Maize Yield Prediction Accuracy Increased By Inclusion of Genetics, Environment, and Management Interactions With Deep Learning

Daniel R Kick[1,2]*, Jacob D. Washburn[1,2]
USDA-ARS, Plant Genetics Research Unit, Columbia, MO[1],
Division of Plant Sciences, University of Missouri, Columbia, MO, USA[2].

Main Questions:

1. Is phenotypic prediction via deep learning more accurate than other models?
2. Does optimization strategy influence performance?
3. Does inclusion of GxExM effects change the importance of input variables?

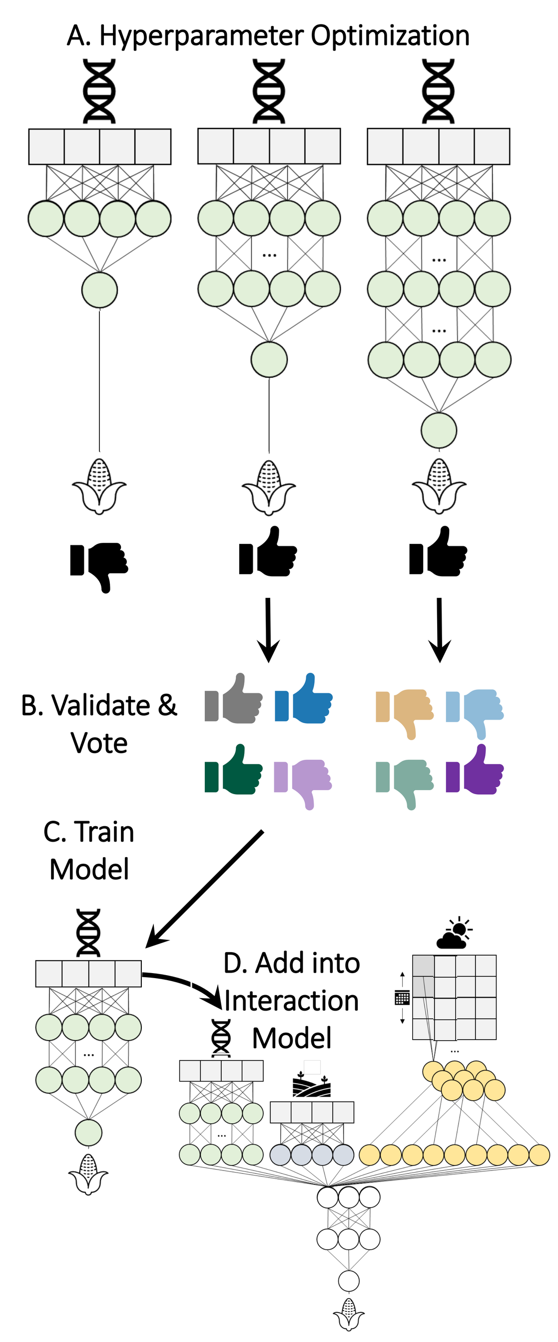
Main Findings:

1. BLUP outperforms Deep Learning, but with greater variability in performance.
2. Optimizing sequentially improves performance.
3. GxExM interactions increase accuracy & alters variables' importance.

Abstract

Phenotypic prediction for polygenic traits and those with interaction effects between genetic, environmental, and management factors remains challenging. We developed a deep learning model with interactions between these factors to better predict maize (*Zea mays*) yield across diverse environments in the continental United States. This model and the modeling approach used is of potential use for genomic selection, management optimization, and forecasting. We find the optimization strategy used influences model performance with the best performance coming from consecutively optimizing submodels, each processing a single data group or interactions between data groups, rather than optimizing all aspects of the model's structure simultaneously. Furthermore, we find deep learning can but is not guaranteed to outperform other model types when all data groups (genomic, soil, and daily weather and management data) are provided. When restricted to a single data group, the model is less accurate. Lastly, we observe that interactions between data groups substantially influence the importance of different variables, reducing the influence of weather events at the end of the season, approximately 200 days following planting.

Methods



Data Preparation

Phenotypic, genotypic, environmental (soil and in field weather data) and management data were provided by the Genomes to Fields Initiative. These data were supplemented with additional weather data from the Daymet database. Using data from 2014-2019, resulted in 96,137 yield measurements taken at 41 sites in the continental United States (158 site-year combinations) and 3,671 unique genotypes. These data were quality controlled with missing and aberrant values being imputed. Inbred genotypic data was PCA transformed with hybrids projected using these PCs. Weather measurements and management applications 75 days before to 204 days after planting (210 days total) were used.

To assess model performance in future years we split these data into a test and training set, stratifying by site-year (observations from one site can be in the training and the test set, but only if they are from different years). Site-year groups were randomly assigned to test set. Groups were approximately balanced by down sampling so that none contained more observations than the smallest group in the test set. After down sampling, if >= 40,000 observations remained and the test set comprised between 10-15% of the total observations the set was accepted. If not, the process was restarted. This resulted in 41,513 observations being retained, 4,240 being in the test set. These included all site-year combinations but not all genotypes. 3,006 unique genotypes were represented. Of these 1,559 occur in both sets, 1,435 in only the training set, and 12 in only the testing set.

Consecutive Optimization (CO)

Three models predicting yield from genomic, soil, or weather and management data are fit. ***(A.)*** Hyperparameter optimization is used to select the network architecture (neurons per layer, layers, etc.). 40 networks were trained on a subset of the training data, the rest used to evaluate the model. Hyperparameters were selected using Bayesian optimization. ***(B.)*** The performance of the top 4 networks was assessed on 10 subsets of the training data for 1000 epochs. Training was divided into 10 bins and model with the lowest mean + standard deviation of loss in the most bins was selected. Epochs to be used was set as the number which minimized the total rolling mean validation loss with a 20-epoch window. ***(C.)*** The selected model was trained on the full training dataset. ***(D.)*** Once steps 1-3 were completed for each data type they were aggregated into a single network. A set of layers was added to allow for interactions between these data types. To select the hyperparameters for this interaction "subnetwork" steps 1-3 were repeated with the genomic, soil, and weather and management subnetworks fixed.

Simultaneous Optimization (SO)

Steps 1-3 above are repeated with interactions between data types allowed at the onset. Using same hyperparameter search space.

Measuring Model Performance

Root mean squared error for the test set is used to assess model performance ($RMSE = \sqrt{\frac{\sum_i^n \hat{y}_i - Y_i}{n}}$). A deep neural network's performance can be sensitive to it's randomly initialized parameter values. To account for this each final architecture was initialized with 10 replicates. To reduce the time required to fit deep learning models, variables were centered and scaled by the training set mean and standard deviation. In the case of yield (bushels/acre) the transformation is: $y = \frac{y_{Original} - 147.397}{48.169}$.

Benchmarking Models: Linear and Machine Learning Models

To assess model performance, linear models with fixed were trained on the same groups of data. These are not amenable to combining variables with one value per season (genetics, soil) with those variable within a season (weather and management). The latter variables were clustered into categories before use. For each variable, a time series k-means with dynamic time warping was fit for values of k from 2-40. k was selected as the value one less than the first k where the silhouette score decreased.

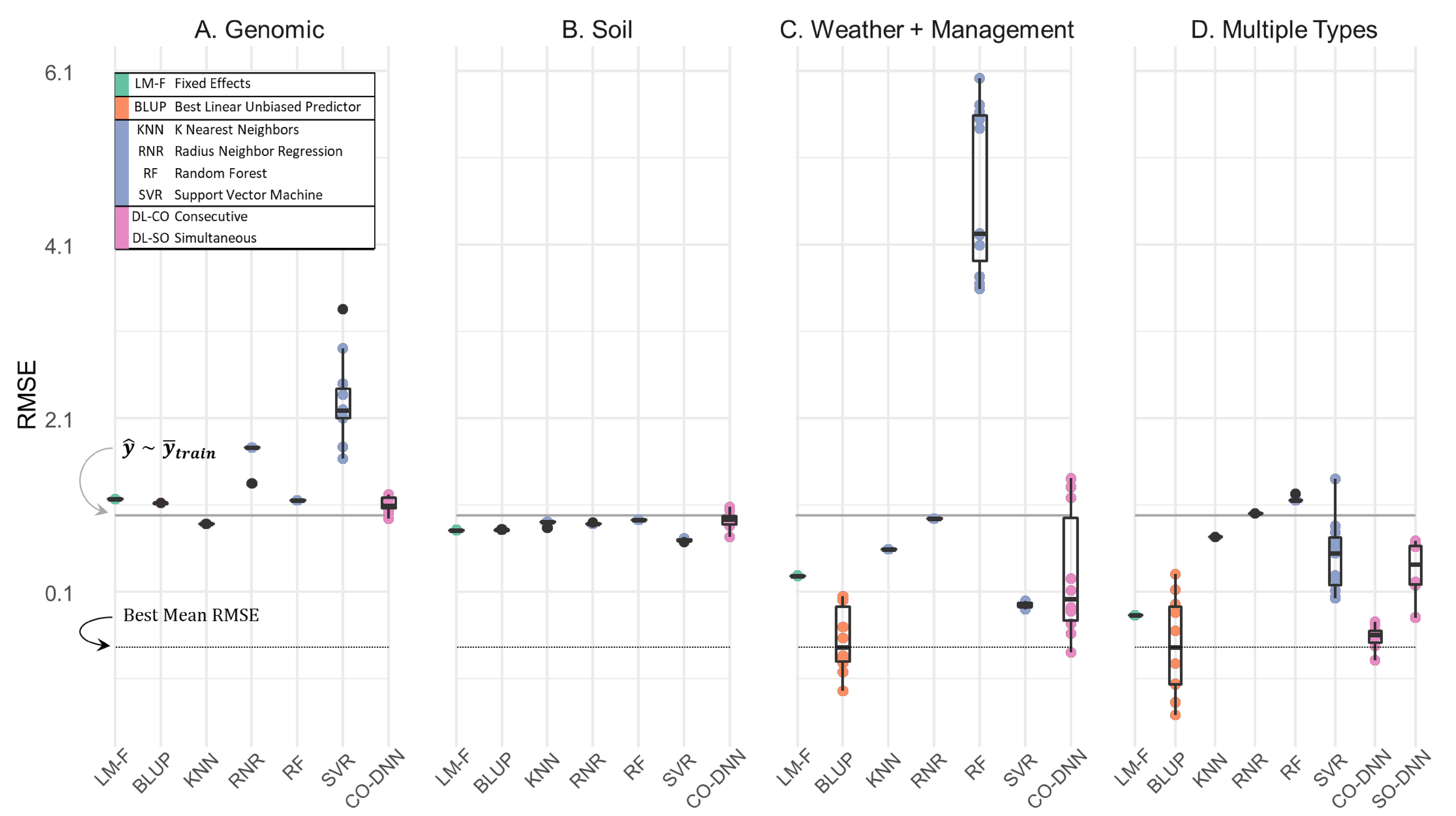
For each neural network's input data (exclusively genomic, soil, weather and management, or all three) Best linear unbiased predictor (BLUP) and machine learning models were created to contextualize model performance. BLUP were modeled on those in Washburn et al. 2021. In the model using multiple data types, genomic by soil and genomic by weather and management interaction effects were included. 4 machine learning models were considered: K Nearest Neighbors (KNN), Radius Neighbors Regression (RNR), Random Forest (RF), Support Vector Machine with a linear kernel (SVR). These models' hyperparameters which were optimized before use.

Feature Importance: Salience

The input variables that are most influential in the predicted yield for a given observation can be determined by calculating each variable's salience. Salience is calculated based off the derivative of each input variable for a given observation. Individual measures of salience are averaged to produce the saliency maps shown here.

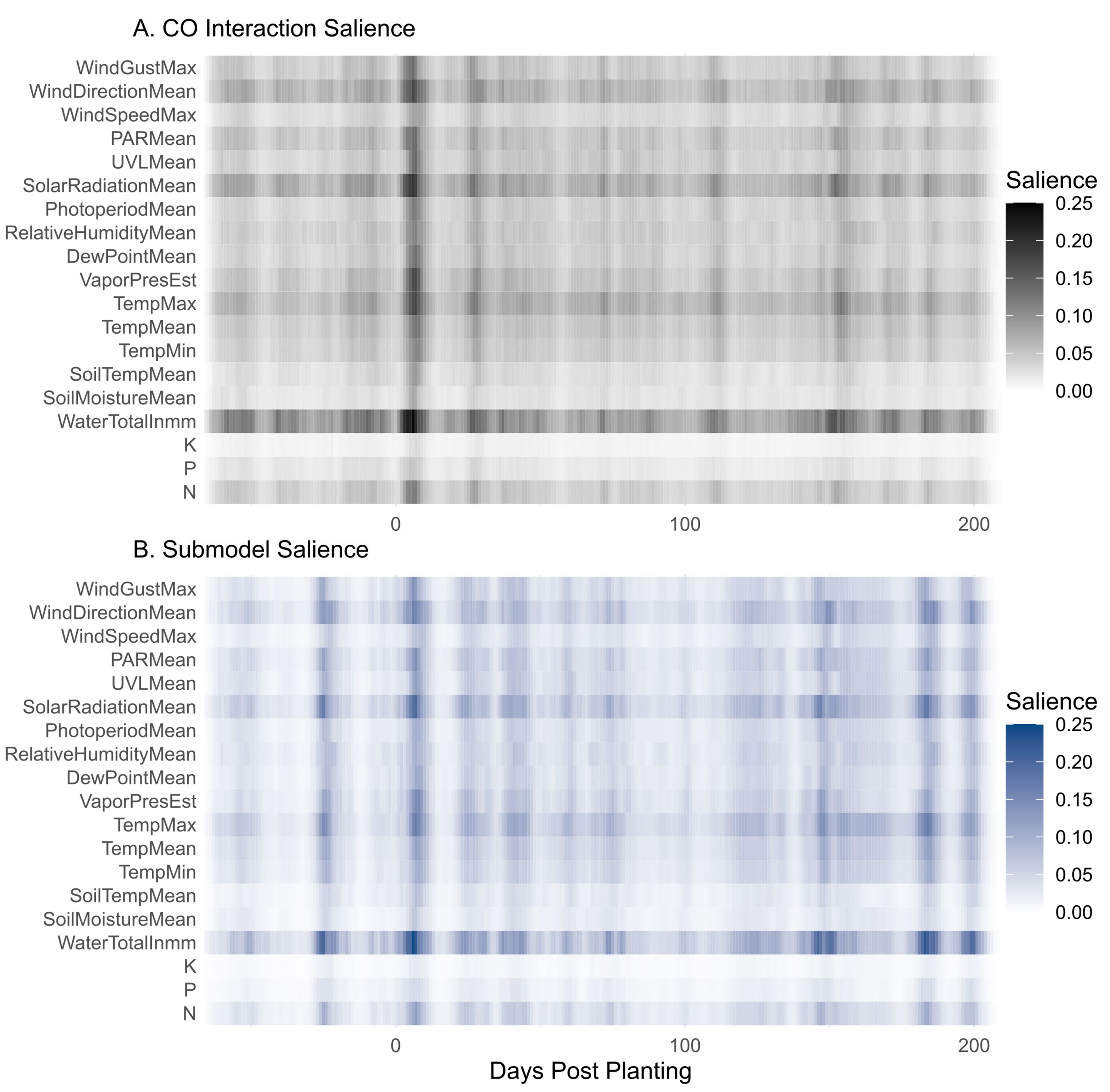
Main Results

Model Performance Across Methodologies and Data Types



The root mean squared error (RMSE) of the testing set is shown for each data group (panels **A. – D.**) and class of model (linear models: green, BLUPs: orange, machine learning: blue, deep learning: pink). Lower values indicate better performance. Deep learning models are divided by whether they were part of the consecutive optimization strategy (DNN-CO) or the simultaneous optimization strategy (DNN-SO). LM-F use all only main effects except in D where interactions between PC1-8 and all weather and management and soil variables are included. BLUPs in D contain genome by soil and genome by weather and management interactions.

Influence of Interaction Effects on Daily Feature Salience



A. Average salience for each day. Interaction model values shown. **B.** The same values for the weather submodel. Salience peaks shortly after planting in both. The submodel contains salient dates prior to planting and near the end of the date range. The interaction-containing model appears to place greater importance for certain features, e.g., irrigation & rainfall, represented as "WaterTotalInmm". The difference between the two saliency maps indicates additional times of sensitivity in the submodel (approximately -25, +180, +195) that the interaction model is relatively insensitive to.

Acknowledgements

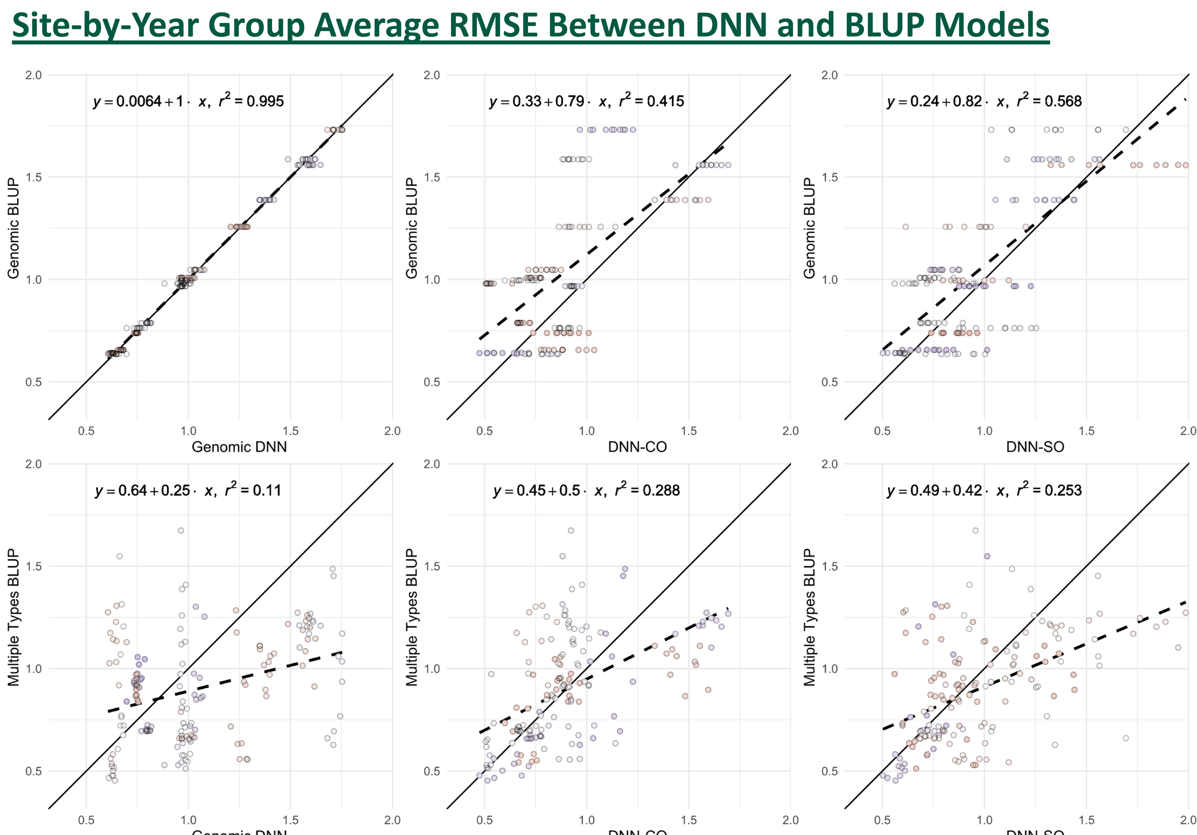
This project was enabled with the mentorship of Dr. Jacob D. Washburn[1,2].
USDA Project: 5070-21000-041-000-D
Provided funding for this study while the Genomes to Fields Intuitive (genomes2fields.org) and the DAYMET database (daymet.ornl.gov) provided the data for this study.

For their contributions in support of this study we would like to thank :
Jason G. Wallace[3], James C. Schnable[4], Judith M. Kolkman[5], Barış Alaca[6,7], Timothy M. Beissinger[6,7], Jode W. Edwards[8], David Ertl[9], Sherry Flint-Garcia[1], Joseph L. Gage[10], Candice N. Hirsch[11], Joseph E. Knoll[12], Natalia de Leon[13], Dayane C. Lima[14], Danilo Moreta[5], Maninder P. Singh[15], Teclemariam Weldekidan[16]

1 United States Department of Agriculture, Agricultural Research Service, Columbia, MO 65211, USA
2 Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA
3 Department of Crop & Soil Science, University of Georgia, Athens GA 30602
4 Center for Plant Science Innovation & Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68588 USA
5 School of Integrative Plant Science, Cornell University, Ithaca NY 14853
6 University of Goettingen, Division of Plant Breeding Methodology, Department of Crop Science
7 University of Goettingen, Center for Integrated Breeding Research
8 United States Department of Agriculture, Agricultural Research Service, Ames, IA 50011, USA
9 Iowa Corn Promotion Board, Johnston, IA 50131 USA
10 Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695, USA
11 Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108
12 USDA-ARS Crop Genetics and Breeding Research Unit, Tifton, GA 31793 USA
13 Department of Agronomy, University of Wisconsin, Madison
14 Department of Agronomy - Plant Breeding & Plant Genetics University of Wisconsin – Madison Plant Breeding and Plant Genetics Program Madison, WI 53706
15 Plant, Soil and Microbial Sciences Dept., Michigan State University, East Lansing, MI 48824
16 University of Delaware Newark, DE 19716 USA

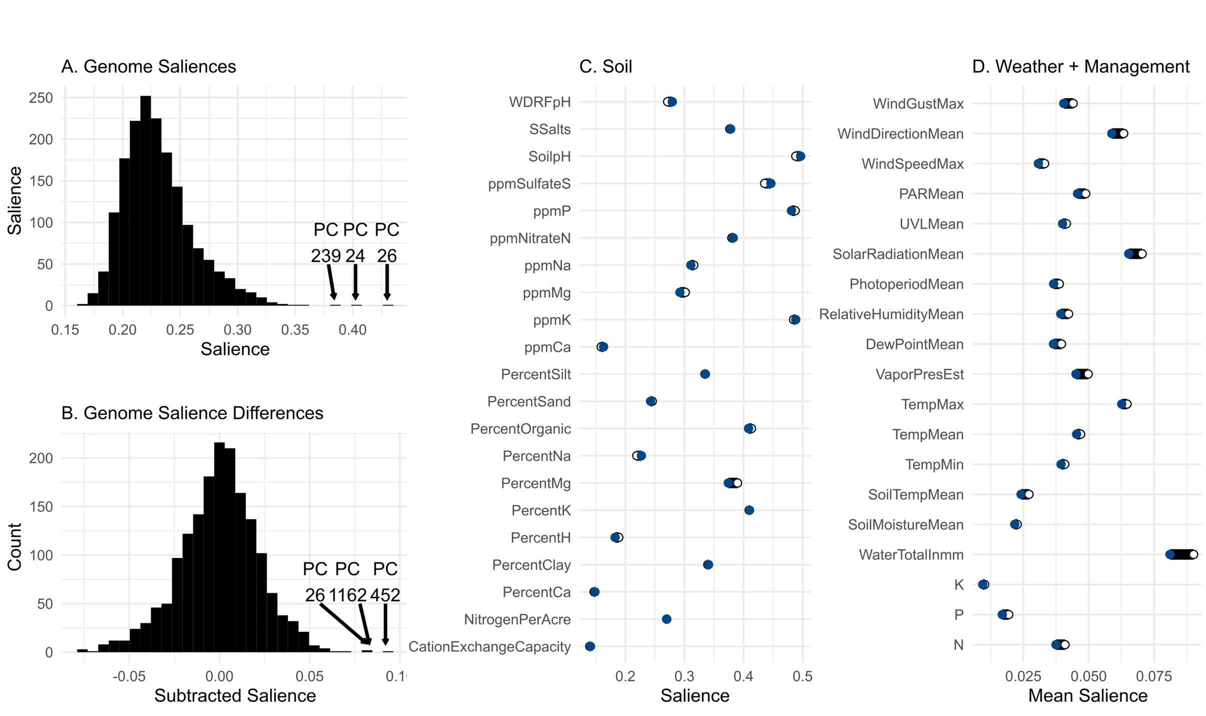
Additional Results

Site-by-Year Group Average RMSE Between DNN and BLUP Models



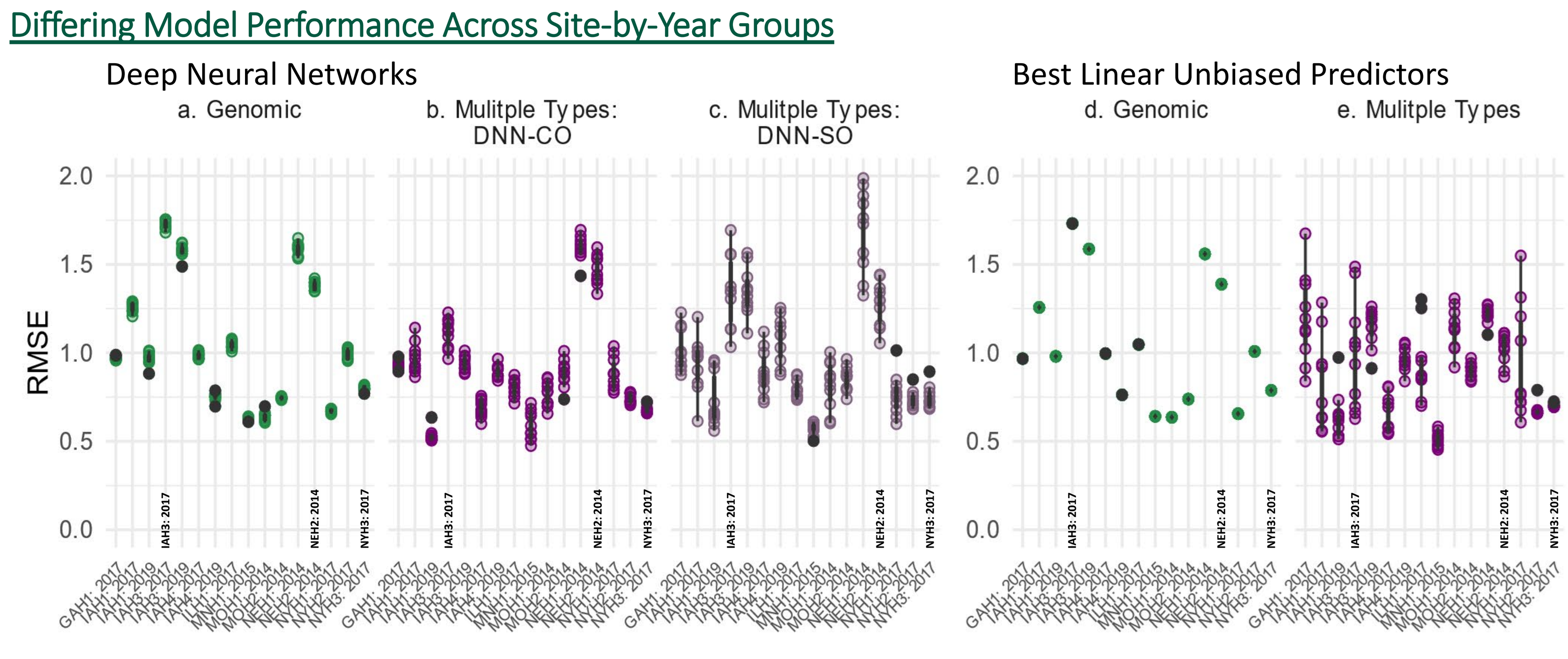
Performance in different site-by-year groups differs with respect to model type (DNN vs BLUP). Lower correlation between model errors (e.g. Multiple Types BLUP vs DNN-CO) suggests potential benefit from ensembling model estimates. Point color indicates correlation within site-by-year group (lavender > 0).

Influence of Interaction Effects on Feature Salience in Aggregate



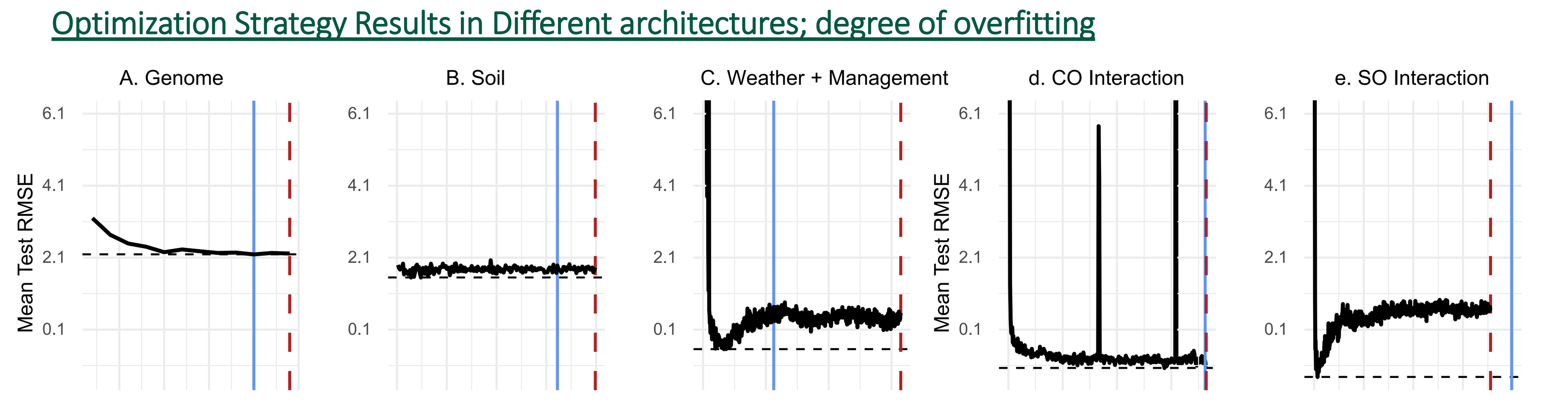
A. A histogram of the saliences for Genomic PCs for the interaction containing DNN-CO model does not indicate a clear preference for PCs that explain more variance. Note that the two most salient PCs were PC 26 and PC 24 which account for 0.350% and 0.392% of the total variance respectively. **B.** The difference between saliences of the full model and genomic submodel are shown. Those with the highest difference, explain little of the total variance. **C.** Salience of soil variables is shown for the interaction model (black open circles) and the submodel (blue filled circles). The difference is shown in black if salience is higher in the interaction model and blue if not. **D.** Salience values for weather and management variables for the interaction model and submodel are shown as average salience over the season. This indicates an overall similarity in salience across features, with the most notable difference being "WaterTotalInmm". Since that these values do not indicate the full effect of the weather and management variables in influencing predictions as these are daily values, not the total for the 210 day time window, resulting in smaller values in D than A - C.

Differing Model Performance Across Site-by-Year Groups



Model performance for site-by-year groups in the testing set for the DNNs and BLUPS. Variability across replicates appears higher in DNN-SO than in DNN-CO. Using multiple data types, RMSE decreases at many sites for DNNs (e.g., IAH3: 2017) and BLUPs but not consistently so (e.g., NEH2: 2014). Between the best two performing models, with high (NYH3: 2017) and low (NEH2: 2014) agreement are seen.

Optimization Strategy Results in Different architectures; degree of overfitting



Mean test set RMSE across 10 replicates (**A. – E.**). The horizontal dashed line is the minimum error. The vertical lines indicate the epochs selected by the chosen heuristic - minimizing total validation error (red dashed line). The mean plus standard deviation of validation error (solid blue line). Both strategies resulted in apparent overfitting in the Weather and Management submodel (C.) and the SO model (E.).