Stepping Towards Causal Learning in Agriculture

Aidan Moller[1], Max Moldovan[1]
[1]Adelaide University

AAGI
Analytics for the Australian Grains Industry

ABSTRACT

Australian grain growers face difficult management decisions in increasingly unpredictable environments, where the effectiveness of a strategy can depend on factors that are only partly observable. Management practices are often evaluated using data-driven models, process-based models, or a combined "hybrid" approach. We take a causal hybrid approach to crop modelling, using an APSIM-derived directed acyclic graph (DAG) to identify which variables to control for when estimating the total effect of nitrogen (N) fertiliser on barley grain yield. Two DAG-informed generalised additive models (GAMs) estimated the same N effects as a baseline GAM with TRIAL (YEAR × LOCATION) random effects, which halved standard errors, while controlling for LOCATION instead of STATE reduced them a further ~5%. Low, high, and very high N increased yield by 0.29, 0.50, and 0.32 t/ha compared to nil N.

1 Causal Question

How will a nitrogen (N) fertiliser treatment (low, high, or very high N rate) change grain yield, relative to nil N?

2 Data

The **Barley Synthesis Project (BSP)** dataset combines 48 trials across WA, Vic, and NSW, from 2010 to 2016. One-way (12 trials) and two-way (36 trials) blocking is used for the treatment factors N, seed rate, and variety. We fit each model to all 10,890 observations, contrasting the estimated effect of N at low (20 or 30 kg N/ha), high (80 or 90 kg N/ha), and very high (150 kg N/ha, in 7 NSW trials) rates against nil (0) N to compute average treatment effects.

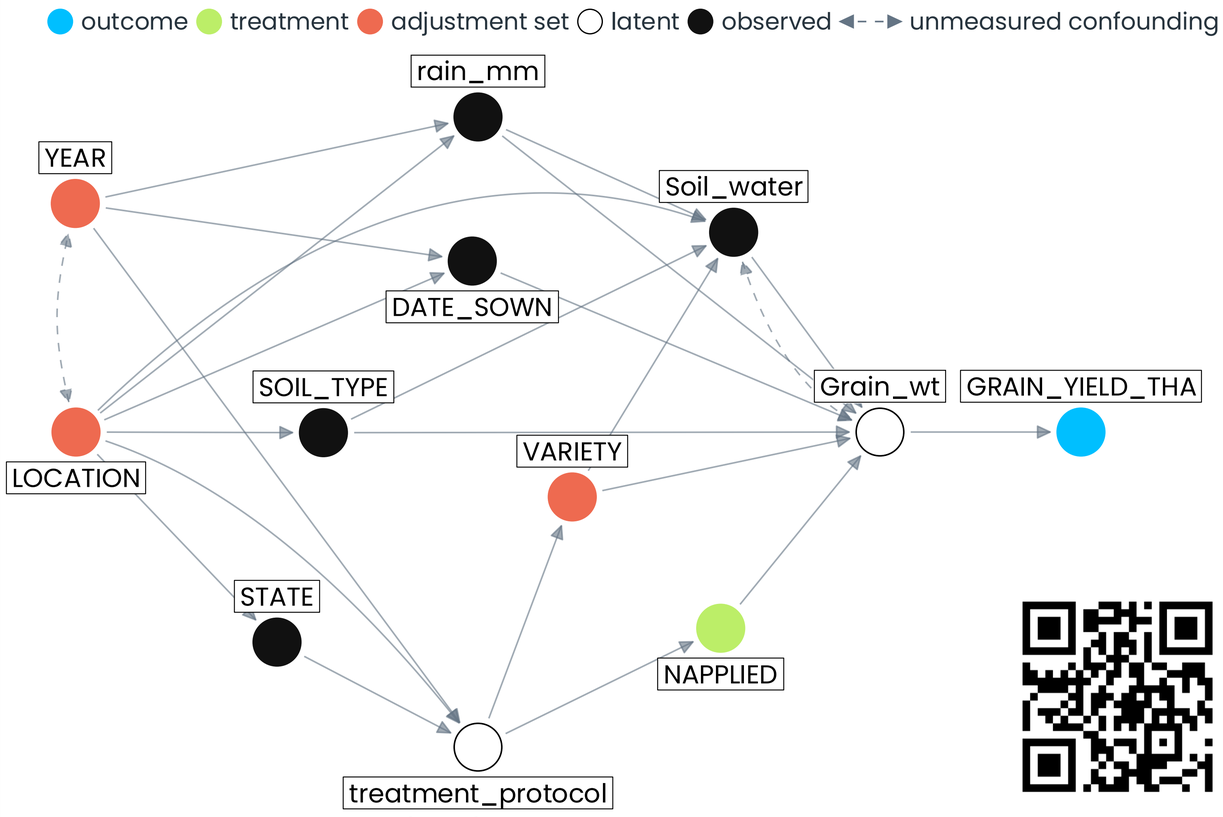
3 Causal Graph

Directed acyclic graphs (DAGs) make explicit statements about the hypothesised causal relationships between variables. Drawing an arrow from one node to another is a statement about both an assumed causal link between two variables and the direction of causality.

The DAG shown below joins two distinct sub-DAGs:

1. **Trial context DAG:** based on an exploratory analysis revealing that NAPPLIED and VARIETY vary by LOCATION, YEAR, and STATE;
2. **Growth process DAG:** derived from a structural causal model of APSIM-embedded mechanisms (Holzworth et al., 2018).

Software demonstration available at dagby.net (Moller, 2026).



A directed acyclic graph (DAG) showing a minimal adjustment set for the total effect of N: **YEAR**, **LOCATION**, and **VARIETY** (DAG-GAM 1). The other minimal set is **DATE_SOWN**, **LOCATION**, **rain_mm**, and **VARIETY** (DAG-GAM 2).

4 Modelling

Generalised additive models (GAMs) with random effects are suitable for agricultural applications where clustered, hierarchical, or longitudinal data are used (Verbyla et al., 1999). We estimate the total effect of N on grain yield using two DAG-informed GAMs (one per minimal adjustment set) and compare the results with two baseline GAMs not informed by a DAG. Baseline models are adapted from Paynter et al. (2024) but omit the interactions between N and the other treatment factors (SEEDRATE × NAPPLIED × VARIETY).

Baseline GAM 1: y ~ STATE + STATE:VARIETY + NAPPLIED + s(rain_mm, k = 4) + s(YEAR, bs = "re")

Baseline GAM 2: y ~ STATE + STATE:VARIETY + NAPPLIED + s(rain_mm, k = 4) + s(YEAR, bs = "re") + s(TRIAL, REP, bs = "re")

DAG-GAM 1: y ~ LOCATION + LOCATION:VARIETY + NAPPLIED + s(YEAR, bs = "re") + s(TRIAL, REP, bs = "re")

DAG-GAM 2: y ~ LOCATION + LOCATION:VARIETY + NAPPLIED + s(rain_mm, k = 4) + s(DATE_SOWN, k = 4) + s(TRIAL, REP, bs = "re")

5 Results

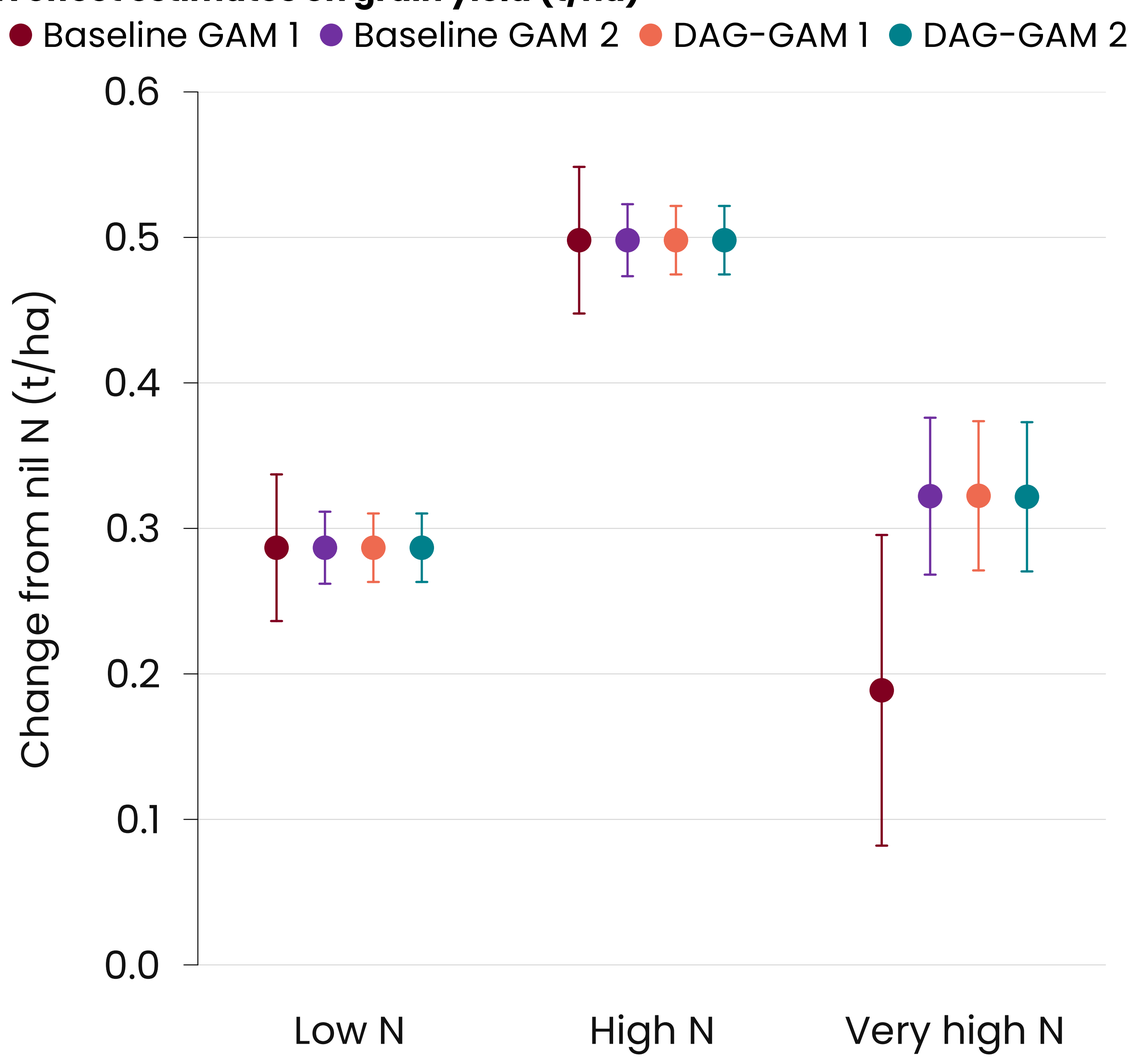
- **Baseline GAM 1** leaves a **backdoor path open** via LOCATION.
- **Baseline GAM 2** effectively closes this path by adding random effects for REP (replicates) within TRIAL (YEAR × LOCATION).
- **Baseline GAM 1** estimated a lower very high N effect (0.19 vs 0.32) but with about 2× larger SEs, mostly attributable to lacking these random effects.
- **DAG-GAM 1** and **DAG-GAM 2** controlled for LOCATION in place of STATE and reduced SEs by about 5% beyond **Baseline GAM 2**, with the same estimated N effects.

Estimated effect of N on grain yield (t/ha)

Model	Low N	High N	Very high N	RMSE	AIC
Baseline GAM 1	0.287 [0.236, 0.337]	0.498 [0.448, 0.548]	0.189 [0.082, 0.295]	1.066	32422.6
Baseline GAM 2	0.287 [0.262, 0.311]	0.498 [0.473, 0.523]	0.322 [0.268, 0.376]	0.520	17063.7
DAG-GAM 1	0.287 [0.263, 0.310]	0.498 [0.475, 0.522]	0.322 [0.271, 0.374]	0.491	16111.7
DAG-GAM 2	0.287 [0.263, 0.310]	0.498 [0.475, 0.522]	0.322 [0.270, 0.373]	0.491	16111.4

Estimates relative to nil N [95% confidence interval]. RMSE calculated on fitted data.

N effect estimates on grain yield (t/ha)



Points and intervals: change from nil N and 95% confidence intervals.

6 Industry impact

This example shows how an APSIM-derived DAG representing plant growth processes can inform statistical models of pooled field trial data. Feedback from the demo (dagby.net) will shape an end product that uses causal methods to support industry-relevant research and improve the reliability of recommendations for Australian grain growers.

7 References

Holzworth, D., Huth, N. I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N. I., Zheng, B., & Snow, V. (2018). APSIM Next Generation: Overcoming challenges in modernising a farming systems model. *Environmental Modelling & Software, 103*, 43–51.

Moller, A. (2026). *Constructing directed acyclic graphs for grain production systems by querying mechanistic knowledge: dagby demonstration* [Computer software]. https://dagby.net

Paynter, B., Moldovan, M., Graham, R., & Khan, H. (2024). *Preserving grain quality: Final report, Theme 4* (GRDC Project No. UCS2301-002RTX). Grains Research and Development Corporation.

Verbyla, A. P., Cullis, B. R., Kenward, M. G., & Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 48*(3), 269–311.

GRDC Grains Research & Development Corporation | Adelaide University | Curtin University | The University of Queensland Australia

The support of the Grains Research and Development Corporation (GRDC) provided through the Analytics for the Australian Grains Industry (AAGI) project is acknowledged.