Maize (Zea mays L.) grain yield is affected by large numbers of genetic and environmental factors (Lee and Tollenaar, 2007). Among them (and excluding diseases), heat and drought stress are considered the most important ones (Wang et al., 2003). These factors affect many aspects of plant physiology, including chlorophyll content, aerial biomass, plant water content, and ultimately grain yield (Sinsawat et al., 2004; Yordanov et al., 2000).

Many of the physiological and agronomical characteristics of a crop that influence grain yield also lead to differences in the reflectance of electromagnetic radiation at different wavelengths.

**ABSTRACT**

Hyperspectral cameras can provide reflectance data at hundreds of wavelengths. This information can be used to derive vegetation indices (VIs) that are correlated with agronomic and physiological traits. However, the data generated by hyperspectral cameras are richer than what can be summarized in a VI. Therefore, in this study, we examined whether prediction equations using hyperspectral image data can lead to better predictive performance for grain yield than what can be achieved using VIs. For hyperspectral prediction equations, we considered three estimation methods: ordinary least squares, partial least squares (a dimension reduction method), and a Bayesian shrinkage and variable selection procedure. We also examined the benefits of combining reflectance data collected at different time points. Data were generated by CIMMYT in 11 maize (Zea mays L.) yield trials conducted in 2014 under heat and drought stress. Our results indicate that using data from 62 bands leads to higher prediction accuracy than what can be achieved using individual VIs. For hyperspectral prediction equations, the shrinkage and variable selection method was the best-performing one. Among the models using data from a single time point, the one using reflectance collected at 28 d after flowering gave the highest prediction accuracy. Combining image data collected at multiple time points led to an increase in prediction accuracy compared with using single-time-point data.

**Abbreviations:** BayesB, a Bayesian shrinkage and variable selection method; CV, cross-validation; CWMI, canopy water mass index; LV, latent variable; MCARI2, modified chlorophyll absorption ratio index; mND, modified normalized difference at 705 nm; MTCI, terrestrial chlorophyll index; NDVI, normalized difference vegetation index; NIR, near-infrared; OLS, ordinary least squares; PLS, partial least squares; PRI, photochemical reflectance index; RR, ridge regression; VI, vegetation index.
These differences can be measured using spectral data collected at different phases of crop development (Hatfield and Prueger, 2010).

High-throughput phenotyping platforms can be used to screen large numbers of genotypes at a relatively low cost (Andrade-Sanchez et al., 2014; Montes et al., 2007) with a nondestructive method. Reflectance data are commonly used to derive vegetation indices (VIs) that are predictive of traits such as total leaf area, chlorophyll content, plant water content, biomass, or yield (Thorpe et al., 2015). Most VIs are based on a few bands of the recorded spectrum. For instance, the normalized difference VI (NDVI) is based on the ratio of the difference between reflectance in the near-infrared (NIR) spectrum, from 700 to 1100 nm, and part of the red color spectrum at 670 nm relative to the sum of both (Tucker, 1979). Other indices commonly used to study agricultural traits include the canopy water mass index (CWWI; Winterhalter et al., 2011), the modified normalized difference at 705 nm wavelength (mND; Sims and Gamon, 2002), the terrestrial chlorophyll index (MTCI; Dash and Curran, 2004), the new modified chlorophyll absorption ratio index (MCAR12; Haboudane et al., 2004), and the photochemical reflectance index (PRI; Gamon et al., 1992).

Hyperspectral cameras can measure reflectance at many wavelengths (also referred to as “bands”). We hypothesize that prediction equations based on whole-spectrum data can lead to higher prediction accuracy than what can be achieved using VIs. Therefore, the first objective of this study was to quantify the prediction accuracy that can be achieved using regression equations that use data from the whole spectrum and to compare this approach with predictions based on VIs.

Prediction equations based on hyperspectral image data can involve large numbers of regression coefficients. Estimating these parameters often requires using methods that can cope with the statistical challenges emerging in models with large numbers of parameters: the so-called “curse of dimensionality” (Bellman, 1957). In the last decades, there have been important developments in statistical methods for high-dimensional regressions (Friedman et al., 2001). Recently, few authors considered using shrinkage and variable selection methods to fit whole-spectrum regression equations for predicting crop yield (Hernandez et al., 2015) or milk components in dairy cows (Ferragina et al., 2015). Therefore, the second objective of this study was to evaluate alternative methods for estimating effects in whole-spectrum prediction equations; three approaches were considered: ordinary least squares (OLS) with all the available bands, partial least squares (PLS, a dimension reduction technique; Kramer, 1998), and a Bayesian shrinkage and variable selection method (BayesB; Meuwissen et al., 2001).

Reflectance data can be collected at multiple phases of crop development. Measurements taken at different time points provide information on different aspects of crop physiology. Given this, we hypothesize that combining data collected at multiple time points in a single prediction equation can increase prediction accuracy relative to what can be achieved using a single snapshot of reflectance. Therefore, the third objective was to quantify the effects of combining spectral data from multiple time points during the crop cycle on prediction accuracy of grain yield. To this end, for each of the methods considered, we compared the predictive performance achieved when using data from a single time point to that of models using data from multiple time points.

**MATERIALS AND METHODS**

The data consisted of 467 experimental and precommercial maize hybrids tested in 11 yield trials performed under heat and drought stress at CIMMYT’s experimental station in Ciudad Obregón, Sonora, Mexico (27°20’ N, 109°54’ W, 38 m asl), during the 2014 summer season (see Supplemental Table S1 for a summary of environmental conditions, number of genotypes, and replicates per trial). Temperatures during the growing season ranged from 22°C at night to 41°C at midday during flowering. Trials were planted on 20 June in two-row plots 4.5 m long, at a density of 6.9 plants m⁻² with 80-cm row spacing, and harvested on 5 October. Trials were laid out in an α-lattice incomplete block design using two (Trials 1 and 5–11) or three replicates (Trials 2–4). During the first 45 d after planting, evaporative losses were fully compensated for on a weekly basis. Thereafter, plant growth relied on residual moisture and two small rains on Day 21 (15 mm) and on Day 31 (10 mm) after flowering (defined as the date on which 50% of plants within a plot were shedding pollen). All trials received two fertilizations: 100 kg ha⁻¹ of (NH₄)₂PO₄ and 500 kg ha⁻¹ of (NH₄)₂SO₄ at sowing and 250 kg ha⁻¹ of (NH₄)₂SO₄ at V5 (Hanway, 1963). Weeds, insects, and diseases were controlled as needed. Plants were hand harvested when all plots had <15% grain moisture. Ears harvested from each plot were shelled, weighed, and subsampled for measuring grain moisture. The trait analyzed in this study was grain yield adjusted to 12.5% moisture and converted to metric tons per hectare.

**Image Data**

Image data were collected using a hyperspectral camera (VNIR Headwall Photonics Micro–Hyperspec ARS3; Headwall Photonics) mounted on a single-engine aircraft Piper PA-16 Clipper. Flights started 55 d after sowing (when most trials had 50% of plots flowering) and were repeated at 62, 69, 75, and 83 d after sowing (hereinafter labeled as T1, T2, …, T5, respectively). To achieve a resolution of 30 cm pixel⁻¹, flights were performed at an altitude of 300 m and a ground speed of ~34 m s⁻¹. The hyperspectral camera had a radiometric resolution of 10 bits. It acquired images from 392 to 850 nm, subdivided into 62 evenly spaced bands at a spectral resolution of 1.9 nm, covering the visible spectrum and part of the NIR spectrum.

A filter was applied to the images to exclude pixels corresponding to a mixture of crop and soil, and to calibrate reflectance intensity. The Atmospheric correction was performed with the SMARTS simulation model developed by...
the National Renewable Energy Laboratory of the USDOE (Gueymard, 2001). This was done using aerosol optical depth measured at 550 nm with a Micro–Tops II sun photometer (Solar LIGHT Company). Hyperspectral camera was radiometrically calibrated with a uniform light source system (integrating sphere, CSTM-USS-2000C Uniform Source System, LabSphere) at four different levels of illumination and six different integration times. Plot images were trimmed by excluding borders of two to three pixels per plot. The plot coordinates were defined based on a grid of polygons representing the trial plots. This grid was adjusted on the map based on the actual location of certain plots in the field, measured with a Trimble R4 GPS receiver. Each of the 62 reflectance bands was measured using a mean value obtained from the central plot pixels. Several VI s (NDVI, CFW MI, mND, MTCI, MCARI2, and PRI) were derived from reflectance data at selected bands (Table 1).

**Single-Time-Point Models**

Single-time-point models were obtained by regressing phenotypes \((y_{ijt})\), where \(i\) indexes trials and \(j\) indexes entries within a trial) on inputs derived from a single time point (T1, T2, ..., T5) using a linear regression model. Inputs were either reflectance at 62 bands, latent variables (LVs) derived from the 62 bands using PLS, or a VI. All inputs were standardized before the models were fitted. The regression equations used to fit single-time-point models were:

62-band model:  
\[ y_{ijt} = \mu_{(i)} + \sum_{k=1}^{62} x_{ikj} \beta_{ij(k)} + \epsilon_{ijt} \]

Table 1. Regression models used to predict maize grain yield from 62 reflectance bands collected with a hyperspectral camera.

<table>
<thead>
<tr>
<th>Model†</th>
<th>Inputs‡</th>
<th>No. of estimation inputs method</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI-based models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDVI ((R_{505} - R_{670})/(R_{505} + R_{670}))</td>
<td>1 OLS</td>
<td></td>
</tr>
<tr>
<td>CFW MI (R_{505}/R_{725})</td>
<td>1 OLS</td>
<td></td>
</tr>
<tr>
<td>mND ((R_{750} - R_{705})/(R_{750} + R_{705}) + 2 R_{446})</td>
<td>1 OLS</td>
<td></td>
</tr>
<tr>
<td>PRI ((R_{531} - R_{570})/(R_{531} + R_{570}))</td>
<td>1 OLS</td>
<td></td>
</tr>
<tr>
<td>MTCI ((R_{735} - R_{708})/(R_{735} + R_{708}))</td>
<td>1 OLS</td>
<td></td>
</tr>
<tr>
<td>MCARI2 (1.5 \left[ \frac{25(R_{705} - R_{670}) - 1.3(R_{705} - R_{531})}{\sqrt{2(R_{705} + 1) - (6 R_{531} - 5 \sqrt{R_{705}}) - 0.5}} \right])</td>
<td>1 OLS</td>
<td></td>
</tr>
<tr>
<td>PLS model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS</td>
<td>Latent variables derived from 62 bands</td>
<td>2–27§</td>
</tr>
<tr>
<td>62-bands regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>62 bands</td>
<td>62</td>
</tr>
<tr>
<td>BayesB</td>
<td>62 bands</td>
<td>62</td>
</tr>
</tbody>
</table>

† VI. vegetation index; NDVI, normalized difference vegetation index; CFW MI, canopy water mass index; mND, modified normalized difference at 705 nm; MTCI, terrestrial chlorophyll index; MCARI2, new modified chlorophyll absorption ratio index; PRI, photochemical reflectance index; OLS, ordinary least squares estimation; PLS, partial least squares regression; BayesB, A Bayesian shrinkage and variable selection method.

‡ For the vegetation indices, formulae are reported with subindex indicating reflectance wavelengths (in nm). For the rest of the models, the input is either the 62 reflectance wavelengths or latent variables derived for them.

§ The number of latent variables was chosen using an internal (to the training dataset) cross-validation. The numbers in the table represent the range of the number of latent variables obtained in the different models considered. For further details, see Supplemental Table S2.

**Multi-Time-Point Models**

Multi-time-point models were similar to the single-time-point models but used inputs from multiple time points (T1:T5, where \(t = 1, 2, 3, 4\). For instance, model T1:T5 included data from all time points, whereas model T2:T5 included data from time points 2, 3, 4, and 5, and so on. Model NDVI\(_{1:5}\) included five predictors, which were NDVIs derived from the image data collected at T1, T2, ..., and T5. For model BayesB, whenever data from multiple time points were used, different regularization parameters (scales and proportions of non-null effects) were estimated for each of the time points.

**Model Assessment**

We compared models according to their ability to predict yield in trials that were not used to train the models. Therefore, we implemented a leave-one-trial-out cross-validation (CV; i.e., an 11-fold CV with entire trials assigned to folds). For each prediction method this analysis yielded 11 estimates (i.e., one per trial) of the within-trial prediction correlation \(r\), and a
weighted average of these correlations was computed for each method using:

$$\bar{r} = \frac{\sum_{i=1}^{n} \frac{v_i}{n_i}}{\sum_{i=1}^{n} \frac{1}{n_i}}$$

where $v_i = (1 - r_i^2)/(n_i - 2)$ is an approximation to the sampling variance of the correlation estimates, which is a function of the observed squared correlation ($r_i^2$) and the number of records ($n_i$) used to compute that correlation (Rahman et al., 1968). The SEs of correlation estimates ($\bar{r}$) were estimated using a bootstrap procedure with 30,000 bootstrap replicates. The results from the bootstrap samples were also used to estimate the proportion of times (across bootstrap samples) that one method yielded a higher $\bar{r}$ than another method.

**Estimation Methods and Software**

All the analyses were performed using R software (R Core Team, 2016). The OLS regressions were fitted using the `lm` function of R. The BayesB model was implemented using the BGLR package (Pérez and de los Campos, 2014). Simplified versions of the scripts used for analysis are provided in the Supplemental Methods. The PLS regression was implemented using the PLS R-package (Mevik and Wehrens, 2007). The number of latent variables used in PLS was chosen using internal validation. For instance, in the 11-fold CV, for each fold, data from the 10 trials that constituted the training set were further partitioned into 70% (training) and 30% (testing) sets, and PLS was fitted with 1, 2, …50 LVs to 70% of the training data. We used each of these PLS models (with 1, 2, …50) to predict yield in the 30% of the data saved for testing and chose the number of LVs that gave the highest prediction correlation in the testing set. Finally, PLS was refitted to data from the 10 trials with the chosen number of LVs and prediction accuracy in the left-out trial.

**RESULTS**

The median yield by trial ranged from 2.28 to 3.99 t ha$^{-1}$ (Fig. 1), and the within-trial standard deviation of yield ranged from 0.59 to 1.12 t ha$^{-1}$. Grain yield had a close-to-normal distribution across trials, except in Trial 2. An ANOVA of yield revealed that 21% of the total variation can be attributed to between-trial differences.

Inspection of the marginal correlations between the 62 bands (Fig. 2) measured at T5 showed that the wavelengths clustered in two groups, one including bands from the visible spectrum (390–720 nm) and the other one including bands of the NIR spectrum (720–850 nm). A similar clustering occurred at other time points (Supplemental Fig. S1–S4). A principal component analysis of the reflectance at T5 indicated that the leading eigenvalue explained 64% of the total variance and that 98% of the variance of reflectance could be explained by the five eigenvectors with largest eigenvalue (Fig. 2, Supplemental Fig. S1–S4).

Table 2 shows estimated CV correlations obtained using image data collected at T5, which was the time point that achieved the highest prediction accuracy. A summary of the results obtained by time point is given in left panel of Fig. 3. The CV correlations varied from 0.29 to 0.41. Among all the methods evaluated, BayesB gave the highest prediction accuracy (0.41), followed by PLS (0.39), PRI, and mND (both 0.38), which were the most predictive indices. However, BayesB gave higher prediction correlation than PRI and mND at 76 and 79% of the 30,000 bootstrap samples, respectively. The results obtained with image data from other time points (T1, …T4) are given in Supplemental Table S3. Overall, prediction accuracy declined for all methods from T5 to T1. However, the reduction in prediction accuracy was more marked for VIs. Overall, BayesB was the best-performing method.
Table 3 shows the results obtained when combining data from T1 to T5 in a single model. Overall, the correlations between predicted and observed yield increased compared with the results obtained with data collected at T5, by 10 to 20% (Fig. 3). Among all the methods evaluated, BayesB was the best-performing one (estimated correlation 0.47), followed by the mND index (0.44). In the bootstrap analyses, BayesB outperformed mND 72% of the time.

The results obtained by combining data from other time points (T4:T5, T3:T5, and T2:T5) are given in Supplemental Table S4. Overall, the best results were obtained when combining data from all the time points available. The method that gave the highest prediction accuracy was BayesB.

DISCUSSION

The continued development of sensing devices (e.g., cameras), aerial equipment (e.g., unmanned aerial vehicles), and image-processing technologies has opened great opportunities for implementing high-throughput phenotyping platforms (Araus and Cairns, 2014). These platforms can be used to scan large numbers of entries (plots, genotypes) using a nondestructive procedure (Cabrera-Bosquet et al., 2012). Hyperspectral cameras can provide reflectance data at potentially hundreds of wavelengths; these data
are of particular interest because plant pigments absorb a large fraction of the radiation within the visible spectrum (400–700 nm) and plant cell structures reflect most of the radiation in the NIR spectrum (700–1100 nm). Traditionally, reflectance data have been used to identify bands that provide information about physiological processes such as photosynthetic activity (Lu et al., 2015); these bands are then combined to form a VI. Based on this approach, several VIs have been proposed and used in largescale phenotyping of biomass, greenness, nitrogen content, pigment composition, and photosynthetic status (Fiorani and Schurr, 2013).

However, reducing hyperspectral data to a VI derived from a few bands may lead to a loss of information. Our results indicate that reducing hyperspectral image data to a VI leads to a loss of prediction accuracy relative to what can be achieved using whole-spectrum data. In some cases (e.g., with data collected at T5 and other individual time points), the loss in prediction accuracy incurred by reducing hyperspectral data to indices was small; however, in other cases (e.g., when using data collected at early time points), the predictive performance of VIs was substantially lower than that of methods using data from all the available bands. Therefore, we conclude that there are clear benefits of using hyperspectral information for prediction of grain yield. However, it is important to highlight that, unlike VIs (which are deterministic functions of reflectance at given bands), the use of hyperspectral prediction equations requires estimating the regression coefficients entering in the equations. These coefficients are likely to vary between traits and (probably) environmental conditions. Therefore, unlike VIs, whole-spectra prediction equations need to be calibrated for particular traits and environmental conditions.

Among the models based on hyperspectral data, the best-performing one was BayesB. Other methods (OLS, PLS, or some indices such as the mND or the PRI) were less stable. In some cases, some of these methods performed close to BayesB; however, in other cases (e.g., VIs in early time points), they performed very poorly.

Previous studies also considered using multi- and hyperspectral data for prediction of agronomic traits. For instance, Hernandez et al. (2015) used ridge regression (RR; Hoerl and Kennard, 1970) to fit a linear model for grain yield using reflectance data from 350 to 2500 nm and reported that the RR method increased goodness of fit to the training data relative to the use of VI. However, Hernandez et al. (2015) compared models based on goodness of fit to the training data and not using CV prediction accuracy, as done here. Montesinos-López et al. (2017) also considered using whole-spectra prediction equations to predict grain yield in wheat (Triticum aestivum L.) and, in agreement with what we found here, they reported higher prediction accuracy using hyperspectral data relative to VIs.

Inspection of the correlation patterns between reflectance at the 62 bands collected indicated that reflectance clustered into two groups (with a threshold of ~730 nm); similar clustering has been reported before in maize (Weber et al., 2012). Moreover, the five leading eigenvalues explained 98% of the total variance among spectra. These patterns suggest that dimension-reduction methods such as PLS with a few latent variables should perform well. However, although PLS outperformed OLS in multi-time-point models, the performance of the PLS regression was not as good as that of BayesB, suggesting that dimension reduction leads to partial loss of information. Ferragina et al. (2015) reported similar findings when comparing PLS with Bayesian models for prediction of fatty acids in milk using mid-infrared transmittance data. Previous studies have also noted the limitations of PLS and suggested that dimension reduction using genetic algorithms may outperform PLS (Kaleita et al., 2006).

We considered six different VIs, including the traditional NDVI and others that were designed to assess features of the crop such as canopy water mass index (e.g., CWMI), chlorophyll content (e.g., mND), or

### Table 3. Estimated within-trial cross-validation correlation by prediction method obtained with image data collected at time points T1:T5.

<table>
<thead>
<tr>
<th>Average within-trial correlation (SE)</th>
<th>Proportion of times the model in row gave higher correlation than the model in column†</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI</td>
<td>0.43 (0.029)</td>
</tr>
<tr>
<td>CWMI</td>
<td>0.32 (0.033)</td>
</tr>
<tr>
<td>mND</td>
<td>0.44 (0.029)</td>
</tr>
<tr>
<td>PRI</td>
<td>0.39 (0.029)</td>
</tr>
<tr>
<td>MTCI</td>
<td>0.36 (0.030)</td>
</tr>
<tr>
<td>MCARI2</td>
<td>0.43 (0.029)</td>
</tr>
<tr>
<td>OLS</td>
<td>0.35 (0.029)</td>
</tr>
<tr>
<td>PLS</td>
<td>0.41 (0.032)</td>
</tr>
<tr>
<td>BayesB</td>
<td>0.47 (0.031)</td>
</tr>
</tbody>
</table>

† Comparisons based on 30,000 bootstrap samples of the vectors containing observed yield and (cross-validation) predictions. NDVI, normalized difference vegetation index; CWMI, canopy water mass index; mND, modified normalized difference at 705 nm; MTCI, terrestrial chlorophyll index; MCARI2, new modified chlorophyll absorption ratio index; PRI, photochemical reflectance index; OLS, ordinary least squares estimation; PLS, partial least squares regression; BayesB, A Bayesian shrinkage and variable selection method.
photosynthesis efficiency (e.g., PRI), as well as two recently proposed indices, MCARI2 and MTCI. Overall, most of the indices have similar predictive performance; however, the mND and PRI indices were the best-performing ones.

Babar et al. (2006) studied the correlation between biomass and NDVI at three growth stages in wheat—booting, heading, and grain filling—and Monteiro et al. (2012) studied the correlation between grain yield and NDVI in beans (Phaseolus vulgaris L.). Both studies reported higher phenotypic correlations in grain-filling stages, which is also in agreement with the results presented for maize in this study.

In environments with severe water limitations, the ability of plants to retain water is key to their survival. Reflectance at bands related to water use efficiency is informative of the crop performance under drought stress (reported for C₃ cereals in Araus et al., 2002). However, the CWMI in this study showed poor performance relative to other VIs tested. Winterhalter et al. (2011) also found low correlations between CWMI and maize grain yield and attributed the relatively poor correlation to the ability of maize to recover from drought stress.

The mND index is a modification of the NDVI that has been reported to be more correlated with chlorophyll content (even when such content is low, the predominant condition at T5) than other indices (Sims and Gamon, 2002). In our study, mND and PRI were the best-performing indices. The latest index has been reported to be correlated with xanthophyll pigments and photosynthetic efficiency in sunflower (Helianthus annuus L.; Gamon et al., 1992) and also showed good predictive performance in maize (Strachan et al., 2002).

Combining data from multiple time points led to moderate gains in prediction accuracy. Reflectance data collected at different time points may provide information about different aspects of crop physiology; this may explain why combining data from multiple time points leads to more stable and more accurate predictions. The benefits of combining data from multiple time points in a single prediction equation were more marked for some VIs (e.g., NDVI) than for OLS, PLS, or BayesB. When data from five time points were combined, the prediction equations of OLS and BayesB involved 310 regression coefficients. With these many coefficients, OLS estimates have large sampling variance, and this may explain why OLS did not benefit from combining multi-time-point data. For this type of model, the benefit of using shrinkage and variable selection methods are clear.

In our study, we assumed that, conditional on the image data, errors were uncorrelated within plots. Further extensions of the methodology may consider including spatial correlations and heterogeneous error variances. These extensions may not have a large impact when it comes to prediction of phenotypes, the focus of this article, but may be more important for prediction of genetic values.

Our results demonstrate that prediction of agronomic traits in future trials (with genotypes and environmental conditions similar to those represented in the training data) is feasible and can be improved, relative to predictions based on VIs, using whole-spectrum prediction equations. Moreover, we show that combining data collected at multiple time points leads to higher prediction accuracy, provided that prediction equations are trained using methods that can cope with the challenges emerging when fitting equations that involve large numbers of predictors.

**Conflict of Interest**
The authors declare that there is no conflict of interest.

**Supplemental Material Available**
Supplemental material for this article is available online. Data is also available at http://data.cimmyt.org/dvn/.

**Acknowledgments**
The authors acknowledge financial support from CIMMYT and from ARVALIS–Institut du vegetal. The development and maintenance of BGLR was supported by NIH grant R01GM101219. The authors thank Drs. Timothy Beissinger and Jean-Luc Jannink for comments provided to earlier versions of the manuscript and Alma McNeb for her editorial suggestions. The authors also acknowledge valuable comments provided by anonymous reviewers and Associate Editor Dr. Jesse Poland.

**References**


