



Agronomic Drivers of Grain Protein in *Triticum durum*: A Structural Equation Modeling Approach

Walker Maess¹, Arthur Fritzke^{2,3}, Ali Lofti¹, Curtis Ponziak¹, Salvador Osuna¹,
Kirstin Bett¹, Steve Shiftliffe¹, Kwabena Nketia¹

¹Plant Science Department University of Saskatchewan, Saskatoon, SK, Canada | ² University of Saskatchewan Computer Science Department, Saskatoon, SK, Canada | ³ Agriculture Agrifoods Canada, Saskatoon, SK

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th
Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.

Grain protein concentration in wheat is influenced by complex interactions among plant development, soil properties, and nitrogen availability. This study evaluated the agronomic drivers of protein using structural equation modeling (SEM) applied to a large dataset ($n \approx 26,000$).

Initial analyses highlighted strong interdependence among predictors, motivating a structural framework to partition direct and indirect effects. The final SEM demonstrated excellent fit (CFI = 0.997, RMSEA = 0.048) and explained approximately 25% of the variance in protein concentration.

Protein was negatively associated with yield (standardized $\beta = -0.26$), indicating a dilution effect, and positively associated with maturity ($\beta = 0.35$) and soil properties ($\beta = 0.27$). Nitrogen availability increased yield but exhibited a weak negative direct effect on protein ($\beta = -0.07$).

These results demonstrate that protein formation is governed by a trade-off between yield and grain filling duration, with soil conditions exerting an additional independent influence. The findings highlight the importance of integrating phenology and soil effects when modeling grain protein, rather than relying solely on input variables such as nitrogen availability.

Keywords.

protein, high-throughput phenotyping, SEM

Introduction

Plant breeding has long been a critical component to the improvement of both production and quality of food grains [1]. Traditional breeding strategies compare the phenotypic outcomes of progeny from known parental plants in the search for favourable traits [2]. The reliance of destructive data sampling in traditional methods is largely undesirable due to the labour and costs associated with these methods [3]. Furthermore, breeding programs must be diligent with trait selection and studies, since too many plant characteristics exist to enable breeding all traits at once [4]. This requires the need for breeders to better understand what factors drive the formation of agronomic traits in plant breeding programs.

Predictive models applied to phenotypic data allows breeders greater efficiencies by relating measurable variables to phenotypic outcomes [5]. These types of models are used to inform programs on cultivar selections. Previous studies have implemented these strategies to understand the yield formation in wheat and how nitrogen plays a role in estimating and modelling the outcomes of the plant [6].

While predictive modelling can provide substantial insight into complex systems, interpretability of the relationships modelled within these systems can become blurry. This study investigates relationships throughout the growing season that influence grain protein in *Triticum durum*. The objective of this study is to inform users of predictive models within High-Throughput Phenotyping on agronomic features best used for protein prediction in *Triticum durum*.

Materials and Methods

Site and Trial Characteristics

Yield-plot trials conducted in the growing season of 2025 provide agronomic data of 101 *Triticum durum* cultivars. These trials are incorporated into a larger study investigating space-time plant interactions for rotational-breeding strategies, known as the ACTIVATE project [7]. This experiment included 100 different cultivars of both *Triticum Durum* and *Lens culinaris*, providing 10,000 varietal combinations (S. Caballero, personal communication, May 20, 2026). Cultivars are seeded using an augmented randomized incomplete block design, with 3 check cultivars planted in each block.

The area of interest is located east of Saskatoon, Saskatchewan comprised mainly of Bradwell Orthic Dark Brown soil expressing a fine sandy loam to loam texture [8]. Durum was planted in May 2025 with a target plant stand of 340 plants/m² at both sites. Nitrogen fertilizer was applied at time of planting at a rate of 14.13 kg N/ha to allow establishment of seedlings. All other nitrogen sources stem from the labile pool, with additional nitrogen mineralized throughout the growing season.

Data

Agronomic data of the cultivars within the trial was taken throughout the season. Data collected include: Seeding Date, Heading Date, Maturity Date, and Height. Date metrics were expressed as Growing Degree Days. Additionally, 249 random plots were soil sampled in the Fall of 2024. These samples provided soil inorganic nitrogen (SIN) supply at 3 depths (0-15-cm, 15-30-cm, and 30-60-cm), with the addition of potentially mineralizable nitrogen (PMN) at the 0-15-cm depth. Nitrogen values were converted into kg/ha to align with grain protein data. SIN was aggregated using the trapezoidal rule [9] to achieve SIN values from 0-60-cm. NDVI segmentation of plots and alignment of data frames on 'PlotID' allowed for the geolocation of sampled plots. Ordinary Krigging of soil sampled values using the 'gstat' and 'likfit' [10], [11], [12] provided spatial interpolation of soil nitrogen values within the trial. Structural Equation Modeling was implemented in R using the 'lavaan' package [13], [14] statistical modeling tools.

Results

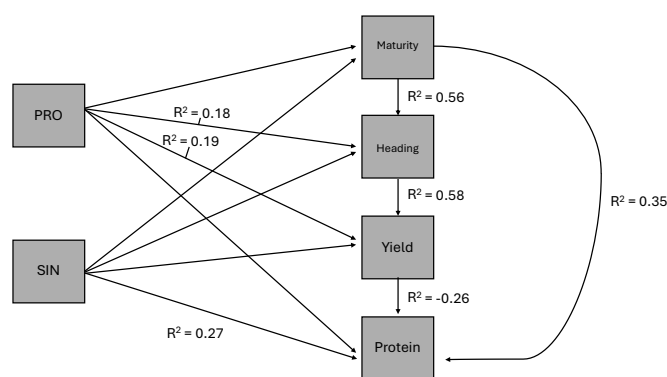


Figure 1: SEM Model of study

Table 1: Fit Indices of SEM Model

Fit Index	Value
Comparative Fit Index	0.997
RMSEA	0.048
SRMR	0.009

Summary

This study investigated the main drivers of grain protein formation in *Triticum durum* from the results of a yield plot trial. Structural Equation Modeling uncovered that diligent measurements of physiological events throughout the growing season paired with background knowledge on the nutrient history of the site can provide insight to what drives grain protein formation.

Acknowledgments

We would like to thank the ACTIVATE team for supplying great amounts of data and support through this project. We have great gratitude for my supervisor, Kwabena Nketia, for his invaluable guidance through these analyses. We would like to acknowledge all those funding this research.

References

- [1] J. M. Alston and P. G. Pardey, "Are Ideas Really Getting Harder to Find?," *Staff Pap. Ser.*, Apr. 2022, [Online]. Available: <https://ideas.repec.org/p/ags/umaesp/320517.html>
- [2] Aswini M. S. *et al.*, "The Role of Genetics and Plant Breeding for Crop Improvement: Current Progress and Future Prospects," *Int. J. Plant Soil Sci.*, vol. 35, no. 20, pp. 190–202, Sep. 2023, doi: 10.9734/ijpss/2023/v35i203798.
- [3] L. Guardia-Velarde, J. E. Cope, H. Metzler, A. Westerbergh, and M. Weih, "Same with less: a method to reduce destructive sampling to estimate nitrogen use efficiency components using allometric relationships in spring wheat (*Triticum aestivum*)," *Funct. Plant Biol.*, vol. 52, no. 5, p. FP24201, May 2025, doi: 10.1071/FP24201.
- [4] P. D. S. Caligari and J. Brown, "Plant Breeding, Practice," in *Encyclopedia of Applied Plant Sciences*, Elsevier, 2017, pp. 229–235. doi: 10.1016/B978-0-12-394807-6.00195-7.
- [5] M. M. Kostić *et al.*, "Comparative Assessment of Remote and Proximal NDVI Sensing for Predicting Wheat Agronomic Traits," *Drones*, vol. 9, no. 9, p. 641, Sep. 2025, doi: 10.3390/drones9090641.
- [6] L. Prey, Y. Hu, and U. Schmidhalter, "High-Throughput Field Phenotyping Traits of Grain Yield Formation and Nitrogen Use Efficiency: Optimizing the Selection of Vegetation Indices and Growth Stages," vol. 10, Jul. 2020, doi: 10.3389/fpls.2019.01672.
- [7] B. Derksen, "Activating genomics to accelerate climate-smart crop delivery," Crop Development Center. Accessed: Jun. 01, 2026. [Online]. Available: <https://cdc.usask.ca/articles/2024/activating-genomics-to-accelerate-climate-smart-crop-delivery.php>
- [8] Agriculture and Agri-Food Canada, "SKSIDv4: Digital Soil Resource Information for Agricultural Saskatchewan," SKSIS Working Group 2026. Accessed: Jun. 01, 2026. [Online]. Available: <http://sksis.ca>
- [9] T. Hengl *et al.*, "SoilGrids250m: Global gridded soil information based on machine learning," *PLOS ONE*, vol. 12, no. 2, p. e0169748, Feb. 2017, doi: 10.1371/journal.pone.0169748.
- [10] E. Pebesma and B. Graeler, "gstat: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation." p. 2.1-4, Feb. 04, 2003. doi: 10.32614/CRAN.package.gstat.
- [11] R Core Team, *R*. (2025). *R*. [Windows]. Available: <https://www.R-project.org/>
- [12] P. J. Ribeiro Jr and P. Diggle, "geoR: Analysis of Geostatistical Data." p. 1.9-6, May 28, 2001. doi: 10.32614/CRAN.package.geoR.
- [13] J. B. Grace, *Structural Equation Modeling and Natural Systems*, 1st ed. Cambridge University Press, 2006. doi: 10.1017/CBO9780511617799.
- [14] Y. Rosseel, "lavaan: An R Package for Structural Equation Modeling," *J. Stat. Softw.*, vol. 48, no. 2, 2012, doi: 10.18637/jss.v048.i02.