

Original Article

Integrating genomic prediction and genotype specific parameter estimation in ecophysiological models: overview and perspectives

Pratishtha Poudel^{1,4}, Bryan Naidenov^{2,†}, Charles Chen^{2,†}, Phillip D. Alderman³ and Stephen M. Welch^{1,†,*}

¹Department of Agronomy, Kansas State University, Manhattan, Kansas 66506, USA

²Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, Oklahoma 74078, USA

³Department of Plant and Soil Sciences, Oklahoma State University, Stillwater, Oklahoma 74078, USA

⁴Present address: Department of Agronomy, Purdue University, West Lafayette, IN 47906, USA

[†]Equal contribution.

*Corresponding author's e-mail address: Stephen M. Welch, Department of Agronomy, Kansas State University, 2104 Throckmorton Plant Science Center, Manhattan, Kansas 66506, USA; E-mail: welchsm@ksu.edu

Citation: Poudel P, Naidenov B, Chen C, Alderman PD, Welch SM. 2023. Integrating genomic prediction and genotype specific parameter estimation in ecophysiological models: overview and perspectives. *In Silico Plants* 2023: diad007; doi: 10.1093/insilicoplants/diad007

Handling editor: Graeme Hammer

ABSTRACT

The Genome-to-Phenome (G2P) problem is one of the highest-priority challenges in applied biology. Ecophysiological crop models (ECM) and genomic prediction (GP) models are quantitative algorithms, which, when given information on a genotype and environment, can produce an accurate estimate of a phenotype of interest. In this article, we discuss how the GP algorithms can be used to estimate genotype-specific parameters (GSPs) in ECMs to develop robust prediction methods. In this approach, the numerical constants (GSPs) that ECMs use to distinguish and characterize crop cultivars/varieties are treated as quantitative traits to be predicted by genomic prediction models from underlying genetic information. In this article we provide information on which GP methods appear favorable for predicting different types of GSPs, such as vernalization sensitivity or potential radiation use efficiency. For each example GSP, we assess a number of GP methods in terms of their suitability using a set of three criteria grounded in genetic architecture, computational requirements, and the use of prior information. In general, we conclude that the most useful algorithms were dependent on both the nature of the particular GSP and the GP methods considered.

KEYWORDS: Ecophysiological model; Genomic prediction model; Genotype-specific parameters.

1. INTRODUCTION

It is widely agreed that agricultural crop production is not on track to meet the production doubling needed by 2050 for humanity to avoid major food security disruption (Godfray et al., 2010). Recognized mitigation strategies include (1) accelerating genetic gains and (2) employing improved in-field production management methods such as precision agriculture. Both of these approaches require an enhanced ability to predict the behaviors of novel genetic lines subjected to innovative management programs in potentially non-analog environments. Acquiring such a predictive ability defines the genome-to-phenome (G2P) problem – declared by the US National Research Council to be one of the highest-priority challenges

in applied biology (Council, 2009). Breeding programs require predictions to efficiently decide on which crosses to make. Farmers need genetically-informed analytics to predict the outcomes of management options among which they have to choose and apply in their unique field environments.

Most generally, a solution to the G2P prediction problem will be a quantitative algorithm, which, when given information on a genotype and environment, can produce an accurate estimate of a phenotype of interest. That said, at present, there are three major and virtually disjoint algorithmic schools for phenotype prediction. These are, from oldest to youngest, quantitative genetics (QG; Fisher, 1919, - see commentary by Moran et al. (1966); Lynch and Walsh, 1998), ecophysiological modeling (ECM; Bouman et al., 1996; Hanks and Ritchie, 1991;

Thornley and Johnson, 1990), and systems biology (SB; Gould et al., 2006; Locke et al., 2005; Voit, 2017; Alon, 2006). Despite the fact that the three paradigms are ostensibly describing the same systems, they strongly separate along various dimensions resulting in very little overlap between their respective scientific communities, thus virtually eliminating any opportunity for broad synergism.

The first such great divide is between those models grounded in differential equations (ECM and SB) and those taking algebraic approaches (QG). ECM and SB methods find it quite easy to calculate time series behaviors like daily values of leaf area or hour-by-hour levels of clock gene expression. QG is typically limited to process endpoint predictions whose accuracy can suffer because time-dependent factors contributing to the results are either glossed over or only represented by average values (e.g. mean growing season temperature).

A second dimension is subject matter focus. QG seeks to relate genotypes to phenotypes directly via general linear models. This works well as heritability increases but QG models have difficulty representing interaction effects – doing so is cumbersome, involves equations with many terms, and results in a loss of statistical power. Because it is the oldest of the three approaches there is a vast literature on different specific techniques, their strengths, and weaknesses (Walsh and Lynch, 2018; Lynch and Walsh, 1998; Falconer and Mackay, 1996). In sharp contrast, ECMs are very good at representing the temporal interactions of mechanistic physiological processes, for example, photosynthesis and growth, but are limited in accommodating genetics. There are now ECMs for all major crops, many minor ones, and even a range of economic weed species (Jones et al., 2003; Hoogenboom et al., 2019; Keating et al., 2003; Debaeke et al., 1997; Kiniry et al., 1992; Williams et al., 1989). Systems biology (SB) models view interacting networks of genes, their products, or other metabolites from the perspectives of biochemical kinetics and/or signal processing circuitry (Voit, 2017; Alon, 2006). Figure 2 in Hammer et al. (2019) set forth a notion wherein SB-level events can have “shorter” or “longer” “distances” to whole-plant phenotypes. Typical examples would be, respectively, flowering time (Gursky et al., 2018; Wang et al., 2014) and yield. The authors suggest that QG and SB models are more likely to be usefully predictive over short phenotypic distances but cannot reach the more distant traits for which models of ecophysiological processes will probably be more helpful. Thus, solving the G2P problems requires melding the skills of models at both end of the distance spectrum.

While there are exceptions (e.g. Chew et al., 2017) most of the example fusions have been between ECMs and QG (White and Hoogenboom, 1996; Raymond et al., 2003; Hammer et al., 2006; Technow et al., 2015; Messina et al., 2018) and we focus herein on this combination – especially the highly effective subset of the latter called genomic prediction, GP. This is complicated by the very different data, object, study types, etc. that separate the two classes of models. ECMs typically seek to predict the time series behavior of an idealized single plant in a specific, circumscribed site. Spatial variation (even within different parcels of the same field) is handled as separate environments. As such, the data requirements calibrating and/or running ECMs require – in the words of a helpful reviewer – comprehensive,

multi-stage measurements. Because of their 40-60 years of development, ECMs can also output hundreds of separate variables, most of which are only of interest to researchers. In contrast, GP models integrate data from dozens or hundreds of genetically different lines, often collected at single points in time such as at harvest but from multi-environment trials spread out over large geographic regions. On their output side, however, GP models typically only attempt to predict a very small number of variables of interest to commercial end-users.

To this point in time, there has been an active community working on combining ECMs and GP, but – in comparison to the world-wide scale of crop breeding and crop modeling efforts – it is rather small. Their approach treats the constant parameters (called “genotype-specific parameters”, GSPs) that ECMs use to distinguish and characterize crop cultivars/varieties as quantitative traits to be predicted by QG models from underlying genetic information. A good entry point into recent (and preceding) literature on this methodology is Technow et al. (2015) who presented an example integration of a maize crop growth model (CGM, a synonym for ECM) with whole genome prediction (WGP) using approximate Bayesian computation and synthetic data. Their work was followed up by Cooper et al. (2016) wherein the CGM-WGP model was applied to a real maize dataset and Messina et al. (2018), where the CGM-WGP model was extended for a wider application in multiple environments using a hierarchical Bayesian approach. Voss-Fels et al. (2019) reviewed the numerous challenges facing genomic selection and suggested that improved predictive potential can be likely be achieved by integrating environmental descriptors and crop models with classic GS models to address deficiencies in the predicting performance in future environments. Diepenbrock et al. (2022) showed that CGM-WGP was more predictively skilled than a commonly used WGP method for untested genotypes evaluated in untested environments. The CGM-WGP approach in Technow et al. (2015) is thoroughly illustrated by Cooper et al. (2021) using a hierarchical crop model combined with an infinitesimal GP model. The article provides a robust conceptual framework of how breeding programs and crop simulation modeling approaches complement each other to achieve shorter breeding cycles and improved genetic gains. Very recently, Messina et al. (2022a) examined the parameters of a simple model to dissect the relationships between maize yield & yield potential vs. yield stability across environments in the context of selection for drought tolerance and Messina et al. (2022b) presented and evaluated a CGM-WGP prediction algorithm for multi-trait prediction.

Because of limited overlap between the ECM and GP communities as well as their modeling approach differences pointed out earlier, the capabilities of ECMs are unknown to the bulk of QG experts as is the range of QG techniques available to extend ECMs into full G2P predictors. Equally unknown are the features of GP approaches among ecophysiological crop modelers. As a move toward closing this information gap, we will, herein, establish GSP categories based on their likely genetic architectures with exemplar GSPs for each category. Using a set of three criteria grounded in architecture, computational requirements, and use of prior information, we shall evaluate a number

of QG prediction methods in terms of their suitability for each category/exemplar.

The intended audience for this paper will have more familiarity with ECMs than QG methods. That said, we will (1) summarize the GSP concept for those not having encountered it before. This will be followed by (2) an explanation of prominent QG methods in some detail, (3) a specification of the criteria to be used in evaluating their applicability, and (4) the GSP examples.

1.1. What are GSPs?

When one fits the regression line $y = mx + b$ using paired (x_i, y_i) data, statisticians refer to the symbols m and b as “parameters”. They are constants (i.e., not dependent on i), where i indexes the observations. In contrast, the symbols x and y are, respectively, the independent and the dependent “variables” – that is, they (at least in principle) take on different values for different i 's. While both elementary and rather abstract, these statements provide a starting point toward understanding GSPs as used within the ECM community.

A crucial distinction between ECMs and models like the above regression equation is that ECMs predict time-varying values. To do so they embody both algebraic equations and differential equations which relate the rates of change of modeled quantities to other variables (i.e. time-varying) and parameters (i.e. constants) present in the ECM. In mathematical vernacular, an ECM is a set of differential algebraic equations, commonly abbreviated as “DAE”.

Over time, the main strategy that has been used to improve ECMs has been reductionism. That is, ECM submodels have been made more elaborate with the addition of more detailed processes, more equations, and more parameters. This begs the question as to whether this reductionism has some natural floor? That is, “Is there some aspect of biology that is intrinsically constant?” For any given individual, the answer is “Yes” – its genetic constitution, i.e. genotype, is fixed at fertilization and does not change although different portions of an individual's genome might be (in)active in different environments and/or time points.

But this does not mean that effective modeling can only be done at that ultimate lowest level. After all, it simply cannot be said that ECMs have been devoid of utility for the 60 years since their inception. Plants are, after all, macroscopic objects that, of necessity, obey the mathematical laws of physics and chemistry at all levels, a fact that most ECMs exploit. The question then becomes how to best couple simplifying regularities existing at the genetic level. A greatly illustrative example points the way. A central component in many ECMs describes the rate of leaf expansion (LER). This can be influenced by a number of processes including the rate at which photosynthesis creates new raw materials, the proportion of that new production allocated to leaf growth, coupled with a proportionality factor (specific leaf area, SLA) relating the latter amount to leaf area (Yin et al., 1999). Another scheme is to assume that new photosynthate is not limiting, in which case leaf expansion is primarily driven by the net storage of incompressible water passing through the plant and the pressures it generates. Note that both of these conceptions are bound by the laws of mass and energy conservation.

The latter approach was used in Reymond et al. (2003) via the equation

$$\frac{dL}{dt} = (T - T_0)(a + b \text{VPD} + c\psi) \quad (1)$$

where $\frac{dL}{dt}$ is the LER; $(T - T_0)a$ is the effect of meristem temperature (T) on LER with T_0 being the base temperature; VPD is the atmospheric vapor pressure deficit; and ψ is soil water potential. Simply stated, the model states that VPD and ψ create a physical gradient causing water to flow through the plant, some of which causes the leaves to expand analogously to water balloons, in a temperature-sensitive fashion.

The quantities a , b , and c are the model parameters. Reymond et al. (2003) used QG methods to find quantitative trait loci (QTL) for these parameters. The most important aspect of this work is that, unlike the QTL for SLA, the ones for these parameters did not depend on the environments used to isolate them – that is, they only depended on plant genotype, which, as noted above, is fixed at fertilization. Thus, these parameters are rightly called *genotype-specific* and are powerful tools for prediction.

Thus, a goal for ECM modeling becomes formulating ECMs whose parameters have this property and then using QG to link them to the genome. For a given variety (even one that does not yet exist), one can then (1) use the markers at the QTL positions to predict the GSPs, (2) insert those values into the ECM, (3) run it with inputs of weather, soils, and production input data characterizing the sites and (possibly alternative) management practices of interest, and (4) predict the performance of that genotype under previously untested conditions (e.g. a specific grower's unique farm or a projected future climate).

Of course, QTLs are only one QG approach and, indeed, far from the best. For example, QTL analysis aims to find sets of individual markers that have statistically significant effects on the trait of interest. In contrast, GP methods are based on the desire to successfully make predictions rather than to identify associations. That is, the assumption that the entire assembly of makers might contain useful information even when individual markers or small subsets of them fall short of statistical significance. Following that tenant and to repeat the aim of this paper in more specific terms, it is to provide information on which GP methods appear favorable for predicting GSPs of different types.

1.2. Genomic prediction models

The capacity of generating rich, multidimensional genomic variants has invited breeders to take advantage of whole-genome approaches to obtain the genomic estimated breeding values (GEBVs) of individuals to facilitate the identification of potential parents for the next generation based solely on genotypic information of existing lines – a method originally called genomic selection (GS; Meuwissen et al., 2001; Crossa et al., 2010; de los Campos and Rodriguez, 2013). At its inception, Meuwissen et al. (2001) assumed that many genes each exerted small phenotypic effects (i.e. what is called an “infinitesimal polygenic genetics architecture”) and provided a mixed model for best linear unbiased prediction (BLUP; together with two other Bayesian estimators). These methods motivate a data-driven approach to estimating total GEBVs from all markers simultaneously without relying on any specific genetic knowledge.

Since then, numerous improvements have been proposed to increase predictability for various crop systems leading to several GP algorithms for GS. For example, [Thavamani Kumar et al. \(2015\)](#), [Wimmer et al. \(2013\)](#), and [Ishwaran and Rao \(2010\)](#) demonstrated that ridge regression best linear unbiased prediction (RRBLUP), also called ridge regression (RR; [Hoerl and Kennard, 1970](#)), can be a good choice in general due to its robustness and computational efficiency. Further, [de los Campos and Rodriguez \(2013\)](#), [Daetwyler et al. \(2010\)](#), and [Hayes et al. \(2010\)](#) presented methods employing variable selection, such as Bayesian LASSO (BL; [Park and Casella, 2008](#)), and claimed superiority to RRBLUP for traits influenced by a small number of genes with large effects (e.g., grain hardness trait in wheat). However, evidence of epistatic effects on polygenic traits is vast and well documented ([Holland, 2010, 2006](#)). For such phenotypes, nonparametric methods, like reproducing kernel Hilbert space (RKHS; [Gianola et al., 2006](#); [Gianola and Van Kaam, 2008](#)), and random forests (RF; [Breiman, 2001](#)), have been considered as alternatives to modeling complex interactions under the assumption that they can capture non-additive effects and the signals from high-order interactions ([Heslot et al., 2012](#); [Pérez-Rodríguez et al., 2012](#)).

The first two models suggested by [Meuwissen et al. \(2001\)](#) have expanded into the Bayesian alphabet ([Gianola et al., 2009](#)) comprising a suite of linear models used in genomic prediction. Conventionally, each model is denoted by a particular letter or combination of letters and symbols (i.e. Bayes A, Bayes B, and Bayes C π) meant to indicate the usage of a specific conditional prior unique to that linear model. The former, (e.g., Bayes A) reduces the effects of some markers toward zero; and the latter (e.g., Bayes B and Bayes C) only retain a subset of markers, assuming the remaining ones to have zero effects. By and large, these Bayesian models can also be divided into two categories along a different dimension - those that exploit variable shrinkage and ones that attempt variable selection.

“Shrinkage” is a global variance reduction technique that can be applied during the fitting phase of an estimator to bias the parameters towards zero by a particular amount, yielding a reduced sensitivity to outliers, robustness to multicollinearity, and more accurate predictions. Reducing the variance of the estimator induces a regularizing effect on the model which can be especially beneficial for overparameterized systems with many predictors. Shrinkage is a prominent component in GP and has seen extensive use in various formulations of the additive model. Charles Roy Henderson originally described the Best Linear Unbiased Predictor (BLUP) model for the estimation of random effects under a shrinkage framework ([Henderson, 1963, 1984](#)). While shrinkage principally operates on the marker effects as a regularization technique, in GP models it also acts as an avenue to impose genetic assumptions about the trait in question. Indeed, in the original work, [Meuwissen et al. \(2001\)](#) expanded on this by introducing two Bayesian linear models, Bayes A and Bayes B, that reflected differing assumptions about the genetic architecture underlying trait variability. Bayes B, for example, operates as a variable selection model by selecting only a subset of markers to be included while shrinking the remaining markers to have zero effect under the assumption that only a small number of genetic factors influence the given trait. A successful application of

this can be seen with the oligogenic architecture of plant height in wheat in a given population, which lends itself to the Bayes B model, where the additive genetic variation for this trait can be mostly represented by only 4 major QTLs ([DeWitt et al., 2021](#)). The flexibility provided by different formulations of shrinkage has led to a large set of applicable models designed to predict traits with a variety of genetic architectures.

Thus, the selection of a model, along with its corresponding prior, is also meant to reflect the underlying genetic architecture of the trait in question and to potentially improve predictive performance. Bayes LASSO, for example, assumes a double exponential prior for marker effects, introducing aggressive shrinkage towards zero for weakly related parameters ([Park and Casella, 2008](#)) under the assumption that the trait is mostly determined by large-effect markers. The Bayes B model, on the other hand, incorporates a mixture distribution to apply strong shrinkage towards a point mass of zero for the effects of a subset of markers, mechanistically suggesting that this marker subset does not impact the predicted trait whatsoever ([Meuwissen et al., 2001](#)). Bayes B has been successfully applied to grain yield prediction in wheat ([Haile et al., 2021](#); [Zhao et al., 2013](#); [Haile et al., 2018](#)), tick resistance in cattle ([Cardoso et al., 2015](#)), and grain yield in maize ([Li et al., 2020a](#)), largely achieving competitive results.

The flexibility provided by the various formulations of the additive model under the Bayesian alphabet framework has made them an especially popular option for predicting complex traits in both livestock ([Wang et al., 2019b](#); [van den Berg et al., 2020](#); [El Jabri et al., 2019](#)) and crops ([Kwong et al., 2017](#); [Deomano et al., 2020](#); [Shikha et al., 2017](#)). Empirically, the original Bayes A and Bayes B models frequently outperform traditional BLUP models in benchmarks ([Hamidi Hay and Roberts, 2017](#); [Santos et al., 2015](#); [Lopes et al., 2021](#); [Palaikostas et al., 2018](#)). Furthermore, the development of more sophisticated Bayesian linear models, like the Bayes C π model ([Habier et al., 2011](#)), has expanded the alphabet and addressed weaknesses found in earlier techniques (in this case the ability to designate an unknown mixture probability in Bayes B). In addition, for traits with higher heritability and fewer QTLs, the Bayes U model that utilizes a “global-local” shrinkage prior (a U shape Horseshoe prior) ([Carvalho et al., 2010](#); [Pong-Wong and Woolliams, 2014](#)) has been considered, owing to the property of the continuous shrinkage global-local prior that can adaptively shrink noise to zero while leaving the effects of important loci unshrunk ([Ge et al., 2019](#)).

While the success of Bayesian alphabet models has been recognized, they remain limited by only modeling additive genetic variance. Recently, interest has shifted to semi-parametric models that can account for non-additive genetic variation, like reproducing kernel Hilbert spaces regression ([Gianola et al., 2006](#)), which has occasionally demonstrated superior prediction accuracy ([Pérez-Rodríguez et al., 2012](#)). In a novel application of genomic prediction, [Wang et al. \(2019a\)](#) provided a framework for predicting genotype-specific parameters in a Brassica rapa leaf growth model. In their evaluation of eight genomic prediction models, including three Bayesian alphabet models, RKHS was found to outperform the other algorithms for all GSPs tested, likely indicating that non-additive models may be required to

accurately reflect the interacting complexities of crop model parameters.

There are a number of interrelated factors commonly discussed in the GS performance literature. These include:

1. the size of the training population, relatedness between training and validation populations (Rincent et al., 2012; Akdemir et al., 2015; Isidro et al., 2015; Michel et al., 2017; Neyhart et al., 2017)
2. marker density (Daetwyler et al., 2010; de Oliveira et al., 2018)
3. trait heritability (Heffner et al., 2009)
4. target trait's genetic architecture and the distribution of linkage disequilibrium (LD) between genetic markers and the underlying QTLs (Desta and Ortiz, 2014)
5. and finally, the genotype-by-environment interaction (Crossa et al., 2017)

In the original GS paper, Meuwissen et al. (2001) considered the numbers of QTLs and relative magnitude of genetic effects to prediction performance and indicated that, without interaction terms, improved accuracy can be expected when the distribution of genetic effects is known. Similarly, including the effect of major genes showed an advantage for GS performance in simulation and empirical studies (Rice and Lipka, 2019; Sarinelli et al., 2019; Zaïm et al., 2020); however, this advantage begins to diminish when the number of QTLs is greater than 10 (Bernardo, 2014). In the cases where full-sib families are used, the performance of the GS algorithm evaluated by cross-validations within the population would be determined by the presence of major genes, or whether or not marker density is enough to capture the LD with QTLs (Howard et al., 2014). Using a double haploid population, prediction accuracy was examined for four phenotypes (grain yield, protein content, grain hardness trait like kernel weight, and adjusted sedimentation value) (Hu et al., 2019). The results showed that BL and RF produce the most stable predictability in all comparisons when considering only the cross-validation by Pearson's correlation as the accuracy measure (Hu et al., 2019). As a classic polygenic trait, the predictability of grain yield varied significantly across different field seasons. Amongst the prediction algorithms examined, the prediction performance of RRBLUP and RKHS varied more among replicates than others.

BL and RRBLUP are both linear models that assume the linearity of marker effects. The difference between these two is the shrinkage of marker effects - RRBLUP assumes equal variance and shrinks all the marker effects to the same level, whereas BL can actually shrink some coefficients exactly to zero, performing as a variable selection method. With a relatively smaller sample size than the DH population addressed in Hu et al. (2019), we suspect that the superior performance of BL was due to its ability to capture the non-uniform distribution of marker effect across the genome (Daetwyler et al., 2010), while avoiding overfitting. Using a large European winter wheat population of 2,325 commercial lines in He et al. (2016), the polygenic nature of grain yield is more likely to be captured by the sample size, and as a result, these associated issues of RRBLUP seemed to be alleviated, thus showing comparable prediction ability.

2. ASSESSMENT CRITERIA

From here forward, we assume that the modeler has an ECM and has selected a group of GSPs to work within the context of some suitable population of genotypes. While, in principle, the researcher might desire to construct GP models for all of the model's GSPs, this might not be necessary. For example, a global sensitivity analysis (e.g. via the variance-based Fourier amplitude sensitivity testing, a.k.a. "FAST") might reveal a subset of the parameters that are quite influential in determining the trait of interest. Were that the case, parsimony (not to mention computational efficiency) would suggest focusing GP modeling on that smaller set. Moreover, we also assume that the GSPs chosen are identifiable. It is not necessarily the case that this is automatically true for any given GSP in any particular ECM, even among widely accepted and broadly used models (Lamsal et al., 2018).

In order to evaluate various GP algorithm types for particular applications, we defined three assessment criteria (AC) to assess each algorithm for its suitability:

- AC1) Handling a spectrum of genetic architectures from oligogenic to polygenic
- AC2) Inclusion into a single-stage versus a multi-stage estimation process
- AC3) Capacity to incorporate prior knowledge/information

In the case of AC1, we recognize that many GP algorithms are designed with assumptions about the genetic architecture underlying a particular trait of interest. Specifically, we are focused here on the question of whether a trait has a few major genes controlling the trait (oligogenic) or a large number of genes most of which have small effects (polygenic). We presume that assumptions about genetic architecture made in the development of these algorithms will make some algorithms more suitable for cases in which the genetic architecture underlying the GSP of interest matches the original assumptions underlying the algorithm.

The second criterion (AC2) is meant to deal with the practical matter of computational complexity and efficiency of the estimation process. Previous efforts in GP and GSP estimation can be classified into two main groups: one-stage and two-stage estimation. In the one-stage approach, GSPs and the trained GP model meant to predict them are simultaneously fitted. In this approach, each iteration of the algorithm requires an evaluation of the ECM for each genotype and environment (Fig. 1). An example of the one-stage estimation method includes Messina et al. (2018). In contrast, the two-stage approach estimates values for the GSPs first and then the GP model is subsequently fitted by the GP algorithm (Fig. 1). A two-stage estimation method is utilized in Toda et al. (2022).

The third criterion (AC3) takes aim at the fact that the levels of knowledge about various processes and the genetics driving them vary depending on the species and process being studied. In cases where much is known about some subset of these, it is inefficient to force the GP algorithm to relearn what we already know. In such cases, it would be preferable to inform the algorithm about what is already known through the use of prior knowledge or information. This might take the form of priors with non-zero expected values for markers or genes where effects

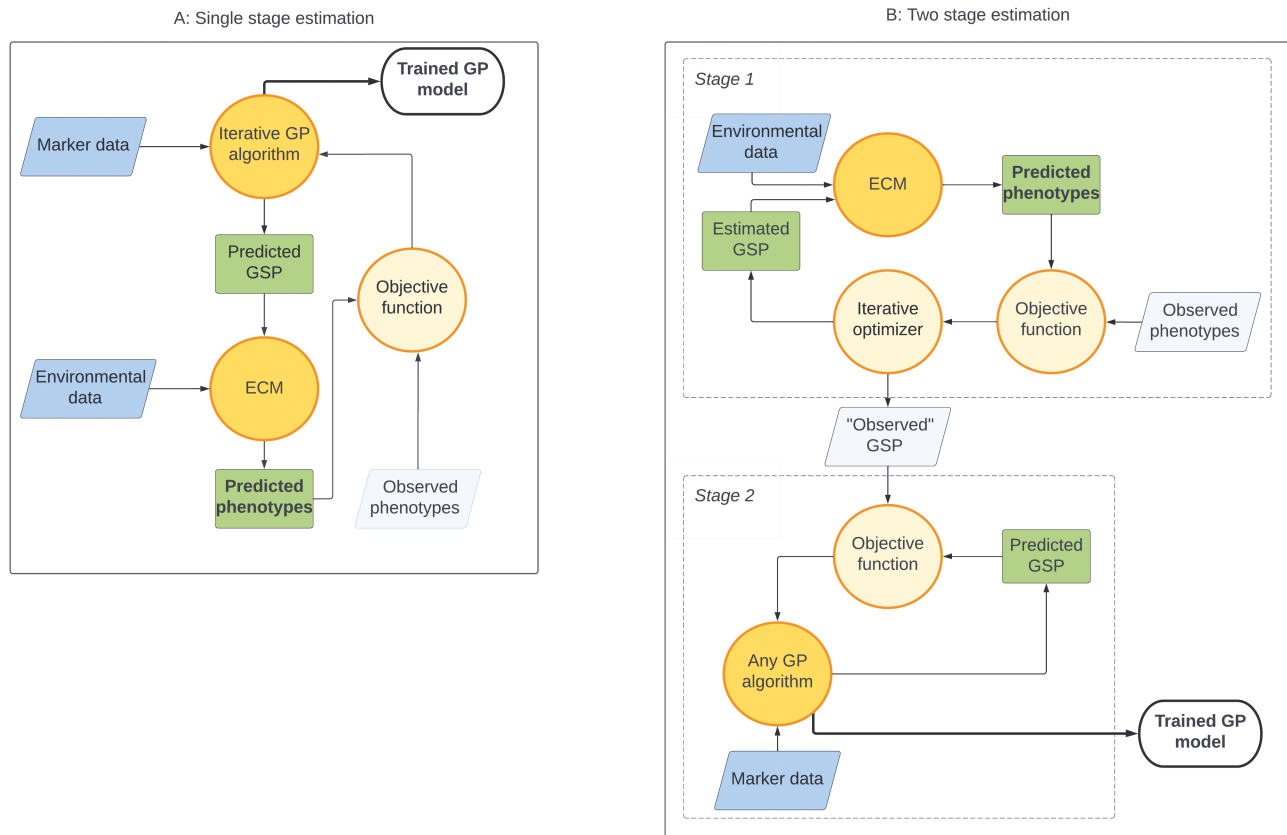


Figure 1. Illustration of one- and two-stage genotype specific parameter (GSP) estimation linked with genomic prediction (GP). Panel A shows the one-stage process in which each iteration of a GP training algorithm requires an evaluation of the ecophysiological crop model (ECM). Panel B shows the two-stage process in which values for GSPs are first estimated in Stage 1. These estimated GSP values are then treated as “Observed” GSPs for subsequent GP model training with one or more algorithms in Stage 2.

are known or in the form of structural priors in which separate prior variances might be specified for markers within vs outside regions previously identified by QTL or other analyses to be regions of importance within the genome. Thus, we deem it important and relevant to assess GP algorithms in terms of their ability to incorporate such prior information.

3. DISCUSSION

The parameters in crop models are usually calibrated for each genotype across all environments by iterative trial and error procedures via different optimization algorithms (Akhavizadegan et al., 2021). In this study, we present genomic prediction models as a potential methodology to estimate the GSPs in crop models. We discuss and characterize the applicability of several GP algorithms under different scenarios as no one GP algorithm is the best fit across different species and traits (Azodi et al., 2019). Below, we present a discussion within each of the three assessment criteria we have established with example GSPs.

3.1. Handling a spectrum of genetic architectures (AC1)

As illustrated in Fig. 2, the success of GP algorithms largely depends on the genetic architecture of a trait. Marker-assisted

selection (MAS) methods lie in the “simpler” end of algorithms being suitable for predicting traits governed by few genes. On the other hand, non-parametric prediction algorithms such as RKHS are flexible and can be more suitable for predicting traits with complex genetic architectures and strong gene-environment interactions (absent by definition in GSPs), as demonstrated in studies by Hu et al. (2023), Costa-Neto et al. (2021), and Gianola (2021). However, in practice, more than one GP models are necessary to fit to determine a method for best prediction. In the following paragraphs, we discuss the potential relations between the range of GP algorithms and trait genetic architecture complexities. This discussion will help to narrow down the range of GP models that need to be fit depending on the phenotypic trait. To infer a trait’s genetic architecture, we use information on number and size of QTLs from QTL studies. For traits with limited QTL studies, we utilize information on trait heritability.

Flowering time: Flowering time in wheat is governed by few major QTLs making it an oligogenic trait (DeWitt et al., 2021). Although considerably influenced by the environment, this trait has a simple genetic architecture governed by a small set of known genes with epistatic interactions. Algorithms that allow variable shrinkages such as Bayesian family and non-parametric methods are suitable in such cases. The non-parametric algorithms potentially capture the complex interactions making

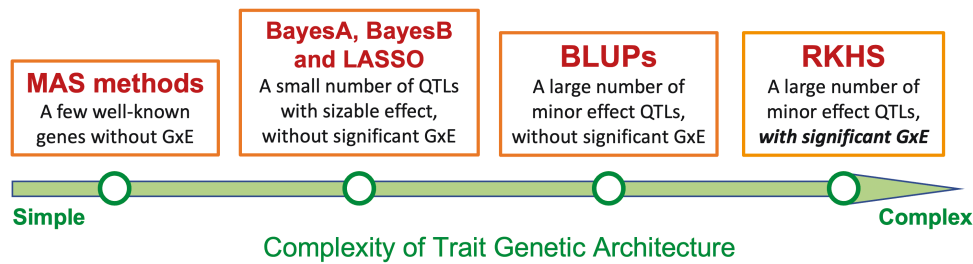


Figure 2. Conventional genomic prediction algorithms using cultivar genomic markers

them suitable to model epistatic genes (Salehi et al., 2021; Momen et al., 2018; Howard et al., 2014). Bayes B, which applies aggressive shrinkage to zero-effect for most markers, was found to be the most accurate model to predict time to young microspore, a flowering time-related trait in wheat (Thavamanikumar et al., 2015). In a study by Pérez-Rodríguez et al. (2012) which compared different GP methods, non-parametric methods such as RKHS were found to more closely predict days to heading in wheat which usually indicates a high degree of genetic interaction. In a commonly used crop model DSSAT-CERES-wheat, the flowering time parameters are P1V and P1D which represent vernalization and photoperiod sensitivity coefficients (Jones et al., 2003). These parameters, when correctly estimated, are genotype-specific and independent of the environment. The genetic architecture of flowering time varies between different species. Flowering time in maize is governed by many QTLs and is considered a polygenic trait with additive marker effects (Li et al., 2016; Buckler et al., 2009). Such a trait would be best predicted by linear algorithms such as RRBLUP since it assumes homogeneous marker effects across loci (Meuwissen et al., 2001; Wang et al., 2018).

Kernel weight: Kernel weight is one of the primary yield components of grain yield. In cereals, the potential kernel weight is considered to have high heritability and thus relatively stable across environments. The trait is governed by many QTLs in the three major cereal crops, wheat (Cao et al., 2020), maize (Liu et al., 2020; Zhang et al., 2020; Liu et al., 2017; Raihan et al., 2016), and rice (Li et al., 2020b; Roy and Shil, 2020; Zhang et al., 2016) with additive gene effects. The prediction accuracies of different genomic prediction models are reported to be similar across parametric and non-parametric models in wheat (Haile et al., 2021), rice (Yu et al., 2022), and maize (Liu et al., 2018) for kernel weight. In crop models, different parameters related to grain characteristics are often used. The GSPs G1 and G2 in DSSAT-CERES-wheat represent kernel number per unit canopy weight at anthesis and standard kernel size under optimum conditions, respectively (Attia et al., 2016).

Leaf area: Leaf area parameters such as potential maximum leaf area and leaf growth rate determine leaf size and growth rate which determines the photosynthetic potential of a plant. In wheat, the uppermost (flag) leaf contributes about 50% of photosynthates for grain filling (El Wazziki et al., 2015; Evans and Rawson, 1970). Fourteen additive QTLs related to flag leaf characteristics were reported for wheat whereby each QTL explained 2.78% to 13.32% (Yan et al., 2020). In a study by Du et al. (2019),

29 QTLs for flag leaf length were identified among which one QTL consistently explained more than 30% of phenotypic variation across two experiment years. For the flag leaf area, the QTL with the greatest effect explained 10.31% and 23.46% across the two years. A similar conclusion of detectable genetic control was also supported in maize (Cui et al., 2017). In another study in maize, 15 QTLs were identified for the leaf area, which jointly explained 65.4% of the total phenotypic variation (Li et al., 2016). Tian et al. (2011) concluded that leaf traits are mostly governed by small effects with little interaction. The QTL studies indicate that these traits are moderately heritable and could be reasonably predicted with GP models with genotypic information for a given population and environment. However, QTLs are usually specific to a population and environment which limits our ability to predict for a novel environment, unless the $G \times E$ effect is explicitly included in the model (Hu et al., 2019). Predicting traits related to leaf characteristics that are environmentally independent and genotype-specific such as maximum leaf size (a GSP) would allow us to utilize the parameter to make predictions across the different environments. Genomic prediction algorithms that address additive genetic variances such as RRBLUP and Bayes LASSO might be adequate to predict GSPs related to leaf traits since they show little epistasis (Tian et al., 2011). On the other hand, since the measurable leaf traits in the field are considerably influenced by the environment, estimating GSPs related to leaf traits with an ECM through a predictive equation for the measurable traits that embody GSPs might be a more successful approach. In addition, due to the cumbersome and destructive nature of data collection, conventionally, leaf traits are measured in a small population of genotypes. Rapid and non-destructive high throughput phenotyping methods can help expand the mapping population for such traits that are challenging to measure in the field. For example, remote sensing with aerial-based platforms have been successfully used to collect data on leaf area index over the growing season in several crops (Wu et al., 2022; Hasan et al., 2019; Simic Milas et al., 2018). Different leaf area parameters are often used by process-based crop models such as SIZELF in CROPGRO-soybean (Boote et al., 2003) and LAI_{mx} in EPIC (Williams et al., 1989) representing maximum leaf size.

Water use efficiency (WUE): WUE is a genetically complex trait considerably influenced by environment and management. While being a genetically complex trait, WUE has been noted to be moderately heritable. WUE is studied at different levels: crop - agronomic WUE, plant - transpiration efficiency (TE), and

leaf - instantaneous water use efficiency (IWUE). At the plant level, carbon isotope elements are often used as surrogate traits for TE and have a high broad sense heritability (Sorgini et al., 2021; Vadez et al., 2014). A study by Sorgini et al. (2021) identified 28 QTLs across different carbon isotope elements in maize. At the leaf level, a study by Lopez et al. (2019) identified four genomic regions associated with IWUE in soybean and suggested the use of Bayesian genomic prediction methods for genotypic selection and prediction for intrinsic WUE. On the contrary, it is challenging to identify the genetic drivers of WUE at the crop level due to the $G \times E$ interactions affecting biomass and water use under field conditions. In a controlled environment experiment in *Setaria*, WUE was reported to be highly heritable and polygenic (Feldman et al., 2018). Genetic manipulations to improve WUE are mostly attempted through manipulation of leaf photosynthesis and stomatal conductance and their relationship as reviewed in Leahey et al. (2019). Prediction of WUE components such as IWUE, TE, and carbon isotope elements is more likely to be successful with genomic prediction than the overall agronomic WUE because of their higher heritability and stability across environments. Examples of parameters related to crop water use in crop models are WP in AquaCrop which represents water productivity or biomass accumulated per unit of transpiration (Steduto et al., 2009) and TE in APSIM-NWheat which represents transpiration efficiency (Keating et al., 2001).

Radiation use efficiency (RUE): RUE is a complex trait affected by many variables such as crop management, climatic conditions, and biotic and abiotic stresses (Ullah et al., 2019). The genetic architecture of RUE is not yet well understood. Complex traits such as yield and biomass are driven by many genes in most crops making them polygenic; a similar phenomenon can be expected for RUE. RUE is typically calculated at the end of the growing season as a ratio of total biomass to intercepted radiation, whereas in reality, RUE is expected to vary over the growing season (Furbank et al., 2019). Broad sense heritability estimates for RUE during different growth stages in wheat were reported in a range of 0.11 to 0.42 with the highest heritability associated with RUE from canopy closure to physiological maturity along with five marker-trait associations (Molero et al., 2019). At the plant level, RUE is a representation of photosynthetic capacity whereby it has a positive relationship with leaf photosynthesis rate until a certain threshold (Cabrera-Bosquet et al., 2016; Sinclair and Muchow, 1999). Understanding the genetic architecture of leaf photosynthesis might be helpful in identifying the genetic mechanisms that drive RUE. Example parameters related to RUE in crop models are PARUE in DSSAT and RUE in APSIM.

3.2. Inclusion into a single-stage versus a multi-stage estimation process (AC2)

When considering one-stage and two-stage estimation approaches there are two primary considerations which distinguish them: 1) the number of evaluations of the ECM and 2) the manner in which estimates/errors interact within/across the stage(s). A primary appeal of the two-stage approach is the potential reduction in the number of ECM evaluations needed for estimation. Assuming that stable, unbiased estimates of GSPs can be generated in the first step, the ECM need not be evaluated

during the training of the GP model by the GP algorithm. Given the computational costs of numerically integrating the ECM, the potential reduction in compute time could be substantial when compared to a one-stage process. This benefit may be especially important considering that the optimal GP algorithm for predicting a given trait is rarely (if ever) known a priori and thus multiple GP algorithms must be trained and evaluated for the same trait of interest. A two-stage process would allow multiple GP algorithms to be applied to a set of GSP estimates with only the computational cost of the additional GP model fitting process.

However, achieving stable, unbiased GSP estimates might be challenging in the presence of, on the one hand, interacting parameters without adequate information from the data leading to model equifinality/non-identifiability or, on the other hand, lack of model expressivity (i.e. the model cannot reproduce regions of the state space within which certain observations lie; Lamal et al., 2018). Either of these cases would introduce error into the GSP estimates upon which the second stage of estimation would be based. This error would add further noise to what is already a noisy signal that GP algorithms must decode. Although a one-stage estimation process would not completely resolve these issues in all situations, it could allow data from all genotypes to simultaneously inform the estimation of GSPs as well as the GP model weights that predict them. This phenomenon is widely known within the field of quantitative genetics (Sorensen and Gianola, 2007), and Bayesian hierarchical modeling more generally (Gelman et al., 2013), where it is referred to as information borrowing.

For example, flowering time in wheat is a trait that is often decomposed into vernalization and photoperiod sensitivity parameters (e.g. P1V and P1D in DSSAT-CERES-wheat or R_v and R_p in APSIM-Wheat). The interacting effects of photoperiod and vernalization sensitivities might make it challenging to obtain stable estimates of these GSPs for any given genotype. However, if the GSPs for all genotypes are estimated simultaneously with the GP model weights, the shared information across genotypes may generate estimates of GSPs for individual genotypes with greater certainty and, as a direct result, more coherence with the degree of genetic similarity between genotypes. As the GP model converges, more closely related genotypes (e.g. individuals with the same alleles for vernalization genes) would tend to have more similar estimates of GSPs (e.g. P1V or R_v) and would thus allow the combined genetic and phenotypic data of the whole population to inform the GSP estimates for a given genotype. Consequently, the estimates for any given genotype would be partially constrained by those of related genotypes and thereby contribute towards obtaining stable estimates.

Figure 3 illustrates this concept. Let the two parameters on the x- and y-axes (α and β) represent hypothetical GSPs for vernalization sensitivity and photoperiod sensitivity, respectively. The contour lines in the figure represent regions of parameter combinations with similar goodness-of-fit given the data available for each genotype. The shaded regions represent the ranges of parameter combinations that have identical or nearly identical goodness-of-fit for that genotype. The orange and red contours represent genetically similar individuals (i.e. they have

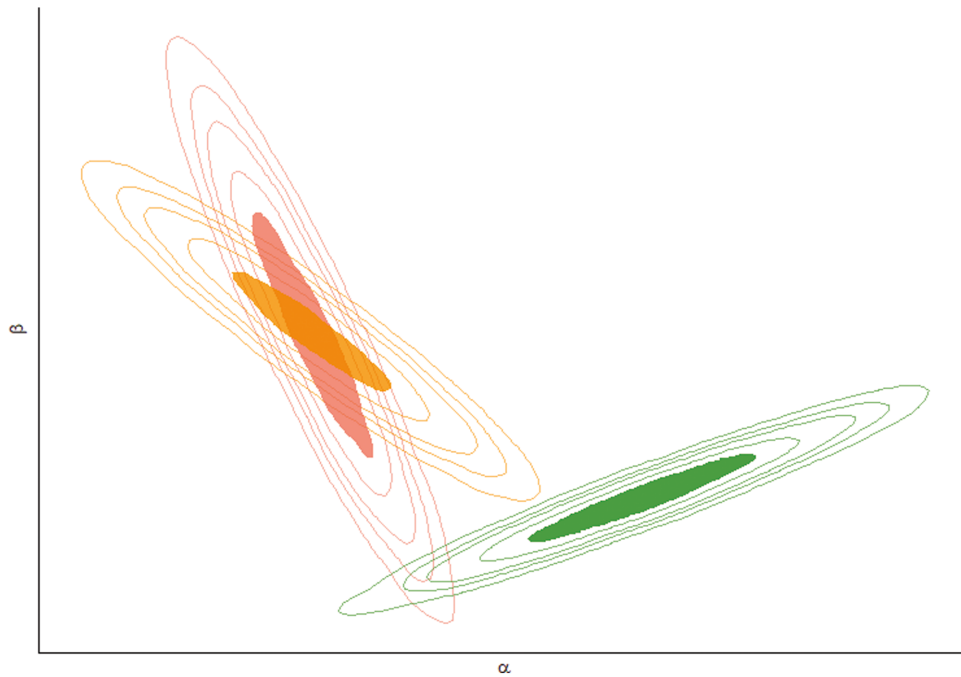


Figure 3. Conceptual diagram showing information borrowing between similar genotypes. The two parameters (α and β) represent hypothetical genotype specific parameters for vernalization sensitivity and photoperiod sensitivity, respectively. The orange and red contours represent genetically similar individuals. The green contours represent a third genotype that is genetically distinct from the other two. The contour lines represent regions of parameter combinations with similar goodness-of-fit within each genotype. The shaded regions represent the ranges of parameter combinations that have identical or nearly identical goodness-of-fit within that genotype. The overlap between red and orange regions represents a region of reduced uncertainty for each genotype (due to higher combined goodness-of-fit) when the close relatedness of red and orange genotypes (as represented by genetic data) are considered.

most or all alleles in common for genes that control vernalization and photoperiod sensitivities). The green contours represent a genotype that is genetically distinct from the other two. In the case of a two-stage process, each genotype would have a range of uncertainty about each parameter constrained only by the information embedded within the data for that genotype. However, in a one-stage estimation process, the genetic data would inform the estimation algorithm of the close relatedness of red and orange compared to other genotypes (e.g. green). That is, in the estimation of the red GSPs, information would be borrowed from data for orange and vice versa. Consequently, the GSP estimates for red and orange would be pulled towards each other into regions consistent with data from both (as represented by the overlap in red and orange shading). The result would be reduced uncertainty in the estimates of α and β for both red and orange genotypes relative to other genotypes.

This general principle of information borrowing between related genotypes is a widely known feature of Bayesian methods in quantitative genetics, wherein various approaches exploit forms of genetic covariance estimates derived from pedigree analysis and genomic relationships (Sorensen and Gianola, 2007). However, in any given case, the choice between the one- and two-stage approaches depends on the tradeoff between the accuracy gains of the former relative to its greater computational requirements.

3.3. Capacity to incorporate prior knowledge/ information (AC3)

The prior knowledge we have about a particular trait's genetic architecture not only depends on the trait but also on the size of the population from which they are measured. If one is working with GSPs or traits that are not commonly measured in large breeding populations, the knowledge we have about the GSP is limited by the small population size.

For example, leaf area traits are typically collected from small-scale experiments with fewer genotypes i.e. a small population, and thus the genetic architecture of leaf area inferred from these data might be apparently simple; but, it may not be an accurate representation of the true genetic architecture of the trait. In such cases of small populations, it's appropriate to use simple prediction algorithms since it is likely that only the prominent major genetic effects will be recoverable from the data. Nevertheless, while dealing with such situations, careful consideration should be given to the inferences as they may not be general enough to extend to other populations. High-throughput phenotyping methods, that can measure leaf area and related traits in large populations in a small amount of time, will prove helpful in further understanding the genetic architecture of GSPs related to leaf area traits.

In contrast, data on flowering time is widely collected in large breeding populations, as a result of which we could have a more comprehensive understanding of its genetic architecture when

genomic information is also available. In addition, pedigree information on the population is also more readily available for such traits that are measured across generations in breeding populations. The pedigree information can be included in some GP models such as single-step genomic BLUP, where the pedigree-based numerator relationship matrix (A matrix) and a genomic relationship matrix (G) are simultaneously used (Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010).

The Bayesian formulations of the GP models allow us to more directly incorporate the genetic information on model parameters through the specification of a prior distribution. The most frequently used GP models with a Bayesian formulation are the Bayes alphabet and non-parametric models such as RKHS. Prior distribution in RKHS is specified for the degrees of genetic similarities, whereas, in Bayesian alphabet models, priors are specified on marker effects and different approaches are taken to formulate the priors. Indeed, as described above in relation to shrinkage and model selection, the differences between the Bayes alphabets are primarily a result of their different prior structures at either the level of data or variance or both. For example, the prior distribution for Bayes A and Bayes B differ at the level of variance. For both models, the prior at the data level is a student-t distribution with mean zero and variance $\sigma_g^2 i$. At the variance level, the same prior (scaled inverse chi-sq distribution with scale parameter S and degrees of freedom ν) is assigned to all marker effects in Bayes A. In Bayes B, however, a mixture prior is assigned such that the prior has a high probability mass at $\sigma_g^2 i = 0$ and the scaled inverse chi-sq prior at $\sigma_g^2 i > 0$. This indicates that Bayes A assumes all markers explain the genetic variance to some effect thus making it suitable to predict traits or GSPs governed by small-effect additive genes. On the contrary, the mixture prior in Bayes B means that a proportion of markers has a non-zero probability of explaining the genetic variance, and the rest marker effects are shrunk to zero. This shows that Bayes B would perform well in predicting traits and GSPs that are mainly governed by a few known genes. A description of priors for different Bayes algorithms has been illustrated in many papers (Meher et al., 2022; Montesinos López et al., 2022; Wang et al., 2018; Howard et al., 2014; Verbyla et al., 2010; Gianola et al., 2009) after being first introduced by Meuwissen et al. (2001).

Thus, while selecting the GP algorithms to predict GSPs in ecophysiological models, different types of prior knowledge of the GSPs or traits related to them should be taken into account. The size of the mapping population and genetic architecture is important to consider not only while selecting algorithms, but also while making inferences. The marker data on GSPs can be directly included in the Bayesian alphabet GP algorithms in the form of a prior distribution of marker effects.

4. CONCLUSIONS

In summary, combining GP estimation of environmentally stable ecophysiological model GSPs can facilitate the prediction of important crop traits that are subject to $G \times E$. The choice of GP algorithms depends on several factors such as trait genetic architecture, the composition of breeding populations, the

estimation process, and the ability to incorporate prior genetic information. High throughput phenotyping methods help expand the mapping population for difficult-to-measure traits such as leaf area and allow us to better understand the genetic architecture of the trait. A two-stage estimation process is more computationally efficient whereas a one-stage estimation process has the ability to leverage genotypic data to obtain better estimates of the GSPs due to information borrowing between genotypes. Some Bayesian GP algorithms can include marker data on GSP directly into the model through a prior distribution. A further research area in Bayesian implementation could extend to specifying priors in such a way that we take into account not only the key genes in related biological pathways but also the knowledge about functional genomic features.

FUNDING

This work was supported by the National Science Foundation [grant number OIA-1826820].

COMPETING INTERESTS

There are no competing interests.

AUTHOR CONTRIBUTIONS STATEMENT

PP, BN, CC and SMW reviewed literature, generated figures, and wrote the final manuscript. PDA contributed to text and figures and revised the final manuscript.

LITERATURE CITED

- I. Aguilar, I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, 93(2):743–752, 2010.
- D. Akdemir, J. I. Sanchez, and J.-L. Jannink. Optimization of genomic selection training populations with a genetic algorithm. *Genetics Selection Evolution*, 47(1):1–10, 2015.
- F. Akhavadegan, J. Ansarifard, L. Wang, I. Huber, and S. V. Archontoulis. A time-dependent parameter estimation framework for crop modeling. *Scientific Reports*, 11(1):1–15, 2021.
- U. Alon. *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2006.
- A. Attia, N. Rajan, Q. Xue, S. Nair, A. Ibrahim, and D. Hays. Application of DSSAT-CERES-Wheat model to simulate winter wheat response to irrigation management in the Texas High Plains. *Agricultural Water Management*, 165:50–60, 2016.
- C. B. Azodi, E. Bolger, A. McCarren, M. Roantree, G. de Los Campos, and S.-H. Shiu. Benchmarking parametric and machine learning models for genomic prediction of complex traits. *G3: Genes[Genomes] Genetics*, 9(11):3691–3702, 2019.
- R. Bernardo. Genomewide selection when major genes are known. *Crop Science*, 54(1):68–75, 2014.

- K. J. Boote, J. W. Jones, W. D. Batchelor, E. D. Nafziger, and O. Myers. Genetic coefficients in the CROPGRO–Soybean model: Links to field performance and genomics. *Agronomy Journal*, 95(1):32–51, 2003.
- B. A. M. Bouman, H. Van Keulen, H. H. Van Laar, and R. Rabbinge. The ‘school of de wit’ crop growth simulation models: a pedigree and historical overview. *Agricultural Systems*, 52(2-3):171–198, 1996.
- L. Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231, 2001.
- E. S. Buckler, J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J. C. Glaubitz, M. M. Goodman, C. Harjes, K. Guill, D. E. Kroon, S. Larsson, N. K. Lepak, H. Li, S. E. Mitchell, G. Pressoir, J. A. Peiffer, M. O. Rosas, T. R. Rocheford, M. C. Romay, S. Romero, S. Salvo, H. S. Villeda, H. S. d. Silva, Q. Sun, F. Tian, N. Upadaya, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M. D. McMullen. The genetic architecture of maize flowering time. *Science*, 325(5941):714–718, 2009.
- L. Cabrera-Bosquet, C. Fournier, N. Brichet, C. Welcker, B. Suard, and F. Tardieu. High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytologist*, 212(1):269–281, 2016.
- S. Cao, M. Hanif, X. Xia, and Z. He. Genetic architecture underpinning yield component traits in wheat. *Theoretical and Applied Genetics*, 133(6):1811–1823, 2020.
- F. F. Cardoso, C. C. G. Gomes, B. P. Sollero, M. M. Oliveira, V. M. Roso, M. L. Piccoli, R. H. Higa, M. J. Yokoo, A. R. Caetano, and I. Aguilar. Genomic prediction for tick resistance in Braford and Hereford cattle. *Journal of Animal Science*, 93(6):2693–2705, 2015.
- C. M. Carvalho, N. G. Polson, and J. G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010.
- Y. H. Chew, D. D. Seaton, and A. J. Millar. Multi-scale modelling to synergise plant systems biology and crop science. *Field Crops Research*, 202:77–83, 2017.
- O. F. Christensen and M. S. Lund. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, 42(1):1–8, 2010.
- M. Cooper, F. Technow, C. Messina, C. Gho, and L. R. Totir. Use of crop growth models with whole-genome prediction: Application to a maize multi-environment trial. *Crop Science*, 56(5):2141–2156, 2016.
- M. Cooper, O. Powell, K. P. Voss-Fels, C. D. Messina, C. Gho, D. W. Podlich, F. Technow, S. C. Chapman, C. A. Beveridge, D. Ortiz-Barrientos, and G. L. Hammer. Modelling selection response in plant-breeding programs using crop models as mechanistic gene-to-phenotype (cgm-g2p) multi-trait link functions. *in silico Plants*, 3(1):diaa016, 2021.
- G. Costa-Neto, R. Fritsche-Neto, and J. Crossa. Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials. *Heredity*, 126(1):92–106, 2021.
- N. R. Council. *A new biology for the 21st century*. National Academies Press, 2009.
- J. Crossa, G. d. I. Campos, P. Pérez, D. Gianola, J. Burgueno, J. L. Araus, D. Makumbi, R. P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Banziger, and H.-J. Braun. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2):713–724, 2010.
- J. Crossa, P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. De Los Campos, J. Burgueno, J. M. González-Camacho, S. Pérez-Elizalde, Y. Beyene, S. Dreisigacker, R. Singh, X. Zhang, M. Gowda, M. Roorkiwal, J. Rutkoski, and R. K. Varshney. Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science*, 22(11):961–975, 2017.
- T.-t. Cui, K.-h. He, L.-g. Chang, X.-h. Zhang, J.-q. Xue, and J.-c. Liu. QTL mapping for leaf area in maize (*Zea mays* L.) under multi-environments. *Journal of Integrative Agriculture*, 16(4):800–808, 2017.
- H. D. Daetwyler, R. Pong-Wong, B. Villanueva, and J. A. Woolliams. The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, 185(3):1021–1031, 2010.
- G. de los Campos and P. P. Rodriguez. BGLR: Bayesian generalized linear regression. R package v. 1.0. 4, 2013.
- A. A. de Oliveira, M. M. Pastina, V. F. de Souza, R. A. da Costa Parrella, R. W. Noda, M. L. F. Simeone, R. E. Schaffert, J. V. de Magalhães, C. M. B. Damasceno, and G. R. A. Margarido. Genomic prediction applied to high-biomass sorghum for bioenergy production. *Molecular Breeding*, 38(4):1–16, 2018.
- P. Debaeke, J. P. Caussanel, J. R. Kinyri, B. Kafiz, and G. Mondragon. Modelling crop: weed interactions in wheat with ALMANAC. *Weed Research*, 37(5):325–341, 1997.
- E. Deomano, P. Jackson, X. Wei, K. Aitken, R. Kota, and P. Pérez-Rodríguez. Genomic prediction of sugar content and cane yield in sugar cane clones in different stages of selection in a breeding program, with and without pedigree information. *Molecular Breeding*, 40(4):1–12, 2020.
- Z. A. Desta and R. Ortiz. Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science*, 19(9):592–601, 2014.
- N. DeWitt, M. Guedira, E. Lauer, J. P. Murphy, D. Marshall, M. Mergoum, J. Johnson, J. B. Holland, and G. Brown-Guedira. Characterizing the oligogenic architecture of plant growth phenotypes informs genomic selection approaches in a common wheat population. *BMC Genomics*, 22(1):1–18, 2021.
- C. H. Diepenbrock, T. Tang, M. Jines, F. Technow, S. Lira, D. Podlich, M. Cooper, and C. Messina. Can we harness digital technologies and physiology to hasten genetic gain in US maize breeding? *Plant physiology*, 188(2):1141–1157, 2022.
- B. Du, L. Liu, Q. Wang, G. Sun, X. Ren, C. Li, and D. Sun. Identification of QTL underlying the leaf length and area of different leaves in barley. *Scientific Reports*, 9(1):1–8, 2019.
- M. El Jabri, M.-P. Sanchez, P. Trossat, C. Laithier, V. Wolf, P. Grosperin, E. Beuvier, O. Rolet-Répécaud, S. Gavoye, Y. Gaüzère, O. Belysheva, E. Notz, D. Boichard, and A. Delacroix-Buchet. Comparison of Bayesian and partial least squares regression methods for mid-infrared prediction of cheese-making properties in Montbéliarde cows. *Journal of Dairy Science*, 102(8):6943–6958, 2019.
- H. El Wazziki, B. El Yousfi, and S. Serghat. Contributions of three upper leaves of wheat, either healthy or inoculated by ‘*Bipolaris sorokiniana*’, to yield and yield components. *Australian Journal of Crop Science*, 9(7):629–637, 2015.
- L. T. Evans and H. M. Rawson. Photosynthesis and respiration by the flag leaf and components of the ear during grain development in wheat. *Australian Journal of Biological Sciences*, 23(2):245–254, 1970.
- D. S. Falconer and T. F. C. Mackay. Introduction to quantitative genetics. Longmans green, harlow, essex, uk. *Introduction to Quantitative Genetics*. 4th ed. Longmans Green, Harlow, Essex, UK., 1996.
- M. J. Feldman, P. Z. Ellsworth, N. Fahlgren, M. A. Gehan, A. B. Cousins, and I. Baxter. Components of water use efficiency have unique genetic signatures in the model C4 grass setaria. *Plant Physiology*, 178(2):699–715, 2018.
- R. A. Fisher. The correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.
- R. T. Furbank, J. A. Jimenez-Berni, B. George-Jaeggli, A. B. Potgieter, and D. M. Deery. Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. *New Phytologist*, 223(4):1714–1727, 2019.
- T. Ge, C.-Y. Chen, Y. Ni, Y.-C. A. Feng, and J. W. Smoller. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, 10(1):1–10, 2019.
- A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. CRC press, Boca Raton, FL, 3rd edition, 2013.

- D. Gianola. Opinionated views on genome-assisted inference and prediction during a pandemic. *Frontiers in Plant Science*, 12:717284, 2021.
- D. Gianola and J. B. Van Kaam. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, 178(4):2289–2303, 2008.
- D. Gianola, R. L. Fernando, and A. Stella. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 173(3):1761–1776, 2006.
- D. Gianola, G. de Los Campos, W. G. Hill, E. Manfredi, and R. Fernando. Additive genetic variability and the Bayesian alphabet. *Genetics*, 183(1):347–363, 2009.
- H. C. J. Godfray, J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin. Food security: the challenge of feeding 9 billion people. *Science*, 327(5967):812–818, 2010.
- P. D. Gould, J. C. W. Locke, C. Larue, M. M. Southern, S. J. Davis, S. Hanano, R. Moyle, R. Milich, J. Putterill, A. J. Millar, and A. Hall. The molecular basis of temperature compensation in the Arabidopsis circadian clock. *The Plant Cell*, 18(5):1177–1187, 2006.
- V. V. Gursky, K. N. Kozlov, S. V. Nuzhdin, and M. G. Samsonova. Dynamical modeling of the core gene network controlling flowering suggests cumulative activation from the FLOWERING LOCUS T gene homologs in chickpea. *Frontiers in genetics*, 9:547, 2018.
- D. Habier, R. L. Fernando, K. Kizilkaya, and D. J. Garrick. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12(1):1–12, 2011.
- J. K. Haile, A. N'Diaye, F. Clarke, J. Clarke, R. Knox, J. Rutkoski, F. M. Bassi, and C. J. Pozniak. Genomic selection for grain yield and quality traits in durum wheat. *Molecular Breeding*, 38(6):1–18, 2018.
- T. A. Haile, S. Walkowiak, A. N'Diaye, J. M. Clarke, P. J. Hucl, R. D. Cuthbert, R. E. Knox, and C. J. Pozniak. Genomic prediction of agronomic traits in wheat using different models and cross-validation designs. *Theoretical and Applied Genetics*, 134(1):381–398, 2021.
- E. Hamidi Hay and A. Roberts. Genomic prediction and genome-wide association analysis of female longevity in a composite beef cattle breed. *Journal of Animal Science*, 95(4):1467–1471, 2017.
- G. Hammer, M. Cooper, F. Tardieu, S. Welch, B. Walsh, F. van Eeuwijk, S. Chapman, and D. Podlich. Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science*, 11(12):587–593, 2006.
- G. Hammer, C. Messina, A. Wu, and M. Cooper. Biological reality and parsimony in crop models—why we need both in crop improvement! *in silico Plants*, 1(1):diz010, 2019.
- J. Hanks and J. T. Ritchie. Modeling plant and soil systems. Soil Science Society of America, 1991.
- U. Hasan, M. Sawut, and S. Chen. Estimating the leaf area index of winter wheat based on unmanned aerial vehicle RGB-image parameters. *Sustainability*, 11(23):6829, 2019.
- B. J. Hayes, J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genetics*, 6(9):e1001139, 2010.
- S. He, A. W. Schulthess, V. Mirdita, Y. Zhao, V. Korzun, R. Bothe, E. Ebmeyer, J. C. Reif, and Y. Jiang. Genomic selection in a commercial winter wheat population. *Theoretical and Applied Genetics*, 129(3):641–651, 2016.
- E. L. Heffner, M. E. Sorrells, and J.-L. Jannink. Genomic selection for crop improvement. 2009.
- C. R. Henderson. Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding*, 1963.
- C. R. Henderson. *Applications of linear models in animal breeding*. University of Guelph, Guelph, ON, Canada, 1984.
- N. Heslot, H.-P. Yang, M. E. Sorrells, and J.-L. Jannink. Genomic selection in plant breeding: a comparison of models. *Crop Science*, 52(1):146–160, 2012.
- A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- J. B. Holland. Theoretical and biological foundations of plant breeding. In *Plant breeding: the Arnel R Hallauer International Symposium*. Blackwell Publishing, Ames, 2006.
- J. B. Holland. Epistasis and plant breeding. *Plant Breeding Reviews: John Wiley & Sons, Inc*, pages 27–92, 2010.
- G. Hoogenboom, C. H. Porter, K. J. Boote, V. Shelia, P. W. Wilkens, U. Singh, J. W. White, S. Asseng, J. I. Lizaso, L. P. Moreno, W. Pavan, R. Ogoshi, L. A. Hunt, G. Y. Tsuji, and J. W. Jones. The DSSAT crop modeling ecosystem. In *Advances in crop modelling for a sustainable agriculture*, pages 173–216. Burleigh Dodds Science Publishing, 2019.
- R. Howard, A. L. Carriquiry, and W. D. Beavis. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes|Genomes|Genetics*, 4(6):1027–1046, 2014.
- X. Hu, B. F. Carver, C. Powers, L. Yan, L. Zhu, and C. Chen. Effectiveness of genomic selection by response to selection for winter wheat variety improvement. *The Plant Genome*, 12(3):180090, 2019.
- X. Hu, B. F. Carver, Y. A. El-Kassaby, L. Zhu, and C. Chen. Weighted kernels improve multi-environment genomic prediction. *Heredity*, 130(2):82–91, 2023.
- H. Ishwaran and J. S. Rao. Generalized ridge regression: geometry and computational solutions when p is larger than n, 2010.
- J. Isidro, J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M. E. Sorrells. Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics*, 128(1):145–158, 2015.
- J. W. Jones, G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. The DSSAT cropping system model. *European Journal of Agronomy*, 18(3-4):235–265, 2003.
- B. A. Keating, H. Meinke, M. E. Probert, N. I. Huth, and I. G. Hills. NWheat: documentation and performance of a wheat module for APSIM. 2001.
- B. A. Keating, P. S. Carberry, G. L. Hammer, M. E. Probert, M. J. Robertson, D. Holzworth, N. I. Huth, J. N. G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J. P. Dimes, M. Silburn, E. Wang, S. Brown, K. L. Bristow, S. Asseng, S. Chapman, R. L. McCown, D. M. Freebairn, and C. J. Smith. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18(3-4):267–288, 2003.
- J. R. Kiriya, R. Blanchet, J. R. Williams, V. Texier, C. A. Jones, and M. Cabelguenne. Sunflower simulation using the EPIC and ALMANAC models. *Field Crops Research*, 30(3-4):403–423, 1992.
- Q. B. Kwong, A. L. Ong, C. K. Teh, F. T. Chew, M. Tammi, S. Mayes, H. Kulaveerasingam, S. H. Yeoh, J. A. Harikrishna, and D. R. Appleton. Genomic selection in commercial perennial crops: applicability and improvement in oil palm (*Elaeis guineensis* jacq.). *Scientific Reports*, 7(1):1–9, 2017.
- A. Lamsal, S. M. Welch, J. W. White, K. R. Thorp, and N. M. Bello. Estimating parametric phenotypes that determine anthesis date in *Zea mays*: Challenges in combining ecophysiological models with genetics. *PLoS One*, 13(4):e0195841, 2018.
- A. D. B. Leakey, J. N. Ferguson, C. P. Pignou, A. Wu, Z. Jin, G. L. Hammer, and D. B. Lobell. Water use efficiency as a constraint and target for improving the resilience and productivity of C3 and C4 crops. *Annual Review of Plant Biology*, 70:781–808, 2019.
- A. Legarra, I. Aguilar, and I. Misztal. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92(9):4656–4663, 2009.
- D. Li, X. Wang, X. Zhang, Q. Chen, G. Xu, D. Xu, C. Wang, Y. Liang, L. Wu, C. Huang, J. Tian, Y. Wu, and F. Tian. The genetic architecture of leaf number and its genetic relationship to flowering time in maize. *New Phytologist*, 210(1):256–268, 2016.

- G. Li, Y. Dong, Y. Zhao, X. Tian, T. Würschum, J. Xue, S. Chen, J. C. Reif, S. Xu, and W. Liu. Genome-wide prediction in a hybrid maize population adapted to Northwest China. *The Crop Journal*, 8(5):830–842, 2020a.
- X. Li, Y. Wei, J. Li, F. Yang, Y. Chen, Y. Chen, S. Guo, and A. Sha. Identification of QTL TGW12 responsible for grain weight in rice based on recombinant inbred line population crossed by wild rice (*Oryza minuta*) introgression line K1561 and *indica* rice G1025. *BMC Genetics*, 21(1):1–10, 2020b.
- J. Liu, J. Huang, H. Guo, L. Lan, H. Wang, Y. Xu, X. Yang, W. Li, H. Tong, Y. Xiao, Q. Pan, Q. Feng, M. S. Raihan, H. Liu, X. Zhang, N. Yang, X. Wang, M. Deng, M. Jin, L. Zhao, X. Luo, Y. Zhou, X. Li, W. Zhan, N. Liu, H. Wang, G. Chen, Q. Li, and J. Yan. The conserved and unique genetic architecture of kernel size and weight in maize and rice. *Plant Physiology*, 175(2):774–785, 2017.
- M. Liu, X. Tan, Y. Yang, P. Liu, X. Zhang, Y. Zhang, L. Wang, Y. Hu, L. Ma, Z. Li, Y. Zhang, C. Zou, H. Lin, S. Gao, M. Lee, T. Lübberstedt, G. Pan, and Y. Shen. Analysis of the genetic architecture of maize kernel size traits by combined linkage and association mapping. *Plant Biotechnology Journal*, 18(1):207–221, 2020.
- X. Liu, H. Wang, H. Wang, Z. Guo, X. Xu, J. Liu, S. Wang, W.-X. Li, C. Zou, B. M. Prasanna, Prasanna, M. S. Olsen, C. Huang, and Y. Xu. Factors affecting genomic selection revealed by empirical evidence in maize. *The Crop Journal*, 6(4):341–352, 2018.
- J. C. W. Locke, M. M. Southern, L. Kozma-Bognár, V. Hibberd, P. E. Brown, M. S. Turner, and A. J. Millar. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology*, 1(1):2005–0013, 2005.
- F. B. Lopes, F. Baldi, T. L. Passafaro, L. C. Brunes, M. F. O. Costa, E. C. Eifert, M. G. Narciso, G. J. M. Rosa, R. B. Lobo, and C. U. Magnabosco. Genome-enabled prediction of meat and carcass traits using Bayesian regression, single-step genomic best linear unbiased prediction and blending methods in Nelore cattle. *Animal*, 15(1):100006, 2021.
- M. A. Lopez, A. Xavier, and K. M. Rainey. Phenotypic variation and genetic architecture for photosynthesis and water use efficiency in soybean (*Glycine max* l. merr.). *Frontiers in Plant Science*, 10:680, 2019.
- M. Lynch and B. Walsh. Genetics and analysis of quantitative traits. 1998.
- P. K. Meher, A. Kumar, and S. K. Pradhan. Genomic selection using Bayesian methods: Models, software, and application. In *Genomics of Cereal Crops*, pages 259–269. Springer, 2022.
- C. Messina, I. A. Ciampitti, D. Berning, D. Bubeck, G. Hammer, and M. Cooper. Sustained improvement in tolerance to water deficit accompanies maize yield increase in temperate environments. *Crop Science*, 62(6):2138–2150, 2022a.
- C. D. Messina, F. Technow, T. Tang, R. Totir, C. Gho, and M. Cooper. Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). *European Journal of Agronomy*, 100:151–162, 2018.
- C. D. Messina, F. Van Eeuwijk, T. Tang, S. K. Truong, R. F. McCormick, F. Technow, O. Powell, L. Mayor, N. Gutterson, J. W. Jones, G. L. Hammer, and M. Cooper. Crop improvement for circular bioeconomy systems. *Journal of the ASABE*, 65(3):491–504, 2022b.
- T. H. E. Meuwissen, B. J. Hayes, and M. E. Goddard. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829, 2001.
- S. Michel, C. Ametz, H. Gungor, B. Akgöl, D. Epure, H. Grausgruber, F. Löschenberger, and H. Buerstmayr. Genomic assisted selection for enhancing line breeding: merging genomic and phenotypic selection in winter wheat breeding programs with preliminary yield trials. *Theoretical and Applied Genetics*, 130(2):363–376, 2017.
- G. Molero, R. Joynson, F. J. Pinera-Chavez, L.-J. Gardiner, C. Rivera-Amado, A. Hall, and M. P. Reynolds. Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential. *Plant Biotechnology Journal*, 17(7):1276–1288, 2019.
- M. Momen, A. A. Mehrgardi, A. Sheikhi, A. Kranis, L. Tusell, G. Morota, G. J. M. Rosa, and D. Gianola. Predictive ability of genome-assisted statistical models under various forms of gene action. *Scientific Reports*, 8(1):1–11, 2018.
- O. A. Montesinos López, A. Montesinos López, and J. Crossa. Bayesian genomic linear regression. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, pages 171–208. Springer, 2022.
- P. A. P. Moran, S. R. A. Fisher, and C. A. B. Smith. *Commentary on RA Fisher's Paper on the Correlation Between Relatives on the Supposition of Mendelian Inheritance: (Transactions of the Royal Society of Edinburgh, 52 (1918), 399-433) By PAP Moran and CAB Smith*. Cambridge University Press, 1966.
- J. L. Neyhart, T. Tiede, A. J. Lorenz, and K. P. Smith. Evaluating methods of updating training data in long-term genomewide selection. *G3: Genes|Genomes|Genetics*, 7(5):1499–1510, 2017.
- C. Palaikostas, S. Cariou, A. Bestin, J.-S. Bruant, P. Haffray, T. Morin, J. Cabon, F. Allal, M. Vandeputte, and R. D. Houston. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in european sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genetics Selection Evolution*, 50(1):1–11, 2018.
- T. Park and G. Casella. The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.
- P. Pérez-Rodríguez, D. Gianola, J. M. González-Camacho, J. Crossa, Y. Manès, and S. Dreisigacker. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3: Genes|Genomes|Genetics*, 2(12):1595–1605, 2012.
- R. Pong-Wong and J. Woolliams. Bayes U: A genomic prediction method based on the horseshoe prior. In *World Congress of Genetics Applied to Livestock Production*, volume 10, 2014.
- M. S. Raihan, J. Liu, J. Huang, H. Guo, Q. Pan, and J. Yan. Multi-environment QTL analysis of grain morphology traits and fine mapping of a kernel-width QTL in Zheng58 × SK maize population. *Theoretical and Applied Genetics*, 129(8):1465–1477, 2016.
- M. Reymond, B. Muller, A. Leonardi, A. Charcosset, and F. Tardieu. Combining QTL analysis and an ecophysiological model to analyse the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiology*, 131:664–675, 2003.
- B. Rice and A. E. Lipka. Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *The Plant Genome*, 12(1), 2019.
- R. Rincón, D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V. M. Rodríguez, J. Moreno-Gonzalez, A. Melchinger, E. Bauer, C. C. Schoen, N. Meyer, C. Giauffret, C. Baulang, P. Jamin, J. Laborde, H. Monod, P. Flament, A. Charcosset, and L. Moreau. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* l.). *Genetics*, 192(2):715–728, 2012.
- S. C. Roy and P. Shil. Assessment of genetic heritability in rice breeding lines based on morphological traits and caryopsis ultrastructure. *Scientific Reports*, 10(1):1–17, 2020.
- A. Salehi, M. Bazrafshan, and R. Abdollahi-Arpanahi. Assessment of parametric and non-parametric methods for prediction of quantitative traits with non-additive genetic architecture. *Annals of Animal Science*, 21(2):469–484, 2021.
- J. P. Santos, H. D. Pereira, R. G. Von Pinho, L. P. Pires, R. B. Camargos, and M. Balestre. Genome-wide prediction of maize single-cross performance, considering non-additive genetic effects. *Genetics and Molecular Research*, 14:18471–18484, 2015.
- J. M. Sarinelli, J. P. Murphy, P. Tyagi, J. B. Holland, J. W. Johnson, M. Mergoum, R. E. Mason, A. Babar, S. Harrison, R. Sutton, C. A. Grifey, and G. Brown-Guedira. Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theoretical and Applied Genetics*, 132(4):1247–1261, 2019.
- M. Shikha, A. Kanika, A. R. Rao, M. G. Mallikarjuna, H. S. Gupta, and T. Nepolean. Genomic selection for drought tolerance using genome-wide SNPs in maize. *Frontiers in Plant Science*, 8:550, 2017.

- A. Simic Milas, M. Romanko, P. Reil, T. Abeysinghe, and A. Marambe. The importance of leaf area index in mapping chlorophyll content of corn under different agricultural treatments using UAV images. *International Journal of Remote Sensing*, 39(15-16):5415–5431, 2018.
- T. R. Sinclair and R. C. Muchow. Radiation use efficiency. *Advances in Agronomy*, 65:215–265, 1999.
- D. Sorensen and D. Gianola. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer, New York, NY, 3rd edition, 2007.
- C. A. Sorgini, L. M. Roberts, M. Sullivan, A. B. Cousins, I. Baxter, and A. J. Studer. The genetic architecture of leaf stable carbon isotope composition in *Zea mays* and the effect of transpiration efficiency on leaf elemental accumulation. *G3: Genes|Genomes|Genetics*, 11(9):jkab222, 2021.
- P. Steduto, T. C. Hsiao, D. Raes, and E. Fereres. AquaCrop—The FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agronomy Journal*, 101(3):426–437, 2009.
- F. Technow, C. D. Messina, L. R. Totir, and M. Cooper. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS one*, 10(6):e0130855, 2015.
- S. Thavamanikumar, R. Dolferus, and B. R. Thumma. Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. *G3: Genes|Genomes|Genetics*, 5(10):1991–1998, 2015.
- J. H. M. Thornley and I. R. Johnson. *Plant and crop modelling*. Clarendon Oxford, 1990.
- F. Tian, P. J. Bradbury, P. J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T. R. Rocheford, M. D. McMullen, J. B. Holland, and E. S. Buckler. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, 43(2):159–162, 2011.
- Y. Toda, G. Sasaki, Y. Ohmori, Y. Yamasaki, H. Takahashi, H. Takanashi, M. Tsuda, H. Kajiya-Kanegae, R. Lopez-Lozano, H. Tsujimoto, A. Kaga, M. Nakazono, T. Fujiwara, F. Baret, and H. Iwata. Genomic prediction of green fraction dynamics in soybean using unmanned aerial vehicles observations. *Frontiers in Plant Science*, 13:828864–828864, 2022.
- H. Ullah, R. Santiago-Arenas, Z. Ferdous, A. Attia, and A. Datta. Improving water use efficiency, nitrogen use efficiency, and radiation use efficiency in field crops under drought stress: A review. *Advances in Agronomy*, 156:109–157, 2019.
- V. Vadez, J. Kholova, S. Medina, A. Kakkera, and H. Anderberg. Transpiration efficiency: new insights into an old story. *Journal of Experimental Botany*, 65(21):6141–6153, 2014.
- I. van den Berg, I. M. MacLeod, C. M. Reich, E. J. Breen, and J. E. Pryce. Optimizing genomic prediction for Australian Red dairy cattle. *Journal of Dairy Science*, 103(7):6276–6298, 2020.
- K. L. Verbyla, P. J. Bowman, B. J. Hayes, and M. E. Goddard. Sensitivity of genomic selection to using different prior distributions. In *BMC Proceedings*, volume 4, pages 1–4. Springer, 2010.
- E. O. Voit. *A first course in systems biology*. Garland Science, 2017.
- K. P. Voss-Fels, M. Cooper, and B. J. Hayes. Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics*, 132:669–686, 2019.
- B. Walsh and M. Lynch. *Evolution and selection of quantitative traits*. Oxford University Press, 2018.
- C. C. N. Wang, P.-C. Chang, K.-L. Ng, C.-M. Chang, P. C. Y. Sheu, and J. J. P. Tsai. A model comparison study of the flowering time regulatory network in Arabidopsis. *BMC Systems Biology*, 8(1):1–12, 2014.
- D. R. Wang, C. R. Guadagno, X. Mao, D. S. Mackay, J. R. Pleban, R. L. Baker, C. Weinig, J.-L. Jannink, and B. E. Ewers. A framework for genomics-informed ecophysiological modeling in plants. *Journal of Experimental Botany*, 70(9):2561–2574, 2019a.
- X. Wang, Y. Xu, Z. Hu, and C. Xu. Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, 6(4):330–340, 2018.
- X. Wang, J. Miao, T. Chang, J. Xia, B. An, Y. Li, L. Xu, L. Zhang, X. Gao, J. Li, and H. Gao. Evaluation of GBLUP, BayesB and elastic net for genomic prediction in Chinese Simmental beef cattle. *PLoS One*, 14(2):e0210442, 2019b.
- J. W. White and G. Hoogenboom. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agronomy Journal*, 88(3):416–422, 1996.
- J. R. Williams, C. A. Jones, J. R. Kiniry, and D. A. Spaniel. The EPIC crop growth model. *Transactions of the ASAE*, 32(2):497–0511, 1989.
- V. Wimmer, C. Lehermeier, T. Albrecht, H.-J. Auinger, Y. Wang, and C.-C. Schön. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics*, 195(2):573–587, 2013.
- J. Wu, S. Wen, Y. Lan, X. Yin, J. Zhang, and Y. Ge. Estimation of cotton canopy parameters based on unmanned aerial vehicle (UAV) oblique photography. *Plant Methods*, 18(1):129, 2022.
- X. Yan, S. Wang, B. Yang, W. Zhang, Y. Cao, Y. Shi, D. Sun, and R. Jing. QTL mapping for flag leaf-related traits and genetic effect of QFLW on flag leaf width using two related introgression line populations in wheat. *Plos One*, 15(3):e0229912, 2020.
- X. Yin, P. Stam, C. J. Dourleijn, and M. J. Kropff. AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley. *Theoretical and Applied Genetics*, 99(1):244–253, 1999.
- P. Yu, C. Ye, L. Li, H. Yin, J. Zhao, Y. Wang, Z. Zhang, W. Li, Y. Long, X. Hu, J. Xiao, G. Jia, and B. Tian. Genome-wide association study and genomic prediction for yield and grain quality traits of hybrid rice. *Molecular Breeding*, 42(4):1–12, 2022.
- M. Zaïm, H. Kabbaj, Z. Kehel, G. Gorjanc, A. Filali-Maltouf, B. Belkadi, M. M. Nachit, and F. M. Bassi. Combining QTL analysis and genomic predictions for four durum wheat populations under drought conditions. *Frontiers in Genetics*, 11:316, 2020.
- H.-W. Zhang, Y.-Y. Fan, Y.-J. Zhu, J.-Y. Chen, S.-B. Yu, and J.-Y. Zhuang. Dissection of the qTGW1. 1 region into two tightly-linked minor QTLs having stable effects for grain weight in rice. *BMC Genetics*, 17(1):1–10, 2016.
- X. Zhang, Z. Guan, L. Wang, J. Fu, Y. Zhang, Z. Li, L. Ma, P. Liu, Y. Zhang, M. Liu, P. Li, C. Zou, Y. He, H. Lin, G. Yuan, S. Gao, G. Pan, and Y. Shen. Combined GWAS and QTL analysis for dissecting the genetic architecture of kernel test weight in maize. *Molecular Genetics and Genomics*, 295(2):409–420, 2020.
- Y. Zhao, J. Zeng, R. Fernando, and J. C. Reif. Genomic prediction of hybrid wheat performance. *Crop Science*, 53(3):802–810, 2013.