

Improving Drought Tolerance in Spring Wheat Through Integration of Aerial Phenotyping and Genomic Selection

Peter Schmuker, Sheri Ryneearson, Tyson Keopke, Michael Pumphrey
Department of Crop and Soil Sciences WSU



Drought Threatens Wheat Production

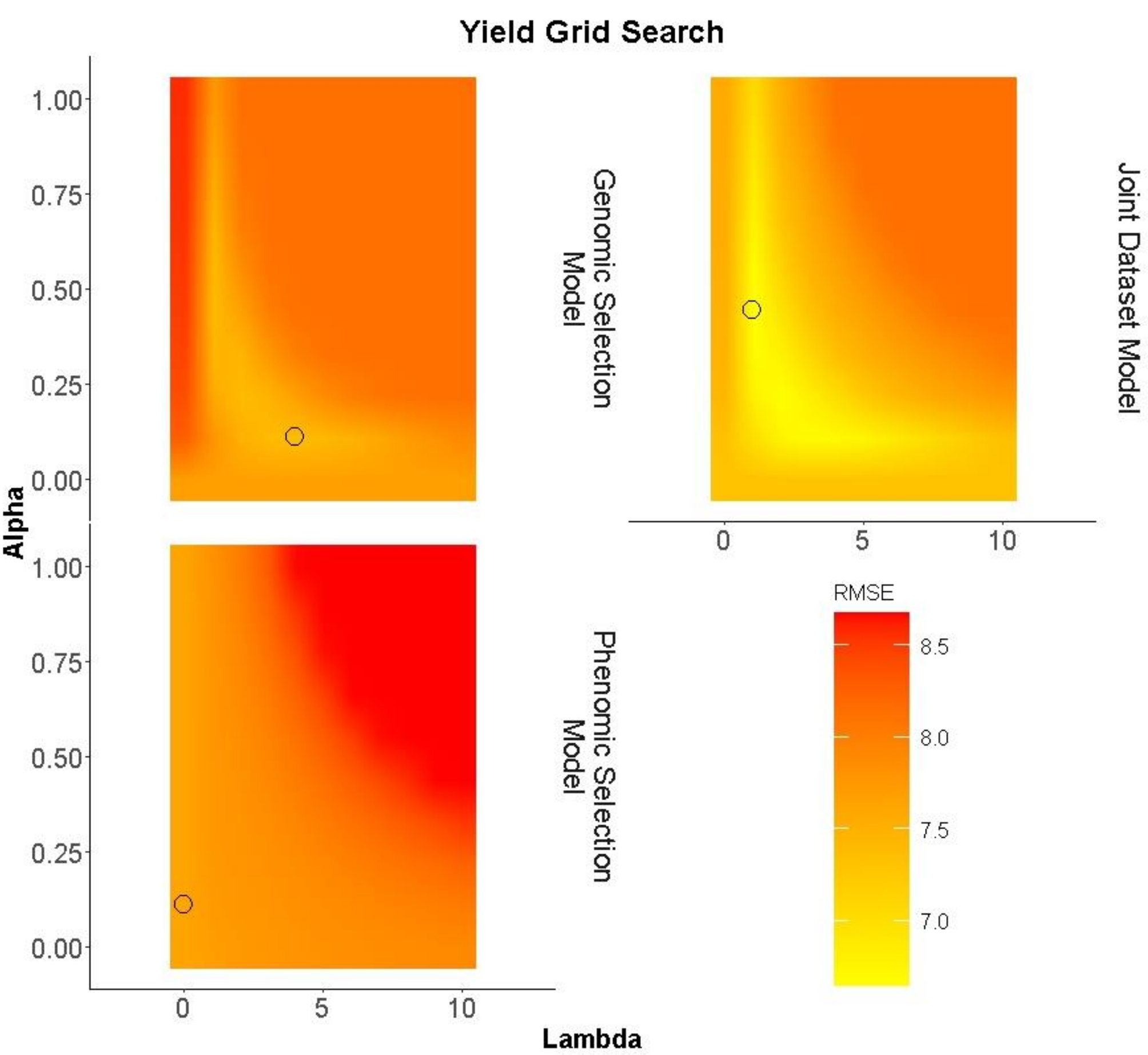
Wheat meets a fifth of global protein and calorie demands across both developed and developing nations. Increasingly dry growing seasons puts spring wheat yields at risk (2). Coupled with a growing population in need of affordable calories, this situation demands rapid progress in the genetic improvement of drought tolerance in spring wheat.

Genomic Selection is an applied breeding method in which genome-wide markers are used to predict plant performance. A predictive model trained under drought conditions can enable plant breeders to evaluate future generations for drought tolerance in absence of drought conditions.

Drone based remote sensing has become commonplace in agriculture. Multispectral cameras mounted to small UAVs provide insight into plant water status and photosynthetic activities through information derived from vegetative indices. Phenomic selection is the process of using information from remote sensors to make breeding decisions.

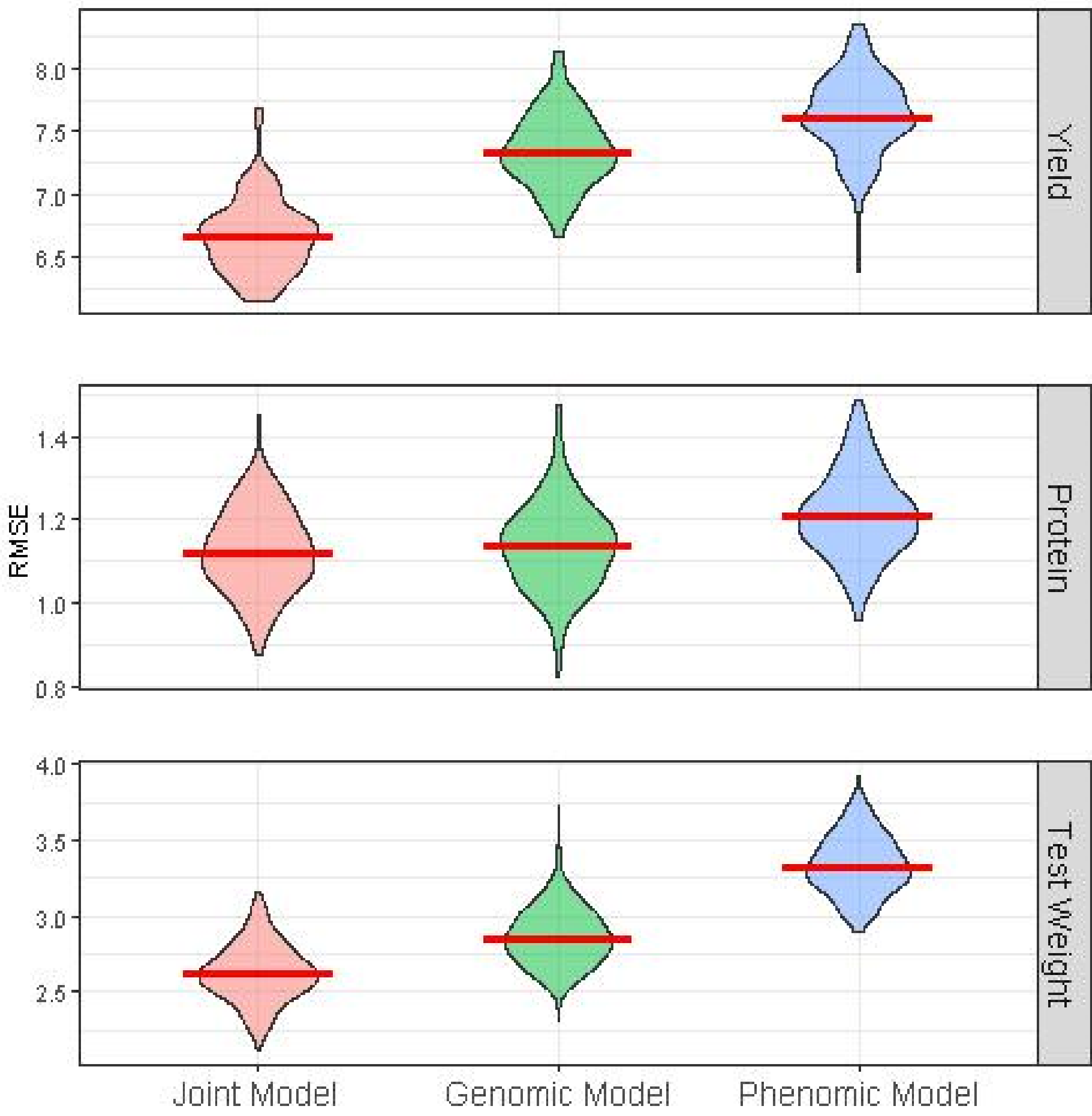
Phenomic and genomic datasets can be combined into a training data for the same predictive models. Studies show the possible improvements in prediction accuracy breeders should expect when this approach is adopted (3). Elastic-Net Regularization is a flexible predictive modeling framework that can be used for prediction of plant performance (4). Researchers have control over two tuning hyperparameters, alpha and lambda. Alpha controls mixing between Ridge Regression and LASSO levels of variable removal. Lambda controls the amount of shrinkage applied to predictor coefficients. Elastic-Net models with a joint phenomic and genomic dataset should outperform model with only individual phenomic or genomic predictive datasets in prediction of yield, test weight, and protein content under drought conditions.

Visual representation of cross-validation grid search for Yield prediction. Lowest average cross-validation error of 100 test folds is circled for each dataset. Best Alpha x Lambda grid combination is shown in the table below. Lambda search values extended out to 500 but the highest Lambda was only 35 for protein prediction with a genomic dataset.



| Best Alpha x Lambda Grid Combination | | | |
|--------------------------------------|---------|-------------|---------|
| Model | Yield | Test Weight | Protein |
| Joint Dataset Model | 0.4 x 1 | 0.1 x 1 | 0 x 28 |
| Phenomic Selection Model | 0.1 x 0 | 0 x 0 | 0 x 0 |
| Genomic Selection Model | 0.1 x 4 | 0 x 58 | 0 x 35 |

Prediction Accuracy by RMSE of Three Datasets



Violin plot of one hundred RMSE values calculated from the twenty rounds of five-fold cross validation at optimal regularization parameters, red bar through the median. Predictions made with a Joint Dataset have the lowest average error out of three datasets in prediction of all three traits.

Methods

- 1,200 wheat lines were planted near Pullman, WA in the summer of 2021.
- A Sentra Multispectral camera was mounted to a DJI Inspire 2 Drone and flown over the field at anthesis.
- Images were processed with Pix4dmapper and plot level indices extracted with R.
- NDVI and NDRE1, both related to photosynthetic activity, and NWI1 and NWI2, both related to plant water status, were calculated for use as Phenomic Predictors.
- Yield, test weight, and protein were recorded on all harvested plots.
- Tissue samples were taken, DNA extracted, and processed for sequencing information with Genotyping-By-Sequencing.
- Tassel Variant Calling GBS Pipeline V2 was used to perform variant calling and VCF Tools for filtering leaving 22,107 genetic markers for use as genomic predictors.

Modeling

- Prediction modeling of protein content was done only in the hard wheat market class.
- A grid of hyperparameter values was created with Alpha values ranging from 0 to 1 by 0.1 and Lambda values from 0 to 500 by 1 for a matrix of 10 by 500 hyperparameter combinations (1).
- To assess realistic model accuracy and to select tuning parameters, five-fold cross validation was repeated twenty times at each hyperparameter grid combinations
- A combination of alpha and lambda values. that produced the lowest average Root Mean Square Error (RMSE) value across individual cross-validation folds was identified per trait and predictor dataset combination.

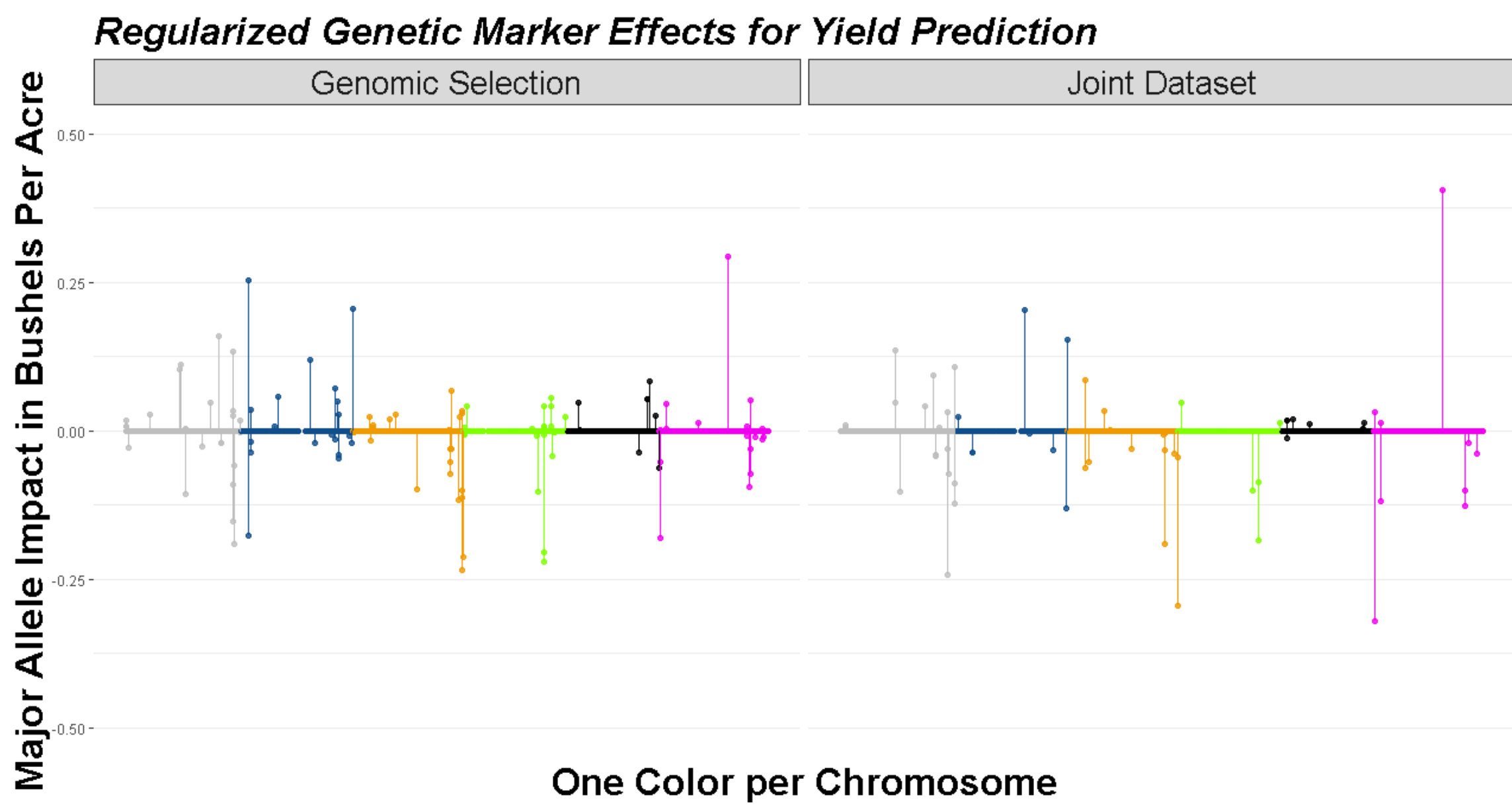
Results

The lowest average RMSE in the testing cross-validation folds for all three traits was observed in models that had a joint dataset of phenomic and genomic information. Genomic prediction models outperformed phenomic prediction models by RMSE values in all traits. Protein content had the smallest improvement in accuracy of the three studied traits with a joint prediction dataset.

| Model | Yield R ² | Yield RMSE | Test Weight R ² | Test Weight RMSE | Protein R ² | Protein RMSE |
|--------------------------|----------------------|------------|----------------------------|------------------|------------------------|--------------|
| Joint Dataset Model | 0.340 | 6.661 | 0.382 | 2.625 | 0.286 | 1.125 |
| Phenomic Selection Model | 0.230 | 7.629 | 0.028 | 3.342 | 0.138 | 1.212 |
| Genomic Selection Model | 0.193 | 7.354 | 0.250 | 2.864 | 0.275 | 1.135 |

| Number of Predictors in Final Regularized Model | | | | |
|-------------------------------------------------|-------|-------------|---------|---------------------|
| Model | Yield | Test Weight | Protein | Possible Predictors |
| Joint Dataset Model | 159 | 572 | 21,034 | 22,111 |
| Phenomic Selection Model | 3 | 4 | 4 | 4 |
| Genomic Selection Model | 304 | 21,805 | 21,030 | 22,107 |

Yield models had fewer non-zero predictors in the final regularized models compared to test weight and protein. This can be viewed in the above table. Predictions made with a joint dataset are still greatly determined by genetic predictors.



The above figure shows a subset of chromosomes and their respective marker effects after optimal regularization. The Genomic Selection model had more, small effect predictors compared to the Joint dataset method which had fewer, larger effects predictors after regularization.

Conclusions

Improvement of drought tolerance for high yield, adequate protein content, and high test will be faster with predictive models with joint genomic and phenomic predictors. Predictive models will empower breeders to make rapid genetic gain in the face of more frequent droughts.

Acknowledgements

Wheat Cap Coordinated Agricultural Project
ARCS Foundation

Sources

(1) Lasso and elastic-net regularized generalized Linear Models. <https://glmnet.stanford.edu/>
(2) Reynolds, M. P., & Braun, H.-J. (2022). *Wheat Improvement Food Security in a changing climate*. Springer International Publishing.
(3) Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., Reynolds, M., & Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes/Genomes/Genetics*, 6(9), 2799–2808. <https://doi.org/10.1534/g3.116.032888>
(4) Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>