Comparing predictive abilities of phenotypic and marker-assisted selection methods in a biparental lettuce population

S. Hadasch\textsuperscript{a,1}, I. Simko\textsuperscript{b,1}, R.J. Hayes\textsuperscript{b}, J. O. Ogutu\textsuperscript{a}, H.P. Piepho \textsuperscript{a}

\textsuperscript{a} University of Hohenheim, Institute of Crop Science, Biostatistics Unit, Fruwirthstrasse 23, 70599 Stuttgart, Germany

\textsuperscript{b} U.S. Department of Agriculture, Agricultural Research Service, Crop Improvement and Protection Unit, 1636 E. Alisal Street, Salinas, CA 93905, USA.

Received \________\ .

\* Corresponding authors (e-mail):

piepho@uni-hohenheim.de

ivan.simko@ars.usda.gov

\textsuperscript{1} The first two authors contributed equally to this work

Abbreviations: AFLP, amplified fragment length polymorphism; BLUP, best linear unbiased prediction; CIM, composite interval mapping; CV, cross validation; DMR, Downy mildew resistance; ES-CV, environmental sampling cross validation; GEBV, genomic estimated breeding values; GP, genomic prediction; GP-MAS, combination of genomic prediction and marker-assisted selection; GS-CV, genotypic sampling cross validation; GSES-CV, genotypic sampling and environmental sampling cross validation; LG, linkage group; MAS, marker-assisted selection; MARS, marker assisted recurrent selection; MLR, multiple linear regression; PS, phenotypic selection; QTL, quantitative trait locus; RIL, recombinant inbred line; RR BLUP, ridge regression BLUP; SL, shelf-life; SNP, single nucleotide polymorphism; T100D, time to 100% decay.
Abstract

Breeding and selection for traits with polygenic inheritance is a challenging task that can be done by phenotypic selection, marker-assisted selection or genome-wide selection. We comparatively evaluated predictive abilities of four selection models on a biparental lettuce population genotyped with 95 SNP and 205 AFLP markers. The four models were based on (i) phenotypic selection, (ii) marker-assisted selection (with QTL-linked markers), (iii) genomic prediction using all the available molecular markers, and (iv) genomic prediction using molecular markers plus QTL-linked markers as fixed covariates. The performance of each model was assessed using data on field resistance to downy mildew (DMR, mean heritability ~0.71) and quality of shelf-life (SL, mean heritability ~0.91) of lettuce evaluated in multiple environments. The predictive ability of each selection model was computed under three cross validation (CV) schemes based on sampling either genotypes, environments or both. For the DMR dataset the predictive abilities of the marker assisted selection model was significantly lower than that of the genomic prediction model. For the SL dataset, the predictive ability of the genomic prediction model was significantly lower than that for the model that uses QTL-linked markers under two of the three CV schemes. Our results show that the predictive ability of the selection models depends strongly on the CV scheme used for prediction and the heritability of the target trait. Our study further shows that molecular markers can be used to predict DMR and SL for individuals originating from this biparental cross that were genotyped but not phenotyped.

Introduction

The use of molecular markers has become widespread in plant breeding since the development of high-throughput genotyping. Genotyping with molecular markers allows
mapping quantitative trait loci (QTLs) (Tanksley, 1993) and estimating their effects. This information can be used in developing assays for marker-assisted selection (MAS). Because a precise assessment of a phenotype is critical for development of MAS assays, at the present time MAS in lettuce is limited to simply inherited traits (Simko, 2013). Selection based on marker information can be performed in several ways (Hospital, 2009): (i) marker-assisted introgression, (ii) screening of populations, (iii) gene pyramiding schemes, (iv) marker-based recurrent selection (MARS), and (v) selection on an index combining molecular and phenotypic scores. Use of markers linked to QTL with significant effects is efficient when the trait of interest is influenced by a limited number of genes only (Heffner et al. 2009). For traits which are influenced by a large number of genes, genome-wide prediction (GP) has been proposed (Meuwissen et al., 2001). In the GP model, all available markers are used to predict breeding values for the unphenotyped genotypes, though the effect of a single marker is usually small. Regardless of the approach used, marker assisted breeding can shorten the time of each selection cycle and accelerate genetic improvement (Varshney et al., 2014). Such selection is particularly advantageous for the traits that are measured using destructive assays, because genotyping can be performed on tissue samples collected from individual plants in early generations. Plants can be further kept for additional analyses and for seed production or other types of propagation.

Marker effects for GP can be estimated using a variety of statistical methods, including Bayesian, Best Linear Unbiased Prediction (BLUP) and machine learning approaches. MAS usually involves multiple linear regression (MLR) of phenotypes on QTL-linked markers (Lorenzana & Bernardo, 2009). GP was reported to be superior to MAS with respect to the expected genetic gain per unit time and cost in winter wheat (Triticum aestivum L.) and
maize (*Zea mays* L.) breeding programs (Heffner et al., 2010). A study performed on maize, barley (*Hordeum vulgare*) and Arabidopsis (*Arabidopsis thaliana* L. Heynh.) datasets revealed that predictive accuracy of GP is usually higher than that of MLR (Lorenzana & Bernardo, 2009). A comparison of the prediction accuracy of the GP, MAS and traditional phenotypic selection (PS) models has also been undertaken for several traits of wheat, indicating that PS outperforms both Bayesian and BLUP approaches of GP, which in turn outperform MLR (Heffner et al., 2011).

To our knowledge no information is available so far concerning empirical comparisons of PS, GP, and MAS models in lettuce (*Lactuca sativa* L.) breeding. Here we compare the predictive abilities of four selection models including PS, GP, MAS and a combination of GP and MAS under three different cross validation (CV) schemes (Fig. 1). A two-stage analysis was applied to each of the three selection methods for two traits. Analyses were performed on a population of recombinant inbred lines (RIL) originating from a biparental cross. The specific objectives of the present study are to compare the predictive abilities of PS, GP, and MAS methods under three different CV schemes and to assess the feasibility of using GP and MAS to guide the development of lettuce breeding lines with improved shelf-life and polygenic resistance to downy mildew.
Material and Methods

Biparental population

A population of 95 RILs was derived from a cross between the cultivars Salinas 88 and La Brillante. ‘La Brillante’ is a Batavia type lettuce which decays rapidly after processing for salad, but has a high field resistance to *Bremia lactucae* Regel (Simko et al., 2012; Hayes et al., 2014a), the pathogen that causes downy mildew of lettuce. ‘Salinas 88’ is a modern iceberg type cultivar that is highly susceptible to downy mildew, but decays very slowly after processing for salad (Simko et al., 2012; Hayes et al., 2014a). Experimental plots were planted using *F*$_8$ seed lots of each RIL plus parents. Genotyping of the population and both parents was performed with 95 single nucleotide polymorphism (SNP) and 205 amplified fragment length polymorphism (AFLP) markers. A detailed description of the genotyping procedure has been published previously (Hayes et al., 2014a).

Resistance to downy mildew

Eighty-nine RILs plus their two parents were tested for resistance to downy mildew (DMR) in three field experiments performed in 2010 and 2011 around the city of Salinas in California, USA. In each experiment, seeds were planted in two or three complete blocks to produce approximately 30 plants of each RIL and parent per plot. Disease assessment was done by scoring disease severity on a scale ranging from 0 (no disease) to 5 (severe disease) for each plot (Simko et al., 2013). Disease assessments were performed at market maturity.
Postharvest decay (shelf-life)

Lettuce plants from the same population were grown in the Salinas area of California (three field experiments, two in 2010 and one in 2011) and the Yuma area of Arizona (one field experiment in 2010) to evaluate decay of salad cut lettuce according to the methods of Hayes and Liu (2008). Plant heads from 90 RILs (86 in one experiment) were harvested, cut into pieces approximately 2.5 cm² in size and stored in plastic bags kept in darkness at 4°C. Up to 9 bags per RIL and parent were arranged in a completely randomized design (CRD). Time to 100% decay (T100D) in days was measured as previously described (Hayes et al., 2014a). Due to environmental effects, population means for T100D ranged from 39 to about 56 days (Hayes et al., 2014a). T100D values were used in all analyses as estimates of shelf-life (SL).

First stage models

A two-stage analysis was implemented to estimate the predictive ability for both traits. In the first stage, adjusted genotype means were estimated. Adjustments were made for the effects of blocks, environments and genotype × environment interaction. In the second stage, the adjusted genotype means from the first stage were used to calculate the genomic estimated breeding values (GEBVs). For the DMR data, we used the following model in the first stage:
$y_{ijh} = \mu + l_h + g_i + (gl)_{ih} + b_{jh} + e_{ijh}$  \hspace{1cm} (1)

where $y_{ijh}$ is the observation for the $i$-th genotype in the $j$-th block ($j=1,\ldots,3$) from the $h$-th experiment ($h=1,\ldots,3$), $\mu$ is the common intercept, $l_h$ is the main effect of the $h$-th experiment, $g_i$ is the main effect of the $i$-th genotype, $(gl)_{ih}$ is the interaction of the $i$-th genotype with the $h$-th experiment, $b_{jh}$ is the effect of the $j$-th block within the $h$-th experiment and $e_{ijh}$ is the residual error. In this model, all effects except $g_i$ were regarded as random effects, each following a normal distribution with zero mean and constant variance. For $e_{ijh}$ experiment-specific variances accounting for heterogeneity of error variances among experiments were fitted.

The SL experiments were laid out as completely randomized designs, i.e. there was no block-effect. Therefore, to analyse the SL data, model (1) was simplified by dropping the block-effects. The resulting model is:

$y_{ijh} = \mu + l_h + g_i + (gl)_{ih} + e_{ijh}$  \hspace{1cm} (2)

where all the terms have the same meanings as in model (1), ($h=1,\ldots,4$), and ($j=1,\ldots,9$). As for the DMR data, experiment-specific error variances were also fitted to the SL data.

**Second stage models**

Adjusted means from the first stage were used as the response variable in the following models in order to estimate genetic variance. Before subjecting the adjusted means to analysis using the models for the second stage, adjusted means for the parents were deleted
to ensure that the estimated genetic variance is based purely on the information from the progenies. The model used in the second stage of GP is:

$$ m = 1μ + Zu + e $$  \hspace{1cm} (3)$$

where $m$ is a vector of adjusted genotype means, $1$ is a vector of ones, $μ$ is the common intercept, $Z$ is the design matrix of marker effects, $u$ is the vector of random marker effects distributed as $N \sim (0, Iσ_u^2)$, where $I$ is the identity matrix and $e$ is the vector of errors associated with $m$, assumed to have zero mean and variance-covariance matrix $R$, i.e., $e \sim (0, R)$. This model yields ridge regression BLUP (RR-BLUP) of the genotypes (Piepho, 2009). In the following text, model (3) will be denoted as a genomic prediction (GP) model.

In all the second stage models, $R$ was replaced with the estimated variance-covariance matrix of the adjusted genotype means from the first stage and was treated as a fixed and known quantity in the second-stage analysis. This ensures that the estimated variance-covariance matrix of adjusted means, which represents the error structure of the phenotypic data, is carried forward from the first to the second stage (Piepho et al., 2012). Fixing $R$ in the second stage is important because we cannot otherwise estimate both the genetic and error variance from $m$ as it contains only one observation for each genotype. Therefore, the error variance has to be fixed in the second stage to avoid confounding with the genetic variance.

For marker-assisted selection (MAS), the markers detected through Composite Interval Mapping (CIM) were used in a multiple regression model with the markers considered as fixed effects. The model can be written as:

$$ m = 1μ + Mv + e $$  \hspace{1cm} (4)$$
where \( m, \mu \) and \( e \) are defined as above, \( M \) represents the known matrix of marker covariates which identify the QTL and \( v \) is a vector of fixed marker effects. Model (4) will be denoted as MAS in what follows.

The GP model can be extended with QTL markers; we call the resulting model the GP-MAS model:

\[
m = 1\mu + Zu + Mv + e \quad ,
\]

where QTL markers are considered as fixed covariates. All the terms in this model are defined analogously as for the GP and MAS models. For the PS model, the adjusted means from model (1) and (2) were used as the training dataset in the cross validation schemes described below.

**Cross validation**

When fitting a model to a given dataset, the model parameters are optimized with respect to the data such that the best fit in terms of the minimum error variance is achieved. The estimated model parameters and the estimated error variance in multi-environment trials are affected by the sample of both genotypes and environments. Therefore a certain degree of uncertainty exists in both the estimated parameters and the estimated error variance due to sampling. The exact amount of the uncertainty cannot be determined as the whole population is not observable. However, cross validation (CV) can be used in order to get a measure for the average amount of uncertainty introduced by sampling genotypes and environments. In a \( k \)-fold CV, for example, the dataset is divided into \( k \) equally or approximately equally sized subsets. One of the \( k \) subsets (validation set) is used in turn to
validate predictions based on a model fit to the other (k-1) subsets (training set). The Pearson correlation coefficient between predictions of the model based on the training set and the validation set can be used to quantify the ability of the model to predict the unobserved phenotypic data.

The GP analysis is used for predicting breeding values of genotypes that have not been phenotyped. This is possible by estimating the effect of each marker and then using the effects to predict the GEBV for an unphenotyped individual for which the marker profile is available. In order to evaluate the ability of a model to predict GEBVs for the unphenotyped individuals, a k-fold CV is commonly used in GP. Here, we use a threefold CV procedure to evaluate the predictive ability of the GP and the GP-MAS model. To increase precision, the splitting of the dataset into three folds was replicated 10 times. The training and validation set consisted of adjusted genotype means generated by analysing the phenotypic data from all the experiments using model (1) or (2). K-1 folds of the adjusted means were concatenated to produce the training set \( (m_{t1}, \text{Fig. 1}) \) which was used to train model (3) to (5). The estimated parameters from this analysis were then applied to predict the GEBVs of the remaining k-th fold \( (m_{v1}) \). As this CV scheme represents genotypic sampling, we denote it by genotypic-sampling cross validation (GS-CV) in what follows.

The prediction of genomic breeding values of the unphenotyped genotypes is not possible with models in which genotypes are treated as fixed effects, such as the PS model, or models in which genotypes are assumed to be uncorrelated or independent. As the major aim of this study was to compare the performances of the four different selection models, the GS-CV scheme could not be used to compare the predictive abilities of all the four models considered here. However, the predictive ability of the PS model can be quantified...
by splitting the data into sets in each of which all the genotypes occur. In the case at hand, all the genotypes (with very few exceptions) occurred in each experiment. Therefore, CV was done across experiments, meaning that the phenotypic data were split into sets on the basis of the available field experiments (environments), resulting in three sets for DMR and four sets for SL. Because this CV procedure samples environments, we denote it as environmental-sampling cross validation (ES-CV). Using this way of splitting the dataset and denoting the number of sets by \( n \) (\( n = 3 \) for DMR and \( n = 4 \) for SL), \( n-1 \) sets were used for training and the remaining \( n \)-th set for validation (Fig. 1). The validation set \( \{v_2\} \) consisted of genotype means generated by analysing phenotypic data from a single experiment using model (1) or (2) and, thus excluding the experiment main effect and the genotype \( \times \) experiment (environment) interaction effects. The training set \( \{t_2\} \) consisted of adjusted genotype means obtained by analysing phenotypic data from \( n-1 \) experiments using model (1) or (2). The training data were analysed using models (3) to (5) and the GEBVs predicted by these models were correlated with the validation set to compute predictive ability. In the ES-CV scheme, the predictive ability quantifies the degree to which the predicted GEBVs approximate the expected phenotypic performance of genotypes in an unobserved experiment.

The two preceding CV schemes thus differ mainly in the sense that the GS-CV scheme accounts for genotypic sampling whereas the ES-CV scheme accounts for environmental sampling. In order to implement genotypic sampling in the ES-CV scheme, the training and validation sets for the ES-CV scheme were further subdivided into \( k = 3 \) subsets by sampling genotypes using the same splitting criterion as that used in the GS-CV scheme. In Fig. 1 and in the following text, we denote the \( k-1 \) folds amalgamated to form the training set by \( t_3 \).
and the held out $k$-th fold by $m_{t_3}$. The corresponding $k$-1 folds of the validation set are denoted by $m_{v_{t_3}}$ and the remaining $k$-th fold by $m_{v_3}$. As for the GS-CV, the splitting of the data into $k$ folds was replicated 10 times. This CV scheme will be called genotypic-sampling environmental-sampling cross validation (GSES-CV). Using the same splits in both the GS-CV and the GSES-CV schemes eliminates all sources of differences between the two schemes other than that due to environmental sampling in the GSES-CV. For models that use marker information, the predictions of $m_{t_3}$ were computed by using the marker effects estimated by applying models (3) to (5) with $m_{v_3}$ as the response variable. Predictions for the PS model were obtained in two steps. In the first step, $m_{t_3}$ was used in a regression relating $m_{v_{t_3}}$ to $m_{t_3}$ (Fig. 1). In the second step, the estimated intercept and slope from this regression were used to build a linear model for predicting the GEBVs for genotypes in $m_{t_3}$.

The training of the PS model involves both $m_{v_{t_3}}$ and $m_{t_3}$ whereas the training of the models that use marker information only involves $m_{t_3}$. Because more data are used for the training of the PS model, this difference needs to be taken into account when the predictive abilities of the models are compared.

Using the GSES-CV scheme enables the estimate of the predictive ability of a model to take both genotypic and environmental sampling into account. A similar comparison of CV schemes under different sampling strategies was used by Utz et al. (2000) to study the genetic variance explained by QTL-linked markers.

< Figure 1 about here >

Choosing markers for MAS models
QTLs were determined for both the population of 90 RILs and for each of the training datasets that were generated in the GS-CV scheme in order to account for variation due to sampling genotypes. QTL mapping was performed in the QGene v. 4.3.9 software package (Joehanes and Nelson, 2008) using CIM with the automatic forward-selection of cofactors. The threshold for significant QTL was set at the genome-wide α value of 0.05 by a permutation procedure with 1,000 iterations. The molecular markers closest to these QTLs were subsequently used as non-random predictor variables in the ES-CV scheme. To account for variation in QTL detection due to sampling genotypes, each of the 30 training datasets of the GS-CV scheme was screened for QTLs. For each training dataset the molecular-markers closest to the significant QTLs detected for the dataset were subsequently used in the MAS and GP-MAS models under the GS-CV and the GSES-CV schemes.

**Calculation of heritability and genetic correlation**

As aids to interpreting the estimated predictive abilities, we computed heritability for each experiment and genetic correlation; the correlations between the performances of all genotypes in all pairs of environments. We calculated heritability for each experiment separately as the squared correlation between the expected value of the phenotypic variance and the expected value of the genetic variance (Estaghvirou et al., 2013). This method was found to consistently perform best in terms of predicting the true heritability compared to other heritability estimation methods under a wide variety of scenarios in several simulation studies (Estaghvirou et al., 2013, 2014, 2015). The genetic variance was estimated using an RR-BLUP model assuming that all the genotypes are correlated according
to the linear variance-covariance structure specified in terms of the marker information of the genotypes.

Genetic correlation between experiments was estimated by using a linear mixed model in which the genetic variances for all genotypes in each experiment as well as the covariances between the performances of all the genotypes in all pairs of the three or four different experiments were estimated. This was achieved by modelling the covariance of the elements in the interaction between genotypes and experiments using an unstructured variance-covariance matrix in the MIXED procedure of the SAS system (Piepho & Möhring, 2011).

**Testing for differences between models**

We tested for significant differences between the mean predictive abilities (correlation coefficients) of the four selection models using a linear mixed model and compared the mean predictive abilities between pairs of the four models using t-tests adjusted for multiplicity using simulation adjustment in the MIXED procedure of the SAS system (SAS Institute 2015). The specific mixed model used differed depending on the particulars of the cross validation scheme. Thus, under the GS-CV scheme the mixed model consisted of a fixed effect for the selection models and random effects for the CV-replicates \((n = 10)\) and CV-folds \((n = 3)\) nested within replicates. Fitting random effects for replicates and folds nested within replicates induces correlations among all the predictive abilities across all the replicates and all predictive abilities across all the folds nested within replicates. The same model as for the GS-CV scheme was used for the GSES-CV scheme except for the inclusion of
additional fixed effects for the validation dataset and the interaction between the selection model and the validation set. Under the ES-CV scheme, there is only one predictive ability estimate for each model and validation set combination. As a result, the mixed model used to compare the predictive abilities under this scheme accounted only for the fixed effects of the selection models and the validation sets.

Results

QTL detection

When all 90 RILs were used for QTL analyses, three significant QTLs for SL were detected on linkage groups (LG) 1, 4, and 9, while four significant QTLs for DMR were located on LG 4 (two QTLs), 7, and 9 (Table 1). The number of QTLs and their effects are somewhat different from previous reports (Hayes at al., 2014a; Simko et al., 2015) because QTL analyses in those reports were performed on data from all evaluations of the traits (i.e. weekly evaluations of disease and decay progress), while the current analyses focused only on DMR resistance at harvest maturity and SL assessed as T100D. QTL mapping performed on the training datasets of the GS-CV determined the same QTLs as were found through analyses of all RILs, with the exception of qSL1 for SL that was not detected on the training datasets, and qDM2.2 that was not detected on complete dataset at harvest maturity, but was significant at earlier evaluations of resistance to downy mildew.

< Table 1 about here >

Heritability and genetic correlation estimates
Heritability of SL, the trait that is conferred by a few genes (one of them having a very large effect), was substantially higher in all experiments (0.807 – 0.971) than heritability of DMR (0.612 – 0.760), the trait is conferred by a larger number of QTLs with smaller effects (Table 2). The estimated genetic correlation for a pair of experiments was similar for both traits (Table 3) and ranged from 0.702 to 0.829 for DMR (mean = 0.776) and from 0.673 to 0.820 for SL (mean = 0.731).

Predictive ability for the DMR data

The predictive ability of the MAS model evaluated under the GS-CV scheme (0.258) was significantly lower than those for the GP (0.607) and the GP-MAS (0.544) models (Table 4). Under the ES-CV scheme the mean predictive ability across all the validation sets obtained for the PS, GP and GP-MAS models were not significantly different from each other (0.676, 0.674, and 0.677, respectively) while the predictive ability of the MAS model was significantly lower (0.471) than those of all the other three models (Table 4). The mean predictive abilities of the GP and GP-MAS models evaluated under the GSES-CV scheme were almost identical (0.487 and 0.488, respectively). These values were significantly lower than that of the PS model (0.680), but significantly higher than the predictive ability of the MAS model (0.344) (Table 4). The MAS model had consistently and significantly the lowest predictive abilities of the four models. The predictive abilities of the GP and GP-MAS models were lowest under the GSES-CV scheme and highest under the ES-CV scheme, while for the
MAS model the lowest predictive ability was observed under the GS-CV scheme. Results of PS were similar under both CV schemes used for evaluation of this model.

Predictive ability for the SL data

Phenotypic data from the SL experiments were analysed according to model (2), accounting for experiment-specific error variances and covariances in the first stage. Diagnostic plots of standardized residuals revealed that the assumptions of homogeneity of residual variance as well as normality of residuals were slightly violated, even though experiment specific error variances were fitted. While the square-root transformation and the logarithmic transformation of the response variable led to higher predictive abilities, they did not improve residual plots (data not shown). Therefore, the untransformed data were used for analysis. The predictive ability of the GP model evaluated under the GS-CV scheme was significantly lower (0.448) than that for the MAS (0.691) and the GP-MAS (0.649) model (Table 5). The predictive abilities of the GP and GP-MAS models substantially increased when evaluated under the ES-CV scheme (0.765 and 0.752, respectively). The values for the GP model in these evaluations were comparable to those for the PS model for all the four validation sets. In contrast, the predictive abilities of the MAS and the GP-MAS models were similar to predictive abilities of the other two models only when experiments 1 and 2 were used for validation but were lower otherwise. The mean predictive abilities of the four selection models, however, were not significantly different from each other.
The predictive abilities of the GP model were lowest from the four models under the GSES-CV scheme for each validation set (Table 5). The inclusion of the QTL-linked markers (the MAS and GP-MAS models) increased predictive abilities under the GSES-CV scheme for the SL data, though even these values were lower than those observed for the PS model. As with the DMR data, predictive abilities of the marker-based models were lower under the GSES-CV than under the ES-CV scheme.

< Table 5 about here >

DISCUSSION

Predictive ability for DMR

The predictive ability of the MAS, GP and the GP-MAS model under the GS-CV scheme were 0.258 ± 0.199, 0.607 ± 0.096 and 0.544 ± 0.152 (mean ± standard deviation), respectively. This indicates that the QTL-linked markers do not contribute to the explanation of the variability in the DMR data. The predictive ability of the GP model under the GS-CV scheme is as high as or higher than that reported for wheat resistance to rust (0.4) (Daetwyler et al., 2014), yellow rust (0.1 to 0.5), and stem rust (0.4 to 0.7) (Ornella et al., 2012). In our study, the standard deviation of the predictive ability of the GP-MAS model was larger than that for the GP model under the GS-CV scheme. A loss of precision when using the GP-MAS model has previously been reported for wheat resistance to Fusarium head blight (Rutoski et al., 2012). The increased variance of the GP-MAS model under the GS-CV scheme may be caused by the fact that the same QTLs were not identified in each of the training datasets. Therefore, the flanking markers differed between the training datasets which likely increased variability in predictive abilities.
The mean predictive abilities of the GP and GP-MAS models were higher under the ES-CV than under the GS-CV scheme. This could be a result of the fact that the ES-CV scheme does not take into consideration genotypic sampling that increases inaccuracy to the larger extent than sampling across experiments. Another factor that may be playing a role in observed differences between GS-CV and ES-CV schemes is that a larger number of genotypes are used to estimate the model parameters in ES-CV than in GS-CV (ca. 90 vs. 60, respectively), leading to more precise estimates. The MAS model had the lowest mean predictive ability compared to the other models in the ES-CV scheme (Table 4). This shows that even in the absence of genotypic sampling, the MAS model explains less variation than the GP model does. Our results are thus in line with previous observations that the GP model typically outperforms the MAS models when polygenic traits are evaluated under the GS-CV scheme (Heffner et al., 2011).

The pattern of variation in predictive ability of all the models under the ES-CV scheme across experiments is similar to that shown by the heritabilities. Notably, if a trait has a low heritability in an experiment then its predictive ability is also low (e.g. Experiment 1). Similarly, variations in predictive abilities under the ES-CV scheme show the same tendency as the average genetic correlations, i.e. the average correlation of one experiment with the other experiments. For example, the average genetic correlation between experiment 1 and the other two experiments is lower than that between experiment 3 and the other two experiments (Table 3). A similar pattern is evident in the variation in predictive ability such that the predictive ability obtained when experiment 1 is the validation set is generally lower than that obtained when experiment 3 is the validation set.
The GSES-CV scheme led to the lowest mean predictive abilities for all models with the exception of the PS model (Table 4). The high predictive ability of the PS model under the GSES-CV scheme is likely a result of the training procedure of the PS model. As mentioned in the materials and methods section, the training of the PS model involves more data than the training of the models that use marker information and further, the data that are used for training of the PS model, namely $m_{st}$, are not independent of the validation set $m_{v3}$. Both of these aspects may artificially increase the predictive ability of the PS model under the GSES-CV scheme, so the comparison of the PS model and the models that use marker information is not completely fair and the significant differences between the mean predictive abilities of the PS model and the other models found here may be caused by the differences in the model training procedures and not directly by a better performance of the PS model.

The key difficulty in deriving a fair CV method is that with the PS model it is not possible to predict the phenotypic performance of unobserved genotypes as the correlation between genotypes is not accounted for in this model. To calculate a predictive ability under the GSES-CV scheme using a simple regression model therefore requires further information about the performance of the genotypes under different conditions (experiments) to estimate the parameters of the regression model. Thus, a comparison of the PS model with the marker-based models as implemented here is not entirely objective with regard to the data being used to estimate model parameters. However, the comparison of the PS model and the other models under the GSES-CV scheme is still valuable because it can be expected that the predictive ability of the PS model under the GSES-CV scheme is overestimated and
therefore, depending on the sign, differences between the PS model and the other models represent lower or upper bounds of the actual differences in predictive abilities.

The results of the GSES-CV show that sampling both genotypes and environments simultaneously reduces predictive ability of the models to a greater degree than separately sampling either genotypes or environments only. Under the GSES-CV scheme only the predictive abilities produced by the GP model exhibit variation across experiments similar to that of the heritabilities. Moreover, under the GSES-CV scheme, variation in the predictive abilities does not reflect the pattern displayed by the average genetic correlations. This possibly reflects the fact that heritability and genetic correlations were calculated based on the entire population of genotypes whereas the predictive abilities were calculated based on a subset of genotypes.

Predictive ability for SL

The MAS and the GP-MAS models had significantly higher predictive ability compared to the GP model for the SL data evaluated under the GS-CV scheme. The higher predictive ability of the MAS and GP-MAS models for the SL data than for the DMR data is likely due to the fact that the QTLs for SL explained more variation than the QTLs for the DMR data. It was reported that the *qSL4* is the major determinant of salad decay (Hayes et al, 2014a) and explained about 74% of the total phenotypic variance of the trait (Table 1). Our results show that in the presence of genotypic sampling the QTL on LG 4 was found in all the 30 training datasets and explained up to 85.4% of the phenotypic variance. In contrast, none of the
QTLs linked to DMR was consistently detected in all the experiments when all genotypes were screened. The largest percentage of the total phenotypic variation explained by a single QTL was estimated to be a little more than 30% (Simko et al., 2015, Table 1). Therefore, the QTL-linked markers for SL also account for more variation in the GP-MAS model and lead to a higher predictive ability. However, the standard deviation of the mean predictive ability for the GP-MAS model was larger than that for the GP model under the GS-CV scheme, as was also the case for the DMR data.

Predictive abilities of the GP and GP-MAS models tested under the ES-CV scheme were higher than those under the GS-CV scheme. Similar results were observed for the DMR data indicating that sampling of genotypes decreases predictive ability of the models to a larger degree than sampling of environments. Predictive abilities calculated for the MAS model showed a large variation among experiments under the ES-CV scheme (Table 5) signalling that the effect of the QTLs varies across experiments. It is also possible that others, yet undetected QTLs, are involved in controlling the variability of SL. Further, in comparison to the DMR data, the larger standard deviation of the mean predictive ability of the MAS probably reflect the fact that the QTL-linked markers explain a larger percentage of the total variation than the DMR data (Table 1). The average genetic correlations for the validation sets show a similar trend as the predictive abilities of all the models under the ES-CV scheme. Heritabilities also show the same trend except when experiment 4 is used as validation set.

For the SL data, the predictive ability of the PS model under the GSES-CV scheme was significantly higher than for the other models (Table 5). As discussed for the DMR data, this may also be caused by the training procedure of the PS model implemented here.
The addition of QTL-linked markers to the GP model (GP-MAS model) significantly increased predictive ability of SL tested under the GSES-CV scheme even though not all genotypes were used for training in this scheme (Table 5). The mean predictive abilities of both models, however, were lower under the GSES-CV scheme than those under the GS-CV or ES-CV schemes. A similar trend of the mean predictive abilities under the different CV schemes can also be seen for the GP and GP-MAS models using DMR data. Under the GSES-CV scheme the predictive ability of the GP model does not reflect the pattern shown by the average genetic correlations of a certain experiment or by the heritability. However, the predictive abilities of the MAS and the GP-MAS models show similar patterns as the average genetic correlations and heritability except when experiment 4 is used for validation. This is likely because the QTL-linked markers capture large fractions of variation (Table 1) in both the training and validation sets.

**Other design and modelling considerations influencing predictive ability**

The analysed experiments were laid out in a randomized complete block design (DMR) or in a completely randomized design (SL). The efficiency of GP models can be increased by using unreplicated designs and simultaneously increasing the number of environments regardless of whether genotypes are assumed to be correlated or independent (Möhring et al., 2014). Similarly, using spatial models with experiments laid out according to augmented designs with checks can also increase the predictive abilities of both the PS and the MAS models.
(Moreau et al., 1999). An increase in the predictive accuracy of the GP model by using spatial adjustments was also reported (Lado et al., 2013). Accordingly, it is likely that the level of predictive abilities we found for the DMR and SL data could be further increased by using spatial models or by allowing a larger number of environments to be sampled with the same total number of plots.

**Choice of selection model**

Lettuce resistance to downy mildew is determined by single, dominant genes conferring qualitative, race specific resistance to the disease and by a combination of multiple genes conferring quantitative (polygenic) resistance. The DMR data analyzed in the present study were collected from evaluations of lettuce resistance in field conditions where only quantitative resistance to *B. lactucae* was observed (Simko et al., 2015). Therefore, the predictive abilities found here are limited to polygenic resistance with each of the detected QTLs explaining only a small fraction of the total phenotypic variation of the trait. Furthermore, neither of the QTLs was detected consistently in all experiments leading to a significant QTL × environment interaction effect detected for all QTLs (Simko et al., 2015). These factors could contribute to the fact that the mean predictive ability of the MAS model was significantly lower than that of the GP model (Table 4). Further, the analyses implemented here do not account for race-specific resistance in general as adjusted genotype means were calculated from multiple environments. A prediction based on race-specific genes would be meaningless anyway as the avirulence genes of the pathogen can vary across locations and growing seasons and multiple races of the pathogen can be
present concurrently. Therefore, the GP model seems to be the most suitable for selecting genotypes with polygenic resistance to downy mildew.

Postharvest decay of fresh-cut lettuce in this population is a relatively simply inherited trait. A single QTL on LG4 ($q_{SL4}$) was detected in all experiments and was a major determinant of the trait explaining up to 74% of the total phenotypic variation (Hayes et al., 2014). The predictive abilities of the MAS model were comparable or higher than those of the GP model (Table 5), indicating that the MAS model could be favored for selection of desirable genotypes. However, the standard deviation of the MAS model was larger compared to the other models under the ES-CV and the GSES-CV scheme. If only currently known QTLs were used for selection, the most optimal model appears to be GP-MAS with the predictive ability similar to the MAS model, but more consistent results.

**Current and envisaged analyses of the ‘Salinas 88’ × ‘La Brillante’ progeny**

‘Salinas 88’ and ‘La Brillante’ are frequently used in our breeding program for introgression of desirable traits. Beside the two traits investigated in the current study (DMR and SL), the cultivars significantly differ in resistance to Verticillium wilt (Hayes et al., 2011), bacterial leaf spot (Hayes et al., 2014b), big-vein, leafminer, lettuce mosaic virus, tipburn, and several economically important horticultural traits such as size, shape, closure, and firmness of a head and leaf color (Simko et al., 2014). RILs of the ‘Salinas 88’ × ‘La Brillante’ mapping population were extensively phenotyped and markers linked to majority of these traits were identified (Hayes et al., 2011, 2014a, 2014b, Simko et al., 2015). Germplasm was recently released from ‘Salinas 88’ × ‘La Brillante’ that combines downy mildew, Verticillium wilt, and
bacterial leaf spot resistance with improved shelf-life, tipburn resistance, and horticultural characters (Simko et al., 2014). While the released line needs further improvements to horticultural quality, we expect that additional progenies originating from ‘Salinas 88’ × ‘La Brillante’ and from matings with this germplasm will be used to develop new breeding lines and cultivars that combine all desirable traits from both parents. However, the horticultural and resistance traits segregating are conferred by numerous QTLs and single dominant genes, thus several thousands of individuals need to be evaluated in multiple experiments to identify the best performing individuals. Because such extensive phenotypic evaluations are very labor intensive and time demanding, use of molecular markers offers an attractive assistance for selecting desirable genotypes in less time.

To use molecular markers for the selection, the markers have to provide a reliable estimate of performance for individuals that would be genotyped, but not phenotyped; similarly as was tested under the GSES-CV scheme. Our study determined that the best prediction models for the trait with relatively lower heritability (DMR) and large QTL × environment interaction effects are GP and GP-MAS, reaching 72% of the ability seen for PS (Table 4), while for the trait with high heritability (SL) the best results were observed for the MAS model with 80% ability compared to PS (Table 5). The GP-MAS model was, however, the most consistent in estimating the performance for both traits. When compared to PS, the relative predictive ability of this model ranged from 72% for DMR to 76% for SL, whereas the ability decreased to 51% for DMR using MAS and to 47% for SL using GP (Tables 4 and 5). When the predictive abilities of the models that use marker information are contrasted to that of the PS model, the predictive ability inflation of the PS model due to the training procedure has to be accounted for, meaning that the real relative performances of the
marker based models may be even higher than observed here. Predictive abilities of the best models indicate that the current set of molecular markers allows estimating breeding values for unphenotyped genotypes of ‘Salinas 88’ × ‘La Brillante’, thus reducing the need for extensive, multi-trial phenotyping. To further increase predictive abilities of the models we are expanding the training population to include other full-sib families that originate from matings of the two parents with additional cultivars.

Conclusions

As expected for polygenic traits with a large QTL × environment interaction, the GP model outperformed the MAS model for the DMR data. For this trait, the MAS model is less suitable for selection than the other two molecular-marker based models. In contrast to the DMR data, the most appropriate selection model for the SL data was less obvious. Including QTL-linked markers in the prediction models generally increased predictive ability. Even so, the predictive abilities of the MAS model varied substantially between experiments. This suggests that the GP-MAS model may be preferred for selecting desirable genotypes for SL in the lettuce population under consideration. Based on the results obtained from DMR and SL datasets, it is advisable to evaluate the predictive ability of a model under all three CV schemes. Analyses performed using the three CV schemes provide a reliable assessment of the influence of both genotypic and environmental sampling on predictive ability.

Our results suggest that the genetic correlation between environments as well as location-specific heritabilities may provide estimates of the expected predictive ability of a model under specific CV schemes. Hence, further exploration of the relationships between
predictive ability and both heritability and genetic correlation may help in optimizing the selection of environments and genotypes likely to yield high predictive abilities. The results of our study show that molecular markers can be effectively used for predicting a line’s resistance to downy mildew and quality of shelf-life, the traits which evaluations are highly demanding for time and labor. Employing molecular markers in ‘Salinas 88’ × ‘La Brillante’ progeny can reduce the need for additional, extensive field experiments that are frequently used for assessing traits with polygenic inheritance.
Literature


Figure1: Overview of the three cross validation (CV) schemes. Numbers in parentheses correspond to the models described in the materials and methods section. In all CV schemes the vector $u$ represents the random effects of SNP and AFLP markers and $v$ represents the fixed effects of QTL-linked markers. In the genotypic sampling CV scheme (GS-CV), phenotypic data from all the experiments were analysed using model (1) or (2) to obtain adjusted genotype means. These means were then split into the training ($m_{t_1}$) and validation ($m_{v_1}$) sets. Marker effects ($u$ and $v$) were estimated using models (3) to (5). Predicted genomic estimated breeding values (GEBVs) for the validation set ($m_{v_1}$) were obtained using the marker effects estimated from the training set. The $m_{v_1}$ were correlated with the GEBVs in the validation set to estimate predictive ability. In the environmental sampling CV scheme (ES-CV) adjusted means generated by analysing phenotypic data from $n$-1 experiments were used as the training set ($m_{t_2}$). The validation set ($m_{v_2}$) consisted of adjusted means obtained from analysing phenotypic data from the $n$-th experiment only. Note that for the phenotypic selection (PS) method $m_{t_2}$ was used directly as the predicted GEBVs. Predictive ability was computed as the correlation between the GEBVs for each of the four models and the GEBVs in the validation set $m_{v_2}$. In the genotypic- and environmental sampling CV scheme (GSES-CV), the training and validation sets derived from the ES-CV scheme were further split into $k=3$ subsets. The training set ($m_{t_3}$) was used to estimate the marker effects based on models (3) to (5). The marker effects were then used to predict the GEBVs for $m_{t_3}$. The $m_{t_3}$ were correlated with $m_{v_3}$ to estimate predictive ability.
Table 1: Maximum percentage of variation explained by QTL-linked markers for resistance to downy mildew (DMR) and shelf-life (SL) calculated by composite interval mapping.

<table>
<thead>
<tr>
<th>QTL</th>
<th>DMR</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All evaluations</td>
<td>Harvest maturity</td>
</tr>
<tr>
<td>qDM2.2</td>
<td>32.8</td>
<td>n.s. #</td>
</tr>
<tr>
<td>qDM4.1</td>
<td>25.4</td>
<td>19.5</td>
</tr>
<tr>
<td>qDM4.2</td>
<td>31.9</td>
<td>31.9</td>
</tr>
<tr>
<td>qDM7.1</td>
<td>32.9</td>
<td>27.5</td>
</tr>
<tr>
<td>qDM9.2</td>
<td>23.0</td>
<td>18.2</td>
</tr>
</tbody>
</table>

† Numeral before decimal point in the QTL designation indicates the linkage group harboring the QTL.
‡ Results from all field evaluations of DMR (Simko et al., 2015) and SL (Hayes et al., 2014a).
§ Results from the evaluations of DMR at harvest maturity and SL as T100D only.
¶ Results from the evaluations of the training datasets of the GS-CV scheme. These datasets consist of 1/3 fewer genotypes than the dataset used for all evaluations. For more details about the datasets used and the detection of QTLs see the sections cross validation and choosing markers for MAS models.
# n.s. QTL not significant at α ≤ 0.05.
Table 2: Heritability resistance to downy mildew (DMR) and for shelf-life (SL) calculated for each experiment separately. Numbering of experiments is independent for each trait; i.e., experiments with the same number were not used to evaluate both DMR and SL.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>DMR</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>0.612</td>
<td>0.931</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>0.743</td>
<td>0.911</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>0.760</td>
<td>0.807</td>
</tr>
<tr>
<td>Experiment 4</td>
<td></td>
<td>0.971</td>
</tr>
</tbody>
</table>
Table 3: Genetic correlations between pairs of experiments for resistance to downy mildew (DMR) and for the shelf-life (SL). Numbering of experiments is independent for each trait; i.e., experiments with the same number were not used to evaluate both DMR and SL.

<table>
<thead>
<tr>
<th></th>
<th>DMR</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exp. 1 Exp. 2 Exp. 3</td>
<td>Exp. 1 Exp. 2 Exp. 3 Exp. 4</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>0.797 1</td>
<td>0.820 1</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>0.702 0.829 1</td>
<td>0.708 0.698 1</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>0.726 0.673 0.759 1</td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Predictive abilities of the four selection models based on three cross validation (CV) schemes for downy mildew resistance (DMR) data.

<table>
<thead>
<tr>
<th>CV scheme ‡</th>
<th>Validation set: genotype means from</th>
<th>Second stage model †</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS-CV</td>
<td>all experiments</td>
<td>PS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>ES-CV</td>
<td>Experiment 1</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>Experiment 2</td>
<td>0.736</td>
</tr>
<tr>
<td></td>
<td>Experiment 3</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td>Mean§</td>
<td>0.676 a (0.067)</td>
</tr>
<tr>
<td>GSES-CV</td>
<td>Experiment 1</td>
<td>0.601</td>
</tr>
<tr>
<td></td>
<td>Experiment 2</td>
<td>0.743</td>
</tr>
<tr>
<td></td>
<td>Experiment 3</td>
<td>0.695</td>
</tr>
<tr>
<td></td>
<td>Mean§</td>
<td>0.680 a (0.072)</td>
</tr>
</tbody>
</table>

† PS, phenotypic selection; MAS, marker assisted selection; GP, genomic prediction; GP-MAS, combination of GP and MAS models.

‡ GS, genotypic sampling; ES, environmental sampling; GSES, genotypic- and environmental sampling.

§ Mean correlation coefficient across all validation datasets. Numbers in parenthesis represent the standard deviation of the mean.

¶ Means within a row followed by different letters are significantly different at p ≤ 0.05.
Table 5: Predictive abilities of the four selection models based on three cross validation (CV) schemes for shelf-life (SL) data.

<table>
<thead>
<tr>
<th>CV scheme</th>
<th>Validation set: genotype means from</th>
<th>PS</th>
<th>MAS</th>
<th>GP</th>
<th>GP-MAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS-CV</td>
<td>all experiments</td>
<td>-</td>
<td>0.691&lt;sup&gt;a&lt;/sup&gt; (0.183)</td>
<td>0.448&lt;sup&gt;b&lt;/sup&gt; (0.109)</td>
<td>0.649&lt;sup&gt;a&lt;/sup&gt; (0.159)</td>
</tr>
<tr>
<td>ES-CV</td>
<td>Experiment 1</td>
<td>0.812</td>
<td>0.865</td>
<td>0.810</td>
<td>0.852</td>
</tr>
<tr>
<td></td>
<td>Experiment 2</td>
<td>0.776</td>
<td>0.723</td>
<td>0.771</td>
<td>0.779</td>
</tr>
<tr>
<td></td>
<td>Experiment 3</td>
<td>0.732</td>
<td>0.526</td>
<td>0.735</td>
<td>0.684</td>
</tr>
<tr>
<td></td>
<td>Experiment 4</td>
<td>0.761</td>
<td>0.625</td>
<td>0.742</td>
<td>0.695</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.770&lt;sup&gt;a&lt;/sup&gt; (0.033)</td>
<td>0.685&lt;sup&gt;a&lt;/sup&gt; (0.145)</td>
<td>0.765&lt;sup&gt;a&lt;/sup&gt; (0.034)</td>
<td>0.752&lt;sup&gt;a&lt;/sup&gt; (0.079)</td>
</tr>
<tr>
<td>GSES-CV</td>
<td>Experiment 1</td>
<td>0.818</td>
<td>0.732</td>
<td>0.417</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>Experiment 2</td>
<td>0.782</td>
<td>0.693</td>
<td>0.315</td>
<td>0.654</td>
</tr>
<tr>
<td></td>
<td>Experiment 3</td>
<td>0.731</td>
<td>0.502</td>
<td>0.367</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>Experiment 4</td>
<td>0.764</td>
<td>0.537</td>
<td>0.357</td>
<td>0.526</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.774&lt;sup&gt;a&lt;/sup&gt; (0.036)</td>
<td>0.616&lt;sup&gt;b&lt;/sup&gt; (0.113)</td>
<td>0.364&lt;sup&gt;c&lt;/sup&gt; (0.042)</td>
<td>0.586&lt;sup&gt;b&lt;/sup&gt; (0.074)</td>
</tr>
</tbody>
</table>

† PS, phenotypic selection; MAS, marker assisted selection; GP, genomic prediction; GP-MAS, combination of GP and MAS models.

‡ GS, genotypic sampling; ES, environmental sampling; GSES, genotypic- and environmental sampling.

§ Mean correlation coefficient across all validation datasets. Numbers in parenthesis represent the standard deviation of the mean.

¶ Means within a row followed by different letters are significantly different at p ≤ 0.05.
**Figure 1:**

<table>
<thead>
<tr>
<th>Splitting of datasets</th>
<th>Parameter estimation</th>
<th>Prediction</th>
<th>Predictive ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS-CV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m_{t1})</td>
<td>(3): (m_{v1} = \mu + Zu + e)</td>
<td>(3): (\hat{m}_{v1} = \hat{\mu} + Z\hat{u})</td>
<td>(\text{cor}(m_{v1}, \hat{m}_{v1}))</td>
</tr>
<tr>
<td>(m_{v1})</td>
<td>(4): (m_{t1} = \mu + Mv + e)</td>
<td>(4): (\hat{m}_{t1} = \hat{\mu} + M\hat{v})</td>
<td></td>
</tr>
<tr>
<td>(5): (m_{t1} = \mu + Zu + Mv + e)</td>
<td>(5): (\hat{m}_{v1} = \hat{\mu} + Z\hat{u} + M\hat{v})</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ES-CV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m_{t2})</td>
<td>(3): (m_{t2} = \mu + Zu + e)</td>
<td>(3): (\hat{m}_{t2} = \hat{\mu} + Z\hat{u})</td>
<td>(\text{cor}(m_{t2}, \hat{m}_{t2}))</td>
</tr>
<tr>
<td>(m_{v2})</td>
<td>(4): (m_{t2} = \mu + Mv + e)</td>
<td>(4): (\hat{m}_{t2} = \hat{\mu} + M\hat{v})</td>
<td></td>
</tr>
<tr>
<td>(5): (m_{t2} = \mu + Zu + Mv + e)</td>
<td>(5): (\hat{m}_{t2} = \hat{\mu} + Z\hat{u} + M\hat{v})</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GSES-CV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m_{t3})</td>
<td>(3): (m_{t3} = \mu + Zu + e)</td>
<td>(3): (\hat{m}_{t3} = \hat{\mu} + Z\hat{u})</td>
<td>(\text{cor}(m_{t3}, \hat{m}_{t3}))</td>
</tr>
<tr>
<td>(m_{v3})</td>
<td>(4): (m_{t3} = \mu + Mv + e)</td>
<td>(4): (\hat{m}_{t3} = \hat{\mu} + M\hat{v})</td>
<td></td>
</tr>
<tr>
<td>(5): (m_{t3} = \mu + Zu + Mv + e)</td>
<td>(5): (\hat{m}_{t3} = \hat{\mu} + Z\hat{u} + M\hat{v})</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parameter estimation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PS: m_{t2})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PS: m_{v2})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PS: m_{t3})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PS: m_{v3})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PS: m_{v3} = a + bm_{t3} + e)</td>
<td>(PS: \hat{m}<em>{v3} = \hat{a} + \hat{b}m</em>{t3})</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(a, b, \mu, \nu, \hat{a}, \hat{b}, \hat{\mu}, \hat{\nu}, \hat{Z}, \hat{M}, \hat{e}\) are parameters to be estimated.