

How do you like them predictions?

My experience with obtaining predictions for GLMMs

Michael Mumford, Kerry Bell

Queensland Department of Agriculture and Fisheries, Leslie Research Facility, Toowoomba, QLD 4350, Australia

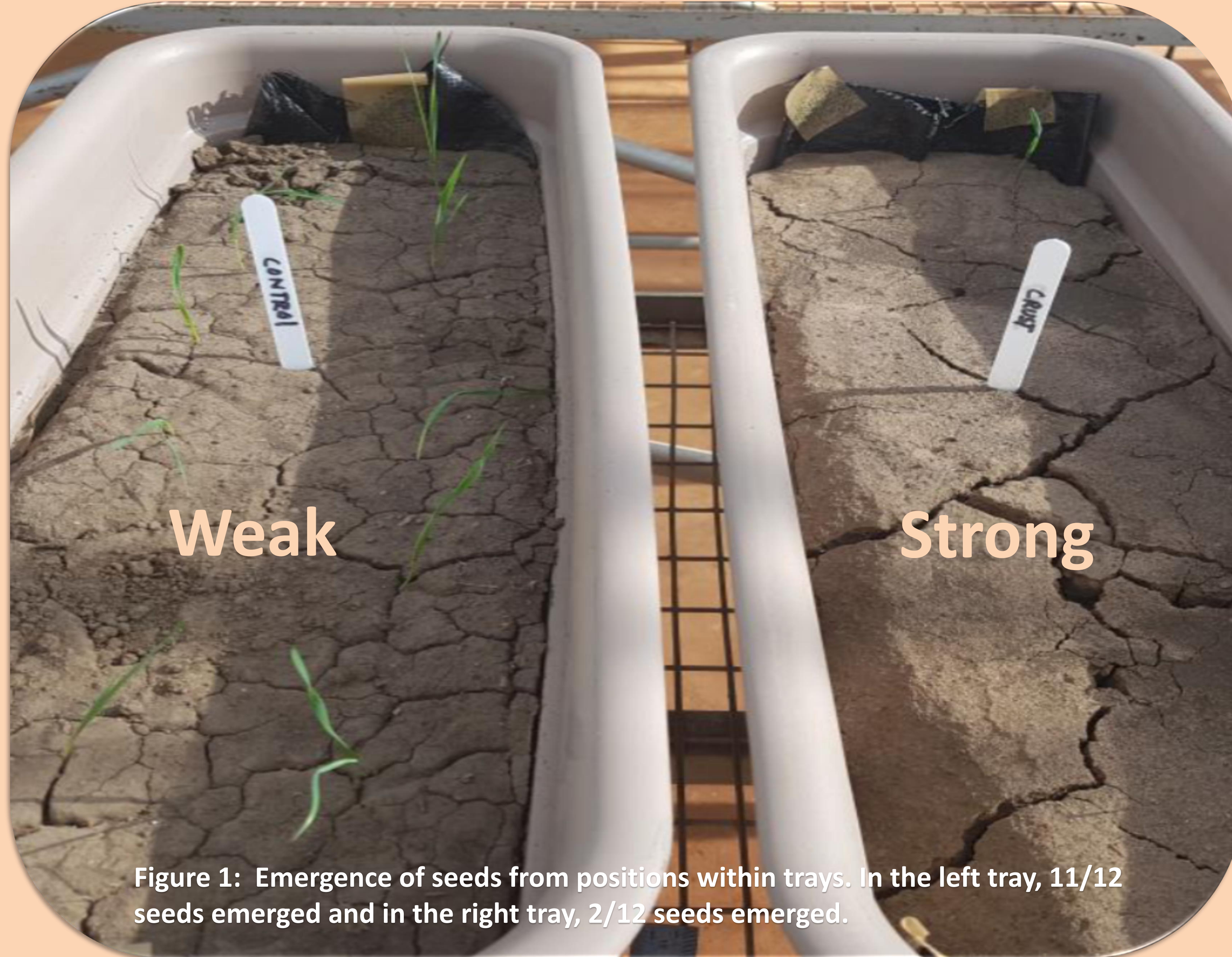


Figure 1: Emergence of seeds from positions within trays. In the left tray, 11/12 seeds emerged and in the right tray, 2/12 seeds emerged.

Introduction

To model treatment effects for a binary response variable, Generalised Linear Models (GLMs) are required. In order to incorporate experimental design terms, an extension to Generalised Linear Mixed Models (GLMMs) is required. Hierarchical Generalised Linear models (HGLMs) are an extension of GLMMs that allow the random effects to be non-normally distributed.

Motivating Example – Sodic soils glasshouse experiment

Many cropping regions in Australia are affected by the surface crusting of sodic soils [1] which results in the reduction of seedling emergence and thus, plant growth. An experiment was performed to investigate seedling germination and emergence proportion of 38 wheat genotypes in strong (i.e. sodic soils) and weak crusts.

Experimental Design

- 48 trays of soil were prepared, and each tray contained 12 positions where one seed was planted.
- Trays were grouped into 6 replicate blocks and 2 crust treatments (strong/weak) were randomly allocated to trays (main plots) (Figure 1).
- 38 genotypes were assigned to trays using an incomplete block design, and then randomly allocated to the 12 positions (sub plots) within a tray.

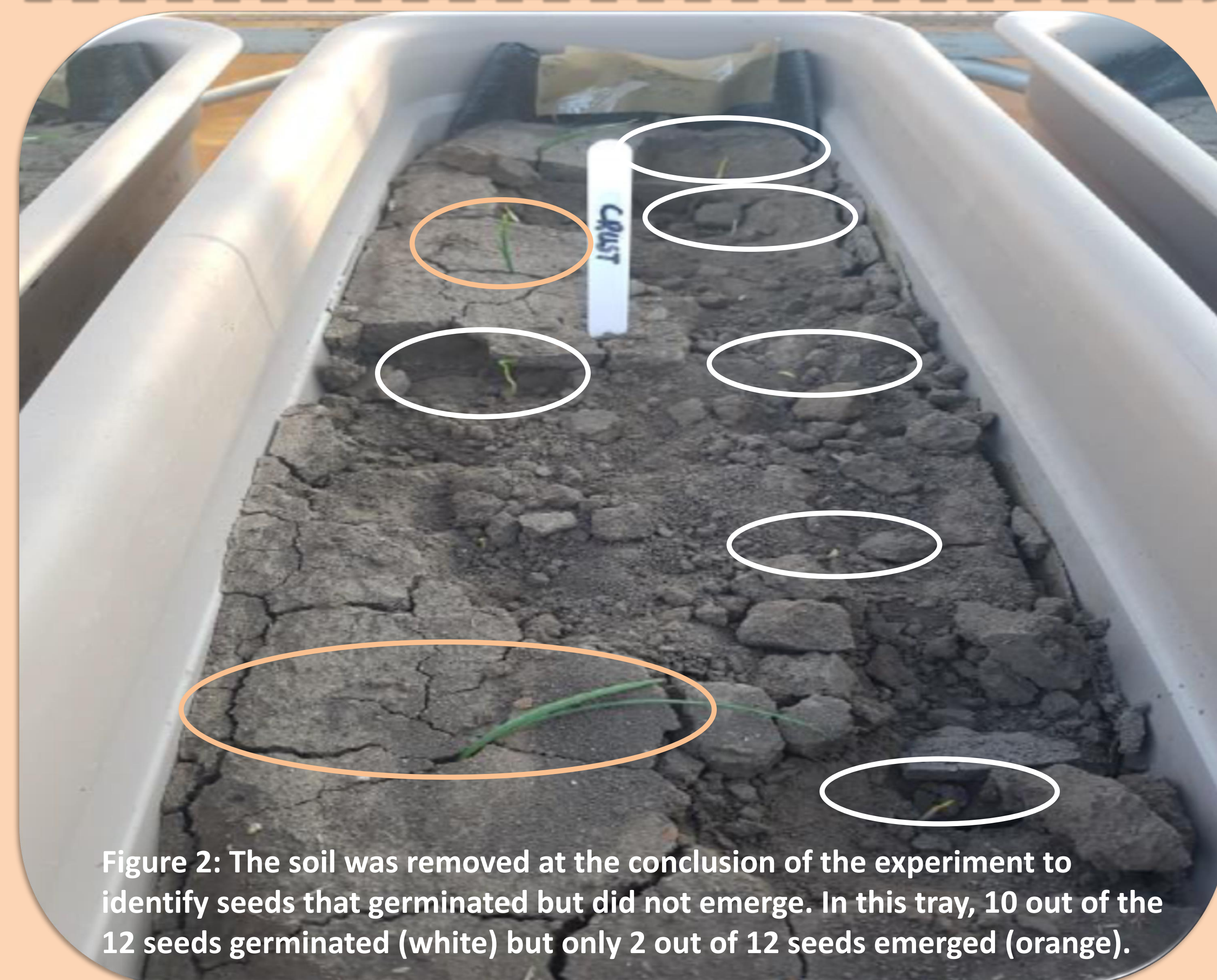


Figure 2: The soil was removed at the conclusion of the experiment to identify seeds that germinated but did not emerge. In this tray, 10 out of the 12 seeds germinated (white) but only 2 out of 12 seeds emerged (orange).

Statistical Methods

- Response variable of interest is emergence proportion (out of the seeds that germinated, Figure 2).
- Two different models used to analyse the data: a) GLMM b) HGLM
- Emergence proportion was assumed to follow a binomial distribution with a logit link function.

$$\log\left(\frac{\theta}{1-\theta}\right) = X\tau + Zu + e, \quad y|u \sim \text{Binomial}(n, \theta)$$

- Models had genotype, crust and their interaction fitted as fixed effects and Rep/Tray as random.
- In the GLMM [3], the random effects were assumed to follow a normal distribution. The dispersion parameter σ_e^2 was also estimated. $u \sim N(0, G)$
- The conjugate HGLM [2,4] assumed that the random effects followed a beta distribution with a logit link function. The dispersion parameter was assumed to follow a gamma distribution with a log link function. $u \sim \text{Beta}(\alpha, \beta)$, $\sigma_e^2 \sim \text{Gamma}(a, b)$
- All analyses were run using GenStat 18th edition [5].

Results

- Genotype x crust interaction effects were non-significant ($P > 0.05$), and thus was removed for both models. There were significant genotype ($P < 0.001$) and crust ($P < 0.001$) main effects in both models.
- For the GLMM, the predictions were generally higher than the observed emergence proportion (Table 1, Figure 3) when the observed emergence proportion is greater than 0.5 and lower when the observed proportion is less than 0.5. This is because averaging of the random effects is performed on the original scale and NOT on the scale of the linear predictor (logit).
- In a conjugate HGLM, random effects are averaged on the scale of the linear predictor.

Table 1: Raw data means and predictions from each of the proposed models. Numbers inside the parentheses are the back-transformed standard errors.

Crust	Observed proportion	GLMM	HGLM
Weak	0.869	0.952	0.885 (0.059)
Strong	0.329	0.296	0.324 (0.034)

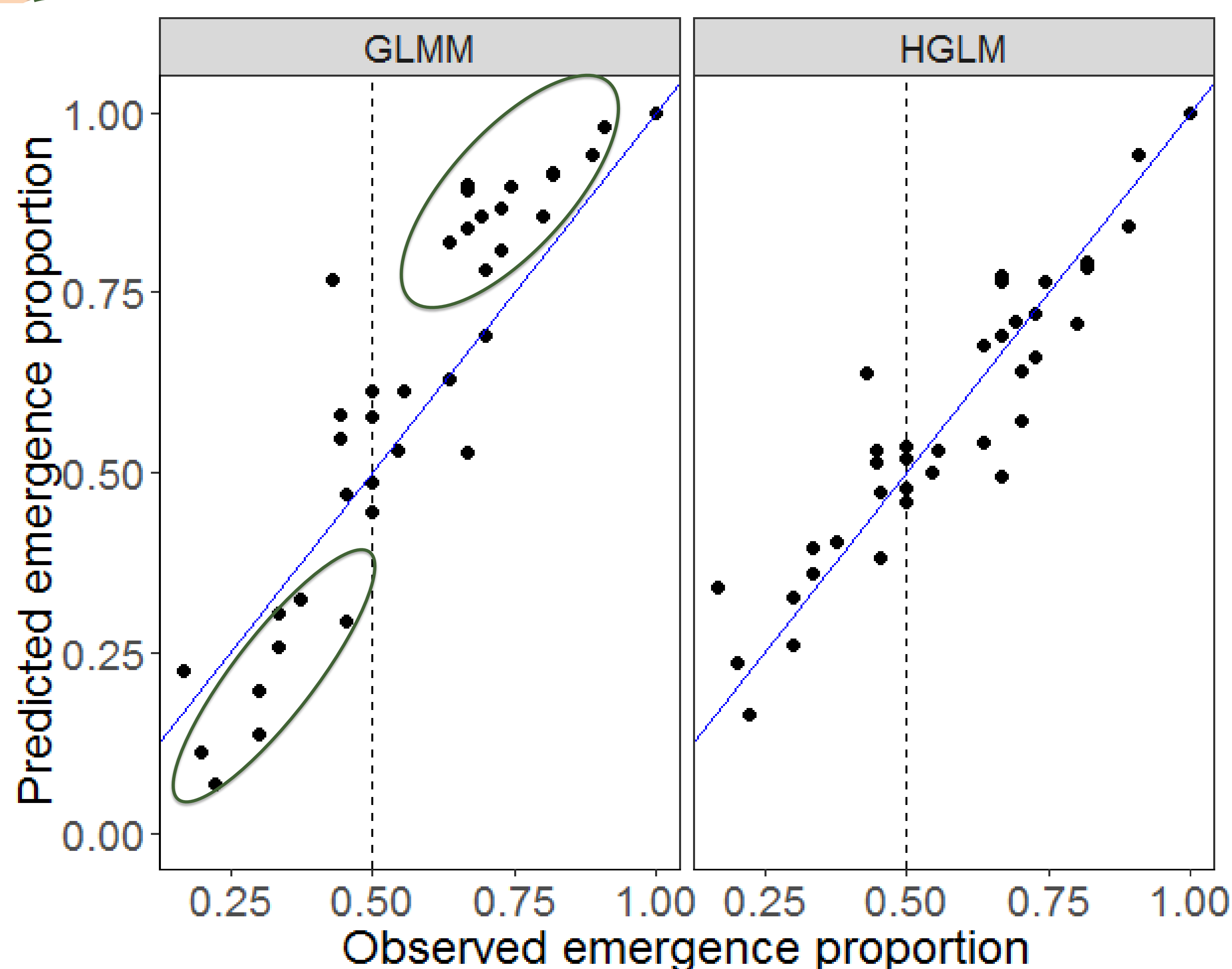


Figure 3: Genotype predictions of emergence from the GLMM model and the HGLM model.

Summary

- HGLMs are an extension of GLMMs that allow the random effects to be non-normally distributed.
- Predictions for binomial GLMMs will tend to be higher than the observed means for proportions greater than 0.5 and lower for proportions less than 0.5.
- Conjugate HGLMs provide a way to obtain predictions using averages of the random effects on the original scale of the data as opposed to the linear predictor scale.
- Both GLMMs and HGLMs are susceptible to estimation biases [2].

References

- [1] Anzoonan, M., Christopher, J., Mumford, M. H., et al (2018). Selection for rapid germination and emergence may improve wheat seedling establishment in the presence of soil surface crusts *Plant and Soil*, 1-13.
- [2] Collins, D. (2008). The performance of estimation methods for generalized linear mixed models. Doctoral dissertation, School of Mathematics & Applied Statistics-Faculty of Informatics, University of Wollongong).
- [3] Lee, Y., & Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, 88(4), 987-1006.
- [4] Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78(4), 719-727.
- [5] VSN International. (2015). *Genstat for Windows 18th Edition*. Hemel Hempstead: VSN International.

Acknowledgements

The authors would like to acknowledge the Grains Research Development Corporation for their investments in the projects UA00159 & DAQ00208. The authors would also like to thank Roger Payne for his ideas and suggestion and Monia Anzoonan for the motivating example.