A Hierarchical Spatial Finlay-Wilkinson Model for Multi-Environment Trial Analysis

Xingche Guo, Somak Dutta, Dan Nettleton

Dept. of Statistics, Iowa State University

Joint Statistical Meetings, 2019

Finlay-Wilkinson (FW) model

- FW model:

$$y_{ijk} = \mu + g_i + h_j + b_i h_j + e_{ijk},$$

- A model for characterizing genotype-by-environment interaction.

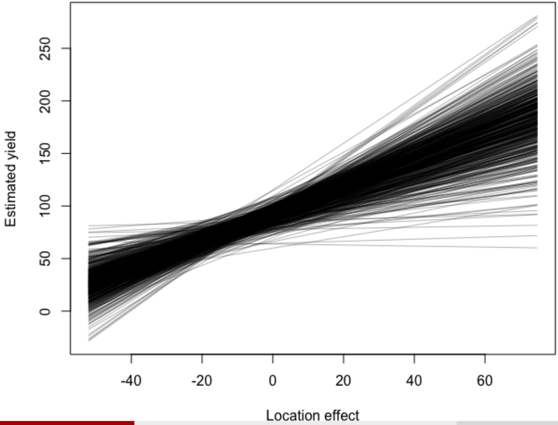
- For genotype i:

$$y_{ijk} = (\mu + g_i) + (b_i + 1)h_j + e_{ijk},$$

  FW model becomes a linear model with intercept $\mu + g_i$ and slope $b_i + 1$.

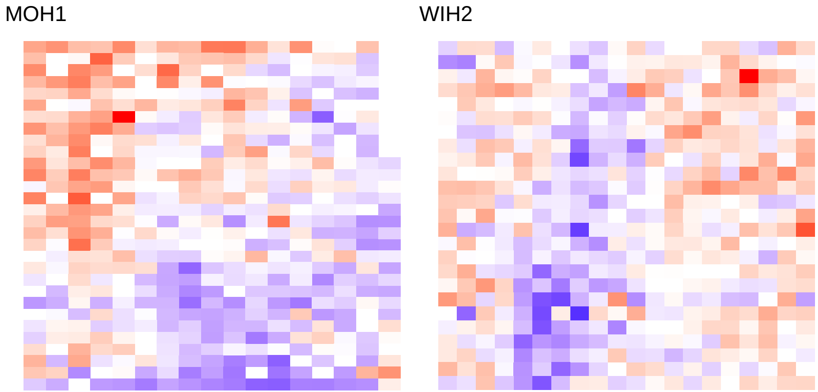
Finlay-Wilkinson (FW) model

This plot gives one example of the fitted FW model (using G2F data). Each line represents the linear relationship between grain yield and location effect for one genotype.



Residuals of FW model for two fields

Problem: the residuals are highly spatially correlated.



Hierarchical spatial Finlay-Wilkinson (SFW) model

- Data model:

$$[y_{ijk}|\mu, \mathbf{g}, \mathbf{b}, \mathbf{h}, \phi] \overset{indep}{\sim} \mathcal{N}(\mu + g_i + h_j + b_i h_j + \phi_{ijk}, \sigma_e^2),$$

- Prior distributions for genotype, slope, and field effects:

$$[\mathbf{g}] \sim \mathcal{N}(\mathbf{0}, \mathbf{A}\sigma_g^2); \quad [\mathbf{b}] \sim \mathcal{N}(\mathbf{0}, \mathbf{A}\sigma_b^2);$$

$$[\mathbf{h}|\gamma] \sim \mathrm{N}(\gamma_1 \mathbf{Z}_1 + \cdots + \gamma_l \mathbf{Z}_l + \cdots + \gamma_L \mathbf{Z}_L, \mathbf{I}\sigma_h^2)$$

A is a known matrix describing the correlation structure of **g** and **b**, $\mathbf{Z}_l$ is the lth environment-specific covariate.

Prior distributions for spatial effects

- A popular model for fertility adjustment in agricultural field trials is the **first order intrinsic autoregression** (Besag and Higdon, 1999; Dutta and Mondal, 2015).
- First order Intrinsic Autoregressive prior (Besag 1995, Besag 1999):

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

where

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

- The distribution of $\boldsymbol{\psi}_j$ is **invariant** to the addition of arbitrary constant.

Prior distributions for spatial effects

- The intrinsic model has an indeterminate overall level.
- The overall levels of the field specific spatial effects are completely confounded with the location effects.
- Not directly applicable for yield prediction.
- **A hard constraint**: set the average of the spatial effects to zero.

Projected intrinsic autoregression (PIAR) prior

Special structure of $\mathbf{W}_j$:

- Then the spectral decomposition of $\mathbf{W}_j$ is given by

$$\mathbf{M}_j\mathbf{W}_j\mathbf{M}_j^{\mathrm{T}} = \theta_j\boldsymbol{\Lambda}_{r_j} \otimes \mathbf{I}_{c_j} + \bar{\theta}_j\mathbf{I}_{r_j} \otimes \boldsymbol{\Lambda}_{c_j}.$$

- where $\mathbf{M}_j = \mathbf{N}_{r_j} \otimes \mathbf{N}_{c_j}$.
- $\boldsymbol{\Lambda}_k$ denote the $k \times k$ diagonal matrix whose uth diagonal entry is $4\sin^2\{\pi(u-1)/(2k)\}$.
- $\mathbf{N}_k$ denotes the $k \times k$ orthogonal matrix whose (u, v)th entry is $1/\sqrt{k}$ if $u = 1$, $\forall v$, and $(2/k)^{1/2}\cos\{\pi(u-1)(v-1/2)/k\}$ otherwise.

Projected intrinsic autoregression (PIAR) prior

- The Gaussian projected intrinsic autoregression (PIAR) on the $r_j \times c_j$ regular array is then defined as:

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

- $\mathbf{B}_j^{\mathrm{T}}$ denotes the $(r_j c_j - 1) \times r_j c_j$ matrix consisting of last $r_j c_j - 1$ rows of $\mathbf{M}_j$.

- $\mathbf{D}_j$ denotes the diagonal matrix consisting of the nonzero eigenvalues of $\mathbf{W}_j$.

Matrix free computation

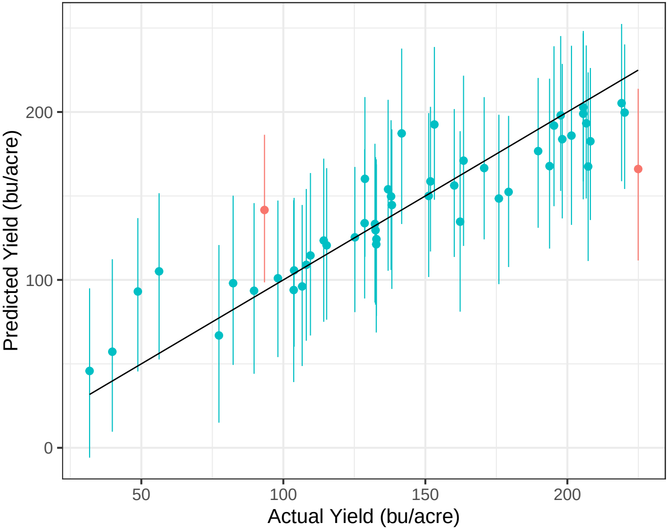
- The covariance matrix of the Gaussian PIAR is a dense singular matrix.
- The computation load for generating ϕ_j from PIAR using knowledge of multivariate statistics is $\mathcal{O}((r_j c_j)^3/3)$.
- Assume small number of missing plots (denote M_j as the number of missing plots).)
- Thus matrix-vector multiplications with $\mathbf{B}_j$ and $\mathbf{B}_j^{\mathrm{T}}$ can also be performed using these discrete cosine transformations (DCT).
- The computation load of matrix free computation algorithm is $\mathcal{O}(r_j c_j M_j \log r_j c_j + M_j^3/3)$.

Within-fields prediction

- Given grain yields values for some plots in fields, predict the yields for the rest plots.
- Implement posterior predictive distributions.
- Important to account for the spatial correlation between plots.
- Kinship information plays a decisive role when there are new crop varieties in the testing sets.

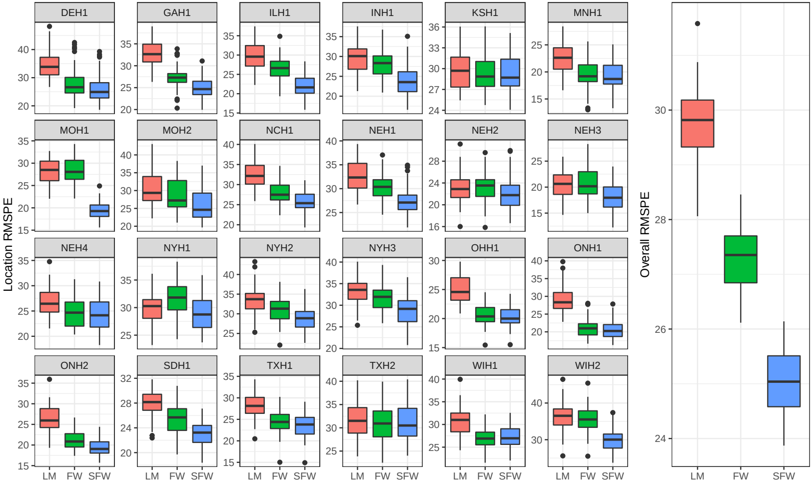
Within-fields prediction

50 Plot Yield Prediction Intervals (95% Credible Level).



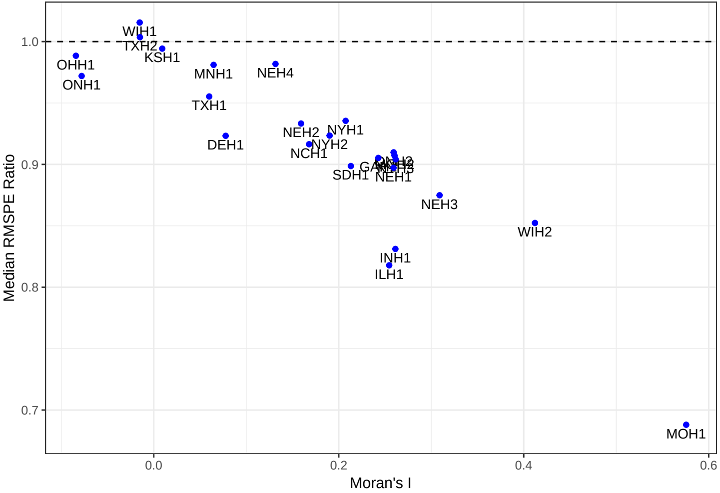
Within-fields prediction

Reduced error for yield prediction.



Within-fields prediction

Level of spatial correlation vs performance of SFW model.

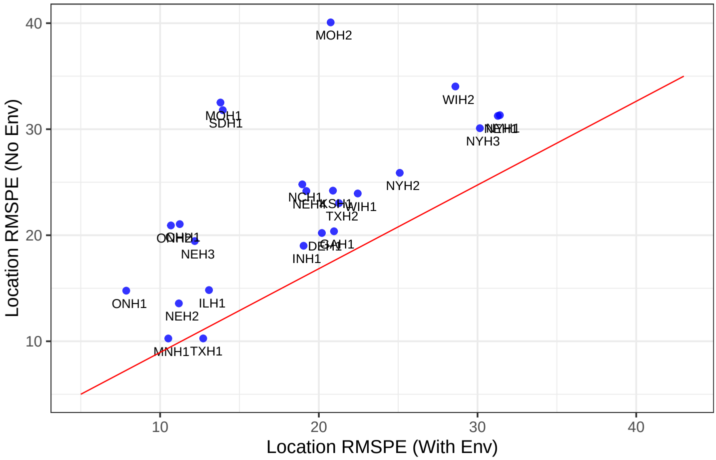


Out-of-fields prediction

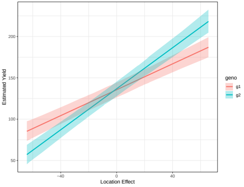
- Predict the grain yields in new fields.
- Need to learn how field effects depend on the weather and soil variables.
- More field trials and environmental data will lead to a more accurate and robust prediction.

Out-of-fields prediction

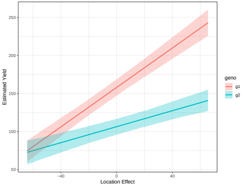
Location-wise RMSPEs computed using temperature and rainfall data (x-axis), versus the location-wise RMSPEs computed not using any environment information (y-axis).



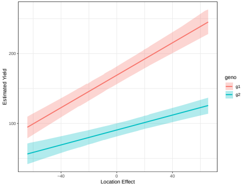
Assessing Uncertainty about FW Regression Lines



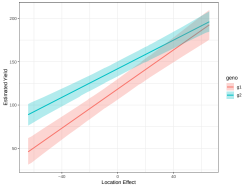
(a)



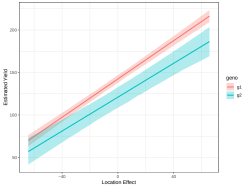
(b)



(c)



(d)



(e)

Figure: Estimated Yield vs Location Effect for pairs of genotypes

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 auto-regressions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(3):699–726.

Thank You!