Fraser Valley *Bombus* Foraging Behaviour Outline

Summary/Background:

**3 groups of questions:**

1. How does **landscape diversity/composition** affect visitation and pollen collection?
2. Is there evidence of **competition** between conspecifics or congenerics for pollen collection?
3. How similar are **pollen and visitation preference**, and what are highly preferred plant species for each behaviour? (conservation related)

In my view, each of these groups of questions could utilize different (but overlapping) datasets and analyses, and could potentially stand alone as a publication or outreach document. Or they could be combined into an incredibly long thesis chapter.

**4 scales of analysis:**

1. **Individual level** – richness, composition, nutritional quality of pollens collected on a single bee
2. **Transect level** – combined metrics across multiple bees collected at a single transect (aim for pool of n = 5 specimens)
3. **Subsite level** – combined metrics across multiple bees collected at a subsite (grouping of ~4 spatially adjacent transects, which are at least 500m from other subsites (aim for pool of n = 5 specimens)
4. **Site level** – combined metrics across multiple bees collected at a site (grouping of ~30 spatially adjacent transect, which are >3km apart)

These scales of analysis can be further divided *temporally*, by round or by month. I will indicate when temporal separation is necessary, and for all cases when individual data is combined (e.g., transect, subsite, and site level metrics) I will only combine data from i) individuals collected during the same time period, or ii) an equal number of individuals across time periods, to avoid introducing temporal confounds.

When I say “pool,” I mean to pool the data **post-sequencing** (not pre-sequencing). This is so that pollen richness/composition can be analyzed at **both** the individual scale and at higher (“population”) levels. Pooling data from multiple individuals is likely more representative of overall pollen collection at a location, in contrast to individual workers which are known to collect a highly conserved # of pollen species per foraging bout.

**Section 1: The effects of landscape composition and diversity on foraging behaviours**

**Q1: How does surrounding landcover impact:**

* **Pollen presence/absence (***data already collected, see below for very prelim analysis***)**
* **Pollen mass**
* **Pollen quality**
* **Pollen richness**
* **Pollen composition**

**Sampling Design:**

Conduct analyses at transect, subsite, and site levels. I believe this could be informative because some combination of local and landscape effects probably drive pollen collection. For example, at the transect level of analysis we may find that landscape has very little effect because pollen collection is mediated by the (very local) availability of flowers. But at a subsite level, landscape may have a stronger effect, as it could mediate turnover of floral availability between grouped transects.

Equally combine individuals across timepoints (e.g., 2 individuals per month) for each location (e.g., transect, subsite, site, depending on scale). Make comparisons irrespective of time. Equal sampling across timepoints is somewhat critical due to changing floral availability which could lead to higher richness in pools where specimens are drawn from different timepoints vs pools where all specimens originate from a single sampling round.

A potential issue here is that because mixtus and impatiens have different phenologies, the months which make most sense for each of the two species may vary (e.g., May-July for mixtus, June-August for impatiens). We could select the two months with the most overlap (June-July) (see fig below), or use different months for different species (but this would limit our comparisons between species a bit).

June 1 = 152, July 1 = 182, August 1 = 213

**A group of graphs with different colored lines

Description automatically generated with medium confidence**

**Proposed Analytical Approach:**

**Q1:** *For pollen presence/absence, weight, richness, and quality, use GLMMs.*

Proposed models, if evenly combining individuals across timepoints:

pollen metric (transect pool) ~ landscape diversity + proportion edge area + proportion blueberry + year + species + (1|transect\*) + (1|site)

pollen metric (subsite pool) ~ landscape diversity + proportion edge area + proportion blueberry + year + species + (1|subsite\*) + (1|site)

pollen metric (site pool) ~ landscape diversity + proportion edge area + proportion blueberry + year + species + (1|site\*)

\*Potential issue here with random effects. Because we are including multiple years and species, this would result in (at most) 4 observations per location. Which is not really enough to fit a random effect. So I’m a bit stumped on how to treat this. Many of the locations (especially transects) would have <4 observations. An alternative approach could be to model years and species separately (I’ve seen some papers where they do this, and then assess whether they get the same result in different years).

For the parasite work, we used transect as random effect (despite having <5 observations for some transects) and the models seemed to run just fine—so possibly the Bayesian models are a little more robust in this way. I will do more reading, but thought maybe you’d have a sense for whether this is a significant issue. Or maybe there’s a totally different approach that I’m just not considering. I’ve read a couple of papers recently in which they used a nonparametric approach (null models to create empirical confidence intervals) which I will look into further.

\*\*Addition: If we consider each bee to be a “site”, my initial thought was to calculate alpha diversity (pollen diversity on a single bee) and gamma diversity (pollen diversity on some fixed number of bees per location/time). However, I am wondering if using some combination of alpha and beta diversity (diversity of pollen on a single bee and turnover of pollen composition between bees) would be an alternative approach? e.g., by calculating a pairwise similarity/dissimilarity metric between specimens at a site. This might even have been something that you suggested during our last meeting, but I didn’t fully process it at the time. I’m also not entirely sure how that would complicate analysis (e.g., each pairwise dissimilarities would not be independent, but maybe random effects could account for that).

*For pollen composition, use ordination:*

Create matrix of presence/absence of each pollen type for each individual, transect, etc. and use NMDS to visualize. Overlay environmental variables (local plant richness, landscape composition/diversity) using envfit function in vegan. Uncertain how/if ordination deals with multiple samples from the same location, so we may need to conduct different ordinations for years/species.

**Q2: Are the above relationships consistent throughout the flight season, or do they vary in time?**

**Sampling Design:** Similar to above, but now including multiple timepoints for each spatial scale.

Select n = 5 specimens from each **month x location x species** combination (where location is each transect, 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 (a colored block indicates 5+ pollen samples available for that condition). An alternative solution could be to reduce the number of sample per spatial scale (e.g., n = 10 samples at site level would be fairly achievable, whereas n = 3 samples at transect level might yield more coverage).

Site Level:

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

Subsite Level:

A screenshot of a computer

Description automatically generatedA screenshot of a graph

Description automatically generated

Transect Level:

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

Example model for transect scale:

pollen metric (transect pool) ~ landscape diversity\*month + proportion edge area\*month + proportion blueberry\*month + year + species + (1|transect) + (1|site)

There are once again going to be issues with fitting random effects, which this time cannot be alleviated by fitting separate models. We could be limited to locations with at least 5 observations across all years, species, and timepoints (out of a possible 2 yrs \* 2 species \* 4 months = 16).

Also—this is a lot of interaction effects, particularly given that month is a categorical variable. Perhaps using AIC for model selection would be useful?

**Q3: How does landcover affect the constancy (pollen richness) of individual foragers?**

**Sampling Design:** Because this is individual level, we can utilize all specimens from the previous section questions.

**Proposed Analytical Approach**: GLMMs

Pollen richness ~ landscape diversity + landscape composition + year + species + month + (1|transect) + (1|site)

Minimum cutoff of 5 observations per transect – I think this will be fairly reasonable since each individual is a separate observation.

**Q4: Does local or landscape-scale floral availability better explain pollen composition and diversity?**

**Sampling Design:**

* Pollen composition at individual and transect levels (as above)
* Floral availability at transect and site levels

Proposed Analytical Approach:

* What proportion of collected pollen species are present in transect or site level vegetation surveys? Do these proportions vary between mixtus and impatiens (e.g., evidence for one foraging more locally than the other)?
* For transect level pools of pollen collection, are the proportion of workers carrying a specific pollen type more closely aligned with the proportional availability of that plant at the transect level, or its proportional availability at the site level. Mantel test? To compare composition of pollen load (matrix) to composition of available flowers (at different scales).
* Could use null model approach from Timberlake et al 2024 (create “null pollen networks” at an individual level, which re-assign the number of links for each forager in proportion to the abundance of each flower available. Repeat this many times to calculate an empirical p-value for the diversity of pollens present in actual pollen loads.

**Section 1 Preliminary Analyses:**

**Q1: Effect of landscape on pollen presence/absence**

Just using 2022 mixtus – not accounting for local floral abundance / diversity A graph with a pink line

Description automatically generatedA graph of pollen presence

Description automatically generated

Model:

Pollen presence ~ julian\_date + julian\_date2 + proportion blueberry + proportion edge + landscape diversity + (1|transect)

R2 = 0.12 (not fantastic)

🡪 note: in this case, some of the points do NOT have 5 observations, but the model runs okay. I think this may be because Bayesian models are more robust to this sort of thing, but I need to do some more reading

🡪 when I re-run the models after filtering out all the transects without <5 observations, I get similar trends but lose support (this could just be because the sample size drops from 750 to approx. 500)

🡪 if I add “site” as a random effect, blueberry no longer has an effect

So my conclusion in that landscape effects may be weak or result from spatial autocorrelation, and that more in depth analysis may be necessary to untangle these. Phenology seems to have a more robust effect: when I convert to a categorical variable (month), workers in May are considerably more likely to be collecting pollen compared to June/July/August. June/July have the lowest pollen collection, but not significantly different than August.

**Q2: How does the effect of landscape on pollen foraging change over time?**

Model:

Pollen presence ~ month\*prop\_blueberry + month\*prop\_edge + month\*landscape\_diversity + (1|transect) + (1|site)

**A graph of different colored lines

Description automatically generated**

This could be extremely overfit, but it seems like proportion edge area has a (weak!!) positive effect on pollen foraging in the early season (May/June), and a negative effect on pollen foraging in the late season (July/August).

A graph of different colors and numbers

Description automatically generated

For landscape diversity—less pollen foraging in more diverse landscapes in May, more pollen foraging in diverse landscapes in August.

Notably, neither of these main effects was significant before adding the interaction with month. Interaction effects could be related to the relative availability of floral resources in different landscape types in different months. It is a bit confusing that configurational (edge density) and compositional (land-use type) diversity have opposite effects in May & August…

However I think the results of both analyses are interesting in combination. The first approach (no interaction with time) seems to tell us that less pollen collection happens in the mid-season (June/July). The second approach (including interactions with month) qualitatively suggests that landscape has more of an effect on pollen collection in the early and late season (when more pollen collection is occurring).

**Section 2: Foraging competition between conspecifics and congenerics**

**Q5) What are the temporal and dietary overlap of *B. impatiens* and *B. mixtus?***

Sampling design: visitation/phenology dataset + pollen data from **section 1.**

**A group of graphs with different colored lines

Description automatically generated with medium confidence**

Much more overlap in 2022, maybe because poor weather delayed development of some mixtus colonies, causing that second peak? (need to convert this plot to bees per sampling minute, right now it’s not standardized)

Proposed Analytical Approach: A bit more qualitative. Compare differences in diet composition, particularly preference. Calculate diet specialization (d’, Bluthgren 2006).

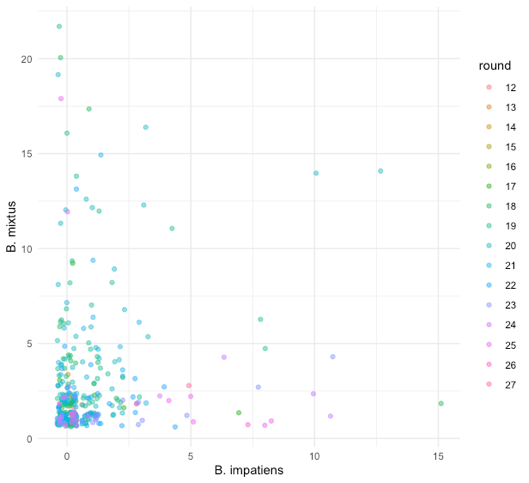
**Q5) Do foragers at sites with high *Bombus* abundance collect:**

* **Less pollen (presence/absence or weight)**
* **Lower richness of pollen**
* **Pollen with lower nutritional quality (P:L ratio)**

Sampling Design: see figures on page below. Draw pollen loads across a variety of sites:

* Mixtus pollen loads at sites with high and low mixtus abundance
* Mixtus pollen loads at sites with high and low impatiens abundance
* Impatiens pollen loads at sites with high and low impatiens abundance
* Impatiens pollen loads at sites with high and low mixtus abundance

Do the same for 2022 – except in 2022, we could incorporate abundance of whole *Bombus* community, if we wanted to. I would expect (potentially) more competition in 2022, due to the seemingly greater temporal overlap in the plots above.



2023, abundance of mixtus and impatiens for each sampling event (only sampling events for which we have at least one **mixtus** pollen load)

A graph with different colored dots

Description automatically generated

2023, abundance of mixtus and impatiens for each sampling event (only sampling events for which we have at least one **impatiens** pollen load)

A graph with different colored dots

Description automatically generated

2023, abundance of mixtus and impatiens for each sampling event (only sampling events for which we have at least one **of each** pollen load)

**Proposed Analytical Methods:**

Individual pollen richness ~ mixtus abundance + impatiens abundance + (1|transect)

🡪 Not enough specimens to screen multiple individuals at a per transect level (maybe 3-ish? Just not a lot of survey events with enough bees), but we could potentially scale up to subsites or sites (e.g., if bee abundance is high at the site level, how does it affect foragers in that site) but I think that would be a less powerful approach because they may not be that concentrated on resource patches

I hypothesize that individual richness could increase in the face of more competition (reduced constancy because there is less pollen available per each floral type). Of course the opposite could also be true if some floral types are completely depleted and are no longer available for collection.

Additionally we could look at composition using ordination—do individuals collected at high/low abundance of each of the two species cluster separately? For example if one species competitively excludes the other from certain pollen types, you might expect to see their pollen composition clustering together when they are found at low abundance, but one species cluster diverging when they are found at high abundance of the other species.

**Q6) Does high *Bombus* abundance cause changes in floral preference, e.g., relaxation of constancy at the individual level (pollen), or reduced collection from preferred plants (pollen & visitation data)?**

**Q7) Are conspecific and congeneric interactions similar? (e.g., do *B. impatiens* abundance *B. mixtus* abundance have similar effects on *B. mixtus* pollen foraging, and vice versa)?**

**Section 3: Floral preference, pollen-nectar overlap, and temporal variation**

**Q8) Which plants are used/preferred for pollen, nectar, or both?**

Sampling design: Use visitation data, and pollen data collected from all specimens in **section 1** – these should provide a good cross-section of pollen usage across space and time.

Proposed Analytical Approach: [need to do more reading]

**Q9) How does pollen/visitation richness vary in time?**

This is effectively the same as Q2 but without the addition of landscape effects – it could be analyzed in a very similar framework.

**Q10) How does preference vary in time? (e.g., do workers quickly detect new plant species and/or quickly abandon species nearing the end of their bloom, or do we observe lag times between availability and visitation/preference? If so, do lag times vary between bee species or floral species?)**

**Q11) Which plants are most used by foraging queens? How does diet breadth / diversity of queens compare to that of workers?**

Sampling design / analytical approach: very little data (~20 pollen samples) so mostly qualitative. However I think it is still useful to assess because this is a vulnerable life stage without much info.

By supplementing with visitation data we could assess preferences as above. Could also take a nonparametric approach to examine diet diversity/breadth. For example, if we have 100 visitation events over a 1 month period for queens, we could assess overall visitation diversity and then draw many permutations of 100 visitations from the worker dataset for each month to compare whether workers or queens visit a higher diversity of flowers. Could then compare to floral availability and determine whether queens are more or less selective than workers.