UNIVERSITY OF HOHENHEIM

Statistical analysis of data from the field phenotyping platform "BreedVision"
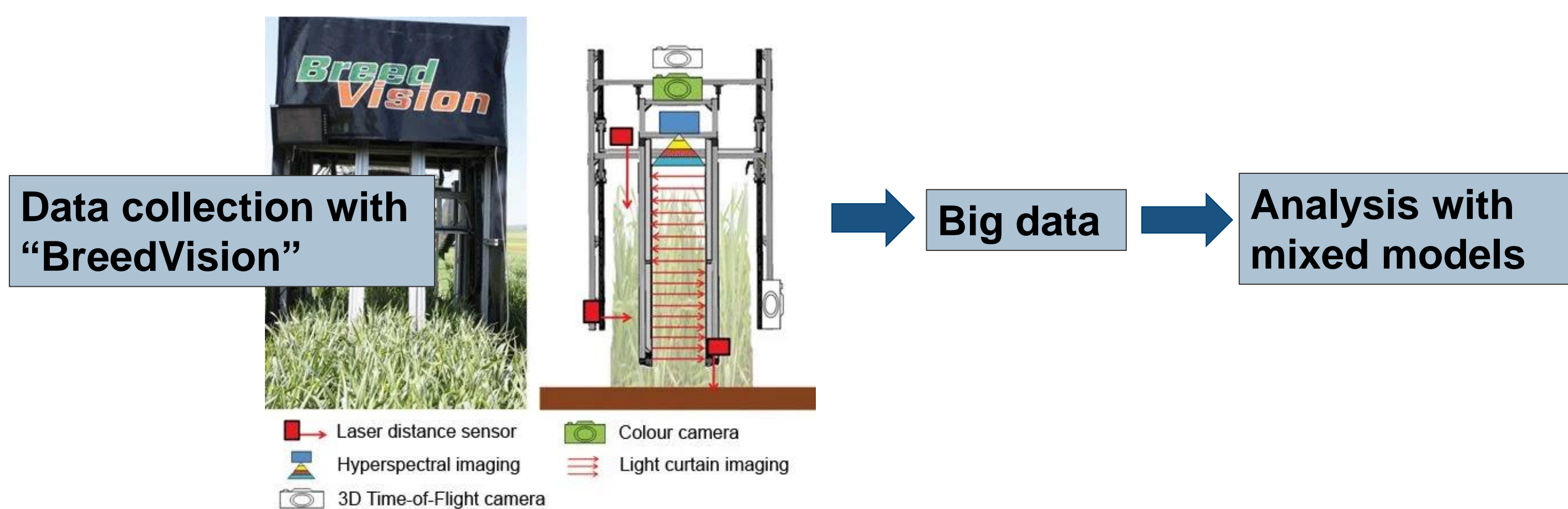
Md Anisur Rahaman[1], Hans Peter Maurer [1], and Hans-Peter Piepho[2]

[1]*University of Hohenheim | State Plant Breeding Institute (720) | Fruwirthstraße 21 | 70599 Stuttgart | Germany*
[2]*University of Hohenheim | Institute of Crop Science | Fruwirthstraße 23 | 70599 Stuttgart | Germany*

Introduction

➢ Phenotyping platform like "Breedvision" is a very useful tool in plant breeding for measuring complex traits (e.g., abiotic stress, growth dynamics).

➢ This phenotyping platform making big and complex data sets readily available for analysis.

➢ Secondary trait must have high heritable correlation with target trait. Adding covariates can reduce error, thus increase prediction accuracy of the model.

➢ Research goal is to i) compare prediction accuracies of univariate and bivariate models. ii) evaluate the effect of adding covariates to the model.

Data collection with "BreedVision" → Big data → Analysis with mixed models

🔴➡ Laser distance sensor 🟫 Colour camera
🟫 Hyperspectral imaging 🔴➡ Light curtain imaging
📷 3D Time-of-Flight camera

Results

Table 1: Comparison between the model without covariates (Eq. 2) and model with covariates (Eq. 3)

Model	AIC	BIC
Eq. 2	1194.5	1368.5
Eq. 3	749.2	926.5

Table 2: Mean correlation and mean RMSE between observed and predicted DMY from univariate (Eq. 1) and bivariate (Eq. 4) analysis, 95% confidence limits (CLs) of means in parenthesis.

Summary statistics	Univariate (Eq. 2)	Bivariate (Eq. 4)
Mean Pearson's correlation	0.8599 [0.8590-0.8607]	0.8537 [0.8528-0.8546]
Mean RMSE	0.7949 [0.7929-0.7968]	0.8103 [0.8092-0.8115]

➢ Model with covariates shows lower AIC and BIC scores (Table 1). So, model with covariates performs better.

➢ Univariate model shows lower mean correlation and RMSE (0.8599 and 0.7949 respectively). Hence, univariate model has better prediction accuracy (Table 2).

Conclusions

➢ Addition of covariates showed better prediction accuracies.

➢ We recommend to include these two covariates (radiation intensity and ambient temperature) in the final model.

➢ There was no gain from the bivariate model in 10000 simulations.

➢ However, the bivariate model with CT can be utilized as a index for indirect selection in adverse situations such as bad weather condition or limited seed supply.

Methods

Univariate model for Dry matter yield (DMY) and Canopy temperature(CT)

1. $DMY_{ijhnt} = \mu_{nt} + \gamma_j + b_{jht} + \tau_{in} + \varepsilon_{ijhnt}$

Here, DMY_{ijhnt} = response of *i*-th genotype in *h*-th block nested within *j*-th replicate, n, t = subscripts added for nitrogen level and trial areas, respectively and ε_{ijh} = residual plot error associated with DMY_{ijh}, $\varepsilon_{ijh} \sim$ N $(0, \sigma_e^2)$.

2. $CT_{ijkhnt} = \mu_{nt} + \gamma_{jt} + b_{jht} + \tau_{in} + \varepsilon_{ijkhnt}$

Here, CT_{ijkhnt} = response of *i*-th genotype in *h*-th block nested within *j*-th replicate with *k*-th measuring replicate and ε_{ijkhnt} = residual plot error of CT_{ijkhnt}. Errors from the same plot piled into a vector $\boldsymbol{\varepsilon_{ijkhnt}}$ and $\boldsymbol{\varepsilon_{ijkhnt}} \sim$ N (0,**R**).

Adding covariates to the model

3. $CT_{ijkhnt} = \mu_{nt} + \gamma_{jt} + b_{jht} + \tau_{in} + \beta_1 x_{1ijkhnt} + \beta_2 x_{2ijkhnt} + \varepsilon_{ijkhnt}$

Here, ($x_{1ijkhnt}$ = ambient temperature corresponding to CT_{ijkhnt} and $x_{2ijkhnt}$ = radiation intensity corresponding to CT_{ijkhnt}, β_1 and β_2 are corresponding slopes for ambient temperature and radiation intensity).

Final model

4. $Y_{ijkhntb} = \mu_{ntb} + \gamma_{jtb} + b_{jhtb} + \tau_{inb} + \beta_1 x_{1ijkhntb} \cdot switch + \beta_2 x_{2ijkhntb} \cdot switch + \varepsilon_{ijkhntb}$

Here, switch= 0 to off covariates in DMY; switch= 1 to include covariates in CT
b= (DMY, CT).

Model evaluation

Data

■ CT ■ DMY | 20% DMY data removed randomly | Univariate Bivariate Estimate dropped DMY
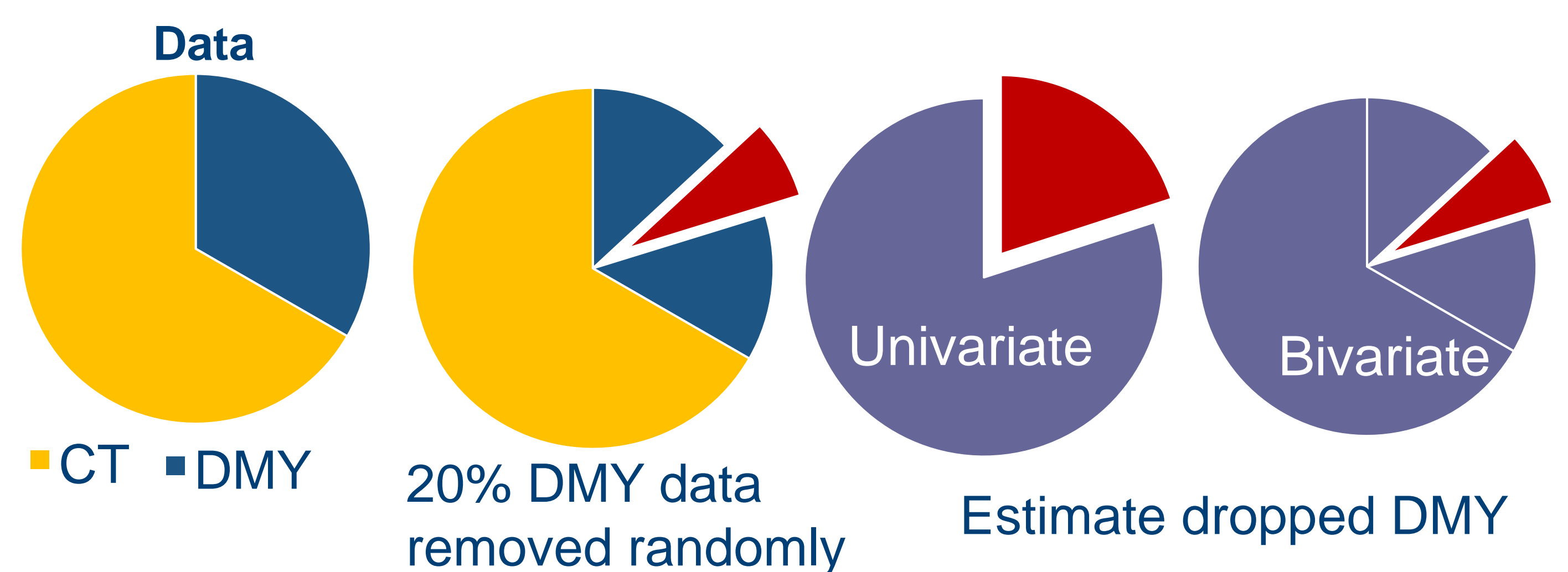
➢ Calculation of Correlation and root mean squared error (RMSE) for 100 and 10000 simulations cross validation for DMY

➢ Calculation of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for evaluation of covariates.