Livin' on the edge: Precision yield data shows evidence of ecosystem services from field boundaries

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

[a]*2500 University Drive NW, Calgary, AB*

Abstract

Field boundaries can provide ecosystem services to crops by creating better abiotic conditions for crop growth, and can also act as habitat for beneficial arthropods. This suggests that crop boundaries may create an intermediate hump-shaped increase in crop yield, where negative edge effects are cancelled out by increased ecosystem services from the field boundary. However, there is little large-scale evidence showing this, largely because plot-scale crop yields are costly and time-consuming to measure. Precision yield data from combine yield monitors has huge potential in this respect, as the equipment is widespread and data is frequently recorded by growers. In this study, we used 252 field-years of yield monitor data from three common crops – wheat (*Triticum aestivum*), canola (*Brassica napus*), or peas (*Pisum sativum*) – recorded across Alberta, Canada, and examined how yield varied with distances from common crop boundary types. Average crop yield tended to increase with distance from crop boundaries before plateauing at about 50 m, and yield variation (SD) tended to decrease with distance. There was evidence of an intermediate increase in yield for wheat away from shelterbelts, and a weak increase in canola, but this was not seen for other crop types or boundary types. This study represents one of the first uses of precision yield data to measure ecosystem service provision at large spatial scales.

Keywords: precision yield; spillover; landscape; shelterbelt;

*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 (Ramankutty *et al.* 2018). This has allowed increases in human population and increased global stability in production (Foley *et al.* 2011). 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 goal of conservationists and agronomists, and holds the potential for win-win scenarios ("multifunctionality" *sensu* Mitchell *et al.* 2014; Frei *et al.* 2018; but see also Karp *et al.* 2018). Key to this goal is the preservation of semi-natural land (SNL) in and around crops, which represents the interface between crops and non-crops within agroecosystems.

SNL in agroecosystems is important for both agricultural production and conservation (Prieto-Benítez & Méndez 2011), and can create better biotic and abiotic conditions for crop growth (Case *et al.* 2019; Weninger *et al.* 2021; but see Lowe *et al.* 2021). SNL acts as habitat for mobile organisms, and can act as a source of ecosystem services, such as pollination or pest control (Albrecht *et al.* 2020; Gardner *et al.* 2021), but this depends on both the crop system (Gray & Lewis 2014; Quinn *et al.* 2017) and landscape context (Marja *et al.* 2019; Boetzl *et al.* 2020). SNL can also create microclimate effects that reduce extreme temperature, trap moisture, and reduce wind speed and soil erosion (Kort 1988; Brandle *et al.* 2004; Wratten *et al.* 2012). Field boundaries (e.g. grass and flower strips, tree or shrub shelterbelts, road right-of-ways) are SNL of particular interest, as they are extremely common and require less land than larger agricultural set-asides.

Semi-natural field boundaries may cause positive or negative effects on crops, depending on the spatial scale at which ecosystem services operate. Low yields near may occur near the edge of the field because of patchy seed emergence, poor microclimate, shading from trees, or greater competition with weeds (Figure 1, Scenario 2). Alternatively, yield may be higher near field boundaries (Scenario 3) if boundaries provide helpful ecosystem services (Duelli & Obrist 2003). For example, crop pollination services from bees that nest in field boundaries drops rapidly with distance (Garibaldi *et al.* 2011) and wind speed is reduced directly downwind of treelines (de Jong & Kowalchuk 1995), so positive effects will be concentrated close to the boundary. Depending on the rate at which positive and negative boundary effects decline with distance, another scenario is also possible. Yield may be maximized at intermediate distances (Scenario 4), suggesting a "Goldilocks" field size where negative edge effects are canceled out by ecosystem services (Kowalchuk & de Jong 1995; Helmers & Brandle 2005; Mitchell *et*

al. 2014; Van Vooren *et al.* 2017).

Scenario 1

Scenario 2

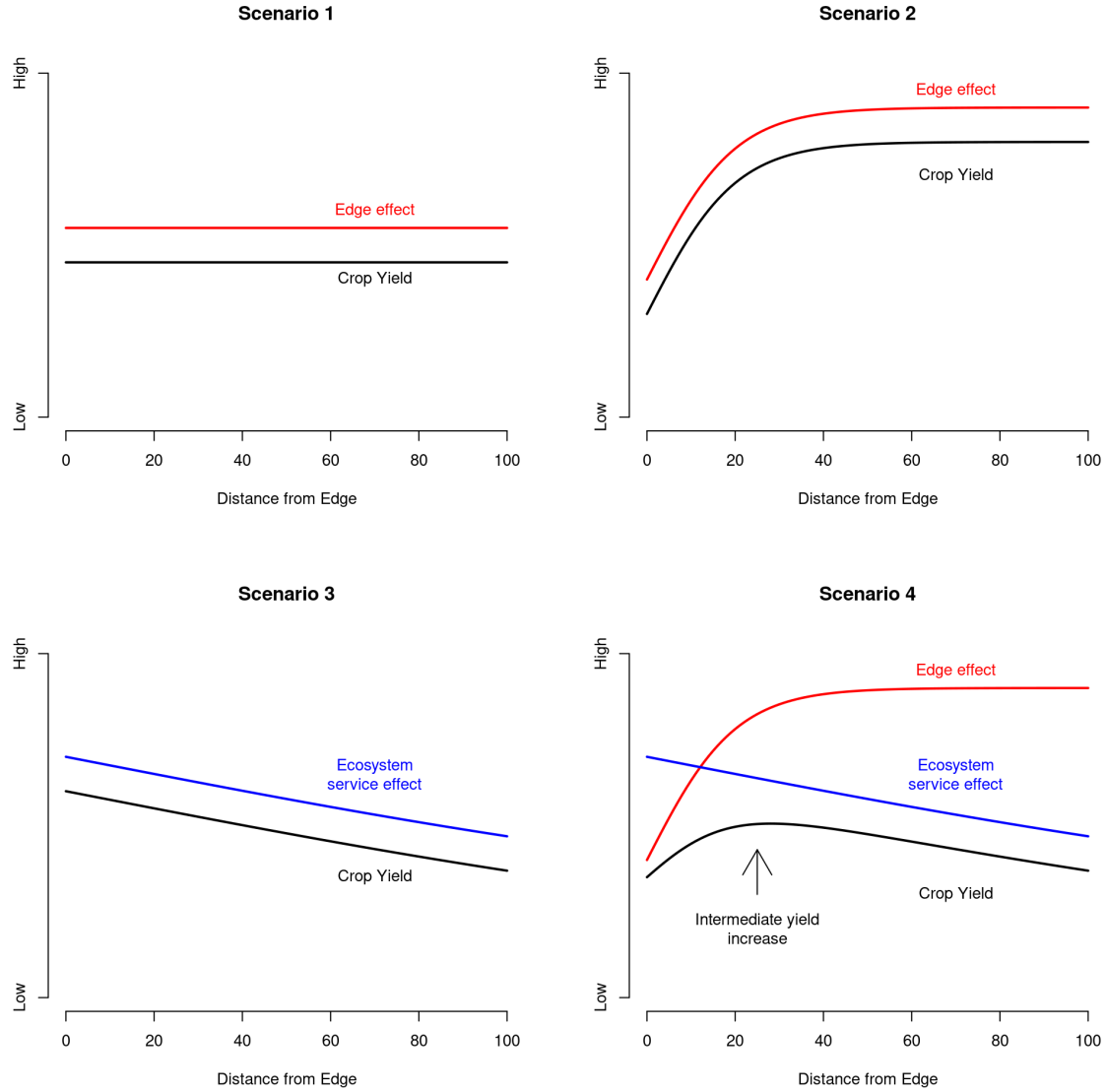**Scenario 3**

Scenario 4

Figure 1: Potential yield patterns, depending on ecosystem service effects with distance. Scenario 1: no effect of boundary. Scenario 2: negative edge effect on yield. Scenario 3: positive ecosystem service effect on yield. Scenario 4: edge effects are shown in red, ecosystem service effect shown blue, leading to an intermediate peak in yield.

Unfortunately, data to inform these decisions are often minimal, as observing boundary effects requires large amounts of expensive and time-consuming crop yield data. This means that studies of SNL effects on crop yield tend to suffer from limited spatial scope and small sample sizes, reducing statistical power and generality (Brandle *et al.* 1984; Baker *et al.* 2018; but see also Redhead *et al.*

2020). Meta-analyses are a common solution to this problem (e.g. Albrecht *et al.* 2020; Lowe *et al.* 2021), but precision yield data holds enormous promise because of its relatively common usage and low operating costs (Steel 2017; Osorio *et al.* 2018). However, its use is limited for several reasons, namely 1) a lack of standardized formats between equipment types, 2) sensor calibration required for field-level accuracy, and 3) low familiarity with spatial statistical methods. Ecosystem services from field boundaries may influence both the mean as well as the variability of yield in agroecosystems (Redhead *et al.* 2020; Hünicken *et al.* 2021); typically only means are considered, but lower variability in yield can also be valuable from a grower's perspective. 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 use a precision yield dataset from Alberta, Canada to ask the following questions: 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? To our knowledge, only (Osorio *et al.* 2018) has studied the effects of field boundaries using precision yield data, making this study one of the first of its kind.

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®. 85% of the crop types where either wheat (*Triticum aestivum*), canola (*Brassica napus*), or peas (*Pisum sativum*), three of the most common crops in rotation in Alberta (Statistics Canada 2014). The remaining crop types had low replication, so we constrained our analysis to only field-years containing wheat (94), canola (119), or peas (39), for a total of 252 field-years of data. Individual fields ranged from 5.9 to 369 ha in area (mean: 72.2 ha), and contained between 1 and 6 years of data (mean: 3.1), for a total of 14.4 million data points.
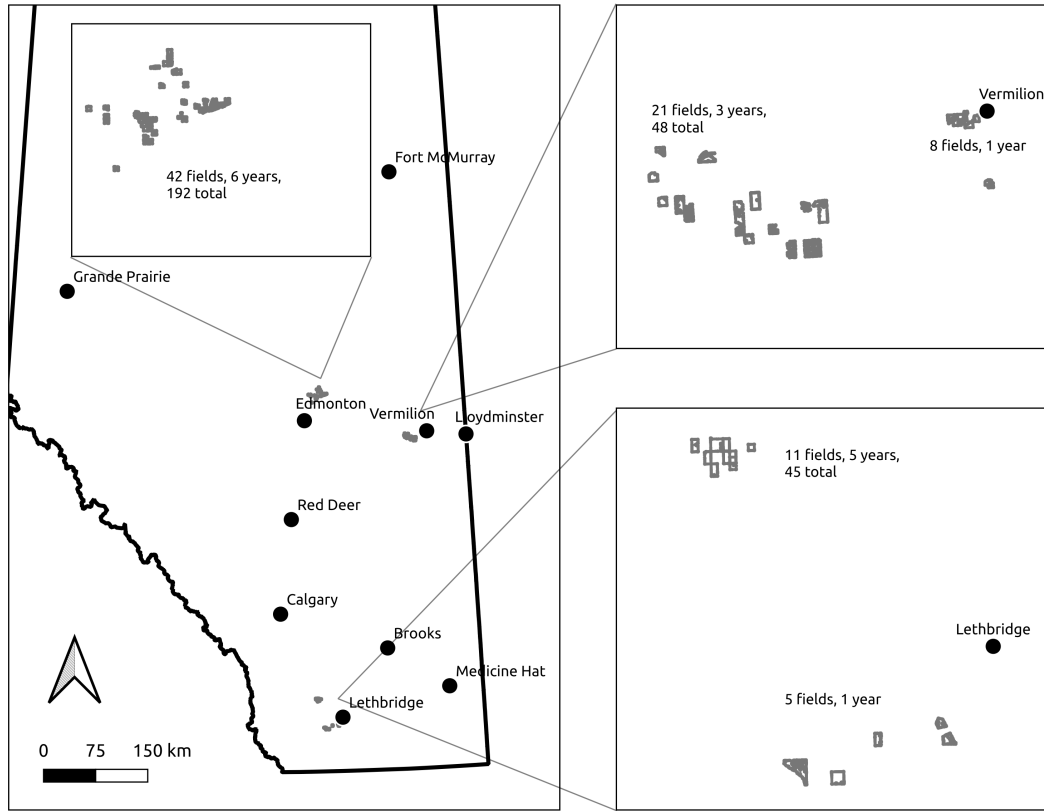
Figure 2: Locations of fields that data were collected from

⁶⁹   Yield data was collected in rectangles of the same length as the data interval (distance = combine

⁷⁰ ground speed × interval, typically 1 second) and the same width as the combine header (5-7 m, adjusted

⁷¹ by the combine to account for overlap with adjacent swaths). We used dry yield (tonnes of seed per

⁷² hectare minus - crop moisture) as our measure of yield, using the size of each rectangle (m^2), dry yield

⁷³ (tonnes), and the spatial location. Because of the large number of yield rectangles per field (30-800

⁷⁴ thousand), we used the centroid of each polygon as its location, treating areal data as point data.

⁷⁵ Seeding and application rates were constant across fields, so we did not consider inputs in our analysis.

⁷⁶ Precision yield data can be highly variable and is prone to extreme outliers, especially when combine

⁷⁷ ground speeds are low and at the beginning of rows (Griffin *et al.* 2007; Whelan & Taylor 2013), or

⁷⁸ when the combine is accelerating or decelerating (Arslan & Colvin 2002). Therefore, we used the

⁷⁹ following filtering procedure to clean the data prior to model fitting (see Appendix A for R code):

1. Removed outliers and "inliers" using the spatial filtering procedure from (Vega *et al.* 2019)

2. Removed unrealistic yield values (10.75 T/ha for wheat, 8 T/ha for canola and peas), and values of zero or less

3. Removed extremely low and high outliers (below 1st percentile, above 99th percentile) for: dry yield, steering angle differences, ground speed

4. Removed points with large changes (>20%) in ground speed

5. Removed extremely low and high changes in position, indicating that the combine had stopped or moved to another part of the field between data collection

6. Sub-sampled filtered data to 50,000 data points per field (reduce computation time)

Field boundaries were digitized using buffers from the yield data locations, then manually checked using satellite imagery from Google Earth and classified land cover data (Agriculture and Agri-Food Canada 2018). Crop boundaries are flexible, and often change yearly depending on planting and emergence conditions. For example, 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, but we used 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

Yield data from each field contains both random and systematic errors that must be modeled in order to reveal underlying changes in yield. Crop yield can vary within a field due to soil conditions, moisture, seeding rates, herbicide application, previous agricultural practices, and other factors not directly related to boundary distance. To account for this, we used additive models of the effect of boundary distance on crop yield while accounting for within-field spatial variation.

We fit the following model to each field-year of data:

$$\sqrt{Yield} \sim Normal(\mu, \sigma)$$

$$\mu = Intercept + f(\text{Boundary Distance}, b = 5)_i +$$

$$f(\text{Easting}, \text{Northing}, b = 80) \tag{1}$$

$$log(\sigma) = Intercept + f(\text{Boundary Distance}, b = 5)_i +$$

$$f(\text{Easting}, \text{Northing}, b = 80)$$

- where:

1. Yield = weight of crop gathered per area of harvest (T/ha)
2. Boundary Distance = distance from field boundary type i (m)
3. Easting, Northing = distance from centre of field (m)
4. 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 within-field spatial variation not related to edges. Spatiotemporal variation can also 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 relationships in the same way as Gaussian processes using a reduced number of dimensions. We used 5 basis dimensions for the distance smoother and 80 for the spatial smoothers, as we hypothesized that boundary effects would not be as complex as within-field variation (as in Figure 1). 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).

The residuals in most of the models displayed both spatial autocorrelation in the residuals, indicating "micro-scale" variation that was not captured by the spatial smoothers (Cressie & Wikle 2011; Wikle *et al.* 2019). This is extremely common in precision yield data, and can bias standard errors of regression terms, but because a) there were a large number of data points per field, b) standard errors were relatively small (i.e. extremely small p-values), and c) results did not change with the addition of more basis functions (120 rather than 80), we decided to leave the models as-is. Temporal autocorrelation in residuals was also present, but we did not include temporal smoothers because spatial and temporal processes are confounded in precision yield data. That is, combining of fields occurs in patterns

7

(e.g. east-to-west) so we cannot tell whether changes in yield in certain areas of the field were caused by deviations in the yield monitor sensors over time, or whether they represent changes in growing conditions. We considered the latter to be a more realistic assumption, so we assigned all random variation as spatial in nature. Non-standard combine driving patterns may be able to gather data in a way that can detect sensor drift, but this is beyond the scope of this research.

To consider results from all field-level models together, we fit models of each field-year independently and then fit overall smoothers as averages of the field-level smoothers. This does not account for uncertainty in the field-level smoothers, so we used an approach similar to bootstrapping:

1. Extract single posterior sample (*rnorm* in R) of smoother parameters from each field-level model using coefficient estimates and standard errors.

2. Use posterior sample to simulate a new smoother from each field.

3. Fit new model of simulated smoothers from all fields (with a field-year random intercept), 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 within-field variability is not directly comparable between fields.

3. Results

3.1. Field-level example

The filtered data from each field 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 boundary distance and spatial variation on both yield mean and yield SD ("patchiness"). Figure 3a shows the filtered data collected from a single field, demonstrating a large amount of variation even after filtering. Yield mean tended to increase with distance from field boundaries while yield SD tended to decrease, but there were no consistent differences between boundary types (Figure 4). Finally, the random spatial smoothers in Figure 3b and c show that yield mean and SD systematically vary within fields, likely due to moisture, elevation, or soil conditions.
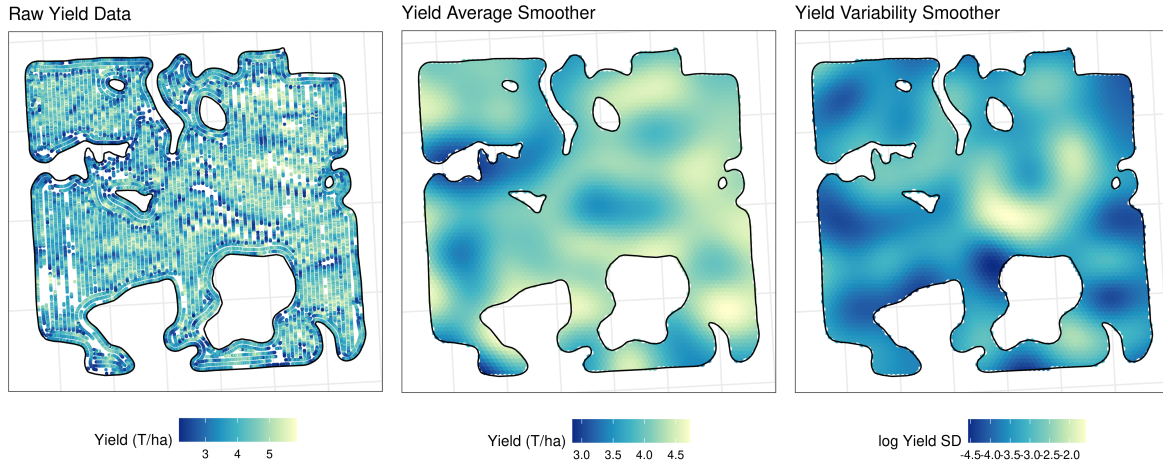
8

Figure 3: Raw data and spatial smoothers from a single field. Yield averages (means) and variability (log SD) were modeled separately, and both show large spatial dependence within the field. Field dimensions are approximately 800 x 800 m, but coordinates are hidden to protect data privacy.
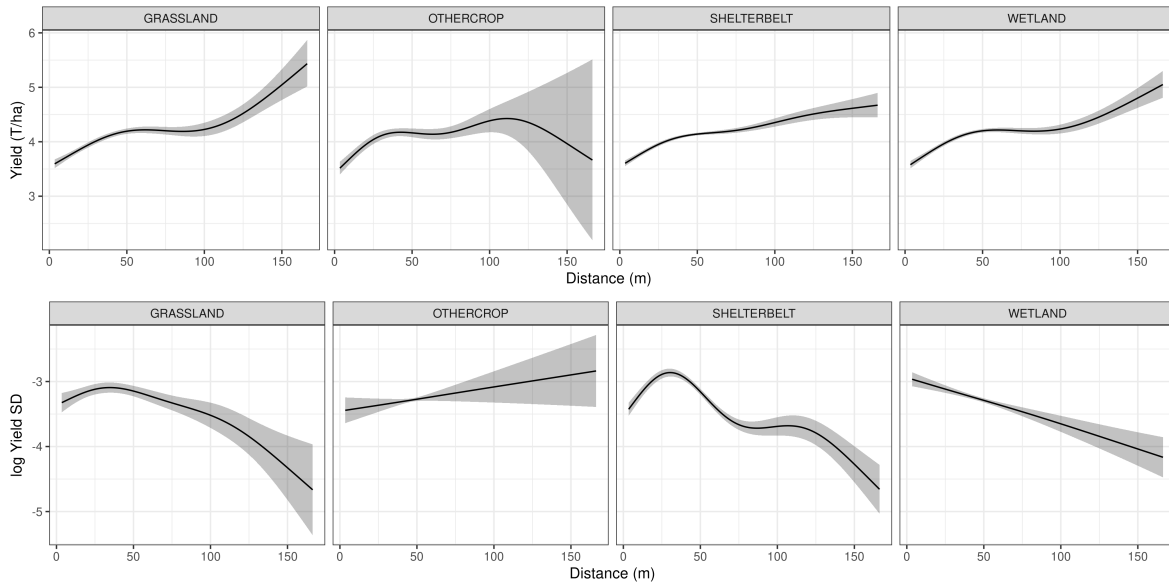

Figure 4: Distance smoothers from a single field, showing a positive saturating effect of distance on average yield (first row), while the variability smoothers show a general decrease in yield SD (second row).

3.2. Overall results

Across all fields, standard field boundaries (grassy borders) were the most common boundary type, followed by shelterbelts and other crops (Figure 5).
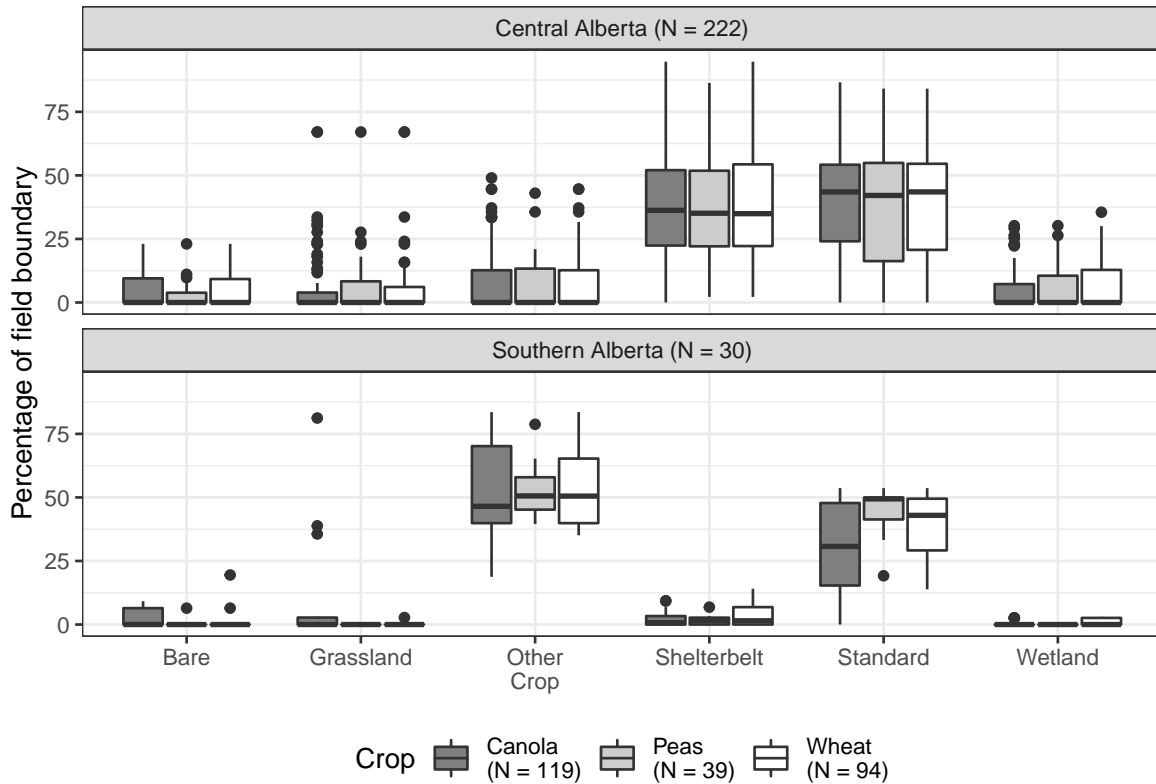
Figure 5: Boxplot of field boundary compositions across sampled fields. The x-axis indicates the boundary type, while the y-axis indicates the percentage of all field boundaries covered by that boundary type.

Overall, both yield average and yield variability were extremely spatially dependent across all fields (average p-values both <0.0001), meaning that within-field variation was the primary driver of yield across canola, wheat, and pea crops (Table 1, row 6). Distances from boundaries were also important, but a lower proportion of the fields had strong effects (rows 1-5, cols 3 + 5).

Average canola yield tended to increase with distance from all crop boundaries, starting at 2.5 T/ha at the field edge and leveling off at about 3 T/ha (Figure 6 top row). Shelterbelt and standard boundaries were well-represented in the data, and average yield quickly increased with distance away from the field boundary, but shelterbelts had a small intermediate maximum at around 50 m before a weak decrease with distance, while average yield away from Standard field boundaries plateaued at 50 m and remained roughly constant with distance. The effect of other boundary types was less consistent,

Table 1: Mean χ^2, p-values, and proportion of smoother p-values less than 0.05 (after Bonneferoni correction) for mean and variability smoothers across all models.

Smoother Type	Mean Yield			Yield Variability		
	χ^2-value	p-value	Prop. <0.05	χ^2-value	p-value	Prop. <0.05
Bare	154.7	0.036	0.877	121.4	0.054	0.860
Grassland	303.1	0.079	0.794	93.2	0.061	0.794
Other Crop	174.9	0.063	0.860	104.5	0.044	0.840
Shelterbelt	328.1	0.014	0.936	137.6	0.058	0.817
Standard	281.9	0.021	0.938	159.6	0.057	0.832
Wetland	148.9	0.045	0.833	106.7	0.013	0.857
Spatial Smoother	17357.5	0.000	1.000	4739.1	0.000	1.000

¹⁷⁴ but average yield tended to increase with distance away from grassland, other crop, and wetland, and

¹⁷⁵ was roughly the same with distance from bare boundary types. Yield variability (Figure 6 bottom row)

¹⁷⁶ was essentially constant with distance for shelterbelt, standard, and other crop field boundaries, but

¹⁷⁷ showed some evidence of decreasing with distance from field boundary at wetlands, bare, and grassland.

¹⁷⁸ Interestingly, there was an intermediate reduction in canola yield variability at about 75 m away from

¹⁷⁹ wetland boundaries, indicating a potential stabilizing effect, but this was offset by the large increase in

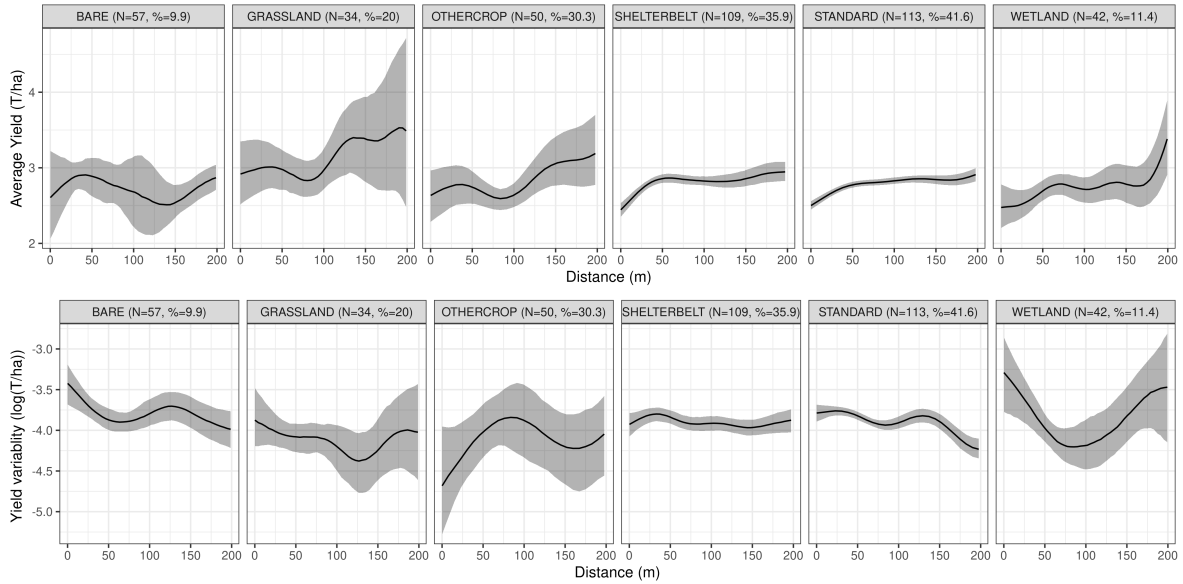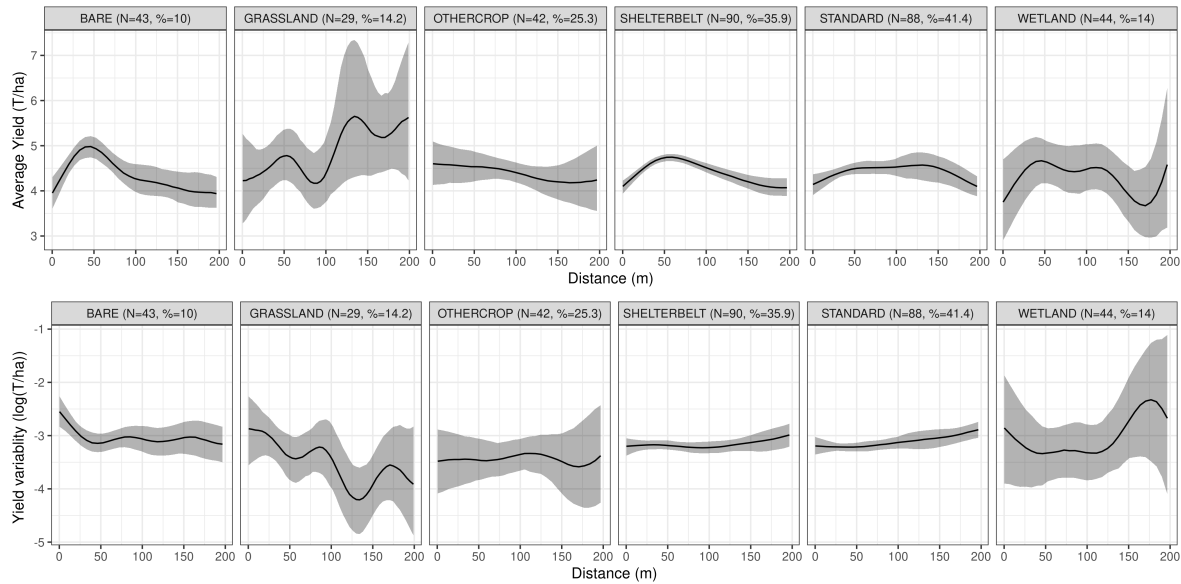¹⁸⁰ yield variability at the edge of wetlands.

Figure 6: Field boundary effect on canola yield, accounting for the effect of spatial variation. Upper panel represents mean yield, while the lower panel represents yield variation. N refers to number of fields containing this boundary type, and % refers to the average percentage of field boundary accounted for by this boundary type.

¹⁸¹ The effect of boundary distance on average wheat yield was much more variable than the patterns

11

seen in canola yield. Average yield (Figure 7, top row) roughly increased with distance away from wetland, standard, and grasslands, but both shelterbelts and bare ground displayed an intermediate maximum at around 50m (similar to canola), and decreased slightly with distance from Other Crops. Yield variability, on the other hand, tended to decrease with distance from bare, grassland, and wetland (before 100 m, at least), increased slightly with distance from shelterbelt and standard boundaries, and was roughly constant with distance from other crops (Figure 7, bottom row).

Figure 7: Field boundary effect on wheat yield, accounting for the effect of spatial variation. Upper panel represents mean yield, while the lower panel represents yield variation. N refers to number of fields containing this boundary type, and % refers to the average percentage of field boundary accounted for by this boundary type.

Pea crops were less well-represented in our data (N = 39 fields), and therefore the effects of boundaries tended to be less consistent. Average pea yields (Figure 8, top row) were roughly 4.25 T/ha at the field boundary and increased to just below 5 T/ha away from shelterbelt, standard, and other crop boundaries, while the effect of bare, grassland, and wetland boundaries was inconsistent. Conversely, yield variability (Figure 8, bottom row) decreased with distance from grassland, shelterbelt, standard, and wetland boundaries, while the pattern with distance from other crop and bare was inconsistent. Unlike canola and wheat crops, there was no obvious evidence of an intermediate increase in yield with distance from the field boundary.

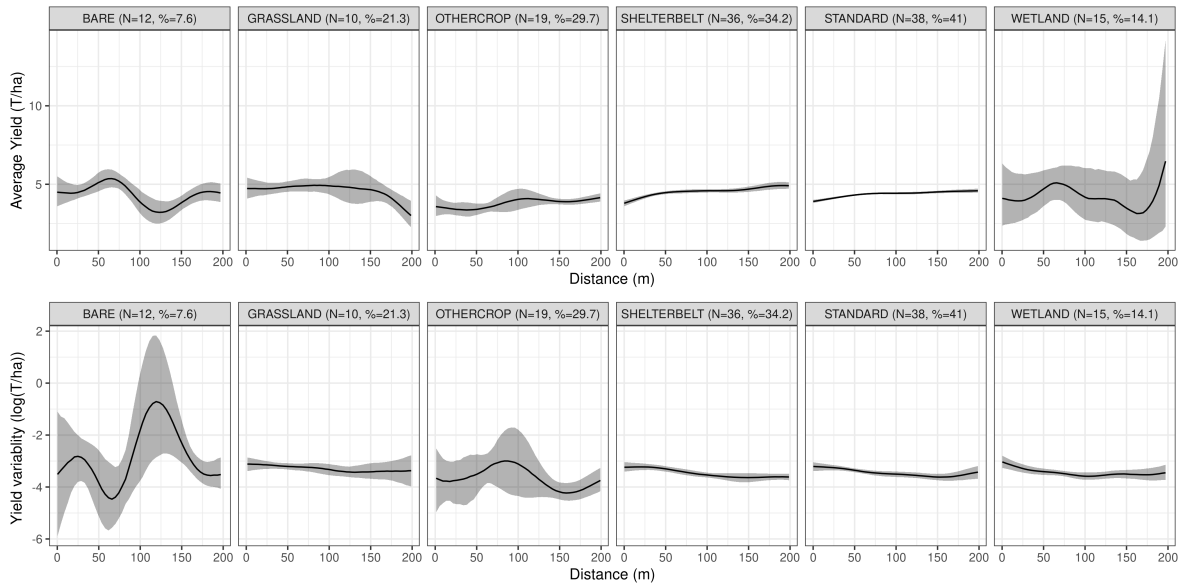Figure 8: Field boundary effect on pea yield, accounting for the effect of spatial variation. Upper panel represents mean yield, while the lower panel represents yield variation. N refers to number of fields containing this boundary type, and % refers to the average percentage of field boundary accounted for by this boundary type.

4. Discussion

Within individual fields, both the average and variability of yield changed with distance from the field boundary, and there were strong spatial relationships within each field, likely related to underlying soil or moisture conditions. Mean yield tended to increase with distance from field boundaries before plateauing, and there was some evidence of intermediate boosts in yield from shelterbelts next to canola and wheat, while peas showed no evidence of this. However, this intermediate increase was small in canola crops, meaning that ecosystem services likely did not vary strongly with distance from field boundaries (Scenario 2, Figure 1). There was a fairly large (0.75 T/ha) intermediate increase in average wheat yields away from shelterbelts, indicating a potential reduction in ecosystem services with distance. However, there was a similar (albeit weaker) pattern associated with bare ground, so this pattern may be caused by something other than the presence of shelterbelts. This indicates that shelterbelts may have a positive effect on wheat yields, and future work should expand this within relevant geographic sub-regions in order to locally inform growers about how best to manage the landscapes surrounding their fields.

Our work showed an intermediate increase in wheat yield with distance from shelterbelts and bare ground, but what is the cause of these patterns? We hypothesize that the potential yield-increasing

13

²¹² mechanisms may be a) microclimate effects, b) soil effects, or c) beneficial animal spillover. Shelterbelts

²¹³ can act to capture snow and moderate downwind microclimate, reducing water stress on crop plants

²¹⁴ and increasing yield (Staple & Lehane 1955; Kowalchuk & de Jong 1995; Bao *et al.* 2012). However,

²¹⁵ Kowalchuk & de Jong (1995) found that positive microclimate effects of shelterbelts on wheat yield

²¹⁶ occur at around 20 m (3-4 times the shelterbelt height), while our work shows a peak at around 50 m.

²¹⁷ This additionally complicated by the fact that water stress varies from year-to-year, and that the impact

²¹⁸ of the shelterbelt also depends on the orientation to prevailing winds and the physical structure of the

²¹⁹ shelterbelt (Pelton 1967; Kort 1988; Singh *et al.* 1999). Unexpectedly, we also found this intermediate

²²⁰ yield maximum with distance away from bare field boundaries, indicating that the patterns we found

²²¹ may be at least somewhat related to water stress. While we consider microclimate to be the most

²²² likely cause of the intermediate increase, it may also be related to soil effects from shelterbelts, where

²²³ erosion is reduced and leaf litter from the shelterbelt can cause an increase in soil deposition (de Jong

²²⁴ & Kowalchuk 1995). Finally, shelterbelts may also act as sources of predators or parasitoids of wheat

²²⁵ crop pests (Dong *et al.* 2015; Ward *et al.* 2021) but their effect depends on pest pressures, which are

²²⁶ typically suppressed by pesticide application. We recommend that follow-up studies a) use precision

²²⁷ yield data to confirm that this pattern exists across broader geographical areas and b) examine the

²²⁸ influence of the three potential mechanisms considered above (as well as their interaction).

²²⁹   Wheat and shelterbelts aside, why did we find few consistent effects of boundary type on crop yield?

²³⁰ Our work examined only the first and last links in the causal chain *Field Boundary → Ecosystem Service*

²³¹ *→ Crop Yield*, meaning that are two mediating relationships where the overall relationship could be

²³² broken or obscured. First, field boundaries may not have any influence on beneficial insect abundance

²³³ if separated from the larger landscape context, or spillover of beneficial organisms may be low. Second,

²³⁴ beneficial insect abundance may not influence crop yield if pest pressure or pollination deficits are

²³⁵ low, which may vary among year or location. Abiotic ecosystem services, such as reduced wind speed

²³⁶ or temperature, also depend on the structure of the boundary and whether wind or temperature are

²³⁷ limiting crop production. Our work suggests that we need to better examine the context in which

²³⁸ a) beneficial organisms spillover from field boundaries and b) under what conditions field boundaries

²³⁹ provide tangible ecosystem services (Karp *et al.* 2018). We consider these reasons below:

²⁴⁰   SNL generally has a positive effect on beneficial insect abundance, but field boundaries only represent

²⁴¹ a fraction of the SNL that surrounds a given field. Our analysis only examined boundary types, without

²⁴² considering the landscape context that fields were embedded in, but some studies show the influence

of landscape composition at larger spatial scales (Nguyen *et al.* 2021) which could be particularly important when beneficial arthropods are disperse very broadly (Öberg *et al.* 2007). There could also be interactions between landscape composition at different scales (Motzke *et al.* 2016; Haslem *et al.* 2021; Robinson *et al.* 2021), but this is not broadly studied. Landscape complexity (Fahrig *et al.* 2011) and connectivity are also associated with higher abundance of beneficial arthropods. This means that field boundaries such as wetlands or shelterbelts may only be helpful if connected to a larger network of non-crop habitat (but see Fahrig *et al.* 2019).

Field boundary characteristics, such as height or species composition, may play an important role in mediating spillover, but our definitions of boundaries were necessarily loose, as we used satellite imagery to manually classify them. Therefore, it may be that other aspects of field boundaries are more important for spillover of ecosystems services into fields. For example, larger or more functionally diverse (Bonifacio *et al.* 2011; Boughey *et al.* 2011; Lecq *et al.* 2018; Stašiov *et al.* 2020) hedgerows tend to support more biodiversity than smaller hedgerows, and some studies have shown that leaf litter and soil compaction influences predatory beetle communities (Sutter *et al.* 2018). These fine-scale boundary characteristics were not measured in our study, but could be used as variables in more biologically-meaningful classifications of field boundaries. Finally, the relevant spatial scale of boundary effects are often related to the height or windward orientation of the shelterbelt (Kort 1988; Kowalchuk & de Jong 1995), and precision yield data has the potential to show this in far greater detail than previous studies.

Our first scenario (Scenario 1, Figure 1) assumes that beneficial organisms disperse from boundaries into the field, and had a generally positive effect on crop yield. However, field boundaries may not have a positive effect if a) ecosystem dis-services – pest insects or weed seeds – also disperse along with beneficials, or if b) beneficials do not disperse from the boundary environment. Pests and weeds can also disperse from field boundaries, which may negate any positive effect of ecosystem services (Zhang *et al.* 2007). However, this is still poorly studied, and parsing out services from disservices is impossible using crop yield data *alone*. Ecosystem services from field boundaries may also be limited if beneficial organisms stay in field boundaries, or disperse at wider spatial scales Robinson *et al.* (2021). Lastly, pest pressure or pollination requirements may be low enough that ecosystem service provision does not significantly change yield. Canola is largely self-pollinating (at least in hybrid canola varieties, Lindström *et al.* 2016) and wheat is wind-pollinated, meaning that slight increases in pollination services may not change yields significantly in either crop (Woodcock *et al.* 2016). Crop pests in most

15

western Canadian fields are suppressed by insecticides (seed-treated or on the crop directly), and insect pest outbreaks tend to be temporally and spatially spotty, meaning spillover of beneficial predators and parasitoids may only matter during outbreak years.

Our work is some of the first to use a large precision yield dataset to answer questions about how landscapes influence yield. Since our sampling represents a limited amount of spatial and temporal coverage, there could be additional confounding variables that obscure signals of yield. Most fields were located in Central Alberta (Dry Mixedwood region, Natural Regions Committee 2006), where boundaries contained a higher proportion of shelterbelts (Figure 5), but differences from Southern AB (Foothills Fescue and Mixedgrass region) may be confounding the signal. Between-year effects may be important, especially during hot dry years (Kort 1988), but this was not considered in our analysis. Future work could examine whether boundaries have differing effects during hot or cool years, as shading or cooling effects may increase yield during hot years but have negative (or no) effect during cooler years. Finally, about 50% of growers in western Canada have collected precision agriculture data, and about 85% have yield monitoring capability (Steel 2017), which means that our work has only scratched the surface of what could be (collectively) one of the richest agroecological datasets ever created. We urge agronomists, growers, and ecologists to adopt more wide-spread use of precision yield data, as it represents a vast source of incredibly detailed information which can be used to answer a variety of important questions.

5. Acknowledgements

6. Funding Sources

References

Agriculture and Agri-Food Canada. (2018). *Annual Crop Inventory*. Available at: https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9. Last accessed 6 February 2019.

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*, 23, 1488–1498.

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

Baker, T.P., Moroni, M.T., Mendham, D.S., Smith, R. & Hunt, M.A. (2018). Impacts of windbreak shelter on crop and livestock production. *Crop and Pasture Science*, 69, 785.

Bao, Y., Li, H. & Zhao, H. (2012). Effect of shelterbelts on winter wheat yields in sanded farmland of north-western Shandong province, China. *Journal of Food Agriculture & Environment*, 10, 1399–1403.

Boetzl, F.A., Schuele, M., Krauss, J. & Steffan-Dewenter, I. (2020). Pest control potential of adjacent agri-environment schemes varies with crop type and is shaped by landscape context and within-field position. *Journal of Applied Ecology*, 57, 1482–1493.

Bonifacio, R.S., Kinross, C.M., Gurr, G.M. & Nicol, H. (2011). The effect of woody plant diversity and other stand and landscape factors on the diversity and abundance of birds using farm shelterbelts. *Pacific Conservation Biology*, 17, 22.

Boughey, K.L., Lake, I.R., Haysom, K.A. & Dolman, P.M. (2011). Improving the biodiversity benefits of hedgerows: How physical characteristics and the proximity of foraging habitat affect the use of linear features by bats. *Biological Conservation*, 144, 1790–1798.

Brandle, J.R., Hodges, L. & Zhou, X.H. (2004). Windbreaks in North American agricultural systems. In: *New vistas in agroforestry: A compendium for 1st world congress of agroforestry, 2004* (eds. Nair, P.K.R., Rao, M.R. & Buck, L.E.). Springer Netherlands, Dordrecht, pp. 65–78.

Brandle, J.R., Johnson, B.B. & Dearmont, D.D. (1984). Windbreak economics: The case of winter wheat production in eastern Nebraska. *Journal of Soil and Water Conservation*, 39, 339–343.

Case, B.S., Pannell, J.L., Stanley, M.C., Norton, D.A., Brugman, A., Funaki, M., *et al.* (2019). Non-production vegetation has a positive effect on ecological processes in agroecosystems. *bioRxiv*, 624635.

Cressie, N. & Wikle, C.K. (2011). *Statistics for spatio-temporal data*. Wiley.

17

de Jong, E. & Kowalchuk, T.E. (1995). The effect of shelterbelts on erosion and soil properties. *Soil Science*, 159, 337–345.

Dong, Z., Ouyang, F., Lu, F. & Ge, F. (2015). Shelterbelts in agricultural landscapes enhance ladybeetle abundance in spillover from cropland to adjacent habitats. *BioControl*, 60, 351–361.

Duelli, P. & Obrist, M.K. (2003). Regional biodiversity in an agricultural landscape: The contribution of seminatural habitat islands. *Basic and Applied Ecology*, 4, 129–138.

Fahrig, L., Arroyo-Rodríguez, V., Bennett, J.R., Boucher-Lalonde, V., Cazetta, E., Currie, D.J., *et al.* (2019). Is habitat fragmentation bad for biodiversity? *Biological Conservation*, 230, 179–186.

Fahrig, L., Baudry, J., Brotons, L., Burel, F.G., Crist, T.O., Fuller, R.J., *et al.* (2011). Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. *Ecology Letters*, 14, 101–112.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., *et al.* (2011). Solutions for a cultivated planet. *Nature*, 478, 337–342.

Frei, B., Renard, D., Mitchell, M.G.E., Seufert, V., Chaplin-Kramer, R., Rhemtulla, J.M., *et al.* (2018). Bright spots in agricultural landscapes: Identifying areas exceeding expectations for multifunctionality and biodiversity. *Journal of Applied Ecology*, 55, 2731–2743.

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*, 58, 2287–2304.

Garibaldi, L.A., Steffan-Dewenter, I., Kremen, C., Morales, J.M., Bommarco, R., Cunningham, S.A., *et al.* (2011). Stability of pollination services decreases with isolation from natural areas despite honey bee visits. *Ecology Letters*, 14, 1062–1072.

Gray, C.L. & Lewis, O.T. (2014). Do riparian forest fragments provide ecosystem services or disservices in surrounding oil palm plantations? *Basic and Applied Ecology*, 15, 693–700.

Griffin, T.W., Brown, J.P. & Lowenberg-DeBoerc, J. (2007). *Yield monitor data analysis protocol: A primer in the management and analysis of precision agriculture data*. SSRN.

Haslem, A., Clarke, R.H., Holland, G.J., Radford, J.Q., Stewart, A. & Bennett, A.F. (2021). Local management or wider context: What determines the value of farm revegetation plantings for birds? *Journal of Applied Ecology*.

Helmers, G. & Brandle, J.R. (2005). Optimum windbreak spacing in great plains agriculture. *Great Plains Research*, 15, 179–198.

18

Hünicken, P.L., Morales, C.L., Aizen, M.A., Anderson, G.K.S., García, N. & Garibaldi, L.A. (2021). Insect pollination enhances yield stability in two pollinator-dependent crops. *Agriculture, Ecosystems & Environment*, 320, 107573.

Karp, D.S., Chaplin-Kramer, R., Meehan, T.D., Martin, E.A., DeClerck, F., Grab, H., *et al.* (2018). Crop pests and predators exhibit inconsistent responses to surrounding landscape composition. *Proceedings of the National Academy of Sciences*, 115, E7863–E7870.

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

Lecq, S., Loisel, A., Mullin, S.J. & Bonnet, X. (2018). Manipulating hedgerow quality: Embankment size influences animal biodiversity in a peri-urban context. *Urban Forestry & Urban Greening*, 35, 1–7.

Lindström, S.A.M., Herbertsson, L., Rundlöf, M., Smith, H.G. & Bommarco, R. (2016). Large-scale pollination experiment demonstrates the importance of insect pollination in winter oilseed rape. *Oecologia*, 180, 759–769.

Lowe, E.B., Groves, R. & Gratton, C. (2021). Impacts of field-edge flower plantings on pollinator conservation and ecosystem service delivery – a meta-analysis. *Agriculture, Ecosystems & Environment*, 310, 107290.

Marja, R., Kleijn, D., Tscharntke, T., Klein, A.-M., Frank, T. & Batáry, P. (2019). Effectiveness of agri-environmental management on pollinators is moderated more by ecological contrast than by landscape structure or land-use intensity. *Ecology Letters*, 22, 1493–1500.

Mitchell, M.G.E., Bennett, E.M. & Gonzalez, A. (2014). Forest fragments modulate the provision of multiple ecosystem services. *Journal of Applied Ecology*, 51, 909–918.

Motzke, I., Klein, A.-M., Saleh, S., Wanger, T.C. & Tscharntke, T. (2016). Habitat management on multiple spatial scales can enhance bee pollination and crop yield in tropical homegardens. *Agriculture, Ecosystems & Environment*, 223, 144–151.

Natural Regions Committee. (2006). *Natural regions and subregions of Alberta.* Government of Alberta, Edmonton.

Nguyen, L.H., Robinson, S.V.J. & Galpern, P. (2021). Medium-resolution multispectral satellite imagery in precision agriculture: Mapping precision canola (*Brassica napus* L.) Yield using Sentinel-2 time series. *Precision Agriculture*, In Review (Preprint: https://www.essoar.org/doi/10.1002/essoar.10507479.2).

Osorio, R.J., Barden, C.J. & Ciampitti, I.A. (2018). GIS approach to estimate windbreak crop yield effects in KansasNebraska. *Agroforestry Systems*, 93, 1567–1576.

Öberg, S., Ekbom, B. & Bommarco, R. (2007). Influence of habitat type and surrounding landscape on spider diversity in Swedish agroecosystems. *Agriculture, Ecosystems & Environment*, 122, 211–219.

Pelton, W.L. (1967). The effect of a windbreak on wind travel, evaporation and wheat yield. *Canadian Journal of Plant Science*, 47, 209–214.

Prieto-Benítez, S. & Méndez, M. (2011). Effects of land management on the abundance and richness of spiders (Araneae): A meta-analysis. *Biological Conservation*, 144, 683–691.

Quinn, N.F., Brainard, D.C. & Szendrei, Z. (2017). Floral strips attract beneficial insects but do not enhance yield in cucumber fields. *Journal of Economic Entomology*, 110, 517–524.

Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., *et al.* (2018). Trends in global agricultural land use: Implications for environmental health and food security. *Annual Review of Plant Biology*, 69, 789–815.

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*, 57, 2180–2190.

Robinson, S.V.J., Edwards, D., Vickruck, J.L., Best, L.R. & Galpern, P. (2021). Non-crop sources of beneficial arthropods vary within-season across a prairie agroecosystem. *Agriculture, Ecosystems & Environment*, 320, 107581.

Singh, H.P., Kohli, R.K. & Batish, D.R. (1999). Impact of *Populus deltoides* and *Dalbergia sissoo* shelterbelts on wheat a comparative study. *International Tree Crops Journal*, 10, 51–60.

Staple, W.J. & Lehane, J.J. (1955). The influence of field shelterbelts on wind velocity, evaporation, soil moisture and crop yield. *Canadian Journal of Agricultural Science*, 35, 440–453.

Stašiov, S., Diviaková, A., Svitok, M., Novikmec, M. & Dovciak, M. (2020). Hedgerows support rich communities of harvestmen (Opiliones) in upland agricultural landscape. *Basic and Applied Ecology*, 47, 73–82.

Statistics Canada. (2014). Table 001-0010 - estimated areas, yield, production and average farm price of principal field crops, in metric units, annual.

Steel, D. (2017). *Analysis of precision agriculture adoption & barriers in western Canada.* Agriculture; Agri-Food Canada.

Sutter, L., Amato, M., Jeanneret, P. & Albrecht, M. (2018). Overwintering of pollen beetles and their predators in oilseed rape and semi-natural habitats. *Agriculture, Ecosystems & Environment*, 265, 275–281.

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.

Vega, A., Córdoba, M., Castro-Franco, M. & Balzarini, M. (2019). Protocol for automating error removal from yield maps. *Precision Agriculture*, 20, 1030–1044.

Ward, S.E., Umina, P.A., Macfadyen, S. & Hoffmann, A.A. (2021). Hymenopteran parasitoids of aphid pests within Australian grain production landscapes. *Insects*, 12, 44.

Weninger, T., Scheper, S., Lackóová, L., Kitzler, B., Gartner, K., King, N.W., *et al.* (2021). Ecosystem services of tree windbreaks in rural landscapes – a systematic review. *Environmental Research Letters*, 16, 103002.

Whelan, B. & Taylor, J. (2013). *Precision agriculture for grain production systems.* CSIRO Publishing.

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

Wikle, C.K., Zammit-Mangion, A. & Cressie, N. (2019). *Spatio-temporal statistics with r.* Chapman; Hall: CRC Press.

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

Woodcock, B.A., Bullock, J.M., McCracken, M., Chapman, R.E., Ball, S.L., Edwards, M.E., *et al.* (2016). Spill-over of pest control and pollination services into arable crops. *Agriculture, Ecosystems & Environment*, 231, 15–23.

Wratten, S.D., Gillespie, M., Decourtye, A., Mader, E. & Desneux, N. (2012). Pollinator habitat enhancement: Benefits to other ecosystem services. *Agriculture, Ecosystems & Environment*, 159, 112–122.

Zamorano, J., Bartomeus, I., Grez, A.A. & Garibaldi, L.A. (2020). Field margin floral enhancements increase pollinator diversity at the field edge but show no consistent spillover into the crop field: A meta-analysis. *Insect Conservation and Diversity*, 13, 519–531.

Zhang, W., Ricketts, T.H., Kremen, C., Carney, K. & Swinton, S.M. (2007). Ecosystem services and dis-services to agriculture. *Ecological Economics*, 64, 253–260.

Appendix A: R Code for Data Filtering

```r
# Helper functions -----------------

#"Inlier" spatial filtering procedure from Vega et al 2019, rewritten to work
#with sf + dplyr. Returns boolean
vegaFilter <- function(data,ycol,pvalCutoff=0.05,nDist=40){
  require(sf)
  require(sp)
  require(spdep)
  require(tidyverse)

  if(!any(class(data) %in% 'sf')) stop('Dataframe must be sf object')

  coords <- st_coordinates(data) #Get coordinates in matrix form

  #Get neighbourhood weights from 0 to ndist meters
  nWeights <- dnearneigh(coords,0,nDist)

  if(any(sapply(nWeights,length)==1)){
    warning('Some points had no neighbours and were removed')
  }

  #Get neighbourhood indices for each point (which other points are in this
  #point's neighbourhood?)
  nIndices <- nb2listw(nWeights, style = "W",zero.policy = TRUE)

  yield <- pull(data,{{ycol}}) #Get yield data column

  #Local Moran's I for each neighbourhood
  LM <- localmoran(yield,nIndices,p.adjust.method="bonferroni",
                   alternative ="less")

  results <- data.frame(LM) %>%
    rename('pval'=contains('Pr.z.')) %>%
    #Filter negative Ii and pvals < 0.05
    mutate(keepThese= Ii > 0 | pval > pvalCutoff) %>%
    #NAs (points had no neighbours)
    mutate(keepThese=ifelse(is.na(keepThese),FALSE,keepThese))

  ret <- pull(results)
  return(ret)
}

#Function to filter anything above a certain z-score. Returns boolean
ZscoreFilter <- function(x,z=3){
  xmean <- mean(x,na.rm=TRUE)
  xsd <- sd(x,na.rm=TRUE)
  zscore <- abs(((x-xmean)/xsd))
  zscore<z
}
```

```r
#Function to filter anything above certain quantiles. Returns boolean
QuantileFilter <- function(x,quant=0.99){
  l <- c((1-quant)/2,1-(1-quant)/2) #Symmetric quantiles
  x>quantile(x,l[1]) & x<quantile(x,l[2])
}


#Function to filter large changes in bearing. Returns boolean unless
#returnDiffs==TRUE
bearingFilter <- function(bearing,q=NULL,z=NULL,returnDiffs=FALSE){
  if(!xor(is.null(q),is.null(z))&!returnDiffs) stop('Input quantiles or Z-score')
  #Difference in compass bearings (in degrees)
  bearingDiff <- function(x1,x2){
    x <- x1-x2
    x <- ifelse(abs(x)>180,x-(360*sign(x)),x) #Angle differences can't be >180
    return(x)
  }

  #Looks 1 point ahead and behind
  bd <- cbind(bearingDiff(lag(bearing),bearing),
              bearingDiff(lead(bearing),bearing))
  #Maximum bearing difference ahead and behind
  bd <- apply(bd,1,function(x) max(abs(x),na.rm=TRUE)*sign(x[which.max(abs(x))]))

  if(returnDiffs) return(bd) #Return bearing differences only, without filtering

  if(!is.null(q)){
    ret <- QuantileFilter(bd,q=q)
  } else {
    ret <- ZscoreFilter(bd,z=z)
  }
  return(ret)
}


#Positional difference filter - filters out very distant and very close points.
#Returns boolean unless returnDiffs==TRUE
posFilter <- function(data,q=NULL,returnDiffs=FALSE){
  require(sf)
  if(is.null(q)&!returnDiffs) stop('Input upper quantile')

  if(units(st_distance(data[1,],data[2,]))$numerator!='m'){
    warning('Position differences not in meters')
  }
  coords <- st_coordinates(data) #Get coordinates
  #Distances between points
  pdiff <- sapply(1:(nrow(coords)-1),function(i) as.numeric(dist(coords[i:(i+1),])))
  pdiff <- cbind(c(pdiff,NA),c(NA,pdiff)) #Forward and backward lags
  pdiff <- apply(pdiff,1,max,na.rm=TRUE) #Maximum distance ahead and behind
  if(returnDiffs){
    return(pdiff)
  } else {
    return(QuantileFilter(pdiff,q)) #Note: uses 2-sided quantiles
  }
```

```r
    }
}

#Filter for (forward and backward) lagged speed differences. Returns boolean
dSpeedFilter <- function(speed,l=c(-1,1),perc=0.2){

  #Matrix of % diffs
  llist <- sapply(l,function(x) (lag2(speed,x)-speed)/lag2(speed,x))

  #Are any lagged speed values > percent change threshold?
  ret <- !apply(llist,1,function(y) any(abs(y)[!is.na(y)]>perc))

  return(ret)
}

#Overloaded lag function that takes negative values
lag2 <- function(x,n){
  require(dplyr)
  if(n==0) {
    return(x) #No lag
  } else if(n>0){
    lag(x,n) #Positive lag
  } else {
    lead(x,abs(n)) #Negative lag
  }
}


#Filtering pipeline  -------------------------

library(tidyverse)
library(sf)

#Note: dat == sf dataframe of yield data with columns: DryYield (T/ha),
#TrackAngle (compass bearing), Speed (km/hr), cropType (Wheat, Canola, Peas)

dat <- dat %>%
  #Vega et al 2019 spatial "inlier" filter - takes a few minutes...
  mutate(vegaFilt = vegaFilter(.,DryYield,nDist = 30))

dat <- dat %>%
  #"BS" filter - removes values >10.75 T/ha for wheat, >8 T/ha for peas & canola
  mutate(noBS = DryYield<ifelse(cropType=='Wheat',10.75,8)) %>%
  #Trim dry yield outliers
  mutate(Qfilt = QuantileFilter(DryYield,q=0.98)) %>%
  #Trim extreme bearing changes (turning)
  mutate(bFilt = bearingFilter(TrackAngle,q=0.98)) %>%
  #Trim absolute speed outliers
  mutate(speedFilt = QuantileFilter(Speed,q=0.98)) %>%
  #Trim speed differences (>20% change 2 steps forward and backward, suggested
  #by Lyle et al 2014)
  mutate(dSpeedFilt = dSpeedFilter(Speed,l=c(-2,-1,1,2),perc = 0.2)) %>%
  #Trim points that are far away from eachother
```

```r
  mutate(posFilt = posFilter(.,q=0.98)) %>%
  #Combine filter criteria
  mutate(allFilt = noBS & vegaFilt & Qfilt & bFilt & speedFilt &
          dSpeedFilt & posFilt) %>%
  #Remove filtered values
  filter(allFilt)
```

Appendix B: R Code for Modelling

```r
# Fit models ----------------------------------------------------------

library(mgcv)

#Note: dat == dataframe of yield data with columns: DryYield (T/ha), dist
#(distance to nearest field boundary, m), boundaryType (type of boundary,
#factor), E,N (easting and northing distance from centre of field, m)

#Uses thin-plate splines with extra shrinkage (bs = 'ts'), with 5 basis
#functions for distance smoothers and 80 basis functions for spatial smoothers
#(see Wood 2017 for more details)

#Variance smoothers
f2 <- '~ s(dist,k=5,bs="ts",by=boundaryType) + s(E,N,k=80)'
#Mean smoothers
f <- 'sqrt(DryYield)~ s(dist,k=5,bs="ts",by=boundaryType) + s(E,N,k=80)')
#Combine into a list
flist <- list(as.formula(f),as.formula(f2))

#Fit model - takes about 45 mins for a 50000 point dataset using an AMD Ryzen 7
#3800X 8-core processor
mod <- gam(flist,data=dat,family=gaulss())
```