009) Ridge regression and extensions for genome-wide selection in maize. *Crop Sci* 49(4):1165–1176. <https://doi.org/10.2135/cropsci2008.10.0595>
- Piepho H-P, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177(3):1881–1888. <https://doi.org/10.1534/genetics.107.074229>
- Piepho H-P, Moehring J, Schulz-Streeck T, Ogutu JO (2012) A stage-wise approach for the analysis of multi-environment trials. *Biom J* 54(6):844–860. <https://doi.org/10.1002/bimj.201100219>
- Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *J Dairy Sci* 95(1):389–400. <https://doi.org/10.3168/jds.2011-4338>
- R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Resende RT, Piepho H-P, Silva-Junior OB, Silva FF, Resende MDV, Grattapaglia D (2020) Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. *Theor Appl Genet*
- Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink J-L, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. *Genetics* 194(2):493–503. <https://doi.org/10.1534/genetics.113.150227>
- Rincent R, Charpentier J-P, Faivre-Rampant P, Paux E, Le Gouis J, Bastien C, Segura V (2018) Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. *G3 Genes Genomes Genet* 8(12):3961–3972. <https://doi.org/10.1534/g3.118.200760>
- Rouse JW, Haas RH, Schell JA, Deering DW (1974) Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351:309–3017
- Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, Reynolds M, Singh R (2016) Canopy temperature and vegetation indices from highthroughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genet* 6(9):2799–2808. <https://doi.org/10.1534/g3.116.032888>
- Sun J, Rutkoski JE, Poland JA, Crossa J, Jannink J-L, Sorrells ME (2017) Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genom.* <https://doi.org/10.3835/plantgenome2016.11.0111>
- Sun J, Poland JA, Mondal S, Crossa J, Juliana P, Singh RP, Rutkoski JE, Jannink J-L, Crespo-Herrera L, Velu G, Huerta-Espino J, Sorrells ME (2019) High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. *Theor Appl Genet* 132(6):1705–1720. <https://doi.org/10.1007/s00122-019-03309-0>
- Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197(4):1343–1355. <https://doi.org/10.1534/genetics.114.165860>
- Thorwarth P, Ahlemeyer J, Bochar A-M, Krumnacker K, Blümel H, Laubach E, Knöchel N, Cselényi L, Ordon F, Schmid KJ (2017) Genomic prediction ability for yield-related traits in German winter barley elite material. *Theor Appl Genet* 130(8):1669–1683. <https://doi.org/10.1007/s00122-017-2917-1>
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (Methodol)* 58(1):267–288
- Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ* 8(2):127–150
- Union E (2010) Communication from the Commission on the practical implementation of the EU biofuels and bioliquids sustainability scheme and on counting rules for biofuels (2010/C 160/02). *Off J Eur Un* 2:8–16
- Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* 154(4):1839–1849
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91(11):4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Verhoeven KJF, Jannink JL, McIntyre LM (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity* 96(2):139–149. <https://doi.org/10.1038/sj.hdy.6800763>
- Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of

- marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC Genom* 15(1):556. <https://doi.org/10.1186/1471-2164-15-556>
- White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, Conley MM, Feldmann KA, French AN, Heun JT, Hunsaker DJ (2012) Field-based phenomics for plant genetics research. *Field Crops Res* 133:101–112. <https://doi.org/10.1016/j.fcr.2012.04.003>
- Wimmer V, Albrecht T, Auinger H-J, Schön C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28(15):2086–2087. <https://doi.org/10.1093/bioinformatics/bts335>
- World bioenergy association (2019) Global bioenergy statistics 2019. [https://worldbioenergy.org/uploads/191129%20WBA%20GBS%202019\\_LQ.pdf](https://worldbioenergy.org/uploads/191129%20WBA%20GBS%202019_LQ.pdf). Accessed 17 Jul 2020
- Würschum T (2019) Modern field phenotyping opens new avenues for selection. In: Miedaner T, Korzun V (eds) *Applications of genetic and genomic research in cereals*. Elsevier, Amsterdam, Netherlands, pp 233–250
- Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. *BMC Genet* 14(1):85. <https://doi.org/10.1186/1471-2156-14-85>
- Xue J, Su B (2017) Significant remote sensing vegetation indices: a review of developments and applications. *J Sens* 2017:1–17. <https://doi.org/10.1155/2017/1353691>
- Yang G, Liu J, Zhao C, Li Z, Huang Y, Yu H, Xu B, Yang X, Zhu D, Zhang X (2017) Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front Plant Sci* 8:1111. <https://doi.org/10.3389/fpls.2017.01111>
- Yue J, Yang G, Li C, Li Z, Wang Y, Feng H, Xu B (2017) Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. *Remote Sens* 9(7):708. <https://doi.org/10.3390/rs9070708>
- Yue J, Feng H, Yang G, Li Z (2018) A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. *Remote Sens* 10(1):66. <https://doi.org/10.3390/rs10010066>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.