**Fraser Valley *Bombus* Foraging Outline – Draft 2 (Landscape Questions)**

**Study Questions:**

1. How does landscape composition/configuration impact pollen collection (presence and mass) and pollen richness and quality? How do these effects vary in time?
2. How does landcover affect pollen richness collected by individual foragers?
3. How does landcover affect pollen compositional turnover between individual workers?
4. How do landscape x time relationships differ between native and introduced bumble bees?

**Levels of Analysis:**

1. **Individual level** – richness, composition, nutritional quality of pollens collected on a single bee
2. **Subsite level** – combined metrics across multiple bees collected at a subsite x month (grouping of 1-4 spatially adjacent transects, subsites spatially separated by >= 500 meters)
3. **Site level** – combined metrics across multiple bees collected at a site x month (grouping of ~30 transects, sites are >3km apart)

**General / Most Important Questions:**

* Should we aim for a balanced sampling design between mixtus and impatiens OR sample in proportion to the total availability of each (e.g., include all the site x time combos where we have enough specimens) OR focus mostly on one species and include a smaller subset of samples from the other species? (e.g., screen mixtus at transect, subsite, site levels, impatiens only at site level)? (currently my sampling scheme is based on the second option, sampling in proportion to availability)
* How many specimens per site x time is “enough” – is it better to take more locations or more specimens per location?
* For questions that involve individual level data, do we need to use transect as a random effect?
* What is the best combination of scales: transect, subsite, site? I don’t think we can do all 3 scales well unless we use only a single species, but I’m uncertain which is the best use of samples. Would we expect to find different patterns across different spatial scales/levels of spatial resolution, and would including transect-level data only create issues with spatial autocorrelation, given how close together transects can be?

**Proposed Sampling Scheme:**

To address all questions, using a “sample in proportion to availability of mixtus and impatiens” approach. We could easily modify this in a couple of ways:

* Reduce number of individuals per location x time to include more location and timepoints
* Include transect level analyses
* Focus primarily on a single species (e.g., mixtus) and include a smaller subset of other species (impatiens, flavifrons)

**Sample Selection:** Select n = 5 specimens from each **month x location x species** combination (where location is each subsite, or site for which we have sufficient specimens). This approach results in temporal and spatial gaps. Site level analyses will have better temporal resolution (more specimens to choose from in each month) while transect/subsite level will have better spatial resolution (a higher number of locations, but typically fewer timepoints). See heatmaps below for sample availability, example at subsite level.

A screenshot of a graph

Description automatically generatedA screenshot of a computer

Description automatically generatedN = 5 individuals per block

N = 3-10 individuals per block (color = number of pollen samples available, starting at n=3)

A screenshot of a computer

Description automatically generatedA screenshot of a graph

Description automatically generated

If following this method, we end up drawing workers from an unequal number of transects per subsite—due partially to sampling effort (more transects in some subsites) but also due to where bees are foraging (all on a single transect or spread across multiple). Below is a distribution of the number of transects represented per subsite x month x species combination.

A graph of a bar

Description automatically generatedThe x axis measures number of distinct transects. I am pondering whether this needs to be accounted for in our models, and if so, how? Different local floral communities could affect diversity of collected pollen. Perhaps including floral diversity (combined across transects from which we screened samples) as a fixed effect would be sufficient to account for this variation.

**Total Sample Size:** following the N = 5 selection method above would yield a total of 910 specimens (390 mixtus, 520 impatiens). We can fit 1140 on 3 sequencing runs, so this would give us some space to do metabarcoding on queens (~25 samples) and an additional ~200 individuals—my plan would be to use these for either flavifrons (see Q4H1) or additional specimens on the mixtus-impatiens abundance gradient (to test questions around pollen competition between species).

**Question 1: How does landcover impact pollen richness and quality (summed across individuals at a location)? How do these effects vary in time?**

**Hypotheses:**

**Q1H1) Pollen collected in more simplified/less seminatural landscapes will be of lower richness and quality.** *Support: Lower resource density will result in fewer dietary options, which will directly decrease pollen richness and may also impact the quality of collected pollens by reducing the ability of individuals workers preferentially select for pollens with optimal protein:lipid ratio.*

**Q1H2) Effects of landscape on pollen richness and quality will be most apparent in the early season (May).**

*A diagram of a flower

Description automatically generated with medium confidence*

*Support from 2022 data:*

*Slightly higher floral abundance in early season is likely driven by sites with blueberry—which is a poor quality pollen source (low protein content) for bumble bees. This could exacerbate pollen quality issues in early season.*

*Lower floral diversity available in the early season + high demand for pollen foraging (provisioning early workers/colony growth) could exacerbate the effect of landscape patchiness on pollen diversity.*

**Statistical Testing:**

Example model at *subsite level:*

pollen richness/quality ~ landscape diversity\*month + proportion seminatural area\*month + floral diversity + year + (1|subsite)

For pollen richness: sum across all sampled individuals. \*\*Optional: account for the number of individuals sampled per sampling event (4-6) using either the built in “offset” option for negative binomial GLMMs (adds sample effort to the model as a multiplicative term) or include effort as a fixed effect (for non-multiplicative relationships).

For pollen quality: Assign P:L ratio value to each pollen type, and sum across all pollen types (in proportion to their presence in pollen loads—here we will be using the “semiquantitative” approach to metabarcoding, because I don’t think this will work well for presence-absence. This value will be on a scale from min(P:L ratio) – max(P:L ratio). Presumably a Gaussian/student’s distribution.

*Optional:* If we include both species (mixtus and impatiens) in the same model, it will increase our number of observations per random level, which will probably be good for model convergence and estimation of random intercepts. However, it will also require us to include many interaction terms (e.g., species\*landscape\_diversity\*time), as I would like to test for differences in landscape x temporal patterns in the two species (see Q4H1). I’m not sure if this will help or hinder the modelling process—my intuition says it would help the random effects, and have no effect on the fixed effects, since we would be fitting the same number of fixed effect parameters overall, but doing it in one model (with twice as much data) instead of two models (with half the data in each). I could be wrong about this, but we could try both.

**Q1H3) Variation in pollen richness and quality between timepoints will be highest in more simplified/less seminatural landscapes.** *Support: Simplified agricultural landscapes may provide resource “pulses”(due to mass flowering crops, weedy species, etc.) but will be less likely to provide diverse, high quality resources through time.*

**Statistical Analysis:** Not sure if we’ll have enough data for this, but for locations with multiple timepoints (most of them) we could calculate the variation in richness/quality between those timepoints and then model that as a function of landscape quality. Perhaps using coefficient of variation as a response variable? Would need to account for differences in the *number of timepoints* incorporated into each estimate. Michelle or Shalene might have additional ideas on how to approach this (or suggestions for alternative methods).

**Question 2: How does landcover affect pollen collection and constancy by individual foragers?**

*Pollen/floral constancy* (definition): *a behavior in which individual workers visit only a single flower type (or collect only a single pollen type) over the course of a foraging bout or multiple foraging bouts. Here I am more interested in the degree of constancy (1/pollen species richness), rather than constancy as a binary variable (1 floral type = constant, >1 floral type = inconstant).*

**Hypotheses:**

**Q2H1) Workers collected in more simplified landscapes or landscapes with a lower proportion of seminatural habitat will be *less likely* to be collecting pollen, particularly during stages of the colony cycle (early and late) when pollen foraging is most common.** *Reasoning: simplified landscapes may induce greater energetic costs wrt foraging, e.g., greater foraging distances. Because pollen is primarily collected for feeding of larvae, while nectar is the primary energy source for workers, I expect that workers in more simplified landscapes will spend more time nectar foraging and less time pollen foraging.*

**Statistical Analysis:** Model pollen collection (0, 1) using a binomial GLMM at the individual level:

Pollen presence ~ landscape diversity\*month + proportion seminatural area\*month + year + (1|transect) + (1|site)

Preliminary analyse using model:

Pollen presence ~ landscape diversity\*month + proportion blueberry \* month + proportion edge \* month + (1|transect) + (1|site)

**A graph of different colored lines

Description automatically generated**A graph of different colors and numbers

Description automatically generated

Some support that edge density is important for pollen foraging in the spring (but decreases pollen foraging in late summer) and the opposite for landscape diversity. These models included proportion blueberry and proportion edge, which I think I’d like to replace with a seminatural habitat variable. These also only include data for *B. mixtus* in 2022, when it was very rainy in May/June. Expanding to include multiple species across two years could help us determine whether this is a consistent effect. Also need to check for spatial autocorrelation!

**Q2H2) Pollen-collecting workers in more simplified/less seminatural landscapes will tend to be carrying a *lower mass* of pollen at the time of collection.** *Support: In landscapes with more patchy resources, locating sufficient abundance of a bee’s preferred/learned floral type may be more challenging, leading to slower accumulation or lower final mass of pollen collected.*

*Counterargument: Previous work shows that foraging bouts can be temporally longer in more fragmented urban landscapes (probably due to patchiness of resources and increased transit time between heterogeneous patches). Bees early in a foraging bout will have lower pollen mass, but more, shorter bouts in good landscapes vs fewer, longer bouts in bad landscapes may result in similar pollen mass distributions among workers.*

**Statistical Methods:** same as Q2H1 but using Gaussian distribution for pollen mass

**Q2H3) Workers foraging in more simplified/less seminatural landscapes will collect a greater number of pollen types (species richness).** *Support: Theory predicts a positive relationship between resource density (landscape quality) and degree of floral constancy (Hayes and Gruter, 2022) regardless of the mechanistic reason behind floral constancy.*

**Statistical Methods:** Similar to Q2H1, but we may have to use a reduced subset of our pollen-screened samples if we are planning to use “transect” as a level of inference. With the proposed sampling scheme above, ~50% of the selected transects have only a single observation, so using transect as a random effect is probably going to be challenging. We could of course just use subsite as the random effect—I think this would be *somewhat* supported based on the literature, especially if we include transect-level floral diversity as a fixed effect. I have read several studies where “site” is defined as a pretty large area (similar or larger than our subsites) and workers are collected randomly throughout that space and considered to come from the same spatial group, even though the local floral assemblages at capture locations are probably quite different. Of course we would lose landscape resolution using this method in comparison to transect-level analyses, but this might be desired since I would expect transect-level analyses could be somewhat spatially autocorrelated/contain overlapping landcover buffers.

**Question 3: How does landcover affect turnover of pollen composition collected by individual workers?**

Q3H1) Pollen loads of individual workers collected in more simplified/less seminatural landscapes will be more similar to one another (lower Jaccard distance). *Support: Patchier resources and/or less seminatural area may lead to a lower availability of abundant and diverse floral resources within the foraging range of individual workers. If workers reduce their constancy as predicted in Q2H3, and there are fewer floral types to choose from a poor landscape, it is likely that the composition of individual pollen loads will become more similar.*

**Statistical Testing:** Calculate pairwise dissimilarity metrics between each pair of pollen-carrying workers at each transect (?), subsite, or site during a timepoint (e.g., no pairwise comparisons across timepoints, years, or species). Dissimilarity values can be confounded by the total number of pollen species per bee, as well as non-independence of pairwise metrics (e.g., if one partner is the same for two pairwise comparisons, those comparisons aren’t independent). Use null models to estimate deviation of observed dissimilarity from what we would expect if bee-pollen interactions were randomly assigned to bees (similar to Lauren’s null models in Global Change Bio paper, but also Timberlake et al., 2024 for pollen null networks). Use corrected dissimilarity values as response variable in GLMM, similar to those described in Q2.Could also test whether dissimilarity is caused by species replacements (different pollen types on different bees) or gain/loss (high nestedness). Nestedness could result from two processes: floral availability being highly nested between sites, *or* dissimilarity driven by how far into the foraging bout the worker is(i.e., if you could sample pollen off of a single worker at two points in time, the earlier sample would be a nested subset of the later sample). I think that a good preliminary analysis here (once we get pollen metabarcoding data) would be to check whether pollen diversity is highly predicted by pollen mass. This could suggest that bees accumulate different pollen types as they progress through the foraging bout.

**Question 4: How do landscape x time relationships differ between native and introduced bumble bees?**

Q1H4) Effects of landscape on pollen quality, composition, and richness will be most apparent in *Bombus mixtus. B. impatiens* will have higher diet richness, lower diet quality, and lower turnover between individuals across all scales of analysis.

*Note: This whole time I’ve been thinking this hypothesis was well supported, but when I calculated the Shannon diversity of floral visitation for each species and plot against number of individuals sampled per species, this is what I got:*

**A graph with numbers and letters

Description automatically generated**

(the one way off to the right at 1400 is impatiens, sorry for the cut off label)

Which raises the question: if this is an interesting line of inquiry, maybe we should be looking at flavifrons as well, which is the next most abundant but has muchlower diet diversity than mixtus and impatiens (this makes sense, because of our bees, flavifrons & californicus are the only ones with long tongues). Diet richness follows a roughly similar pattern but is more linear. Some of this could be temporal: longer flight season = exposure to more floral species