**USE OF HIGH-RESOLUTION IMAGE DATA OUTPERFORMS VEGETATION INDICES IN PREDICTION OF MAIZE YIELD**

Fernando Aguate1, Sam Trachsel2, Juan Burgueño3, Jose Crossa3, Monica Balzarini1 & Gustavo de los Campos4

1: Facultad de Ciencias Agropecuarias (FCA), Universidad Nacional de Córdoba (UNC), Cátedra de Estadística, 5000, Córdoba, Argentina; 2: Global Maize Program - Physiology, International Maize and Wheat Improvement Center (CIMMYT), Carretera México Veracruz, Km 45, Texcoco, Estado de Mexico, 56237, México; 3: Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Carretera México Veracruz, Km 45, Texcoco, Estado de México, 56237, México;

4: Epidemiology & Biostatistics and Probability & Statistics departments, Michigan State University 909 Fee Road, East Lansing, Michigan 48824, USA.

**ABSTRACT**.Modern high-resolution cameras can provide reflectance data at potentially hundreds of wavelengths. This information can be used to predict physiological agronomic and disease traits. Traditionally, image data were used to derive vegetation indices (VI), which are predictive of traits. However, the data generated by high-resolution cameras contains more information than what can be summarized using VI. The objective of this study was to compare the predictive performance of regression methods using information from 62 bands to that of VI derived from the same reflectance data. We considered both ordinary-least square regressions and a Bayesian shrinkage/variable selection procedure. The data were generated by CIMMYT in 12 maize yield trials conducted in 2014 under irrigation and combined heat and drought stress. The trait analyzed was grain yield (ton/ha) and inputs were either VI or normalized reflectance at 62 bands, all collected at five different time points from flowering to pre-harvest. We show that using data from all bands leads to higher cross-validation prediction accuracy than using VI. Among the models that used data from a single time point, the ones using data collected at pre-harvest gave the highest prediction accuracy. Combining image data collected at different time points led to a small increase in prediction accuracy relative to models using data from a single time point.

**H**igh-throughput phenotyping platforms (HPP) can be used to screen large numbers of genotypes at a relatively low cost (Montes et al., Trends Plant Sci, 2007). Canopy absorbance of electromagnetic power at different wavelengths generates patterns that can be used as indirect measurements of agronomic and physiological traits (Babar et al., Crop Sci, 2006). Usually, reflectance data have been used to derive vegetation indices (VI) that are predictive of traits such as total leaf area, chlorophyll content, biomass or yield. Most VI are based on relatively low-resolution data. For instance, the Normalized Difference Vegetation Index (NDVI) is based on the ratio of the difference between reflectance in the near-infra red (NIR) spectrum and part of the red color spectrum at 670 nm relative to the sum of both (Tucker, Remote Sens. Environ., 1979). Other commonly used indices include the Canopy Water Mass Index (CWMI, Winterhalter et al., Eur. J. Agron., 2011), the Modified Normalized Difference at 705 nm wavelength (mND, e.g., Sims and Gamon, Remote Sens. Environ, 2002) and the Photochemical Reflectance Index (PRI e.g., Gamon et al., Remote Sens. Environ., 1992).

Modern high-resolution cameras can collect reflectance at a large number of bands simultaneously. These data are potentially more informative than what can be summarized with a VI. We hypothesize that the use of high-spectrum data can lead to higher prediction accuracy for grain yield than the accuracy achieved using VI. Prediction equations based on high-resolution image data can involve large numbers of regression coefficients. Estimating these parameters may require using methods that can cope with the statistical challenges emerging in models with large numbers of parameters: the so-called “curse of dimensionality”. In the last few years there has been an important development of statistical methods for high dimensional regressions (Friedman, Hastie and Tibshirani, Springer series in statistics, 2001). Some of these methods have been adopted for genomic prediction (Meuwissen et al., Genetics, 2001). More recently, a few authors have considered using high dimensional regression methods for prediction of crop yield (Hernandez et al., Remote Sens., 2015) or milk components (Ferragina et al., Dairy Sci, 2015) using spectra data. In this study, we aimed to evaluate the prediction accuracy that can be achieved using high-spectrum data for predicting maize yield and to compare the predictive power of these models with the power achievable using VI.

**Materials and Methods**

**Data** was generated in 12 maize experiments, carried out at CIMMYT’s experiment station in Ciudad Obregon, Sonora, Mexico (27°20′ N, 109°54′ W, 38 masl) during the 2014 summer season. The genotypes used in the trials included experimental and pre-commercial hybrids. Eleven of the trials were carried out under combined heat and drought stress, and one under well-watered conditions. Experiments were laid out in an alpha lattice incomplete block design using two (trials 5, 7, 8, …, 12) or three (trials 1, 2, 3, 4 and 6) replicates. Trials were planted at a density of 6.9 plants m-2. All trials received two fertilizations: 100 kg/ha of (NH4) H2PO4 and 500 kg/ha (NH4)2 SO4 at sowing and 250 kg/ha of (NH4)2 SO4 at V5. Weeds, insects, and diseases were controlled as needed. Image data were collected using a hyperspectral camera (ARS3, Headwall Photonics, Inc., Fitchburg, MA, USA) mounted onto a Piper PA-16 Clipper (Piper, Vero Beach, FL, USA). Flights were carried out around noon at 55, 62, 69, 75 and 83 days after sowing (hereinafter labeled as T1, T2,…,T5, respectively). The hyperspectral camera had a radiometric resolution of 10 bits, configured at 62 bands wavelength from 400 to 850 nm. A filter was applied to the images to exclude pixels corresponding to a mixture of crop and soil. The plots were trimmed to avoid borders. The mean reflectance values for each band and plot were obtained from the pixels with centers inside the plots. From these 62, bands NDVI, CWMI, mND index and PRI were derived. All inputs, including reflectance and indices, were centered and scaled to a null mean and unit variance. The trait analyzed in this study was grain yield in ton/ha at 10% moisture level and shelling percentage of 80% was assumed.

**Single time point models** were obtained by linearly regressing phenotypes (, yield for the *i* trial and the *j* data-point within a trial) on inputs derived from a single time point (T1, T2,..,T5). Inputs were either standardized reflectance at 62 bands, 5-PC derived from reflectance at all 62 bands, or VI. The regression equations for these three methods were:

62-bands regression: ,

PC- regression: , and

Index-regression:

where represents reflectance at the *k*th band collected at time point *t* (t=1,…,5), are the loadings of the *lth* PC on the *ijth* entry derived from spectra collected at time *t* and is the *mth* index (NDVI, CWMI, mND and PRI were considered) derived from data collected at time point *t* at the *ijth* entry. The error terms () were assumed to be independent with null mean and trial-specific variance.

**Multi time point models** were similar to the single time point models but incorporated in the regression inputs collected at multiple time points (T*r*:5 where *r* was 1, 2, 3 or 4; for instance, model T2:5 included data from time points 2, 3, 4 and 5).

**Parameter estimation**. Index and PC regressions were fitted using OLS. For regressions involving data from 62 bands we considered both OLS and a Bayesian shrinkage-variable selection procedure using a prior with a point of mass at zero and a t-slab (the so-called model BayesB) as implemented in the BGLR R-package (Pérez and de los Campos, Genetics, 2014).

**Model assessment**. Models were first fitted to the entire data set (full-data analysis) to evaluate goodness of fit to the training data. Subsequently we carried out a cross validation (CV) with trials assigned to the testing fold. This “leave-on-trial-out” CV gives the appropriate design for evaluating the ability of a fitted equation to predict yield in a future trial. Prediction accuracy in the training data and in CV was assessed using the correlation between observations and predictions across and within trials. The across-trial correlation was obtained by correlating the entire yield vector with the entire prediction vector. This correlation reflects the ability of a model to predict both differences between and within trial. The within-trial correlation reflects the ability of a prediction equation to rank plots within a trial. The average within-trial correlation was computed by weighting the observed correlation within trials by the square root of the precision (inverse of the SE) of the estimated correlation.

**Results**

**Figure 1** shows a boxplot of grain yield by trial. Grain yield had a relatively symmetric distribution across trials, except in trial # 2. An analysis of variance revealed that 51% of yield variance can be attributed to between-trial differences. Trial # 12 was WW and had the highest mean and slightly larger variance. The 62 reflectance bands clustered in two clear groups (data not shown) one including wavelengths below the NIR spectrum and one above it. This clustering occurred at all time points but was clearer at T5 as a result of lower water availability in the soil. 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 first five principal components.

**Table 1** shows the estimated correlations and the corresponding 95% confidence intervals (CI) obtained using image data from pre-harvest (T5). Goodness-of-fit. The across-trial correlations obtained within training data ranged from moderate (0.35, PRI) to high (0.76 for OLS and BB). As expected, the within-trial correlations were lower, ranging from 0.24 (CWMI) to 0.47 (OLS). Overall, and as one would expect, models using multiple predictors (PC, OLS, BB) fitted the data better than index-based models. Prediction accuracy. The cross-validation correlations were slightly lower than the correlations obtained in the training data and ranged from very low (0.04, PRI) to moderately high (0.64, OLS). Models using data from the 62 bands (both OLS and BB) had higher prediction accuracy than index-based or PC models. Among the VI, the mND was the one that had best predictive performance; the NDVI performed only slightly worse than mND. Since the correlation is invariant under linear transformations, the within-trial correlations for simple linear regressions (i.e., the models based on VI in our case) are equal in training and CV. Relative to the best performing index, model BB gave gains in correlation of 58% (across-trials correlation) and 18% (within-trial correlation). The OLS method performed similarly as BB, with only a slightly smaller correlation in CV.

**Figure 2** shows the average within trial CV correlation obtained using data from different time points. The right panel (A) gives results for models using data from a single time point; results obtained with multi-time points are given in the right panel (B). For simplicity, we only present results for models NDVI, OLS, BB and PC. Models based on other VI performed either similarly or worse than NDVI. Prediction accuracy increased as time increased from T1 to T5. The only exception was T3, which for some methods gave lower prediction accuracy than the one achieved using data collected at T2– rain that fell the night before image data were collected at T3 may have affected the quality of the measurements at that time point. Overall, model BB had the best predictive performance across time points. OLS had a similar, but slightly worse, predictive performance than BB in most time points, except T3 where BB gave a clearly higher prediction accuracy. NDVI and PC regression had clearly worse predictive performance than BB. Combining data from multi-time points gave marginal gains in prediction accuracy for BB, no clear gains for OLS, and was not beneficial for PC-regressions.

**Discussion and Conclusions**

We compared the performance of models using high-resolution image data with that of index-based models and PC-regression. Our results indicate that there are clear advantages in using reflectance intensities from all available bands and that reducing high-resolution image data to a single index or a few PC leads to a loss of prediction accuracy; this, despite the fact that the leading 5 PC explained 98% of the variance. Interestingly, OLS and BB performed similarly; however, BB was more stable across time points and less prone to over-fitting.

Among models using reflectance data from a single-time point, prediction accuracy increased from T1 to T5, suggesting that data collected closer to pre-harvest are more predictive than early measurements. Combining data from multi-time points may lead to small gains in prediction accuracy, but this can only be achieved when using regularized regressions such as BB. With other methods, using data from multi-time points may lead to over-fitting and loss of prediction accuracy.

The overall levels of CV correlations achieved with BB (~ 0.6 across trials and ~0.4 within trial) are still moderate; more research is needed to establish which factors (sample size, homogeneity of trials, number of bands) affect prediction accuracy and how it can be further improved.

Our study combined data from eleven drought and one irrigated trial. An additional analysis based on drought trials only (data not shown) did not yield higher prediction accuracy than the combined analysis presented here. This suggests that regression on reflectance captures differences induced by irrigation. However, further research is needed to establish whether separate equations need to be fitted for contrasting environmental conditions.