Living on the edge: crop boundary effects on canola and wheat yields in Alberta

Samuel V. J. Robinson[a,*], Lan H. Nguyen[a], Paul Galpern[a]

[a]2500 University Drive NW, Calgary, AB

Abstract

Abstract goes here

Keywords: precision yield; spillover; landscape;

*Corresponding Author

Email addresses: samuel.robinson@ucalgary.ca (Samuel V. J. Robinson), hoanglan.nguyen@ucalgary.ca (Lan H. Nguyen), paul.galpern@ucalgary.ca (Paul Galpern)

1. Introduction

Intensive agricultural production has increased over the last 100 years, and agricultural land now makes up over a third of ice-free land on Earth This has allowed increases in human population and increased (global) stability in production However, this is not without cost, as higher-diversity non-crops are converted to lower-diversity crops, resulting in loss of habitat and overall biodiversity of non-target organisms Maintaining both biodiversity and production in agroecosystems represents a seldom-considered goal of conservationists and agronomists, and hold the potential for win-win scenarios Key to this is the preservation of semi-natural land (SNL), which represents the interface between crops and non-crops within agroecosystems

SNL in and around crops is important for both agricultural production and conservation They are habitat for mobile organisms, and can therefore act as sources of ecosystem services such as pollination or pest control (Albrecht *et al.* (2020); Gardner *et al.* (2021)) They also can create microclimate effects that reduce extreme temperature, trap moisture, and reduce wind speed (Kort (1988)) Unfortunately, most of the research is concentrated in Europe, and they tend to be less-studied on other continents In particular, North American agroecosystems have larger fields, different crop varieties, and agronomic practices, all of which could negate or confound potential effects of SNL

SNL may affect yields at intermediate distances, depending on the spatial scale at which ecosystem services operate Edge effects cause low yields at the edge of crops because of sparse or late seedling emergence, poor microclimate, and competition with weeds At the same time, the centre of large fields will not receive ecosystem services if they decay with distance from edges (Kowalchuk & Jong (1995)) For example, pollination services from central place foragers that nest in SNL but forage in crops drops rapidly with distance Therefore, yield may be maximized at intermediate distances, where the ecosystem services cancel out negative edge effects (e.g. Van Vooren *et al.* (2017)) This suggests a “goldilocks" field size, where negative edge effects are canceled out by ecosystem services

Studies of SNL effects on crop yield also suffer from limited scope (e.g. few crop types) and small sample sizes, limiting inference and reducing generality (but see wheat study) Meta-analyses have been the standard solution to this problem, but precision yield data holds enormous promise. However, its use is limited for several reasons: 1) Lack of standardized formats between equipment types, 2) Sensor calibration required for field-level accuracy, and 3) unfamiliarity with spatial statistics. Ecosystem services can influence both the mean and variability of yield in agroecosystems (Redhead *et al.* (2020)). Typically only averages (means) are considered, but higher stability (lower variance) in yield can also be valuable. There are few studies of yield variability (only those with large datasets), but precision yield data opens up the possibility of modeling within-field variability, as well as average yield.

In this paper we ask: 1. How does crop yield change with distance from the edge of field? 2. Does this depend on type of field edge? 3. Is there an intermediate distance where yield is maximized or variance is minimized?

2. Methods

2.1. Data collection

Precision yield data were collected directly from farmers across Alberta. Farmers were solicited for yield data through local agronomists, and we received 298 field-years of data from 5 growers across a total of 7 years (2014-2020). We converted raw data to a standard csv format using Ag Leader SMS. 72% of the crop types where either wheat (*Triticum aestivum*) or canola (*Brassica napus*), two of the most common crops in rotation in Alberta. The remaining crop types had low replication, so we constrained our analysis to only field-years containing wheat (94) or canola (119). Individual fields contained between 1 to 5 years of data (mean: 2.7), containing a total of 18.4 million data points.

Yield data is collected in rectangles of the same length as the data interval (distance = combine ground speed × interval, typically 1 second) and the same width as the combine header (5-7 m, adjusted to account for overlap with adjacent swaths). We extracted the size of each polygon (m^2), dry yield (tonnes), and the spatial location, and the sequence of collection (1 - end of harvest) Because of the large number of yield rectangles per field (30-800 thousand), we used the centroid of each polygon as its location, treating areal data as point data Seeding and application rates were constant across fields, so we did not consider inputs in our analysis We used dry yield (tonnes of seed/hectare minus recorded crop moisture) as our measure of crop yield Precision yield data can be highly variable and is prone to extreme outliers (both positive and negative), especially when combine ground speeds are low and at the beginning of rows Therefore, we filtered data that lay outside the 95th percentiles of yield within each field, values less than or equal to zero, and values outside of the 95th percentiles of ground speed. Due to the large number of data at each field, we sub-sampled to 50,000 data points per field to reduce computation time.

Field boundaries were automatically digitized using buffers from the yield data locations, then manually checked using satellite imagery from Google Earth and classified land cover data from AAFC. Crop boundaries are flexible, and often change yearly depending on planting and emergence conditions (e.g. flooding during some years). Ephemeral wetlands are flooded during some years, but consist mainly of grasses during dry years, and grass boundaries can change if fields are used for as haying or pasture during crop rotation. This makes accurate and consistent classification of field boundaries difficult. We defined the following general categories for field boundaries: 1. Standard: grassy field edge, staging yard, or road right-of-way (grassy strip typically 5-10m wide) 2. Wetland: permanent wetland; borders are largely unchanged from year-to-year 3. Shelterbelt: permanent windbreaks, shelterbelts, remnant forests, or shrublands 4. Other crop: annual crop or pasture with little or no visible boundary between planted areas 5. Bare: unplanted, fallow, flooded area, temporary wetland (only present for a single season), staging yard, oil and gas equipment, or road without a planted boundary 6. Grassland: permanent seminatural grassland or pasture (not in rotation)

<!-- STANDARD -->

<!-- WETLAND -->

<!-- SHELTERBELT -->

<!-- OTHERCROP -->

<!-- BARE -->

<!-- GRASSLAND -->

2.2. Analysis

Yield data from each field contains both random and systematic errors that must be modeled in order to reveal underlying changes in yield. We fit an additive model of the effect of boundary distance on crop yield while accounting for within-field spatial variation and temporal variation in the data. Crop yield can vary within a field due to soil conditions, moisture, seeding rates, herbicide application, and previous agricultural practices. Ground speed is extremely important to yield monitor accuracy (Arslan & Colvin 2002), with low ground speed registering higher yields. Sensor calibration can reduce combine-level bias (such as a combine recording consistently higher/lower yields across fields), this does not address sensor drift that occurs over time. This may be caused by sensors accumulating debris during harvest (especially in canola, pers. comm. Trent Clark), leading to changes in accuracy and bias over time. To model this, we fit the following model to each field-year of data:

$$
\begin{aligned}
\sqrt{yield} \sim & Normal(\mu, \sigma) \\
\mu = & Intercept + log(\text{Polygon Size}) + f(\text{Edge Distance}, b = 12)_i + \\
& f(\text{Easting}, \text{Northing}, b = 60) + f(\text{Sequence}, b = 60) \\
log(\sigma) = & Intercept + log(\text{Polygon Size}) + f(\text{Edge Distance}, b = 12)_i + \\
& f(\text{Easting}, \text{Northing}, b = 60) + f(\text{Sequence}, b = 60)
\end{aligned} \tag{1}
$$

- where:

1. Polygon Size = distance traveled × width of header bar (m^2)
2. Edge Distance = distance from field edge type i (m)
3. Easting, Northing = distance from centre of field (m)
4. Sequence = order of harvest within field (1–N points)
5. f(x,b) = penalized thin-plate regression spline, where x is the predictor and b is the number of basis dimensions

In addition to modeling edge effects (our variable of interest), this model also accounts for a) differences in harvested area, b) within-field spatial variation not related to edges, and d) shifts in combine accuracy during harvest. Spatial or temporal variation can be modelled using a Gaussian Process Model (e.g. Kriging, SPDE Approximations) but this was computationally unfeasible with 50,000 data points per field. Penalized splines offer a compromise, as they account for nonlinear “wiggly" relationships in the same way as Gaussian processes but using a reduced number of dimensions and computation time. The number of basis dimensions was checked with the *gam.check* function from *mgcv*. The relationship between polygon size and yield was modeled with a log-linear relationship with a single slope term, as this closely matched smoothed versions. Because polygon size is not intuitive, we converted it into ground speed by regressing speed on distance traveled per polygon, then using this to back-calculate the ground speed for a given polygon size (given a fixed width of header). All models were fit in *R* using the *mgcv* library (version 1.8.36, Wood 2017), and figures were created with *ggplot2* and *ggpubr* (versions 3.3.3, Wickham 2016; and 0.4.0, Kassambara 2020).

Despite the use of spatial and temporal smoothers, the residuals in most of the models displayed a large amount of spatial and temporal autocorrelation. *Not really sure how we can get around this. In practice, this means that SEs for all coefficients are too small (i.e. we gathered 1000 points, but because* *they're spatially related, it's only equivalent to 100 points).*

To consider results from all field-level models together, we fit models independent of each other and fit an "overall" smoother as averages of the field-level smoothers. However, this does not account for uncertainty in the field-level smoothers, so we used an approach similar to bootstrapping of hierarchical mixed effects models: 1. Extract single posterior sample (*rnorm* in R) of smoother parameters from each field-level model using coefficient estimates and standard error 2. Use posterior sample to create a new smoother from each field 3. Fit new model of new smoothers from all fields, and save this "meta-smoother" 4. Repeat 1000 times, and calculate coverage intervals (5-95% percentiles) on saved meta-smoothers This gives coverage intervals (CIs) for the "average" smoother while accounting for field-level variability, and is conceptually similar to a meta-analysis. Spatial variability was not averaged between fields because differences in field geometry make it impossible to compare fields.

3. Results

3.1. Field-level example

The filtered data from each field still contained a large amount of variation, but the non-stationary additive models were able to reveal underlying yield patterns. To illustrate this, we show the results from a single wheat field, where we show the relative impact of ground speed, harvest sequence, boundary distance, and spatial variation on both yield mean and yield SD ("patchiness"). Figure 1 shows the raw data collected from a single field, demonstrating a large amount of variation even among the filtered data. Combine ground speed had a large effect on both yield mean and SD, and there was also a strong effect of harvest sequence on yield mean and SD (Figure 2. Yield mean and SD tended to decrease with distance from field boundaries, but there were no consistent differences between boundary types (Figure 3. Finally, the random spatial smoothers in Figure 4 show that yield mean and SD systematically vary within fields, likely due to moisture and soil conditions (Figure 4.

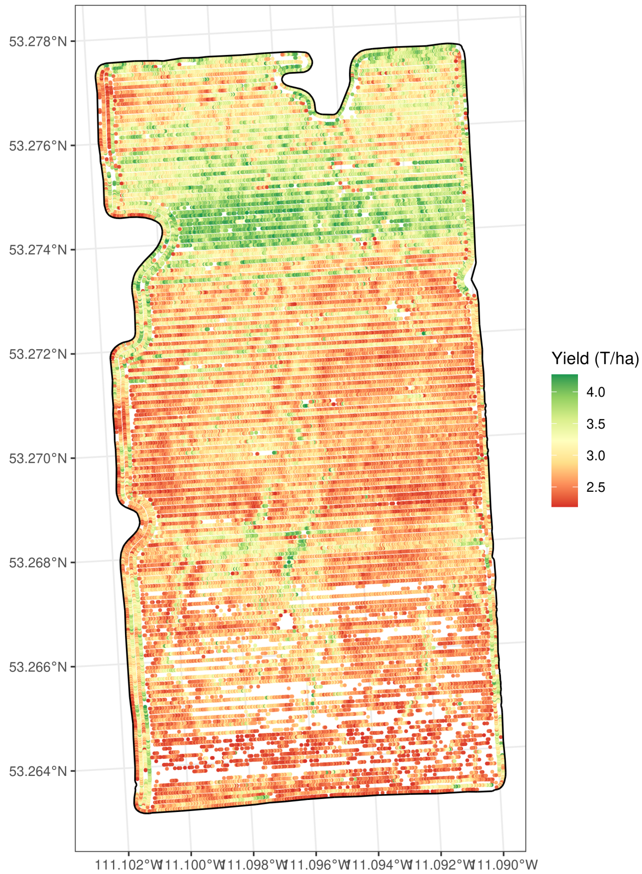

Figure 1: Example output from a single field.

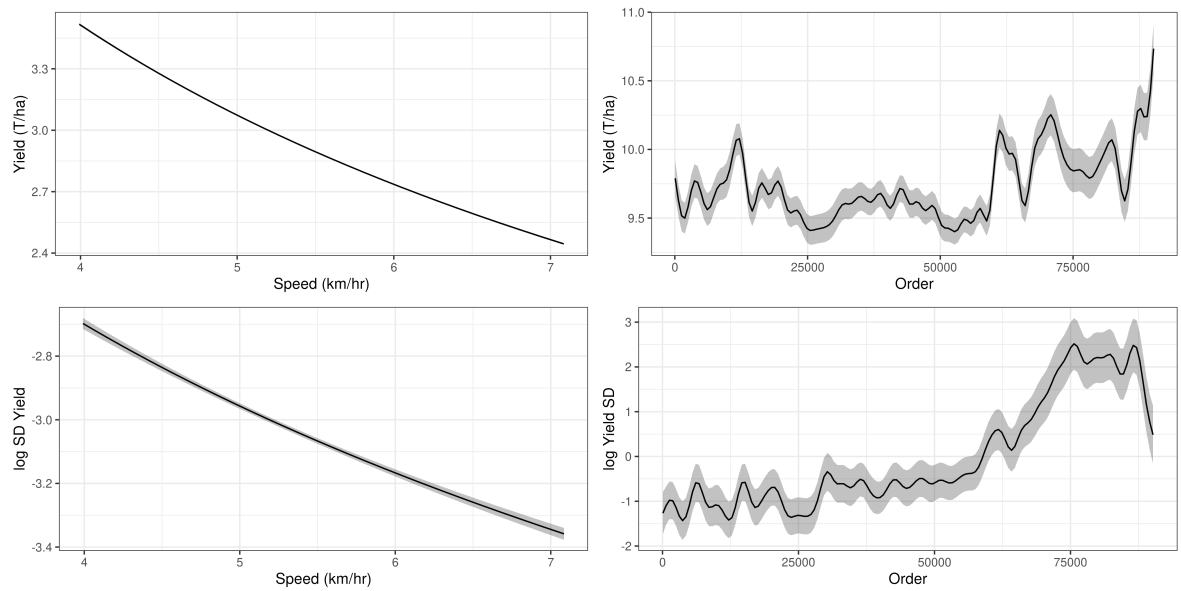

Figure 2: Ground speed and harvest sequence (random temporal variation) smoothers from a single field.

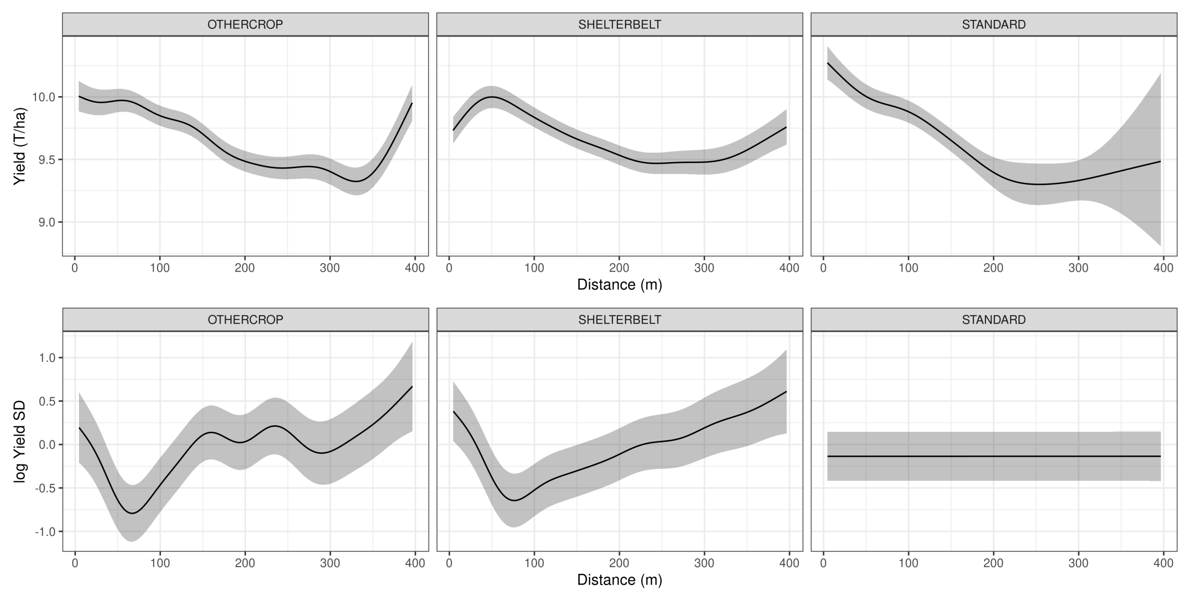

Figure 3: Distance smoothers from a single field.

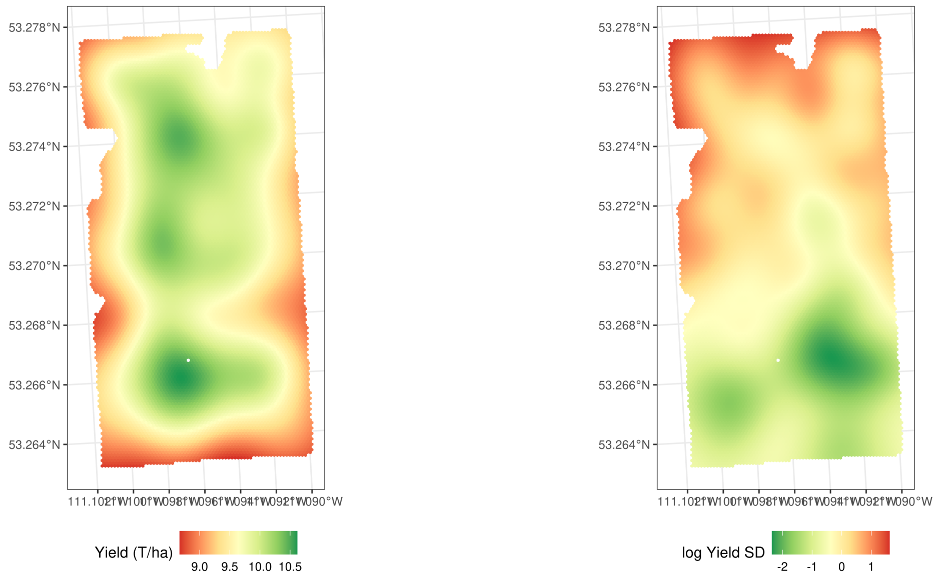

Figure 4: Spatial (random spatial variation) smoothers from a single field.

3.2. Overall results

Across all fields, “Standard” (grassy border) was the most common boundary type, followed by Shelterbelts and Other Crops.

3.2.1. Canola fields

Overall, there was a strong negative effect of combine ground speed on both the average and variability of canola yield. At the field level, both yield average and yield variability (SD of yield) of yield were extremely spatially (average p-values: 2e-16, 1.77e-05) and temporally dependent (2e-16, 0.000243), meaning that average and variability changed with location in the field, as well as sequentially during the harvest. However, there was no difference in average or variability with harvest sequence when all fields were considered, meaning that combines did not appear to distort yield data from start to finish in any systematic way.

Average canola yield tended to increase with distance from Standard and Other Crop boundaries, leveling off at about 25 m (Figure 5 top row, Other crop and Standard panels); yield was approximately 0.1 T/ha (1.5 bu/ac) lower at the edge for both of these boundary types. Interestingly, Shelterbelts were well-represented in the data (Table **??**) but the negative edge effect was not strong (Figure 5 top row, Shelterbelt panel), indicating that they may provide better micro-climates for crops located near them. Other boundary types were less common in the data and had correspondingly weaker effects on mean yield, but all showed a small negative edge effect. Yield variability tended to decrease with distance from the edge of the field, indicating that edge effects contribute to yield “patchiness” as well as averages. At Standard, Bare, and Wetland field boundaries, variability decreased until about 100 m, after which variablity was constant, while the pattern was less consistent in Grassland, Other Crop, and Shelterbelt.

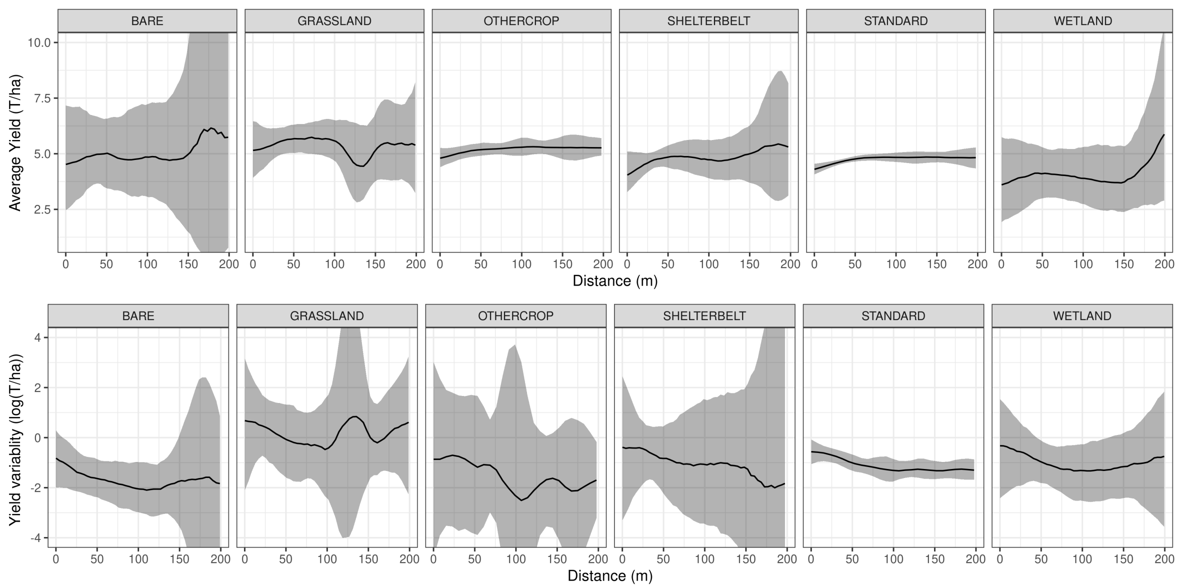

Figure 5: Field boundary effect on canola yield, accounting for the effect of combine speed, spatial variation, and harvest sequence. Upper panel represents mean yield, while the lower panel represents yield variation (i.e. "patchiness").

3.2.2. Wheat fields

Similar to canola, there was a strong negative effect of ground speed on both yield and variability of wheat yield (Figure 9, left panels). At the field level, both yield average and yield variability (SD of yield) of yield were spatially (average p-values: 2e-16, 2e-16) and temporally dependent (2e-16, 2e-16), but harvest sequence did not distort the average or SD of wheat yield in a systematic way (Figure 9, right panel).

Average wheat yield increased slightly with distance from the field boundary, but this was much more variable than the patterns seen in canola yield (Figure 6). There was also a slight increase in variability away from the field boundary, at least for Standard and Other Crop boundaries, but this was inconsistent among other boundary types.

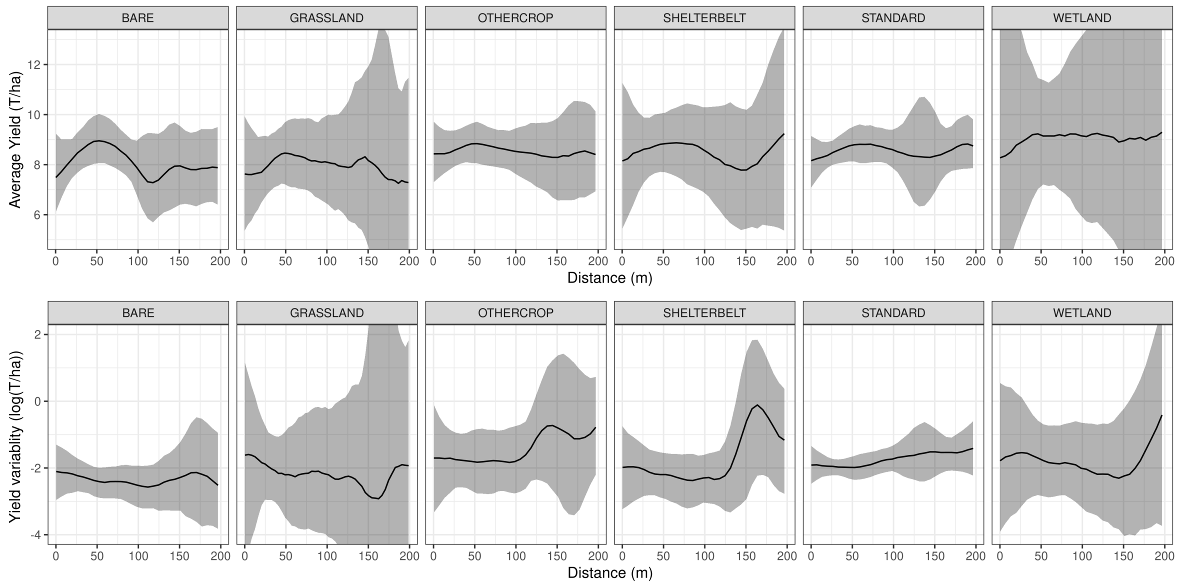

Figure 6: Field boundary effect on canola yield, accounting for the effect of combine speed, spatial variation, and harvest sequence. Upper panel represents mean yield, while the lower panel represents yield variation (i.e. "patchiness").

4. Discussion

- Summary:
  - At the level of an individual field, there were

5. Author Contributions

SVJR and PG conceived of the project. SVJR collected data, conducted analysis, and wrote the manuscript.

6. Acknowledgements

We thank Trent Clark, Dean Hubbard, Alvin French, and Kristina Polziehn for providing yield data for us, and providing insight on their yield data. We also thank Autumn Barnes and Keith Gabert of the Canola Council of Canada, and Laurel Thompson and JP Pettyjohn of Lakeland College for connecting us with farmers and agronomists. Funding for this research was provided by Ducks Unlimited Canada's Institute for Wetland and Waterfowl Research, the Alberta Canola Producers Commission, Manitoba Canola Growers Association, SaskCanola, the Alberta Biodiversity Monitoring Institute, and the Alberta Conservation Association.

References

Albrecht, M., Kleijn, D., Williams, N.M., Tschumi, M., Blaauw, B.R., Bommarco, R., *et al.* (2020). The effectiveness of flower strips and hedgerows on pest control, pollination services and crop yield: A quantitative synthesis. *Ecology Letters*.

Arslan, S. & Colvin, T.S. (2002). An evaluation of the response of yield monitors and combines to varying yields. *Precision Agriculture*, 3, 107–122.

Gardner, E., Breeze, T.D., Clough, Y., Smith, H.G., Baldock, K.C.R., Campbell, A., *et al.* (2021). Field boundary features can stabilise bee populations and the pollination of mass-flowering crops in rotational systems. *Journal of Applied Ecology*, In press.

Kassambara, A. (2020). *ggpubr: 'ggplot2' based publication ready plots.*

Kort, J. (1988). 9. Benefits of windbreaks to field and forage crops. *Agriculture, Ecosystems & Environment*, 22-23, 165–190.

Kowalchuk, T.E. & Jong, E. de. (1995). Shelterbelts and their effect on crop yield. *Canadian Journal of Soil Science*, 75, 543–550.

Redhead, J.W., Oliver, T.H., Woodcock, B.A. & Pywell, R.F. (2020). The influence of landscape composition and configuration on crop yield resilience. *Journal of Applied Ecology*.

Van Vooren, L., Bert, R., Steven, B., Pieter, D.F., Victoria, N., Paul, P., *et al.* (2017). Ecosystem service delivery of agri-environment measures: A synthesis for hedgerows and grass strips on arable land. *Agriculture, Ecosystems & Environment*, 244, 32–51.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York.

Wood, S.N. (2017). *Generalized additive models: An introduction with R.* CRC press.

Appendix A: Supplementary Material

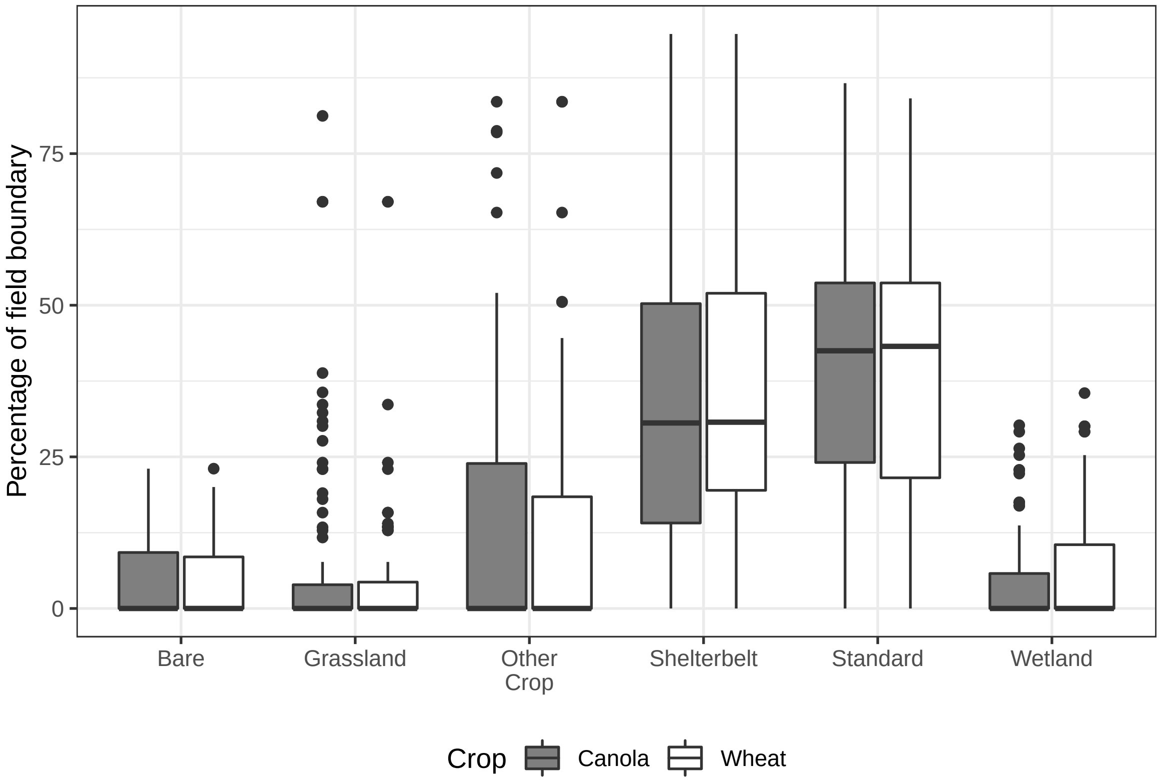

Figure 7: Percentage of field boundaries across sampled fields.

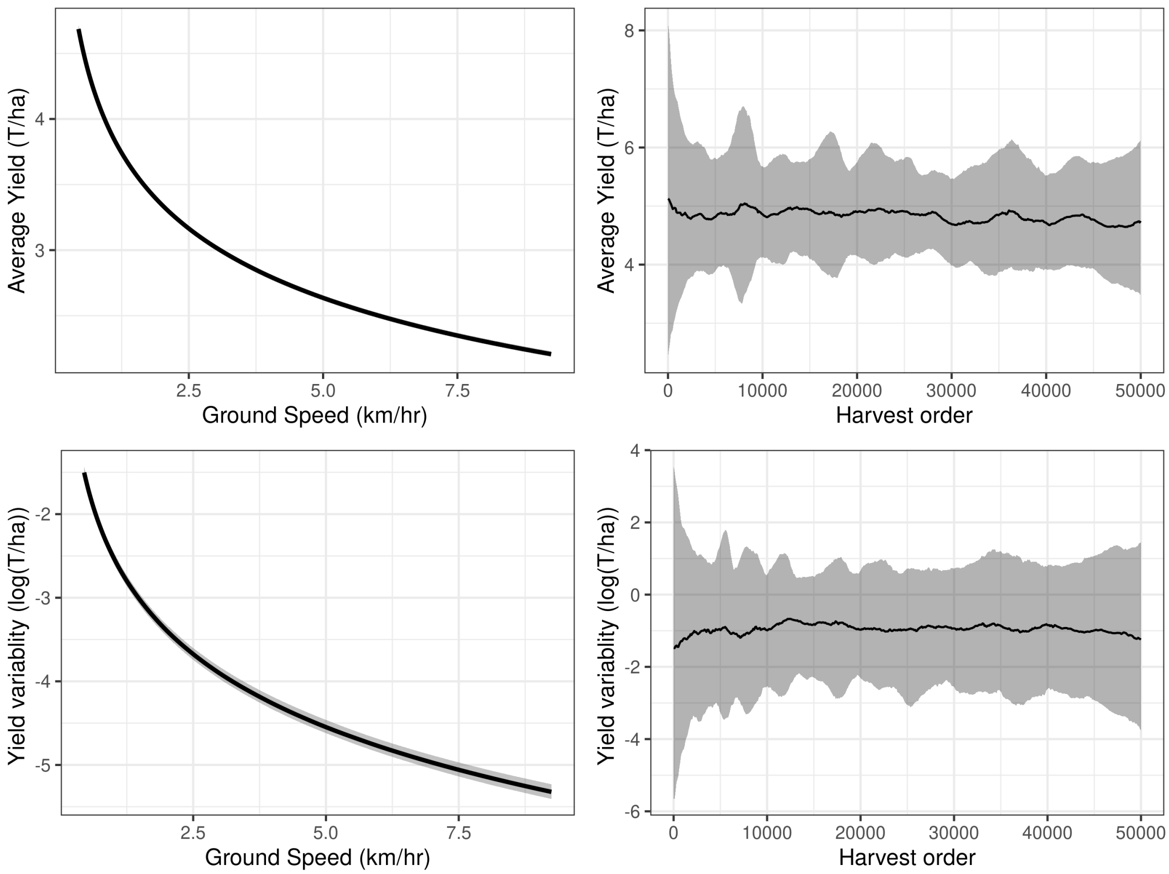

Figure 8: Effect of ground speed and order (sequence of harvest) on canola yield.

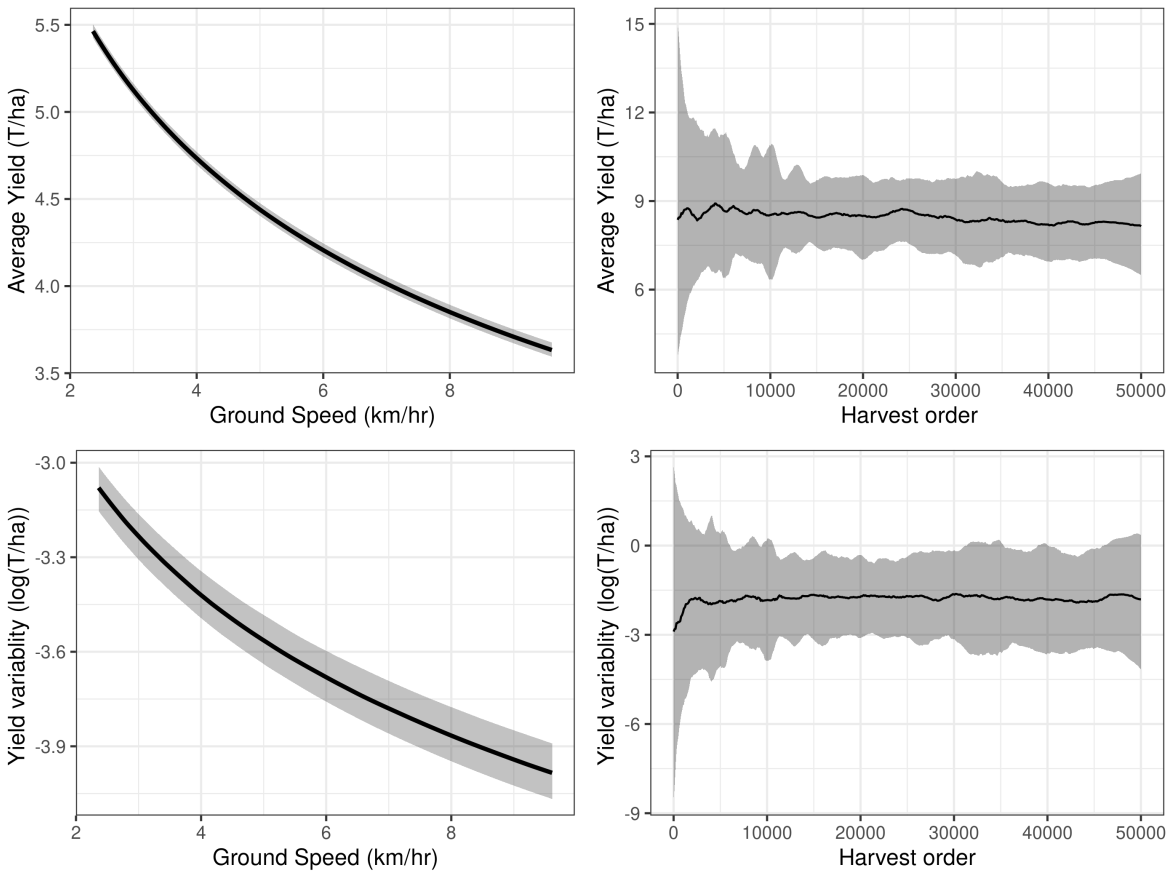

Figure 9: Effect of ground speed and order (sequence of harvest) on wheat yield.