A Hierarchical Spatial Finlay-Wilkinson Model for Analysis of Multi-Environment Field Trials

Xingche Guo*, Somak Dutta, Dan Nettleton
Department of Statistics, Iowa State University

Objective

- Create a multidisciplinary framework for understanding and managing crop performance over diverse environmental conditions.
- Incorporate spatial effects, genetic, and environmental information to enhance understanding of genomes-by-environment (GxE) interaction.
- Distinguish corn varieties that are highly adaptable to changing environments.
- Predict the yield of a (possibly novel) corn variety in a (possibly new) environment.

Data Introduction

We only focus on 2015 G2F dataset with:

- A subset of 24 environments (field trials).
- Yield recorded for 10,971 field plots with known spatial locations.
- A total of 1,105 hybrid genotypes (varieties).
- SNPs sequence data at ∼1M genomic loci are available.
- Time-indexed measurements for weather variables (temperature, rainfall amount, solar radiation, etc), and several soil variables (pH value, soil organic matter, etc).

Future Work

- Allow more complex models (non-linear models, time series models, functional data models, etc) for environmental covariates.
- Formulate better kinship matrix to improve estimation and further accelerate the algorithm.
- Extend to generalized HSFW model to account for discrete value responses.

References

Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(4):691–746.

Dutta, S. and Mondal, D. (2015). An h-likelihood method for spatial mixed linear models based on intrinsic autoregressions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(3):699–726.

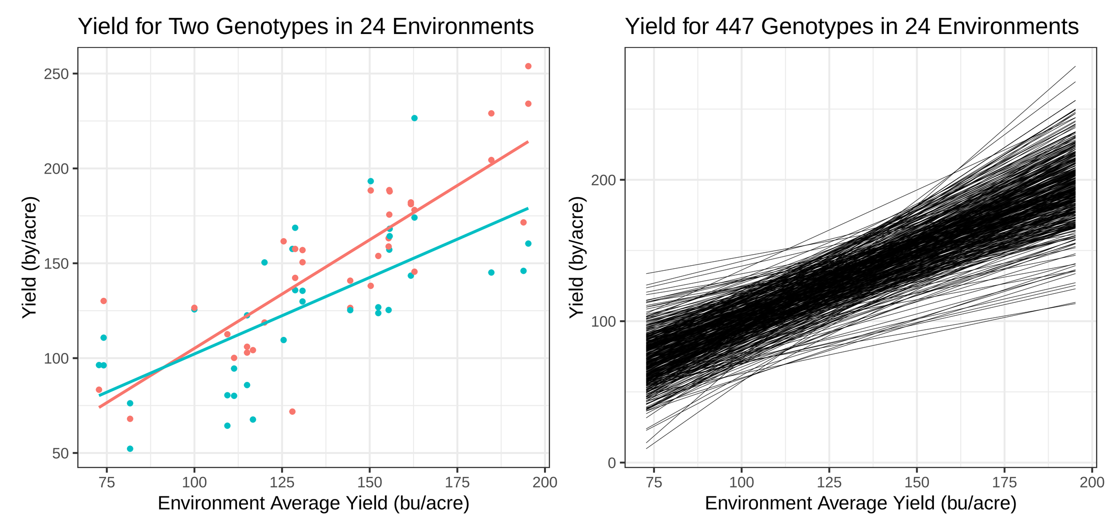
Finlay, K. and Wilkinson, G. (1963). The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research, 14(6):742–754.

Acknowledgements

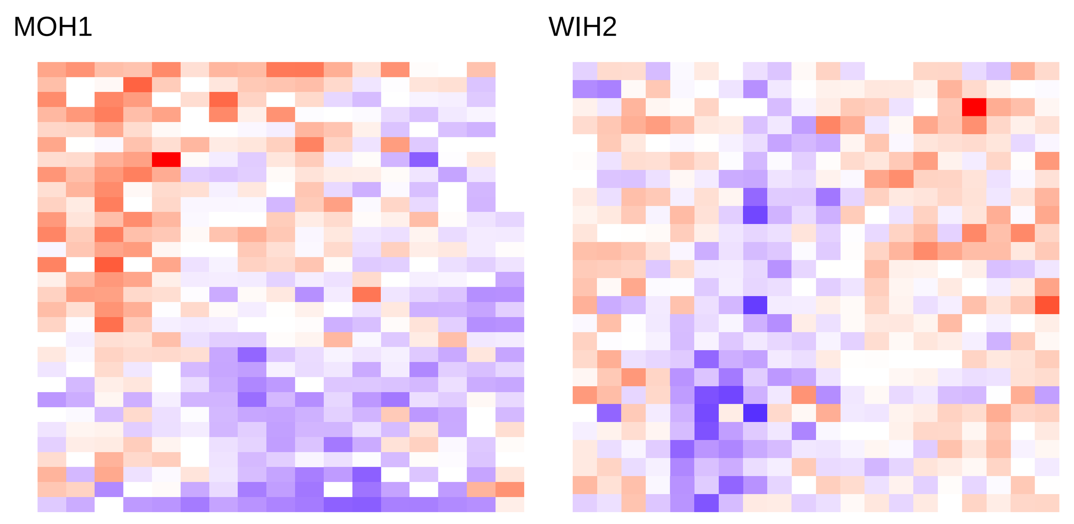
The authors acknowledge financial support of Iowa State University Plant Sciences Institute Scholars Program, the Baker Center for Bioinformatics and Biological Statistics, and the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. IOW03617, which is supported by USDA/NIFA and State of Iowa funds.

Finlay-Wilkinson (FW) Model

- FW model (Finlay and Wilkinson, 1963): $y_{ijk} = \mu + g_i + h_j + b_i h_j + e_{ijk}$, where μ is the overall mean, g_i is the genotype effect, h_j is the environment effect, $b_i h_j$ is the FW-type multiplicative interaction effect.



(a) ***FW Model Mechanism Diagram:*** *Left panel: the scatterplot of yield versus environment average yield for 24 field trials and the fitted FW lines for two example corn varieties using 2015 G2F data; Right panel: the fitted FW lines for 447 example corn varieties.*



(b) ***Residuals of FW Model for Two Fields:*** *The rectangular plots in the fields are colored based on the value of the residuals. We find that there are clear color patterns in the fields, which means the residuals are highly spatially correlated.*

Hierarchical Spatial Finlay-Wilkinson (HSFW) Model

- Data model: $[y_{ijk} | \mu, \mathbf{g}, \mathbf{b}, \mathbf{h}, \boldsymbol{\phi}] \overset{indep}{\sim} \mathcal{N}(\mu + g_i + h_j + b_i h_j + \phi_{ijk}, \sigma_e^2)$.
- Prior distributions for genotype, slope, and environment effects:

$$[\mathbf{g}] \sim \mathcal{N}(\mathbf{0}, \mathbf{A}\sigma_g^2); \qquad [\mathbf{b}] \sim \mathcal{N}(\mathbf{0}, \mathbf{A}\sigma_b^2); \qquad [\mathbf{h}|\boldsymbol{\gamma}] \sim \mathrm{N}(\gamma_1 \mathbf{Z}_1 + \cdots + \gamma_L \mathbf{Z}_L, \mathbf{I}\sigma_h^2).$$

  $\mathbf{A}$ is the kinship matrix describing the correlation structure between different hybrid corn varieties; $\mathbf{Z}_l$ is the lth standardized environmental covariate.

- First order intrinsic autoregression (IAR) prior for spatial effects: (Besag and Higdon, 1999; Dutta and Mondal, 2015).

$$[\boldsymbol{\psi}_j | \theta_j, \sigma_j^2] \propto |\sigma_j^{-2} \mathbf{W}_j|_+^{1/2} \exp\left(-\tfrac{1}{2}\sigma_j^{-2} \boldsymbol{\psi}_j^{\mathrm{T}} \mathbf{W}_j \boldsymbol{\psi}_j\right),$$

  where

$$\boldsymbol{\psi}_j^{\mathrm{T}} \mathbf{W}_j \boldsymbol{\psi}_j = \theta_j \sum\sum (\psi_{u,v} - \psi_{u-1,v})^2 + \bar{\theta}_j \sum\sum (\psi_{u,v} - \psi_{u,v-1})^2.$$

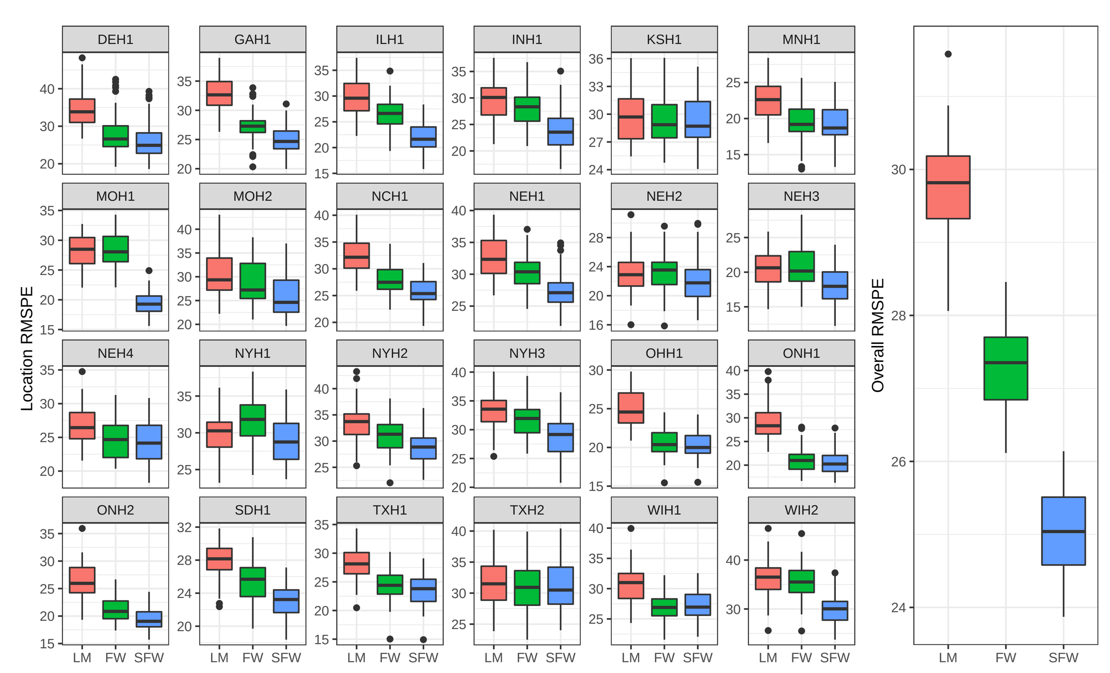
- Projected intrinsic autoregression (PIAR) prior (A sum-to-zero constrained version of IAR prior):

$$\boldsymbol{\phi}_j = \mathbf{B}_j \boldsymbol{\varphi}_j, \qquad \boldsymbol{\varphi}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_j^{-1}),$$

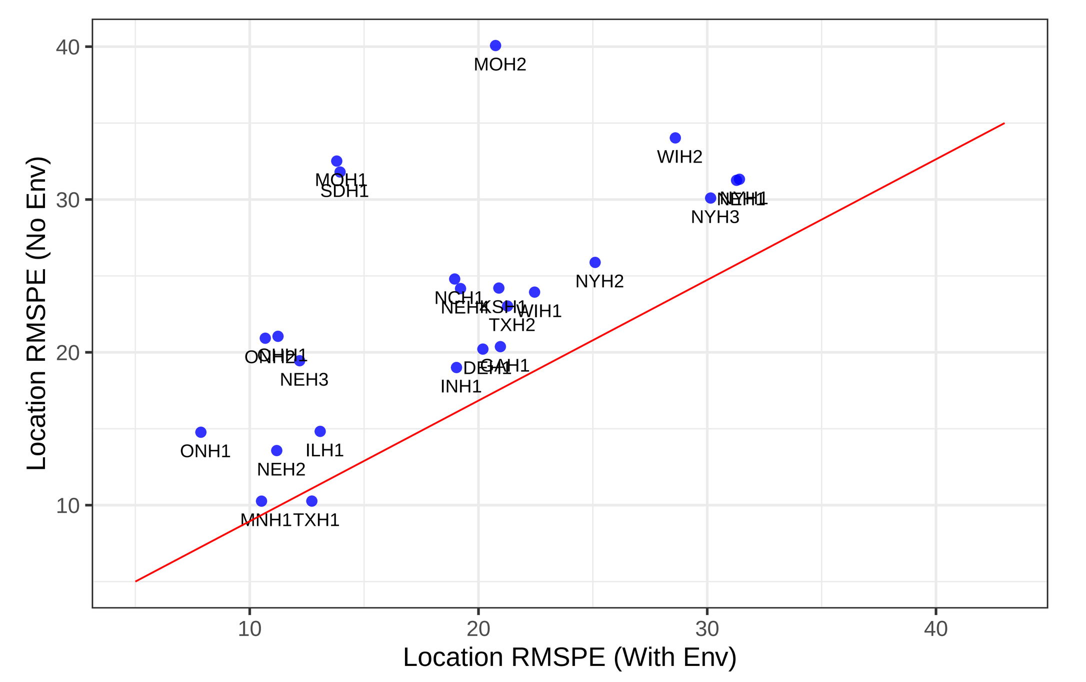
  $\mathbf{D}_j$ is the diagonal matrix with its diagonal entries to be all the nonzero eigenvalues of $\mathbf{W}_j$; $\mathbf{B}_j$ is the corresponding eigenvector matrix. We develop matrix free algorithm for the fast computation.

- Implement posterior predictive distributions for: (i) within-field prediction; (ii) prediction in new environments. Prediction intervals can also be obtained.

Data Analysis



(a) ***Model Evaluation via Within-Field Prediction:*** *Left panel: location-wise RMSPE computed by linear additive model (LM), Finlay-Wilkinson model (FW), and Finlay-Wilkinson model with spatial effects (SFW); Right Panel: overall RMSPE computed by these three models.*



(b) ***Predict in New Environments:*** *The scatterplot of the 24 location-wise RMSPEs computed using temperature and rainfall data (x-axis), versus the location-wise RMSPEs computed not using any environment information (y-axis).*

	90% CL			95% CL		
	LM	FW	SFW	LM	FW	SFW
Coverage Percentages	90.3%	89.9%	90.1%	95.3%	94.9%	94.5%
Median Interval Widths	98.4	90.3	80.1	117.3	107.5	95.76

Compare Prediction Interval Widths: *the median credible interval widths of LM, FW, and SFW models at 90% and 95% credible levels are provided. SFW model has a more precise interval prediction given that SFW model has the shortest interval widths at the same coverage levels.*