Edge effects 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: beetles; spiders; harvestmen;

*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 allows 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
  - They also can create microclimate effects that reduce extreme temperature, trap moisture, and reduce wind speed
  - 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, and different varieties and agronomic practices, all of which could negate 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
  - 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

  - 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)

  - For this reason, large-scale precision yield data holds enormous promise for agronomy, as it allows
  - 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
  - Typically only averages (means) are considered, but higher stability (lower variance) in yield can also be valuable
  - Size examples: wheat study from the UK
  - 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 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 were poorly replicated in our sample, 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)
    - 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 after accounting for crop moisture) as our measure of crop yield
    - 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)
    - Additionally, seminatural features often change yearly
        * Ephemeral wetlands are flooded during some years, but consist mainly of grasses during dry years
        * 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)

2.2. Analysis

- At each field, 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 combine yield monitor
  - 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
  - While 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 within fields
  - This may be caused by sensors accumulating debris during harvest (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:

$$
\begin{aligned}
sqrt(yield) \sim & Normal(\mu, \sigma) \\
\mu = & Intercept + log(PolygonSize) + f(DistancefromEdge_i, b = 12)_i + \\
& f(Easting, Northing, b = 60) + f(Sequence, b = 60) \\
log(\sigma) = & Intercept + log(PolygonSize) + f(DistancefromEdge, b = 12) + \\
& f(Easting, Northing, b = 60) + f(Sequence, b = 60)
\end{aligned} \tag{1}
$$

  - where:

1. Polygon Size = distance traveled × width of header bar (m^2)
2. Distance from Edge = 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 combine speed, b) within-field spatial variation not related to edges, and d) shifts in combine accuracy during harvest
    - Spatial or temporal variation is typically modelled using a Gaussian Process Model (Kriging) or approximations such as Stochastic Partial Differential Equations (e.g. INLA), but this was computationally infeasible 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 with substantially reduced computation time
    - The number of basis dimensions was checked with the *gam.check* function from *mgcv*
    - The relationship between polygon size (i.e. ground speed) and yield was modeled with a log-linear relationship with a single slope term, as this closely matched smoothed versions
    - 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).
- To consider results from all field-level models, we fit models independent of each other, and “overall” smoothers were taken 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 new simulated smoother from each field
3. Fit new model of simulated 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

3. Results

Results here

4. Discussion

Discussion here

5. Authors' contributions

Author's contribution

6. Acknowledgements

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

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

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

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

Supplemental materials here