Examining the movement patterns of the European common blue butterfly (*Polyommatus icarus*): A study on short-term and long-term movement of an invasive species

Allegra A. Spensieri

A Thesis in
The Department
of
Biology

Presented in Partial Fulfillment of the
Requirements For the Degree of Master of Science (Biology)
at Concordia University
Montreal, Quebec, Canada

August 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Allegra Spensieri

Entitled: Examining the movement patterns of the European common blue butterfly (*Polyommatus icarus):* A study on short-term and long-term movement of an invasive species

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

________________________________ Chair
Dr. Emma Despland

________________________________ External Examiner
Dr. Robert Weladji

________________________________ Examiner
Dr. Sarah Turner

________________________________ Examiner
Dr. Emma Despland

________________________________ Thesis Supervisor
Dr. Eric Pedersen

Approved by ____
Dr. Robert Weladji, Graduate Program Director

____
Dr. Pascale Sicotte, Dean of Arts and Science

Date ____

Abstract

Examining the movement patterns of the European common blue butterfly (*Polyommatus icarus):* A study on short-term and long-term movement of an invasive species

Allegra A. Spensieri

Understanding how individuals interact with biotic and abiotic factors in their habitats, and how these interactions influence movement through landscapes is a key step in understanding dispersal events. The world is experiencing increased introduction of non-native species into new regions, however without an understanding of how species use the space they occupy, it is not possible to understand how dispersal events occur. We studied the short- and long-term movement patterns of a non-native butterfly species in Montreal, Canada: the European common blue butterfly, *Polyommatus icarus*, to determine whether the movement patterns of these butterflies over multiple days can be predicted by individual movement behaviours measured over short time periods. We asked what factors can predict short-term movements in females of this species, and whether these short-term movements can be used to predict long-term movements. We examined short-term movement by following individuals over short time periods and compared these movements to long-term movements observed via a mark-release-recapture study. In doing so, we found that flowering, host-plant species presence, increased vegetation height, and road edges can predict short-term movement, and that the short-term movement model produced can be used to accurately predict long-term movements when they are less than 100 m. These findings suggest that an understanding of ground cover characteristics is important in being able to predict the dispersal of most individuals, but that further work will be required to accurately predict long-distance dispersal events, which are what appear to be driving range expansion of *P. icarus* in North America.

Acknowledgments

First and foremost, I'd like to thank my supervisor, Eric. Thank you for being exactly the type of MSc supervisor that I needed. You provided me, and the rest of the lab, not only with scientific and statistical insight, but also with an unlimited supply of coffee and lots of your time. You encouraged me to do things outside of my comfort zone, and (potentially most importantly) let me get away with calling butterflies "the fish of the sky". Thank you. To my committee members, Emma and Jochen, thank you for your insight on this project, and thank you to Sarah for stepping in for my defense. To Sonia and Robert from the Concordia Biology Department, thank you for your help and advice on growing plants in the greenhouse. To the cities of Verdun, Cote-des-Neiges, and the groundskeepers of the MUHC, thank you for the permission to catch butterflies in your fields. This research was funded through an NSERC grant to Eric Pedersen at Concordia University and by a Concordia Merit Scholarship. This project was also supported by BIOS2 and QCBS, via travel funding and workshops.

To the Pedersen Lab, including the lab members who passed through SP 404.03 before me, thank you for building a lab environment where silly polls and biology memes are welcome along with code de-bugging and writing feedback. Natalie, thank you for helping me write so many lines of code this past year. You have been an amazing lab mate, office mate, and friend to have over the past 2 years – I truly don't think I'd have those road crossings figured out yet without your help. Spencer, thank you for all the work you put into the field season with me, you taught me many things over the course of the summer, and the data collection would not have been nearly as successful without you. Your help, feedback, and ideas were invaluable to me. To Alienor, Alex and Fonya, thank you for listening to me talk about butterflies at so many lab meetings and for all your feedback over the past 2 years. To the Ziter lab, but most notably Bella, Chloe, and Nicole, thank you for involving me in many of your lab activities, and for reminding me that in a department full of very smart people, it's way more important to just be kind.

The last 2 years didn't occur in isolation though, and this research project was also made possible by all the people who influence my life daily. To all the parental figures in my life, whether by blood, love, or choice, but especially my mum, Heidi, thank you for the years of encouragement and support, I simply would not be me without you. To my sister, Olivia, you told me to "just write a perfect thesis", while that isn't possible, thank you for thinking that I could, and for always being there. To Nikki, Julia, Maddie, Emilie, Casey, Ryan, Declan, Oscar, thank you for the years of friendship, laughs, and recently, distractions from butterflies. To my boyfriend, Riley, thank you for all the above, for always believing in me, and especially for reminding me of it on the days that I didn't.

Finally, to anyone who ever asked me how the butterflies were, thank you for caring (or at least pretending to), I may finally have answers for you.

Contribution of Authors

Conception and design of analysis: Allegra, Eric

Data collection: Allegra

Contribution of additional data or analysis tools: Eric

Analysis of data: Allegra, Eric

Drafting of manuscript: Allegra

Critical revision: Allegra, Eric

Table of Contents

List of Figures ..... *vii*

List of Tables ..... *ix*

List of Supplementary Material ..... *x*

Chapter 1: Literature Review *1*

Movement of organisms **1**

Dispersal and metapopulations **4**

Butterflies as a model organism to study the relationship between short- and long-term movement patterns **5**

Chapter 2: Examining the movement patterns of the European common blue butterfly (Polyommatus icarus): A study on short-term and long-term movement of an invasive species *8*

Introduction **8**

Methods **13**

Summary of methods 13
Study areas 14
Data collection 16
Statistical Analysis 22

Results **34**

Short-term movement 34
Mark-recapture results 36
Comparing estimated movement model with mark-recapture results 38

Discussion **41**

Vegetation 42
Road crossing 44
Distance 46
Implications 51

Conclusion **51**

Chapter 3: General Conclusion *53*

References *56*

Supplementary Material *61*

List of Figures

Figure 1: Framework of methods used throughout this study, demonstrating the ways which different elements were used to understand whether habitat-based short-term movement predictors, collected by observing individuals, can be used to predict long-term movements of *P. icarus*, based on a mark-release-recapture study. Each box represents a step in the methodological process, and arrows indicate the start of a new step.

Figure 2: Map of study sites and their location in Montreal (a), with the MUHC above (b) and the aqueduct below (c). Outlined areas on the study sites represent sampling zones and are numbered based on how they were designated during sampling. Zone 6 at the MUHC site is outlined in red, as it was planned in a location that was not accessible for sampling, and Zone 14 at the MUHC site does not exist as it was absorbed into Zone 13.

Figure 3: Description and images of techniques used while collecting data on short and long-term movement patterns. Techniques include alternate step production and measurements (a), ground cover sampling (b), vegetation height measurement (c) and *P. icarus* marking (d).

Figure 4: Graphic explanation of how steps less than 1 m were aggregated into one step during data analysis. Panel (a) depicts the observed steps as they were measured during short-term movement data collection, with each circle representing the 30 cm radius plot within which ground cover information was collected in the field, and arrows depict the *P. icarus* individual's sequence of steps, with the distance between each step noted in meters. Panel (b) depicts how multiple steps were aggregated into one step if they were less than 1 m apart from each other, where the circled areas represent how steps less than 1 m from each other were grouped together. In both panels, images of plants represent different species, and in panel b the plant images circled represent that the aggregated ground cover data used the highest ground cover value from all the individual steps for the aggregated step.

Figure 5: Estimated smooth, using thin plate regression splines (TPRS), and parametric terms of various habitat quality variables on the relative preference of a female *P. icarus* to move to a given location. Y-axes represent the partial effect of preference to move to a location and is represented on an exponential scale. X-axes represent (a) percent host-plant coverage, (b) number of host-plant species in bloom, (c) maximum vegetation height in centimeters, (d) the percent coverage of non-vegetation, (e) whether a road crossing is required, and (f) the log-2 transformed distance to the next landing location, with a supplementary scale above depicting the corresponding distance in meters. Gray bands around the smooth terms and bars around the parametric terms represent the 95% confidence intervals of the estimated function.

Figure 6: Map of study sites, with the MUHC on the left (a) and the aqueduct on the right (b), with the marking and recapture locations of *P. icarus* individuals that were recaught in the MRR study. Green points represent initial marking location and brown points represent recapture locations. Overall, 48 individuals were recaptured at the MUHC, and 142 individuals were recaptured at the aqueduct.

Figure 7: Image of study sites, with the MUHC on the left (a) and the aqueduct on the right (b), representing the predicted probabilities of where a *P. icarus* individual would end up if given an infinite amount of time, regardless of where in the study site its movement began. Blue areas indicate that an individual would be less likely to end up there, while red values indicate that an

individual would be more likely to end up there. Note the difference in probability values between the two study sites.

Figure 8: Plot of the log-2 transformed likelihood ratio for each recaptured individual. Where likelihood ratio is the probability of observed movement divided by the long-run probability of movement. Panels a and b depict likelihood ratio values for individuals at the MUHC site, while panels c and d depict likelihood ratio values for individuals at the aqueduct site. Panels (a) and (c) have x-axis values of log-2 (hours since capture) to show the relationship between the likelihood ratio and the amount of time moved, while panels (b) and (d) have x-axis values of log-2(distance) to show the relationship between the likelihood values and the distance moved. Each point represents a different recaptured *P. icarus* individual, with females shown in blue and males shown in green. Trend lines represent the estimated smooth curve for the sex and predictor of interest. The horizontal dashed line on the 0 line indicates equal probability under the two models for individuals to have ended up in the locations they did. Note the difference in x-axis scale difference between hours and distance between mark and recapture.

List of Tables

Table 1: Results of mark-release recapture study on *P. icarus* at two sites in Montreal. Results shown include total number of marked individuals, number of recaptures, percentage of recaptures, maximum displacement distance, mean displacement distance, and the third quantile of displacement distances. Results shown for only females, only males, and both sexes combined. Values in parentheses in max displacement column are distance values for two individuals that moved exceptionally long distances compared to other individuals.

List of Supplementary Material

Figure S1: Observations of *P. icarus* over time in North America as reported on the community science platform “iNaturalist” (Quebec, Ontario and Vermont). Red points indicate where a *P. icarus* individual was reported being seen. Images are grouped by observations that occurred within five years of each other, other than the most recent image which only contains four years.

Figure S2: Dates of short-term data collection at the MUHC (a) and aqueduct (b) sites during the 2023 summer field season. Dates are formatted as day/month. Note that Zone 6 at the MUHC does not exist, as it was planned in a location that was not accessible for sampling, and Zone 14 at the MUHC site does not exist as it was absorbed into Zone 13. Short-term movement sampling did not occur in Zones 15 or 16 at the MUHC site as there were too many hazards to safely do so.

Figure S3: Histogram of all step lengths (observed and alternate steps) recorded during short-term movement data collection from all paths. Panel (a) shows the distribution of distances between steps based on original measurements. Panel (b) shows the distribution of distances between steps after aggregating sequential observed steps shorter than 1 m from the same path into one step and calculating the distance of all observed steps to all alternate steps within a path. Red vertical lines represent the lengths of the three pre-measured alternate step lengths, 1.5 m, 3.5 m, and 7.5 m. Note the difference in scales on the x and y axes.

Figure S4: Visual outputs of spatial GAMs generated by equation (4) for the MUHC site. Panels represent the predicted values of (a) percent host-plant coverage, (b) number of host-plant species in bloom, (c) percent of non-vegetative coverage, and (d) maximum vegetation height. Darker colouring represents smaller predictions, while lighter colouring represents larger predictions. Empty pixels represent areas with insufficient information to extrapolate a prediction, due to insufficient sampling in that area. Note that roads and buildings were not removed from this image but were in the matrix created from it.

Figure S5: Visual outputs of spatial GAMs generated by equation (4) for the aqueduct site. Panels represent the predicted values of (a) percent host-plant coverage, (b) number of host-plant species in bloom, (c) percent of non-vegetative coverage, and (d) maximum vegetation height. Darker colouring represents smaller predictions, while lighter colouring represents larger predictions. Empty pixels represent areas with insufficient information to extrapolate a prediction, due to insufficient sampling in that area. Note that roads and buildings were not removed from this image but were in the matrix created from it.

Figure S6: Example of a short-term movement path of a *P. icarus* female. Green points represent landing locations. Depicted on the left (a) are all the short-term movement paths recorded at the MUHC study site, and on the right (b) is a zoomed in path with the sequence of steps connected to each other with the measured distances reported. Note that distances may not appear to be to scale as the GPS used was only accurate up to three meters.

Figure S7: Histogram of distribution of displacements of recaptured *P. icarus* individuals collected via a mark-release-recapture study.

Figure S8: Maps of both study sites, with MUHC on the left (a) and aqueduct on the right (b). Points represent all capture locations of *P. icarus* individuals, with males represented in green and females represented in blue. Locations shown outside of study sites are due to GPS error.

Figure S9: Scatterplot comparing the log-2 transformed likelihood ratios *P. icarus* individuals who crossed roads and those that did not. The left panel (a) depicts the MUHC site, and the right panel (b) depicts the aqueduct site. Note the difference in x-axis scale between sites.

Figure S10: Histogram of observed step lengths of *P. icarus* females collected during short-term movement data collection. The red distribution curve represents the Gamma distribution used to calculate the distance parameter in $Pref(x, y)$. The Gamma distribution has a shape parameter of 1.41 and a scale parameter of 1.0.

Table S1: Information about distance moved and time passed between initial capture and recapture of *P. icarus* individuals at two field sites, collected via a MRR study.

Chapter 1: Literature Review

Movement of organisms

Movement is responsible for interactions between individuals from the same populations as well as different populations, and the spread of species into novel habitats along with the new types of interactions that occur because of this spread. From passive forms of movement, such as wind and water dispersal of pollen, to the motivated movements made by animals, movement is an important component in the make up of populations, communities, and ecosystems (Nathan et al., 2008). Births, deaths, and movements are the three components that add and remove individuals from a given area. All three of these components are important in understanding how individuals and populations interact with each other. Of these three components, movement may be the hardest to study, because of how movement processes differ depending on the individual, their environment, and the timescale at which the movement is studied (Turchin, 1998). Without understanding movement, we do not have a complete picture of species or population interactions.

Many species disperse only via passive forms of movement (e.g., many plant seeds, some jellyfish), which allow them to allocate very little energy towards movement but result in no choice as to where they end up - these species cannot choose to land in an optimal spot (Nathan et al., 2008). On the other hand, species that use active forms of movement have the choice to move from one spot to another but must allocate energy towards their movement. The energy expenditure required by active movement means that we expect individuals engaging in active movement to choose to move towards spots that are better suited to them, and away from those that are less suited to their current needs (Nathan et al., 2008). The factors that make a spot desirable or undesirable will depend on the moving individual and its current needs. Where they choose to move, and their movement patterns, may be motivated by internal (e.g., hunger, reproductive cues) or external (e.g., weather, presence of others) factors (Nathan et al., 2008).

The ideal free distribution theory (IFD) explains that when individuals are in a position to behave ideally, they will choose to move to the spots around them which will maximize their fitness. When individuals are all equally suited to an environment, a spot which will maximize their fitness is one where the ratio of resources, including energy conservation, to other individuals is the highest (Fletcher & Fortin, 2018). Individuals engaged in active movement are expected to behave according to the expectations of IFD. They are limited by the resources around them and within their perceptive range though, which often means that there is a trade-off between moving a long distance to find optimal resources and moving a shorter distance to find good-enough resources (Abrahams, 1986).

Step selection functions (SSFs) are used to understand how animals move in relation to resource availability and habitat selection (Avgar et al., 2016). They link an individual's known consecutive locations to each other (observed steps) and compare the characteristics of these locations to a set of other locations that were also available to the individual at comparable distances (alternate steps) (Thurfjell et al., 2014). These functions are an extension of resource selection functions, but with the added step of taking an individual's perceptive range into account, both spatially and temporally (Avgar et al., 2016). SSFs have been used in a variety of ways to examine animal movement patterns under varying conditions, as they are more flexible and account for autocorrelation in animal tracking data better than resource selection functions do (Klappstein et al., 2024). SSFs are designed to take the habitat characteristics of the species of interest into account and calculate an SSF score for known locations and a subset of other locations available to the moving individual. They can also be used to simulate biased random walks across landscapes, which allows SSFs estimated across a short period of time to be used to predict long-term movement paths for individuals (Duchesne et al., 2015). By comparing individuals' known locations to other locations available to them, which they did not choose to move to, SSFs can be used to infer the habitat preferences of a species and predict where they will move in their landscape over time (Duchesne et al., 2015; Fortin et al., 2005).

Movement can be examined at different spatial scales, known as short-term and long-term paths. Short-term paths are movement paths that can be observed over a short period of time and may only involve a handful of relocations; they occur on a relatively fine scale both spatially and temporally. The movements observed in short-term paths are typically motivated by only one or two factors, which may be internal or external. Long-term paths are observed over a longer period of time (days, years, or a lifetime) and space, and are a combination of many short-term paths. Long-term paths are composed of many relocations and will include most possible movement motivations relevant to the individual in motion (Nathan et al., 2008). These paths are more difficult and time consuming to measure than short-term paths because of the higher temporal and spatial scales associated with them. An understanding of the short-term movements of a species in different contexts, and the movement motivations relevant to that species, can be used to infer the movement steps that sum up to an entire long-term movement process though.

Rare events, caused by uncommon behaviours or chance (e.g., strong winds, human-mediate dispersal) can cause individuals to move much further distances than they would be expected to based on short-term movement knowledge alone. These events are difficult to observe and difficult to account for in movement models (Clark et al., 1998; Gippet et al., 2019). Depending on the species, these rare, difficult to observe movements may be more, or less, important to long-term movements than observable short-term movements are. To completely understand long-term movements, the relative importance of difficult to observe and observable movements to overall long-term movement should be understood. With an understanding of long-term movement patterns, we have the capacity to understand how individuals and populations use their entire landscapes, and the rate at which they may move through them and into different areas.

Dispersal and metapopulations

The availability and location of resources is a main driver of movement for individuals engaged in active movement. The overall layout of resources in space make up a mosaic of hospitable and inhospitable areas, where hospitable areas have the necessary resources in them to support a population of a given species, and inhospitable ones do not (Nowicki et al., 2014). Moving forward, hospitable areas will be referred to as “patches” and inhospitable areas will be referred to as the “matrix” (Hanski & Gilpin, 1991; Nowicki et al., 2014). Individuals will spend most of their time in patches, but will avoid or pass through the matrix (Forman & Godron, 1981, 1986). The distinction between how individuals move through each of these areas is crucial in understanding movement patterns in both the short- and the long-term.

Highly fragmented landscapes are made up of many separate patches, sometimes connected by corridors, in which each support their own populations. In theory, these patches are delineated by characteristics of the habitat they are in, but practically they are determined by the general movement of the individuals living in them. Although a single area of contiguous habitat may appear to be its own patch upon observation, if two groups of the same species live within that area but do not interact with each other it is actually two patches, and if individuals spill outside of what appears to be a patch and move through the area surrounding it frequently, the patch is larger than it appears to be (Hanski, 1982). Understanding movements made over short time periods are therefore not sufficient to truly understand patch structure for a species. Long-term patterns of movement must be understood as well to be able to accurately identify patches for a given species (Girvetz & Greco, 2007). Some individuals do leave the patches that their original populations occupy though, through dispersal (Nathan et al., 2003). These movements can result in new populations being established in unoccupied patches, in individuals never encountering another patch and dying in the matrix, or in individuals encountering already occupied patches and interacting with the population already established there. When individuals from different populations interact with each other on a regular basis via displacement of individuals, the interacting populations make up a metapopulation (Hanski & Gilpin, 1991). These movements between populations are as important to population dynamics as birth and death rates are, and metapopulations can contribute to the persistence of a species in a geographic area when habitat fragmentation and destruction occur (Hanski, 1998).

Butterflies as a model organism to study the relationship between short- and long-term movement patterns

The European common blue butterfly *Polyommatus icarus* Rottemburg, [1775] (Lepidoptera: Lycaenidae), is a butterfly native to Europe and Asia which was first detected in Canada in 2005 (Hall, 2007). *P. icarus*' lifecycle consists of four stages: egg, larva, pupa, adult. The eggs are laid on host-plants and after hatching, larvae feed on host-plant leaves. After going through the larval and pupal stages, an adult butterfly emerges from its chrysalis and will typically live from two to four weeks, depending on environmental conditions and resource availability. Male and female *P. icarus* adults feed on the nectar of a variety of Fabaceae flowers, most notably, *Lotus corniculatus*. *P. icarus* is a multivoltine species, producing multiple broods (generations) per summer. Depending on the local climate conditions, the species is known to produce 2-3 broods per year (Dexheimer, 2021; *iNaturalist*, 2024).

Given *P. icarus'* recorded preference for *L. corniculatus* as a host-plant and this plant's abundance throughout Canada, often in highly fragmented and urbanized areas where other plants cannot grow, it is possible that *P. icarus* will rapidly expand its range in North America (Dexheimer & Despland, 2023; Fischer et al., 2022; Gutiérrez et al., 2001; Rivest & Kharouba, 2021). *P. icarus'* relatively recent introduction to North America, the fact that it is known as a low-moderate disperser but has expanded its range in North America already, and the species' expansive native range, spanning many different temperature and weather gradients make it an interesting study species to examine whether long-term movements can be predicted by short-term movements (*iNaturalist*, 2024; Rivest & Kharouba, 2021).

The Lepidoptera Order of insects includes butterflies and moths. In their mature forms, the majority of these winged insects feed on nectar from flowers, while immature forms feed on leaves (Altermatt & Pearse, 2011). Like many insect species, males seek out females for mating opportunities, and females oviposit (lay) eggs, but do not care for their young other than deciding where to oviposit. Male and female butterflies have different reasons to move; males move primarily to find mates and feed, while females move to feed and oviposit in suitable habitat. The sexes exhibit different movement patterns within the same species in many cases, with females being less active overall, but being more likely to cross boundaries and move through non-habitat than males (Altermatt & Pearse, 2011; Evans et al., 2020; Kallioniemi et al., 2014).

Although mature and immature butterflies both feed on plants, the leaves that immature butterflies eat and the flowers that mature butterflies use for nectar do not always come from the same plant. In some species, immature and mature butterflies do not feed on the same plant, while in other species, including *P. icarus*, they do, and some butterfly species, such as *Pieris rapae* and *P. icarus*, will feed on many different plants (generalists), while other species, such as *Callophrys rubi* and *Pyrgus malvae*, will only feed on a few plants (specialists)(Altermatt & Pearse, 2011; Dennis et al., 2017). In species where immature and mature forms do not feed on the same plants, or only share some plants in common, females may choose to allocate their energy towards ovipositing on plants that will be suitable for her offspring as larvae, even though seeking out those plants will not benefit her directly in any way (Evans et al., 2020). In many species, females will therefore exhibit different short-term movement patterns when they are moving to feed and when they are moving to oviposit (Evans et al., 2020). The differences in short-term movements between male and female butterflies and between females in their feeding phase compared to their oviposition phase are important to consider when examining long-term movement in butterflies. While many metapopulation models view dispersal as random movements, dispersal is made up of short-term movements, which we know are not random in butterflies and many other individuals engaged in active movement (Altermatt & Pearse, 2011; Heinz et al., 2006). It is therefore important to consider the motivators and predictors of short-term movement when trying to understand the dispersal of individuals engaged in active movement.

Chapter 2: Examining the movement patterns of the European common blue butterfly (*Polyommatus icarus*): A study on short-term and long-term movement of an invasive species

Introduction

Different species have different resource and spatial needs, and will move through an area differently, depending on these needs. While an area may look the same on the surface, that same geographic area can be divided into different habitable patches depending on the species of interest (Forman & Godron, 1981). To understand how an individual engaged in active movement uses the space it occupies, and what makes an area a habitatable patch for that species, we must understand how individuals of that species move in response to the presence, or absence, of resources. Given that movements made over a longer time period are made up of short-term movements, in theory a complete understanding of short-term movements should allow for an understanding of long-term patterns of movements across landscapes when no unusual movements contribute to long-term movement (i.e., human-mediated dispersal, strong gusts of wind veering an individual off their usual path) (Nathan et al., 2003, 2008).

Short-term movements can be used to understand the habitat preferences and short-term movement motivators, or short-term movement predictors, of a species. While long-term movements are made up of many short-term movements, some types of short-term movements are rare and/or difficult to observe (Clark et al., 1998). The ability to infer the long-term movements of a species based solely on that species' short-term movements depends in part on the proportion of that species' short-term movements that are easily observable. If many individuals take part in difficult to observe short-term movements, inferences made about long-term movement based on short-term movements will be less reliable than if most short-term movements are observable. For many species, long-term movements include rare, difficult to observe movements; whether we can infer long-term movement from many short-term movement observations depends on our knowledge of these rare movements, and how frequently they occur.

Dispersal is the type of movement that describes the movement of individuals away from their natal locations, and is associated with the spread of a species through space (Nathan et al., 2003). Dispersal is a long-term movement type, as it typically takes place over a long period of time, which makes it a difficult movement process to observe directly (Turchin, 1998). This type of movement influences population interactions and species range expansions, which makes it a movement of particular interest as invasion ecology, the study of human-mediated introduction of organisms, becomes a more prominent area of study (Diniz et al., 2019; Richardson & Pyšek, 2008; Van Dyck & Baguette, 2005). Within invasion ecology, there are different definitions for the term “invasive”. The term “invasive species” can be used to describe a species that is not native or endemic to an area it is found in that has negative economic or ecological effects on the new area it is found in (e.g., spongey moths in North America) (Gippet et al., 2019), or a non-native species that spreads at a high rate in its newly established area (e.g., goldfish in Canadian waterways) (Ricciardi & Cohen, 2007). For the purposes of this study, “invasive” will be used to describe a species that spreads at a high rate through non-native areas.

Step-selection functions (SSFs) are a tool used to predict patterns of short-term movement in animals based on habitat quality and preferences; they model the probability of an individual choosing to move to a given location based on its habitat quality (see Methods for more detail) (Avgar et al., 2016). SSFs can be used to understand short-term movements through a landscape based on the species’ preferences by modelling the probability of an individual moving from one location to another based on the quality of the environment (e.g., food preferences or availability, preferred ground cover) and the species' movement patterns (e.g., distance between locations, turning angles) (Thurfjell et al., 2014). SSFs do so by comparing the attributes of locations used by an individual to those of nearby locations not used by an individual to estimate the relative importance of habitat features and barriers for a given species (Panzacchi et al., 2016). The scale which SSFs can be used at depends on the species being studied, but they are most commonly used to understand the short-term movements influenced by individuals' decisions and preferences as they move throughout their landscape (Avgar et al., 2016).

The European common blue butterfly (*Polyommatus icarus* Rottemburg, [1775]) (Lepidoptera: Lycaenidae), is a butterfly native to Europe and Asia which has been introduced to Canada. The first noted sightings of this species in North America occurred in 2005 near the Mirabel airport, about 50 km north of the city of Montreal, Quebec (Hall, 2007). Since then, it has become the most common butterfly species in Montreal and has spread throughout Quebec, and into Ontario and the Northeastern United States (Figure S1) (*iNaturalist*, 2024; Rivest & Kharouba, 2021, 2024). Although there is no evidence to suggest that *P. icarus* has had any negative effects on North American ecosystems so far, its rapid rate of spread over the past 20 years classifies it as an invasive species.

P. icarus is a generalist species and uses a variety of Fabaceae plants for oviposition and feeding. These host-plants are native to the same areas as *P. icarus* but have long been naturalized in North America, their populations sustain themselves without human intervention (Pyšek & Richardson, 2006). They are weedy species in North America, meaning that they are persistent, reproduce in large numbers and grow in disturbed areas (Aarssen et al., 1986; Turkington et al., 1978; Turkington & Burdon, 1983; Turkington & Cavers, 1979; Turkington &

Franko, 1980; USDA NRCS Idaho Plant Materials, n.d.; Wang & Sakiroglu, 2021). As such, they tend to grow in disturbed areas with no clear ownership, such as parking lots, parks, and along roadways. The distribution and lack of regulation of these host-plants throughout North America indicates that *P. icarus* could continue to expand its range rapidly across its invasive range. *Lotus corniculatus*, bird's foot trefoil, is the most important host-plant for *P. icarus* in both its native and invasion ranges (Dexheimer & Despland, 2023; Fischer et al., 2022; Gutiérrez et al., 2001; Rivest & Kharouba, 2021). *P. icarus* also has a preference for oviposition on short vegetation in both ranges (Dexheimer & Despland, 2023; Habel et al., 2016; León-Cortés et al., 1999). Relatively short vegetation and *L. corniculatus* presence can both be used as predictors of *P. icarus* presence in both its native and invasion ranges. These are typical characteristics of heaths and grasslands, such as traditional grazing pastures in Europe and Asia, but also of infertile, saline and acidic areas, such as fragmented landscapes and cities, in North America (Pellissier et al., 2012; Turkington & Franko, 1980).

The difference in presence of these characteristics in each range is reflected in the difference in distribution of *P. icarus* in its invasive range compared to its native range. Although *P. icarus* is found widely throughout Europe, rural habitats are more suitable for this species in its native range, and its populations have been found to be in decline in Western Europe at a higher rate in cities than in rural areas (Cowley et al., 1999; Dennis et al., 2017). In North America however, records of *P. icarus* show its populations to be found mostly, though not exclusively, in and around urban areas, and its abundance has been found to increase in increasingly urbanized parts of Montreal, Quebec with moderate levels of disturbance (*iNaturalist*, 2024; Rivest & Kharouba, 2021). A study conducted on *P. icarus* in and around Montreal in 2021 found its abundance to be highest in areas with moderate levels of habitat disturbance, and for its abundance to increase in areas with higher urbanization and less forests. They also performed a MRR study, where they recorded a 2% recapture rate and a maximum displacement of 860 m and a mean displacement of 75 m (Rivest & Kharouba, 2021). This study did not examine the short-term movements of this low-to-moderate dispersing non-native species though. Nor did it examine the behaviours that may be contributing to its overall dispersal since its arrival to North America in 2005, or the specific movement patterns that may allow it to continue to expand its invasive range. Unlike with larger organisms with longer lifespans, these butterflies cannot be tracked with existing telemetric approaches (e.g. GPS or acoustic tags) to see exactly how they move through space in real-time (Thurfjell et al., 2014). So, we aim to fill this gap by studying both short- and long-term movement patterns of *P. icarus* populations in Montreal in relation to habitat characteristics and ground cover to understand if short-term movements can be used to predict long-term movements.

To do so, we have asked what habitat characteristics can predict short-term movement in female *P. icarus* individuals? As well as whether observable, short-term, female *P. icarus* movements can be used to predict the species' long-term movements, including dispersal. We predict that short-term movements will be predicted by the habitat characteristics mentioned above: increased host-plant presence, most notably *L. corniculatus*, and short vegetation. We also predict that the short-term movements observed in females will be reflected in the long-term movements measured of both sexes, when patch characteristics are considered, demonstrating potential for short-term movements to be accurate at predicting long-term movements for this butterfly species in North America. If short-term movements can be used to predict long-term movements in this species, we will know that we understand its movement patterns well and can predict dispersal. If not, we will have evidence that other factors other than easily observable short-term movements contribute to this species' long-term movements.

Methods

Summary of methods

This study occurred in two separate but connected parts, which are described in more detail below. While determining the short-term movement predictors of *P. icarus* did not require information collected from long-term movements, determining whether short-term movements can be used to predict long-term movements required information from both the short-term and long-term data collected. Figure 1 shows how these pieces of information came together, to estimate a movement matrix based on habitat qualities and *P. icarus* preference and compare the predictions of this matrix to mark-release-recapture data. In this study, the term "preference" refers to revealed preference, the behaviours that demonstrate an individual's choices based on the available options available, not preference related to a conscious awareness of the choices being made (Sen, 1973). The statistical analysis section goes into further detail.

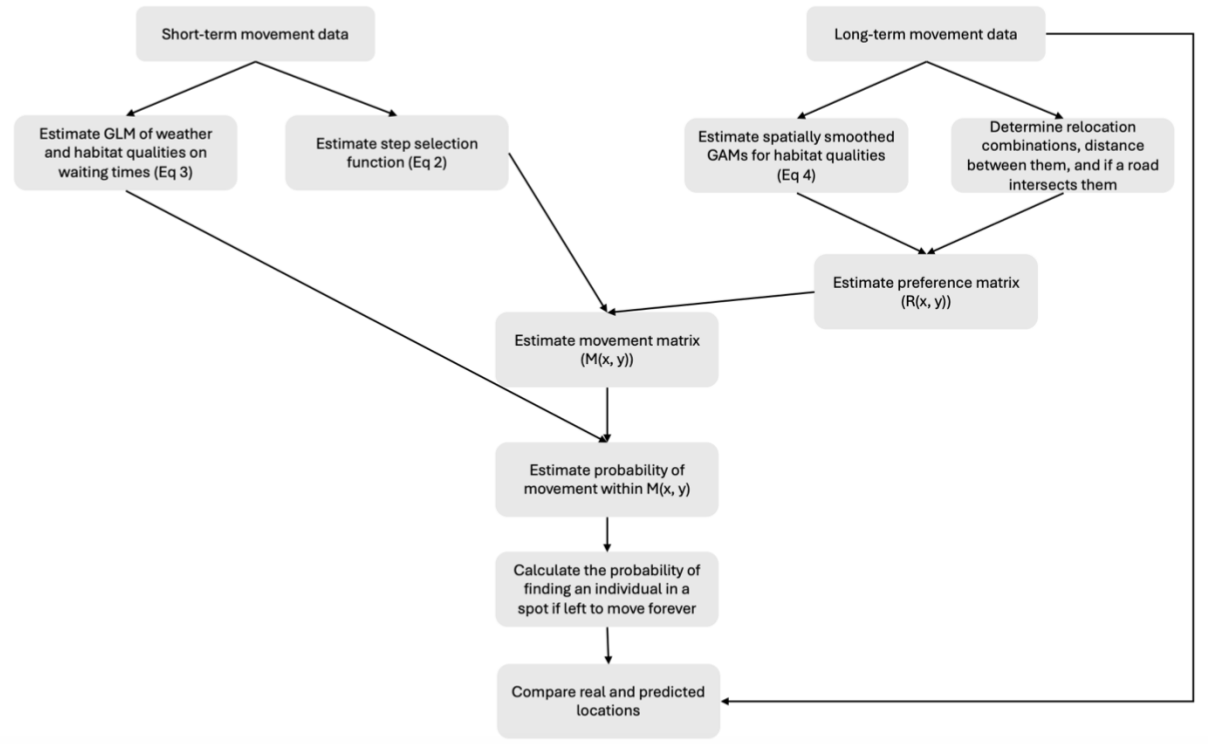


Figure 1: Framework of methods used throughout this study, demonstrating the ways which different elements were used to understand whether habitat-based short-term movement predictors, collected by observing individuals, can be used to predict long-term movements of *P. icarus*, based on a mark-release-recapture study. Each box represents a step in the methodological process, and arrows indicate the start of a new step.

Study areas

We collected data from May 2023 to August 2023 in Montreal, Quebec at two study sites: the grounds of the McGill University Health Centre (MUHC) (45.47ºN, -73.60ºW) and a segment of 1.4 km along a canal (45.45ºN, -73.58ºW). We chose sites based on *P. icarus* sightings in the Montreal area posted on the iNaturalist platform (*iNaturalist*, 2024) as well as based on vegetation coverage and site layout observations made in April 2023 prior to the start of butterfly data collection. These sites are approximately three kilometers from each other but are separated by two major highways and two canals, leading to a very low chance of inter-site movement of the individuals studied, based on maximum displacement distances recorded in previous studies (Gutiérrez et al., 2001; Kuussaari et al., 2014; Rivest & Kharouba, 2021). The MUHC and the aqueduct are both within six km of Montreal's downtown core (Figure 2a), but these sites are made up of grassy areas, some of which are mowed regularly, while others are left to grow for approximately a month at a time. We identified the seven host-plants of interest (*Lotus corniculatus, Trifolium pratense, Trifolium repens, Melilotus albus, Medicago sativa, Medicago lupulina, Vicia cracca*) at both sites. Both sites have bike paths, sidewalks, and roads separating vegetative areas from each other.

We chose these study sites because they represent two different spatial arrangements of habitat; MUHC's vegetative areas are clustered around the hospital building (Figure 2b), while the aqueduct's vegetative areas are long and linear (Figure 2c). We chose sites with different spatial arrangements to see if *P. icarus* may move differently in areas with different patch arrangements. We divided both sites into multiple zones to ensure that equal effort was given to all parts of each site. These zones ranged in size from 2500 m^2 to 10 000 m^2 and were separated based on geographic features which could be easily identified without having to take exact GPS location data, including roads and distinct landmarks. The preferred geographic features for dividing the sites into zones were roads, because we considered these to be part of the matrix for *P. icarus*. This makes roads a relevant geographic feature to the species being studied, and not just to humans. However, to keep the zones within a comparable range of sizes, we used specific trees and electrical poles to delineate zones when no roads were conveniently placed. We collected both short- and long-term data at both these sites to answer our research questions.

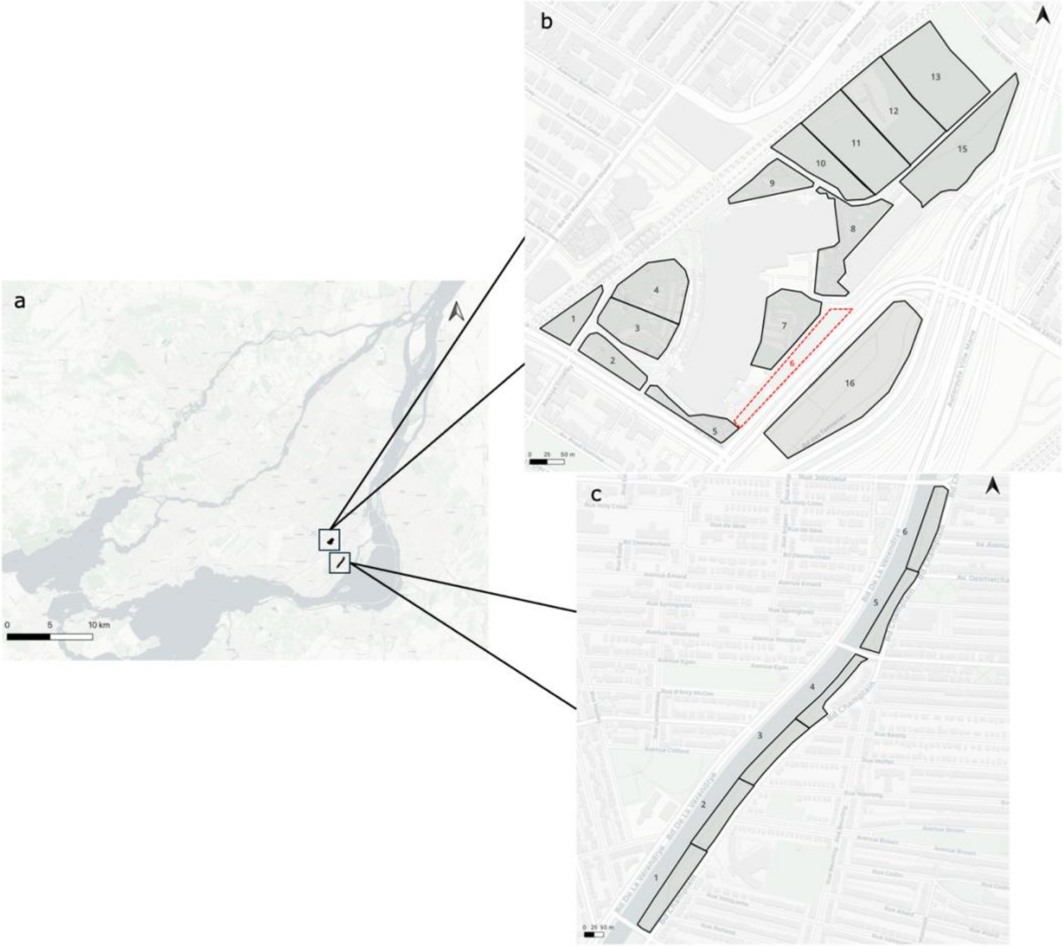


Figure 2: Map of study sites and their location in Montreal (a), with the MUHC above (b) and the aqueduct below (c). Outlined areas on the study sites represent sampling zones and are numbered based on how they were designated during sampling. Zone 6 at the MUHC site is outlined in red, as it was planned in a location that was not accessible for sampling, and Zone 14 at the MUHC site does not exist as it was absorbed into Zone 13.

Data collection

Short-term movement

We measured short-term movement by following female *P. icarus* individuals and making note of their landing locations for up to 15 minutes or ten landing locations (steps), whichever came first. We chose ten steps as this created long enough paths to understand short-term movement predictors, while also lowering the chances of losing track of the individual being followed in that time. We set the 15-minute time limit to decrease the amount of time spent waiting for an individual to fly that simply was not going to. Since we were interested in how the plants surrounding an individual may impact its movement, waiting for more than 15 minutes for an individual to move was not an efficient use of time. Approximately the same number of paths were followed in each site between May and August, with 38 paths followed at the MUHC over 17 days and 42 paths followed at the aqueduct over 14 days (Figure S2).

We chose to follow only females because of the differences in habitat characteristics known to predict female and male movement, and because of the role of female movement in spatial population dynamics (Kallioniemi et al., 2014; Reim et al., 2019).While females primarily move in the interest of feeding or laying eggs, males will move to feed, to defend their territory, and to court females. Given these differences in movement predictors between sexes, it would have made the data collected more difficult to interpret if the paths of both sexes had been followed (Kallioniemi et al., 2014). Further, we only followed females because for a population to form in a new area there must be at least one female in that new area to lay eggs; as such, only female movement will lead to dispersal at the population scale.

After locating a female *P. icarus*, we followed the individual from a distance, marked each spot on which the individual landed with a cone, and recorded the plant species on which the individual landed, and the length of time spent at each step. Once 15 minutes, or ten steps, had passed we recorded the coordinates of each step using a Bad Elf GPS Pro (BE-GPS-2200), which had a horizontal precision of three meters. We then generated alternate steps for each observed step to use for selection (Figure 3a). We created three alternate steps from each observed step, where the observed steps were the spots where the individual landed. To do so, we measured an alternate vector from each landing spot. We used three pre-measured rope lengths of 150 cm, 358 cm, and 750 cm to create three alternate steps from every observed step in the path. The alternate step distances used were based on the distribution of distances of step lengths observed in a pilot study conducted in 2021. The alternate step distances chosen were equal to the 25th, 50th, and 75th quartiles of observed step lengths from the pilot data, excluding outlying distances over 2000 cm. As we did not observe any correlations in turn angles from one step to the next in the pilot data, we determined the direction of each alternate step from each starting location randomly by using a spinner and measuring the rope length out in the direction the arrow pointed to. For each observed and alterative step location, we measured the distance to the nearest cm and absolute angle of the path from the starting point of the step-to-step location.

We recorded information on the ground cover within a 30 cm radius of the center of each observed and alternate step (Figure 3b). The ground cover categories we recorded were the coverage of host-plants in the radius, the coverage of other vegetation, and the coverage of non-vegetation. We characterized ground cover for each type using a discrete scale ranked from 0-3. Where zero represented that 0% of the ground in the 30 cm radius was covered by the variable of interest, one represented 1% - 25%, two represented 25% - 75% and three represented 75% - 100%. We recorded when a host-plant had any flowers within the 30 cm radius. In the case where an alternate step ended up somewhere that could not be safely reached, such as over a fence, or on a very steep slope, we recorded its ground cover as either three "other vegetation" or three "non-vegetation", depending on what we could see from afar. We photographed each observed and alternate step in front of delineated height chart, to determine the maximum vegetation height within the 30 cm radius of each step (Figure 3c).

Long-term movement

We measured long-term movement via a mark-release-recapture (MRR) study. To do so, *P. icarus* individuals were caught using 30 cm diameter sweep nets. Each captured individual was marked with an alphanumeric ID on their outer hind wings using a fine tip permanent, marker (Figure 3d), a method which has been used in MRR studies on *P. icarus* and other butterfly species (Gutiérrez et al., 2001; Janz et al., 2005; Morton, 1982; Rivest & Kharouba, 2021). We noted the sex and the GPS location where each marked butterfly was captured. If it was determined that the ID marking may have been unclear, we photographed the marked wing before the individual was released. After release, we recorded the ground cover information within a 30 cm radius of the catch location using the same ground cover recording system as in the short-term movement study (Figure 3c). We characterized the disturbance state of the location of each capture by noting whether the area had been (1) mowed in the last 3 weeks (approximated by eye), (2) unmowed, or (3) mowed but not in the last 3 weeks. We took ground cover photos periodically throughout the day to use as reference for general ground cover qualities in different locations.

During the process of catching and marking individuals, we recorded the ID number of any butterfly that had been previously marked (i.e. recaptured). To avoid handling individuals too frequently and underestimating dispersal, we only noted recaught butterflies if they had been marked on a different day (Gutiérrez et al., 2001; Morton, 1982). We recorded the same information when an individual was re-caught as during its initial capture, as well as an indication that the butterfly was a recapture. We photographed the individual's marked wing if unclear what the alphanumeric ID was. We then compared these photos to those taken when individuals were initially marked, to confirm ambiguous IDs.

To ensure equal sampling effort across sites, we controlled for MRR sampling effort within each zone of both study sites by setting a maximum sampling time per zone. We did so by setting the sampling time for a zone equal to the area of each zone (measured in m^2) divided by

70 in minutes, and rounding this value the nearest 15 minutes. We determined this time by estimating the amount of time needed to sample the smallest zone, and then scaling up from that time for the larger zones. For all zones, if a *P. icarus* individual was not caught within a 15-minute interval, we moved on to the next sampling zone for the day.

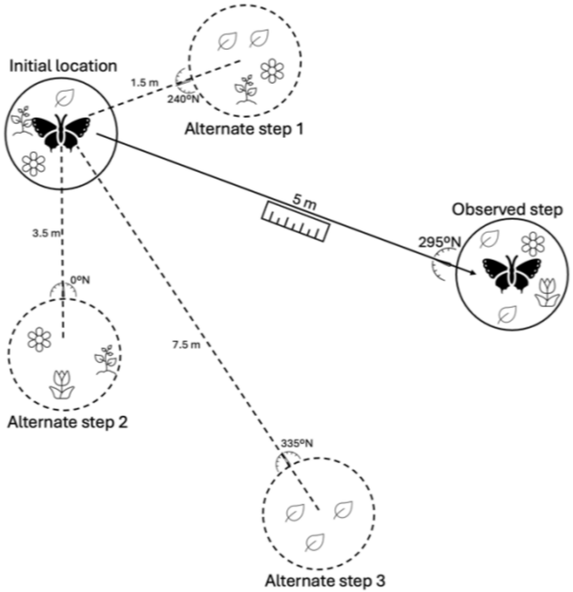
<table>
<tr><th>Description</th><th>Image</th></tr>
<tr><td>(a)
Example of observed and alternate step setup (2/10 observed steps pictured). Observed steps are locations where an individual landed, and alternate steps are locations we determined semi-randomly. Observed movement is represented by a solid line and observed step are represented by a solid-lined circle with a butterfly within them (initial location and observed step), while alternate steps are represented by a dashed-lined circle (alternate step 1, 2, 3) and alternate movement by a dashed line. Angles shown are absolute angles from the initial step, and distances shown are measured from the center of the initial step.</td><td>
</td></tr>
<tr><td>(b)
Example plot illustrating ground cover. The white circle is a 30 cm radius hoop centred on the observed landing location This plot would be characterized as having a score of 3 for host-plants, a score of 1 for non-host-plants, and a score of 0 for non-vegetation.</td><td></td></tr>
<tr><td>(c)
Vegetation height measurement. The white circle is a 30 cm radius hoop centred on the observed landing location. The paper behind the circle is delineated at every 10 cm and was used to estimate the height of plants in the plot.</td><td>
</td></tr>
<tr><td>(d)
A marked *P. icarus.* We marked individuals with permanent markers on their outer-hindwing and made note of their sex and location.</td><td>
</td></tr>
</table>

Figure 3: Description and images of techniques used while collecting data on short and long-term movement patterns. Techniques include alternate step production and measurements (a), ground cover sampling (b), vegetation height measurement (c) and *P. icarus* marking (d).

Weather

While collecting data, temperature (ºC) and wind (km/hr) conditions were recorded based on hourly reporting from accuweather.com (*AccuWeather*, 2023). We estimated cloud coverage based on the percentage of the sky that was concealed by clouds at the beginning of each hour. We noted the intensity and length of time of precipitation when it occurred, and sorted precipitation types into numbered groups, where 1 represented misty conditions, 2 represented sprinkling rain, 3 represented light rain, and 4 represented constant rain. Precipitation did not occur often enough during data collection to be used in any analyses.

Statistical Analysis

All statistical analyses were conducted in R statistical software version 4.2.1 (R Core Team, 2021). We re-classified the ground cover rankings from 0, 1, 2, 3 (as described in the short-term data collection section) to the midpoints of each of the percentage bins each of these values represented, or 0, 12.5%, 50%, 87.5% respectively. For each location of interest (observed steps, alternate steps, and capture locations), we added these percentage values to estimate the overall percent coverage of host-plants in that location, setting a 100% maximum coverage. We added the total number of host-plant species that were flowering to determine the number of host-plant species in bloom in each location.

Short-term movement

We used the short-term movement data collected to build a step selection function with penalized smooths, which can be used in step selection analyses to model non-linear relationships between habitat characteristics and movement without overfitting the data by incorporating penalties (Klappstein et al., 2024). In doing so, we compared the characteristics of observed steps and alternate steps using an SSF. These functions are estimated using conditional logistic regression models and usually determine the lengths and turn angles of the alternate steps from distributions established from observations of tracked individuals. The habitat characteristics of the observed and alternate steps are then determined based on geospatial data to understand what habitat characteristic the species of interest prefers, based on the comparison between the characteristics of the observed and alternate steps using an equation like equation (1) (Thurfjell et al., 2014). Given that the habitat characteristics and the movement distances of interest in this study are very fine scale, we instead collected the distance, angle, and habitat characteristics as described above, equation (1) still represents the basic form used to compare observed and alternate steps to each other.

Equation (1)

$$\Pr\left(L_{t,j} \mid L_{t,1}, L_{t,2}, L_{t,3}, \dots, L_{t,n}\right) = \frac{\exp\left(Pref_{t,j}\right)}{\sum_{j=1}^{n} \exp\left(Pref_{t,j}\right)}$$

This equation calculates the probability of an individual ending up at location $L_{t,j}$ given the alternate possible locations $L_{t,1}, L_{t,2}, L_{t,3}, \dots, L_{t,n}$, where $Pref_{t,j}$ is the preference of an individual to move to that location given its habitat qualities, as estimated by a conditional logistic regression model (Fortin et al., 2005; Klappstein et al., 2024). In doing this, SSFs compare the habitat characteristics of interest in locations used by an individual to those in locations not used, but available, to that individual to understand the habitat preferences of a species and predict where they will move over time (Thurfjell et al., 2014).

We first aggregated observed steps (Figure 4a) from the same flight path into one step if they were less than 1 m away from the previous observed step ($s_{n+1} - s_n < 1$ m). We did this to account for human error in ground cover reporting between steps in very close proximity to each other which should have had the same ground cover recordings, and because our shortest alternate step distance was greater than 1 m. We assigned the highest value of each piece of ground cover information from the steps that were grouped together to form the newly created aggregate step and the geographic information (latitude and longitude) of the first step in the grouped steps to the aggregated step that was kept (Figure 4b). Each aggregated observed step was compared to every alternate step within the same path, not only to the three alternate steps built from that observed step during data collection. This was done because the shortest of the three pre-determined alternate step lengths ended up being longer than 70% of the observed step lengths measured, which made the alternate step lengths difficult to compare to the observed step lengths (Figure S3). Given that all alternate locations in the same path would have been possible landing locations for the individual being followed, and to have more alternate step lengths comparable to the observed step lengths, we decided it was fitting to compare all alternate steps to each observed step.

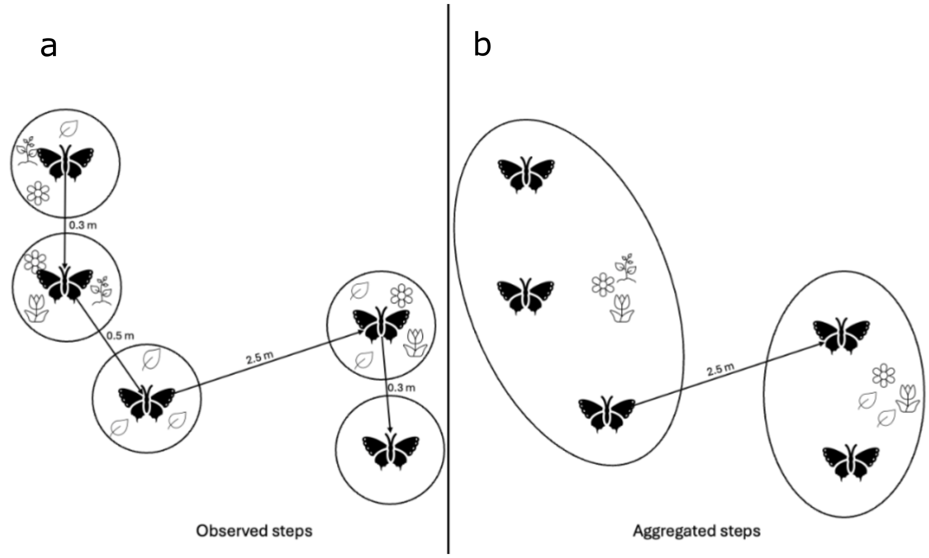


Figure 4: Graphic explanation of how steps less than 1 m were aggregated into one step during data analysis. Panel (a) depicts the observed steps as they were measured during short-term movement data collection, with each circle representing the 30 cm radius plot within which ground cover information was collected in the field, and arrows depict the *P. icarus* individual's sequence of steps, with the distance between each step noted in meters. Panel (b) depicts how multiple steps were aggregated into one step if they were less than 1 m apart from each other, where the circled areas represent how steps less than 1 m from each other were grouped together. In both panels, images of plants represent different species, and in panel b the plant images circled represent that the aggregated ground cover data used the highest ground cover value from all the individual steps for the aggregated step.

To do so, we fit a Cox proportional hazard generalized additive model (GAM) using the *mgcv* package with non-linear smooth functions for habitat qualities. We used the Cox proportional hazard family here because it has recently been shown that it is mathematically equivalent to using a conditional logistic regression model, when hazards are stratified by whether a location represents an observed or alternate step, and performs equally as well when used in an SSF context (Klappstein et al., 2024). The smooth functions used each represent a flexible, non-linear relationship between the local habitat predictor variables and the preference (response). Each smooth function is made up of multiple basis functions, which determine how flexible, or wiggly, the smooth functions will be. We generated these smooth terms using thin plate regression splines (TPRS) and set the basis sizes (k), which define the number of basis functions to be used in the smooth, to values large enough to capture the non-linear relationships between the predictor and response variables, while not overfitting the data provided to the model (Pedersen et al., 2019). The habitat qualities we included in the model were percent host-plant coverage (k = 6), number of host-plant species flowering (k = 5), maximum vegetation height (k=10) and the log-2-transformed distance to the next step, (k=10) and linear effects for whether a step crossed a paved area and the percentage of non-vegetation in a given step (Wood, 2011, 2023).

We experimented with multiple versions of the final model used (Equation (2), Figure 5), including models incorporating the ground cover of each host-plant species as separate predictors, separate models for the two study sites, and modelled with different subsets of the final variables used. We also varied the basis sizes, and transformations of the distance variable used. To evaluate the different models, we considered how well each model's outputs aligned with existing knowledge of this butterfly species and movement. We considered the magnitude and direction of each covariate's coefficient, and made sure that the final model's outputs were realistic and supported by field observations. After doing this, and finding comparable AIC values between the different models estimated, we found that equation (2) fit the data best by producing results consistent with what was observed in the field and in the literature.

Equation (2)

$$Pref_i = f_1(HostCover_i) + f_2(HostFlowers_i) + f_3(VegHeight_i) + f_4(\log 2(distance_i)) + \alpha_1 \times NonVegCover_i + \alpha_2 \times RoadCrossing_i$$

The response variable, preference, is on an exponential scale and indicates the likelihood of an individual to take a step towards a location with x amount of the covariate in question, when the rest of the covariates are held constant. This is more intuitively understood if we compare the preferences for two possible locations for a single step. If one location has a preference value of $Pref_1$, and the second location has a value of $Pref_2$, then the probability that an individual chooses location 1 would be: $\Pr(L_1 \mid L_1, L_2) = \frac{\exp(Pref_1)}{\exp(Pref_1)+\exp(Pref_2)}$. If $Pref_1 > Pref_2$,the first step will be preferred to the second step, if the values are equal, the butterfly is equally likely to travel to either location in that step, and $Pref_1 < Pref_2$, the butterfly is more likely to end up in the second location rather than the first (Klappstein et al., 2024). For two locations that have equal habitat quality values for all but one variable, the functions $f_1, f_2, \dots f_n$ from equation (2) estimate the relative probability of an individual selecting one location over the other, based on the relative preference for the differing variable. This makes it possible to isolate the effect of only the differing variable on the short-term movement pattern observed.

To estimate the distribution of waiting times (that is, how long each butterfly waited in a given location before moving to a new one), we fit a generalized linear model (GLM) of waiting time as a function of local ground cover characteristics and the weather information recorded during the same hour to examine if the weather had an impact on the length of time between consecutive movements. We used a log-link Gamma distributed GLM and set the time spent in total at each aggregated step as the response variable, and temperature (ºC), cloud cover (%), wind (km/hr), percent host-plant coverage, number of host-plant species flowering, and maximum vegetation height as the predictor variables.

Equation (3)

$$Waiting_i \sim Gamma(\mu_i, \phi)$$

$$\log(\mu_i) = \beta_0 + \beta_1 \times HostCover_i + \beta_2 \times HostFlowers_i + \beta_3 \times VegHeight_i + \beta_4 \times Temperature + \beta_5 \times Wind + \beta_6 \times CloudCover$$

Where $\sim Gamma(\mu_i, \phi)$ indicates that we assumed that the waiting time for a butterfly to leave location i was distributed following a Gamma distribution with a mean value of μ_i and a scale parameter of ϕ (Avgar et al., 2016; Klappstein et al., 2024). We used this equation format as recommended by Zuur & Ieno, 2016.

Long-term movement

To estimate the long-term movement of *P. icarus* individuals in the landscapes studied, we used the habitat preferences and waiting time distributions estimated (Equation (2), Equation (3)) to estimate a model of long-term movement as a continuous time Markov chain movement model (Hanks & Hughes, 2016). This is a stochastic process model defined in continuous time on a discrete space, which incorporates habitat characteristics to understand the effect of directional bias, directional persistence, and movement rates on movement behaviour. It does so by considering the sequence of locations that individuals move through, in a discrete gridded space, and the time spent in them, following an exponential distribution, to model the movement of individuals overtime between areas with different habitat qualities. These models consider the known location of individuals, they do not consider past possible locations, and their habitat preferences to understand where individuals may move through time (Hanks et al., 2015; Hanks & Hughes, 2016). Using a continuous time Markov chain movement model allowed us to account for the probability of movement between different locations of each study site over time, considering both the quality of each location and the time spent in each location. By incorporating location-specific waiting times based on habitat quality, the model takes the non-linear nature of movement into account, where individuals may spend more time in high-quality patches and avoid low-quality areas altogether.

We used the ground cover information collected over the course of the field season to build spatially smoothed GAMs using the *mgcv* package to predict the quality of every 2 m x 2 m location from each study site to account for exact locations of the sites which we did not sample. We modeled the spatial variation of each ground cover variable in each study site (k) as a normally distributed variable with a local mean value of $\mu_{k,i}$ and standard deviation σ_k as shown in equation (4). We modeled the mean ground cover variable using a two-dimensional thin plate regression spline with 30 basis functions (Wood, 2023).

Equation (4)

$$GroundCover_{k,i} \sim Normal\left(\mu_{k,i}, \sigma_k\right)$$

$$\mu_{k,i} = g_k(northing_i, easting_i)$$

We then used these spatially smoothed GAMs to generate maps of estimated ground cover at a 2m x 2m spatial resolution for both the MUHC and aqueduct sites (Figure S4, Figure S5). This resulted in estimated ground cover variables in a lattice of 40142 cells for the MUHC and 11652 cells for the aqueduct. Using these ground cover maps, and the model estimated in equation (2), we calculated an estimated latent local “quality” function variable Qual(x) for each location *x*, using only the local ground cover and height variables for each lattice cell. This was done using the predicted values estimated by the GAM model described by equation (4) and using equation (2) to estimate a preference value for each location *x*, with the distance and road-crossing effects in the model set to zero, so that only the local preference value of each location *x* was calculated. We estimated a distance factor $\left(D(x,y)\right)$ to account for the preference to move the distance from a start to an end location using a displacement distribution estimated from the distribution of observed step distances, assuming that log-2-transformed step-lengths were distributed following a Gamma distribution. We used a distance kernel function $K\left(Dist(x,y)\right)$

to represent the probability of moving a certain distance between *x* and *y*. As the distance between locations increases, the number of locations available to an individual increases as well though, so the probability of ending up at any specific location decreases. To account for this, in equation (5) the kernel function is divided by the distance (Klappstein et al., 2024). Similarly, we considered whether a road would have to be crossed ($Road(x, y)$) to move from the start to the end cell by assigning values of zero to combinations which did not have to cross a road, and values of -3.7 to combinations that did have to cross a road, which was the coefficient value for road-crossing calculated by equation (2).

We estimated a latent preference matrix $Pref(x, y)$ for each combination of cells *x* and *y*, where $Pref(x, y)$ indicates the latent preference score for location *x*, starting from location *y*. Entries for this matrix were calculated using equation (5) and the distance ($D(x, y)$), road crossing ($Road(x, y)$), and quality values ($Qual(x)$) calculated in the previous step.

Equation (5)

$$Pref(x, y) = Qual(x) + \log \frac{K\big(Dist(x, y)\big)}{Dist(x, y)} + f_7(Road(x, y))$$

Following this, we generated an n-by-n matrix of relative step preferences, R(x,y), where the value R_{xy} in the matrix represented the estimated probability of ending up in cell x, given the individual started in cell y. To calculate R_{xy} we translated these preferences into entries in the relative preference matrix R(x,y) using equation (6):

Equation (6)

$$\underset{x \neq y}{R(x, y)} = \frac{e^{Pref(x,y)}}{\sum_{z \neq y} e^{Pref(z,y)}}$$

We then translated this relative preference matrix R(x,y) into a movement matrix M(x,y) that gives the estimated instantaneous rate of movement from location y to location x in each unit of time. We used mean waiting-time values calculated by equation (3) to predict how long an individual would spend in each 2 m x 2 m cell before leaving that cell based on the habitat quality of that cell, to give an estimated waiting time for each starting location W(y). To generate M(x, y), we multiplied each column of the preference matrix R(x, y) by the corresponding waiting-times W(y) calculated for each starting cell y, to understand the likelihood of an individual moving from one 2 m x 2 m cell to the cells surrounding it. Finally, we set the diagonal of M(x, y) to the negative values of its column sums, to ensure that the column sums of the matrix were all equal to zero. This added an assumption that an individual would not choose to leave a location and then come back to the same location. Matrix M(x, y) cell values then represented an individual's preference to move from one location to every other accessible location within the study site.

The matrices we created were sparse because only *x, y* cell combinations which represented locations that were not roads or buildings and were within 40 m of each other were filled out, due to computing capabilities. 40 m is approximately twice as long as the longest observed step recorded during the short-term movement data collection, which is why 40 m was chosen here, to keep all locations that could realistically be reached within one relocation. We excluded roads and buildings given that these are non-habitat for butterflies, and the preference values calculated on a road or building would be 0.

If the movement of individual butterflies across long periods of time is just due to the cumulation of multiple displacements due to habitat selection with waiting times between steps predicted by equation (3), then the matrix M(x, y) can be used as a generator matrix M of a

Continuous Time Markov Chain (CTMC) random walk model (Hanks et al., 2015; Hanks & Hughes, 2016) that can be used to predict the probability $Pr(x|t, x_0)$ that a butterfly that started at location x_0 might be found at any given location x on the landscape after any given length of time t. As noted in Hanks and Hughes (2016), the transition probability matrix $P_t(x, y)$ (i.e. the probability of moving to location y after time *t* having started from location x) for a CTMC model characterized by a generator matrix M is given by equation (7):

Equation (7)

$$P_t = \exp(Mt)$$

Where the function exp(Mt) represents the matrix exponential of the matrix M scaled by the length of time passed (Otto & Day, 2007). While calculating the exponential for a large matrix is prohibitively computationally expensive (Hanks & Hughes, 2016), it is possible to calculate an approximate value for P_t by using a relatively small number *k* of the left ($\psi_{1:k}$) and right ($\phi_{1:k}$) eigenvectors of the matrix M, corresponding to the *k* eigenvalues $\lambda_{1:k}$ that have the largest real parts (Eric Pedersen, personal communication, 2024; Noé & Clementi, 2015). The approximate probability of moving from location x to location y after time t is then given by $\tilde{P}_t$ in equation (8):

Equation (8)

$$\tilde{P}_t = \phi_{1:k} \exp(diag(\lambda_{1:k}) \times t)\, \psi_{1:k}$$

We used equation (8) and the 50 leading eigenvectors from M(x,y) to calculate the estimated approximate probability Pr(y|x,t) = $\tilde{P}_t(x, y)$ of moving from the cell that a given butterfly was first captured (y) in to the cell where the butterfly was recaptured (x), setting t equal to the number of minutes that had passed from when the butterfly was first observed to when it was recaptured. The time (t) between capture and recapture events was divided by two, assuming individuals only moved during 12 hours of the day, not 24, to account for the low movement of butterflies overnight. The effect of weather on movement was not considered in this model because data collection could not occur on cool or cloudy days, given that butterflies do not move much in these conditions (Evans et al., 2019). As a result of this, movement distances may be somewhat over-estimated, but likely not by a large enough extent to have any major effect on the result given that there were not many cool and cloudy days.

We used the value of the eigenvector of M(x, y) associated with the eigenvalue with the largest real value as the estimated probability of a *P. icarus* individual being found in a given cell over time, and the probability of finding an individual at a given location within the cell if left to move for an infinite amount of time (Otto & Day, 2007; Wilson et al., 2018). Finally, we compared these predictions to the observed long-term movement collected via the MRR study to see if the short-term movement collected would accurately predict where individuals may end up. We did this by dividing the probability of an individual moving from the location it was marked at to the location it was recaptured in during the amount of time between mark and recapture by the probability of finding an individual at a given location within the site if left to move for an infinite amount of time. This resulted in a value known as the likelihood ratio.

Log-likelihood ratio values greater than zero indicated that the movement model predicted that this individual was more likely to have ended up in the recapture location given its starting location and time between captures. Log-likelihood ratio values equal to zero indicated that it was equally likely to find a butterfly in a given location under the time-dependent movement model or the long-term distribution model, and log-likelihood ratio values less than zero indicated that the probability of recapturing a butterfly in a given location was lower in the time-dependent model compared to the long-term model predictions.

Results

Short-term movement

We followed 80 female *P. icarus* paths to understand their short-term movement predictors. See Figure S6a for all followed path locations at the MUHC, and Figure S6b for an example followed path layout. Of these 80 females, only two of them crossed a road during the monitoring period. Using the model described in equation (2) described above, we found that increased host-plant coverage (Figure 5a), number of host-plant species in bloom (Figure 5b) and vegetation height (Figure 5c) increase the preference of an individual to move to a given location, with host-plant coverage and host-plant species in bloom having a larger effect than vegetation height. Non-vegetation coverage (Figure 5d) increased the preference of an individual to move to a given location up to a certain point, after which the preference to move decreased, but remained higher than when no non-vegetation was present in a location. Locations that were a further distance (Figure 5f) from each other or required a road to be crossed (Figure 5e) were less likely to be moved to than nearby locations with no road to be crossed, indicating that individuals do not tend to cross roads, or move long distances between landing locations. All relationships had p-values less than 0.0001, relative to a null hypothesis of no preference effect for each variable. Similar trends were found when each host-plant species was modelled separately in terms of percent ground cover as when all seven were modelled together.

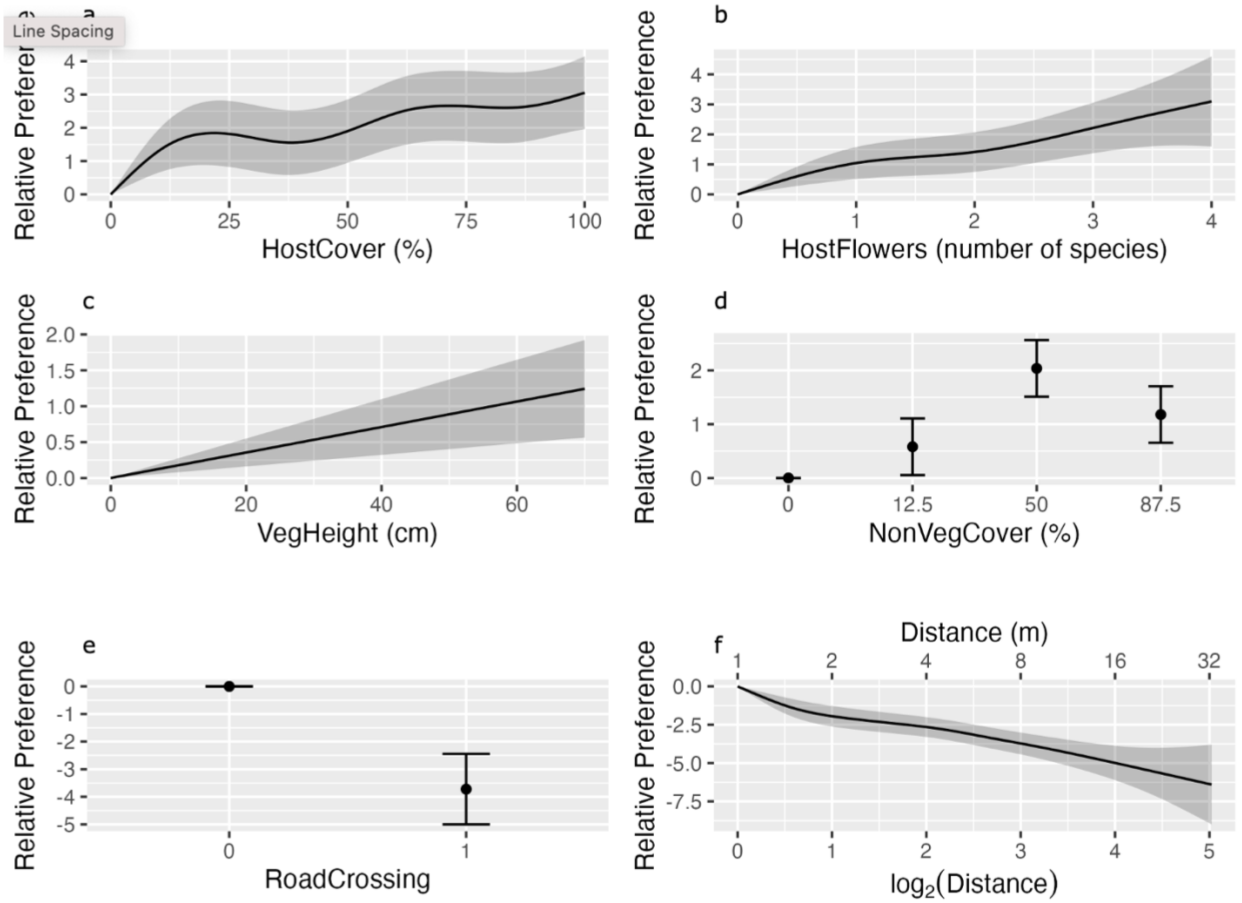


Figure 5: Estimated smooth, using thin plate regression splines (TPRS), and parametric terms of various habitat quality variables on the relative preference of a female *P. icarus* to move to a given location. Y-axes represent the partial effect of preference to move to a location and is represented on an exponential scale. X-axes represent (a) percent host-plant coverage, (b) number of host-plant species in bloom, (c) maximum vegetation height in centimeters, (d) the percent coverage of non-vegetation, (e) whether a road crossing is required, and (f) the log-2 transformed distance to the next landing location, with a supplementary scale above depicting the corresponding distance in meters. Gray bands around the smooth terms and bars around the parametric terms represent the 95% confidence intervals of the estimated function.

We used equation (3) to examine the effects of weather and ground cover characteristics on the time *P. icarus* females spent between landing locations. We estimated that waiting time before leaving cells increased with temperature, wind, host-plant species in bloom, and vegetation height, and decreased with percent cloud cover and percent host-plant coverage. None of the estimated weather or ground cover estimates were statistically significant at the alpha=0.05 level though. It should be noted that sampling occurred during optimal butterfly movement conditions, as females did not fly in cool or windy conditions, and sampling took place during the day, so there was not much variation in the temperature or wind strength throughout the sampling process. Based on personal observation and the literature, butterflies do not move frequently during cold and overcast conditions (Evans et al., 2019).

Mark-recapture results

We recorded 3755 *P. icarus* catching events over the course of the study period in two study sites. Of these, 203 were recaptures of 190 distinct individuals, as we recaptured some individuals more than once. The locations of each recaptured individual and its initial location are shown in Figure 6, with individuals caught at the MUHC in Figure 6a and those caught at the aqueduct in Figure 6b. The distribution of displacement distances is depicted in Figure S7.

Recaptured females had a maximum displacement of 3100 m, but this was due to a single outlying butterfly that was tagged at the MUHC site and detected at the aqueduct site. Excluding this individual, the maximum displacement was 770 m, the mean displacement was 140 m, and the third quartile of displacement was 180 m (Table 1). Females were recaptured between 2 and 14 days after their initial capture. The maximum displacement of recaptured males was 3200 m, but this was again due to a single male that was tagged at the MUHC and recaptured at the aqueduct. Excluding this outlying male, the maximum displacement observed was 1130 m, the mean displacement of males was 150 m, and the third quartile of observed displacements was 160 m. Males were recaptured between 1 and 42 days after their initial capture, or between 1 and 22 days with an outlier excluded. The outlier displacement values for both sexes are from individuals that moved from the MUHC to the aqueduct study site. A Wilcoxon rank-sum test showed that there was no difference in median displacement distances of the males compared to the females (W = 2814, $p = 0.89$).

Table 1: Results of MRR study on *P. icarus* at two sites in Montreal. Results shown include total number of marked individuals, number of recaptures, percentage of recaptures, maximum displacement distance, mean displacement distance, and the third quantile of displacement distances. Results shown for only females, only males, and both sexes combined. Values in parentheses in max displacement column are distance values for two individuals that moved exceptionally long distances compared to other individuals.

	Marked	Recaptured	Recapture %	Max displacement	Mean displacement	75th percentile displacement
Female	1003	33	3.3	770 (3100) m	140 m	180 m
Male	2747	170	6.2	1130 (3200) m	150 m	160 m
Total	3755*	203	5.4	1130 (3200) m	150 m	160 m

*5 individuals' sexes not confirmed

Figure 6: Map of study sites, with the MUHC on the left (a) and the aqueduct on the right (b), with the marking and recapture locations of *P. icarus* individuals that were recaught in the MRR study. Green points represent initial marking location and brown points represent recapture locations. Overall, 48 individuals were recaptured at the MUHC, and 142 individuals were recaptured at the aqueduct.

Comparing estimated movement model with mark-recapture results

We used the leading eigenvector of the movement matrix M(x,y) to calculate results about where individuals would be likely to end up given an infinite amount of time to move about each study site, known as the long-run occurrence probability (Figure 7) (De Alfaro, 1998). Areas estimated to have a higher probability of occurrence in Figure 7 are the areas that based on matrix M(x,y) would be the most likely for *P. icarus* individuals to be found in. Areas estimated to have a higher probability of occurrence in Figure 7a correspond with areas of the MUHC site that generally had high host-plant coverage (Figure S4a), many host-plants species in bloom (Figure S4b), low non-vegetation coverage (Figure S4c), and high vegetation (Figure S4d). Areas estimated to have a higher probability of occurrence in Figure 7b correspond with areas of the aqueduct site that had high host-plant coverage (Figure S5a), many host-plant species in bloom (Figure S5b), mid-range non-vegetation values (Figure S5c), as they are right next to a bike path, and a mix of short and tall vegetation (Figure S5d).

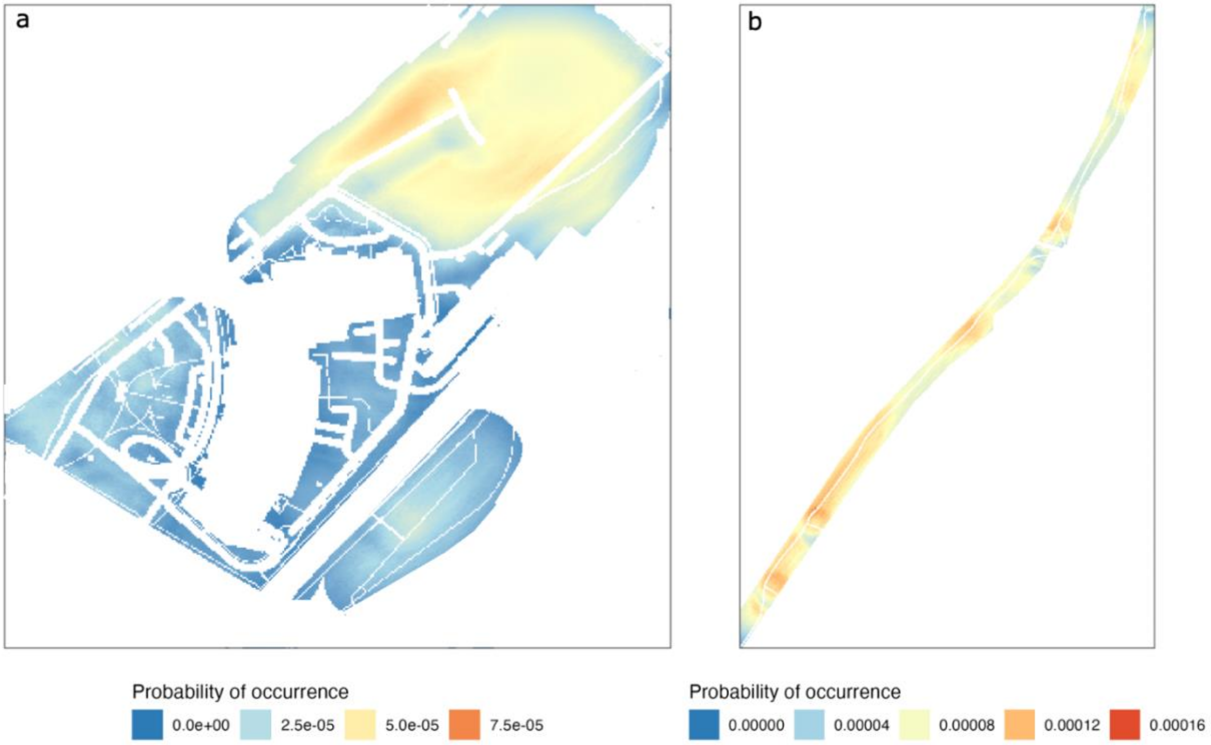


Figure 7: Image of study sites, with the MUHC on the left (a) and the aqueduct on the right (b), representing the predicted probabilities of where a *P. icarus* individual would end up if given an infinite amount of time, regardless of where in the study site its movement began. Blue areas indicate that an individual would be less likely to end up there, while red values indicate that an individual would be more likely to end up there. Note the difference in probability values between the two study sites.

We examined whether the values predicted in our movement matrix corresponded to the real displacements observed through the MRR study conducted by calculating the log-likelihood ratios (Figure 8). Log-likelihood ratio values of 0 correspond to an average probability for an individual to have moved to its recapture location given the individual's starting location, the time passed between mark and recapture, and the habitat qualities in the areas it would have to move through. Log-likelihood ratio values greater than 0 indicate that an individual had an above average probability to have ended up in its recapture location, and log-likelihood ratio values below 0 indicate a below average probability to have ended up in its recapture location.

The log-likelihood ratio values ranged from -20.2- 4.5 at the MUHC site (Figure 8a, Figure 8b) and from -23.0 – 3.0 at the aqueduct (Figure 8c, Figure 8d). As depicted in Figure 8a, at the MUHC site there is no trend between the amount of time passed between mark and recapture events and the probability of an individual to have ended up in the location they did based on our predictions. Figure 8b shows that at this site, further distances travelled between mark and recapture locations do coincide with a lower probability for an individual to have ended up in the location they did, based on our predictions. Similar trends exist at the aqueduct site as depicted in Figure 8c and Figure 8d. This indicates that a longer period between mark and recapture times does not coincide with further displacement at either site, even though more individuals moved further distances at the aqueduct.

A Wilcoxon rank-sum test showed no differences in the likelihood ratio values of the males compared to the females overall or at either study site ($W = 2925$, $p = 0.6$), ($W_{MUHC} = 241$, $p_{MUHC} = 0.1$), ($W_{aqueduct} = 1471$, $p_{aqueduct} = 0.7$). Although not all the likelihood ratio values are above 1, this lack of difference supports our hypothesis that the short-term movement patterns of females should be able to predict the long-term movement patterns of both sexes. There was no difference in the likelihood ratio values from one study site compared to the other ($W= 3752$, $p = 0.7$).

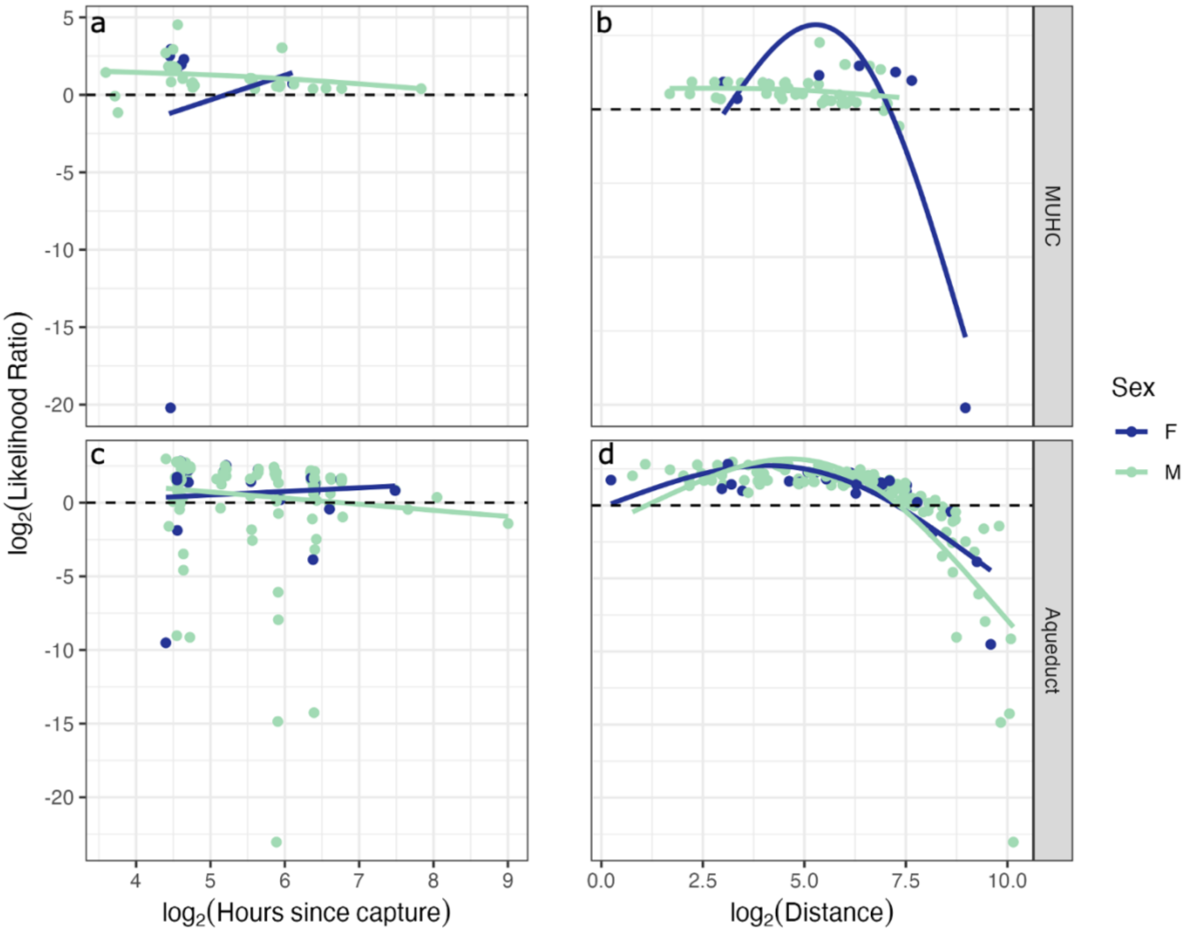


Figure 8: Plot of the log-2 transformed likelihood ratio for each recaptured individual. Where likelihood ratio is the probability of observed movement divided by the long-run probability of movement. Panels a and b depict likelihood ratio values for individuals at the MUHC site, while panels c and d depict likelihood ratio values for individuals at the aqueduct site. Panels (a) and (c) have x-axis values of log-2 (hours since capture) to show the relationship between the likelihood ratio and the amount of time moved, while panels (b) and (d) have x-axis values of log-2(distance) to show the relationship between the likelihood values and the distance moved. Each point represents a different recaptured *P. icarus* individual, with females shown in blue and males shown in green. Trend lines represent the estimated smooth curve for the sex and predictor of interest. The horizontal dashed line on the 0 line indicates equal probability under the two models for individuals to have ended up in the locations they did. Note the difference in x-axis scale difference between hours and distance between mark and recapture.

Discussion

Our study examined both the short-term and long-term movements of a non-native butterfly species in Montreal, *Polyommatus icarus*. Using both collected data and statistical modelling, we aimed to understand what factors motivate *P. icarus* individuals' short-term movements, and whether these motivators could be used to predict movement over time. We predicted that increased host-plant presence and short vegetation would increase movement to a given location, and that short-term habitat selection movements observed in females would predict the long-term patterns of movement of individuals.

We conducted a mark-release-recapture study and recorded a 5.4% recapture rate, which was higher than anticipated, based on a previous MRR study on *P. icarus* in North America and the sizes of the areas we surveyed (Rivest & Kharouba, 2021). However, it is important to note that findings presented are based on a small portion of the populations studied.

Vegetation

In our analysis of short-term habitat-selection behaviour of female *P. icarus*, we found that increased host-plant coverage, increased number of host-plant species flowering, increased vegetation height, and non-zero non-vegetation coverage increased the probability of a *P. icarus* female to land in a given spot. Increase in distance from one spot to the next and having to cross a road to move from one spot to the next decreased the probability of landing in a given spot. These findings support our prediction that host-plant presence is a movement motivator for this species, but do not support our prediction *P. icarus* prefers to stay in habitats with shorter vegetation.

We predicted that *P. icarus* would prefer shorter vegetation, based on studies conducted on *P. icarus* in its native and introduced ranges (Dexheimer & Despland, 2023; Habel et al., 2016; León-Cortés et al., 1999). We found the opposite effect to be true. However, the studies consulted from this species' native range examined the relationship between vegetation height and egg presence or egg oviposition. We based our prediction on these findings assuming that we would frequently witness *P. icarus* females landing to oviposit. Females of the lepidoptera family are known to move through and land in areas with different vegetative characteristics when their larval host-plants are not the same as their sources of food at maturity (Evans et al., 2020; Reim et al., 2018). *P. icarus* has been documented to oviposit and nectar on the many of the same Fabaceae plants though, and even oviposit right after feeding, so we do not believe that the preference to land on taller vegetation was based on this mismatch (Dexheimer & Despland, 2023; Emmet & Heath, 1989; Janz et al., 2005).

Another study conducted in Montreal found that adult *P. icarus* abundance was also negatively associated with increased vegetation height. This prior study quantified abundance based on number of individuals observed, while we were interested specifically in landing locations (Dexheimer & Despland, 2023). Given that *P. icarus* females prefer to oviposit on plants with flowers than without flowers, the difference in the effect of vegetation height which we found compared to past studies may be related to a confounding relationship between vegetation height and host-plant species in our sampling sites (Janz et al., 2005). Areas with very short vegetation were mowed weekly and were less likely to have any host-plant species in bloom than those with medium to long vegetation were. These mowed areas had lower host-plant species richness than areas that were not mowed as well, frequently containing *L. corniculatus*, *T. pratense*, *T. repens* and *M. sativa*, but almost never containing *M. lupulina*, *M. albus* or *V. cracca*. Given that frequently mowed areas had fewer host-plant species overall and in bloom, and that the areas we studied had a mix of short, medium and tall vegetation heights in close proximity to each other, females would have had less of a reason to land on short vegetation, even if many females were present in that area. We believe that this explains why we do not see shorter vegetation as a strong predictor of step selection in our model.

We predicted that increased host-plant species presence would increase the likelihood of a given individual to land in a spot, given the known relationship between *P. icarus* and its host-plants (Dexheimer & Despland, 2023). This was the case for both the percent coverage of host-plant species and the number of host-plant species flowering. However, we found that having some non-vegetation in a spot made that spot more likely to be landed on as well (Figure 5d). *P. icarus* has previously been found to be most abundant in moderately disturbed sites containing some dirt and gravel paths in its invasive range (Rivest & Kharouba, 2021). These findings are likely both driven by the presence of *L. corniculatus*, *P. icarus*' preferred host-plant, which grows extensively on roadsides in Canada, as do its other host-plants (Emmet & Heath, 1989; Turkington & Franko, 1980).

Road crossing

For the purpose of this study, a road is considered any paved linear area separating vegetative areas from each other, including sidewalks, bike paths, residential streets, boulevards, and highways. We observed that *P. icarus* were very unlikely to cross even narrow roads, with only two individuals crossing a road during the short-term data collection process. Our SSF predicted a road crossing coefficient of -3.7, corresponding to an individual only crossing a road 2.5% of the time if given the choice to land in another equidistant location with equal habitat quality without having to cross a road. This, and the non-vegetation coverage finding tell us that although *P. icarus* females may be likely to move through areas near a road, they are not likely to choose to cross that road. This is further amplified by the below average likelihood calculated for individuals to relocate to new locations further than three meters away, which is narrower than an average road (Figure 5f).

Given that we observed anecdotally (but not frequently during the short-term data collection process) both male and female *P. icarus* individuals crossing roads anecdotally but not frequently during the short-term data collection process, we began specifically watching for road crossings near the end of the data collection process. To do this, we watched a segment of road for 15 minutes at a time and noted the number of males and females that crossed the road in that time. We only counted a road as being crossed if the individual completely crossed the road; individuals that moved into the road and then turned back were not counted as having crossed a road. Throughout ten 15-minute periods of watching for individuals to cross roads at both study sites in optimal weather conditions when many *P. icarus* individuals were in flight, we recorded eight males and seven females cross a road. These observations are not included in our model, but show that while crossings are rare, *P. icarus* individuals do cross roads more often than expected by our short-term movement model.

One hundred and fourteen of the recaptured *P. icarus* individuals crossed at least one road or path during their displacement, based on the straight-line movement between marking and recapture locations. Upon initial data exploration, we thought that the difference in road crossing behaviours observed in the short-term and long-term movement data recorded would cause the likelihood ratios calculated to be lower for recaptured individuals that did cross roads compared to those that did not. This would have been due to the combination of the negative road crossing coefficient produced by the short-term movement model (Figure 5e), and of the relatively long distance required to cross most roads (15 m – 20 m for a standard two-way road) which is also disfavored by the short-term movement model (Figure 5f). However, there was no difference at the alpha 0.05 level in the likelihood ratios calculated for individuals that crossed a road between their initial marking and recapture location at the MUHC site, but at the aqueduct site there was (Figure S9). We believe this is driven by the lower occurrence probability estimated on one side of the bike path that runs along the aqueduct site, which is the road that many individuals crossed, and by the further distances moved by individuals that crossed a road between mark and recapture.

Road crossings occurred more frequently over the long-term than we expected them to, based on short-term movement observations and past studies. Given that crossing a road did not have the same effect on the likelihood ratios calculated at both study sites, it seems like the effect of the ground cover and distance variables are what is causing the likelihood ratio values to be lower for individuals that crossed a road at the aqueduct, and not by the road crossing effect itself. These findings, and the fact that *P. icarus* continues to inhabit roadsides in its native region when motorways are built, support the theory that *P. icarus* may continue to expand its range in North America using roadsides as corridors (Dexheimer, 2021; Munguira & Thomas, 1992).

Distance

Most recaptured individuals (82%) were found in locations that our long-term movement matrix (M(x,y)) predicted above average probability values for, given their starting location and surrounding habitat, represented by a log-likelihood ratio above 0 for the observed relocation (Figure 8). These findings reflect that the real movements observed by recaptured *P. icarus* individuals would have been predicted by our step selection model in some cases but would not have been in others. While the time between mark and recapture did not influence the log-likelihood ratio values calculated, overall, the log-likelihood ratios calculated were higher for individuals that moved shorter distances between initial marking and recapture (Figure 8b, Figure 8d). This indicates that our predictions of long-term movement based on *P. icarus* individuals' observable short-term movement patterns are most accurate for individuals that move less than 100 m over time, and that the distances moved by recaptured individuals was not dependent on the length of time between mark and recapture. Our log-likelihood ratio values for individuals that moved further than 100 m over multiple days fall below 0, which indicates that we would not have predicted those individuals to move the distance they did over time based on our understanding of habitat quality in the area. This implies that the range of possible long-term distance combinations were not completely observed during short-term movement observations.

We predicted that although the short-term movements information was based on only female *P. icarus* individuals, that the long-term predictions would be equally accurate for males and females even if the sexes move through space on the short-term differently, given that populations cannot disperse without female dispersal (Evans et al., 2020). Our findings agree with this prediction, as the log-likelihood ratios for males and females did not differ from each other (Figure 8). While at the MUHC site, the log-likelihood ratio values appear to follow different trends for males and females (Figure 8a, Figure 8b), these trends are largely influenced by a female outlier who moved a far distance between mark and recapture events within a short period, and by the fact that so few females were recaptured at the MUHC field site. The trends for males and females at this field site do still both indicate that further distances travelled have lower log-likelihood ratio values, but the low number of females recaptured at the MUHC site make this difficult to visualize.

The distances travelled between males and females did not differ from each other either, and of the 190 the individuals which were recaptured, there were two *P. icarus* individuals, one female and one male, that moved much further than the other 188 recaptured individuals. These two individuals were caught at the MUHC site and then recaptured at the aqueduct site, with a recorded displacement of over 3 km. While these movements are surprising, our MRR protocols called for taking images of individuals that were not marked clearly, both during marking and recapture. These individuals were recorded as being the same sex when marked and when recaptured and were not photographed during data collection, so we are quite sure that these observations do represent true long-distance displacements, rather than misidentifications of locally marked butterflies.

The two individuals that moved these long distances, while the mean and third quartile of displacement of all the other individuals recaptured was less than 200 m, are examples of long-distance dispersal. While long-distance dispersal is rare, these events are often the ones that drive species range expansions and ignoring them can result in an incomplete understanding of dispersal (McCaslin et al., 2020; Nathan, 2006). Like short- and long-term movements, long-distance dispersal distances are species dependent, and what is considered a long-distance dispersal event depends how far individuals of a species move on average (Nathan et al., 2003). In this case, two out of 190 recaptured individuals underwent a long-distance dispersal, which our long-term movement model would not have predicted even if we had expanded the model to include the areas outside of each study site, given that they were distances of over 100 m. This implies that long-distance dispersal is more common in *P. icarus* populations than anticipated based purely on habitat selection behaviours. Future movement models estimated for similar species should consider including components that allow for more frequent long-distance dispersal, such as a distance distribution that does not converge to 0 as quickly as a Gamma distribution does.

On a population level, this proportion of individuals undergoing long-term dispersal events can result in rapid range expansion, and large metapopulation structures. In this case, it is unlikely that the two individuals that underwent long-distance dispersal did so by choice, given that this displacement would have required crossing multiple large highways and two canals. We expect that wind events contributed to this displacement, not only including natural wind direction and strength, but also airflow created by vehicles. This highlights how environmental factors can allow for movement between patches, and allow individuals from different populations to interact, creating metapopulations. Whether this movement occurred by choice or wind influence, capturing these long-distance dispersal events in the small portion of the population which we sampled indicates that these long-distance dispersals likely occur frequently in *P. icarus* populations, and contribute to metapopulation dynamics. These interactions between populations can contribute to gene flow and the persistence of a species in a geographic area when habitat fragmentation and destruction occur (Hanski, 1998).

As previously mentioned, individuals who moved farther distances between their marking and recapture location had lower likelihood ratio values overall. This indicates that the distance effect included in our long-term model was not optimally set. This was anticipated, given that we used a Gamma distribution to estimate the preference for an individual to move from one spot to the next, but that a Gamma distribution did not fit the distribution of observed short-term step lengths very well, underestimating the probability of steps longer than 3 m (Figure S10). Using a Gamma distribution on the log-2 transformed distances had a better fit than not transforming the distances or using an exponential distribution though, which is why we chose to use the Gamma distribution to estimate preference even though it was not optimal. So, our findings reflect that the real movements observed by recaptured *P. icarus* individuals can be predicted by our step selection model in cases where individuals move less than 100 m. Although this short distance cannot be used to infer future range expansion, it is a step in the right direction, as we can see that the variables chosen to include in both the short-term and long-term models, other than distance, were understood and applied well overall. It is unlikely that a one size fits all approach to predicting distance in the long-term will ever be accurate, given that the majority of *P. icarus* individuals tend to move short distances while some, which drive the majority of range expansion, will move long distances (Nathan et al., 2003).

Our comparison of two study sites, the MUHC and the aqueduct, with different habitat configurations provides insight into the role of landscape structure in shaping long-term movement patterns (Revillaa & Wiegand, 2008). The recaptured individuals from the MUHC moved smaller distances than those from the aqueduct site (Table 1) and had higher likelihood ratio values overall (Figure 8). The MUHC site has a scattered, broken up patch configuration (Figure 7a), while the aqueduct has a long, linear, more continuous patch configuration (Figure 7b). This implies that individuals at the MUHC left their patches less than those at the aqueduct did, likely because there was less suitable habitat for them to move through between patches, with roads and parking lots in between them (Figure 7a), whereas the aqueduct site is made up of long stretches of vegetation (Figure 7b), which makes movement less risky for a butterfly (Evans et al., 2020). We recaptured more individuals at the aqueduct, which supports this idea, as recapturing more individuals even though they moved further distances points towards individuals living for longer, which seems to be the case given that the hours passed between marking and recapture are higher at the aqueduct than the MUHC (Table S1). The ground cover being more consistent throughout the aqueduct compared to the MUHC resulted in lower likelihood ratio values at the aqueduct because there were more high-quality locations within the study site for *P. icarus* individuals to move to, making the long-term movement model less certain about where individuals would end up overtime.

Implications

Our results provide insight towards modelling future range expansions of *P. icarus* in North America. Although this non-native species has not been deemed a risk to native species thus far, it is spreading rapidly and could still pose a risk to native species persistence in areas which it has not yet reached (i.e., via resource competition). Due to the potential for non-native species to have negative impacts on native ecology (Ricciardi & Cohen, 2007), it will be important to understand the characteristics of the corridors through which *P. icarus* is spreading, and how to limit that spread. Our short-term models indicate that host-plant species on the edges of roads could be preferred landing locations for *P. icarus* females, and our long-term observations and models indicate that both male and female individuals are likely to cross these roads. While dispersal over 100 m was not modelled accurately through this study, we did find evidence of long-distance dispersal, which is believed to be the first evidence of this behaviour in *P. icarus* in North America, and points towards *P. icarus* living in larger metapopulations than anticipated. Future work could focus on estimating long-distance dispersal more accurately, as well as using genetic testing to determine the extent of metapopulations of this species, and even if the expansion of *P. icarus'* range from Montreal to Toronto (> 500 km) was caused by long-distance dispersal events, or by a separate human mediated introduction.

Conclusion

Overall, we found that non-native host-plants and road edges are key factors to consider when examining *P. icarus* movement, and that a small portion of individuals undergoing long-distance dispersal may be responsible for the rapid range expansion which this butterfly species has undergone in North America. Since non-native plants and road edges are abundant both in and between major cities in Canada and the United States, these findings point towards further range expansion in the coming years.

Chapter 3: General Conclusion

Based on this study's findings, our research suggests that the long-term movement patterns of low dispersing male and female *Polyommatus icarus* individuals can be predicted based on female short-term movement patterns and habitat quality. However, *P. icarus* long-distance dispersal events appear to be more common than previously believed and are more difficult to predict based on short-term movement patterns alone. We found this to be the case at two study sites, which both showed that predicting long-term movements over 100 m would not be accurate given the methods we used, and that many individuals move further than 100 m over a period of multiple days.

Individuals with long-term movement displacements of less than 100 m were found to have above average probabilities of ending up in their recapture locations based on their initial location, the time passed between mark and recapture, and matrix M(x, y) which represented an individual's preference to move from one location to every other accessible location within the study site based on habitat qualities and waiting time. This indicates that the habitat qualities, percent host-plant coverage, number of host-plant species in bloom, percent non-vegetative coverage, and vegetation height, which influence movement of this butterfly species are understood, and that the effect of distance and crossing roads is what should be further studied. These variables are more difficult to study but are important pieces in understanding how long-distance dispersal events may continue to drive the range expansion of this species in North America.

P. icarus' abundance is greater in moderately disturbed areas with a higher proportion of urban land, and in North America its host-plant species are widespread and thrive in fragmented, urbanized areas (Rivest & Kharouba, 2021). We found that likelihood ratio values were higher in areas with less continuous habitat, likely because of the clearer indication of which parts of the landscape could be patches and which parts could not be, which is typical for many areas where butterflies would be found in cities. Long stretches of vegetation, like the aqueduct study site, are typical layouts for the sides of highways in many parts of North America though, which appear to be the routes by which *P. icarus* has spread so far (Figure S1). The diversity of patch structures within which this species can live and disperse highlights the importance of incorporating different landscape configurations into movement models like the one we have described. The layout of patches, not just their quality, has an impact on the accuracy of predicting long-term movement based on short-term movement patterns.

Our findings support those of Rivest & Kharouba, that *P. icarus* is expanding its range through fragmented, urbanized landscapes, and our evidence of long-distance dispersal points towards metapopulation structures forming in these types of landscapes. These metapopuation interactions will further support range expansion by facilitating gene-flow, which can increase genetic diversity and population resilience. As *P. icarus* individuals continue to undergo long-distance dispersal, they may continue to expand the species' range into currently unoccupied areas, or increase their poulation sizes in currently occupied areas. There is currently no evidence to suggest that *P. icarus* will have any negative impacts on native species or the native ecosystem, but continuing to monitor the species' range expansion, and its movement corridors, will be important in case this butterfly species is found to have any negative effects.

The effects of weather on *P. icarus* movement were not thoroughly examined in this study because of the need for warmth and sunlight to be able to observe frequent short-term movements. We believe that accounting for lower movement on cloudy days would influence the likelihood ratios calculated, as it would take longer for individuals to move to their recapture locations than our model predicts, since our model is using a constant waiting time for each cell, and not slowing down during lower movement periods. Taking this information into account would likely make the distance effect less well understood and point towards an even stronger need to better understand the effect of distance on long-term movement.

Moving forward, the effects of long-distance displacement, weather, and habitat quality and layout, will all be important to consider to completely understand the dispersal mechanisms and potential range expansion of *P. icarus* in North America, as well as the long-term movement patterns of species with similar long-distance dispersal capabilities.

References

Aarssen, L. W., Hall, I. V., & Jensen, K. I. N. (1986). The biology of Canadian weeds.: 76. *Vicia angustifolia* L., *V . cracca* L., *V . sativa* L., *V . tetrasperma* (L.) *Schreb* . and *V . villosa* Roth. *Canadian Journal of Plant Science*, *66*(3), 711–737. https://doi.org/10.4141/cjps86-092

Abrahams, M. V. (1986). Patch choice under perceptual constraints: A cause for departures from an ideal free distribution. *Behavioral Ecology and Sociobiology*, *19*(6), 409–415. https://doi.org/10.1007/BF00300543

AccuWeather. (2023). https://www.accuweather.com/

Altermatt, F., & Pearse, I. S. (2011). Similarity and specialization of the larval versus adult diet of european butterflies and moths. *The American Naturalist*, *178*(3), 372–382. https://doi.org/10.1086/661248

Avgar, T., Potts, J. R., Lewis, M. A., & Boyce, M. S. (2016). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, *7*(5), Article 5. https://doi.org/10.1111/2041-210X.12528

Clark, J. S., Fastie, C., Hurtt, G., Jackson, S. T., Johnson, C., King, G. A., Lewis, M., Lynch, J., Pacala, S., Prentice, C., Schupp, E. W., Webb, T., III, & Wyckoff, P. (1998). Reid's paradox of rapid plant migration: Dispersal theory and interpretation of paleoecological records. *BioScience*, *48*(1), 13–24. https://doi.org/10.2307/1313224

Cowley, M. J. R., Thomas, C. D., Thomas, J. A., & Warren, M. S. (1999). Flight areas of British butterflies: Assessing species status and decline. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1428), 1587–1592. https://doi.org/10.1098/rspb.1999.0819

De Alfaro, L. (1998). How to specify and verify the long-run average behaviour of probabilistic systems. *Proceedings. Thirteenth Annual IEEE Symposium on Logic in Computer Science (Cat. No.98CB36226)*, 454–465. https://doi.org/10.1109/LICS.1998.705679

Dennis, E. B., Morgan, B. J. T., Roy, D. B., & Brereton, T. M. (2017). Urban indicators for UK butterflies. *Ecological Indicators*, *76*, 184–193. https://doi.org/10.1016/j.ecolind.2017.01.009

Dexheimer, E. (2021). *Myrmecophily and habitat use of the European common blue (*Polyommatus icarus*: (Lycaenidae Rottemburg, 1775)) in Quebec, North America* [Concordia University]. https://spectrum.library.concordia.ca/id/eprint/988770/1/Dexheimer_MA_F2021.pdf

Dexheimer, E., & Despland, E. (2023). Newly introduced butterfly species' urban habitat use driven by shorter vegetation and exotic plants. *Biological Invasions*, *25*(6). https://doi.org/10.1007/s10530-023-03009-3

Diniz, M. F., Cushman, S. A., Machado, R. B., & De Marco Júnior, P. (2019). Landscape connectivity modeling from the perspective of animal dispersal. *Landscape Ecology*, *35*(1), 41–58. https://doi.org/10.1007/s10980-019-00935-3

Duchesne, T., Fortin, D., & Rivest, L.-P. (2015). Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. *PLOS ONE*, *10*. https://doi.org/10.1371/journal.pone.0122947

Emmet, A., & Heath, J. (1989). *The moths and butterflies of Britain and Ireland* (pp. 157–160). Harley Books.

Evans, L. C., Sibly, R. M., Thorbek, P., Sims, I., Oliver, T. H., & Walters, R. J. (2019). Integrating the influence of weather into mechanistic models of butterfly movement. *Movement Ecology*, *7*, 7–24. https://doi.org/10.1186/s40462-019-0171-7

Evans, L. C., Sibly, R. M., Thorbek, P., Sims, I., Oliver, T. H., & Walters, R. J. (2020). The importance of including habitat-specific behaviour in models of butterfly movement. *Oecologia*, *193*(2), Article 2. https://doi.org/10.1007/s00442-020-04638-4

Fischer, C., Hanslin, H. M., Hovstad, K. A., D'Amico, M., Kollmann, J., Kroeger, S. B., Bastianelli, G., Habel, J. C., Rygne, H., & Lennartsson, T. (2022). The contribution of roadsides to connect grassland habitat patches for butterflies in landscapes of contrasting permeability. *Journal of Environmental Management*, *311*, 114846. https://doi.org/10.1016/j.jenvman.2022.114846

Fletcher, R., & Fortin, M.-J. (2018). Space use and resource selection. In R. Fletcher & M.-J. Fortin (Eds.), *Spatial ecology and conservation modeling: Applications with r* (pp. 271–320). Springer International Publishing. https://doi.org/10.1007/978-3-030-01989-1_8

Forman, R. T. T., & Godron, M. (1981). Patches and structural components for a landscape ecology. *BioScience*, *31*(10), 733–740. https://doi.org/10.2307/1308780

Forman, R. T. T., & Godron, M. (1986). Part II: Landscape Structure. In *Landscape ecology* (1st ed., Vol. 1). John Wiley & Sons.

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., & Mao, J. S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in yellowstone national park. *Ecology*, *86*(5), 1320–1330. https://doi.org/10.1890/04-0953

Gippet, J. M. W., Liebhold, A. M., Fenn-Moltu, G., & Bertelsmeier, C. (2019). Human-mediated dispersal in insects. *Current Opinion in Insect Science*, *35*, 96–102. https://doi.org/10.1016/j.cois.2019.07.005

Girvetz, E. H., & Greco, S. E. (2007). How to define a patch: A spatial model for hierarchically delineating organism-specific habitat patches. *Landscape Ecology*, *22*(8), 1131–1142. https://doi.org/10.1007/s10980-007-9104-8

Gutiérrez, D., León-Cortés, J. L., Menéndez, R., Wilson, R. J., Cowley, M. J. R., & Thomas, C. D. (2001). Metapopulations of four lepidopteran herbivores on a single host plant, *Lotus corniculatus* . *Ecology*, *82*(5), 1371–1386. https://doi.org/10.2307/2679996

Habel, J. C., Teucher, M., Ulrich, W., Bauer, M., & Rödder, D. (2016). Drones for butterfly conservation: Larval habitat assessment with an unmanned aerial vehicle. *Landscape Ecology*, *31*(10), 2385–2395. https://doi.org/10.1007/s10980-016-0409-3

Hall, P. (2007). The European common blue *Polyommatus icarus* : New alien butterfly to Canada and North America. *Lepidopterists Society Newsletter*, *49*(4), 111.

Hanks, E. M., Hooten, M. B., & Alldredge, M. W. (2015). Continuous-time discrete-space models for animal movement. *The Annals of Applied Statistics*, *9*(1), 145–165. https://doi.org/10.1214/14-AOAS803

Hanks, E. M., & Hughes, D. A. (2016). *Flexible discrete space models of animal movement*. arXiv preprint arXiv:1606.07986. http://arxiv.org/abs/1606.07986

Hanski, I. (1982). Dynamics of regional distribution: The core and satellite species hypothesis. *Oikos*, *38*(2), Article 2. JSTOR. https://doi.org/10.2307/3544021

Hanski, I. (1998). Metapopulation dynamics: Review article. *Nature*, *396*, 41–49.

Hanski, I., & Gilpin, M. (1991). Metapopulation dynamics: Brief history and conceptual domain. *Biological Journal of the Linnean Society*, *42*(1–2), Article 1–2. https://doi.org/10.1111/j.1095-8312.1991.tb00548.x

Heinz, S. K., Wissel, C., & Frank, K. (2006). The viability of metapopulations: Individual dispersal behaviour matters. *Landscape Ecology*, *21*(1), 77–89. https://doi.org/10.1007/s10980-005-0148-3

iNaturalist community. Observations of *Polyommatus icarus* from Canada and United States of America observed on/between May 1, 2005 – July 1, 2024. Exported from https://www.inaturalist.org on July 1, 2024

Janz, N., Bergström, A., & Sjögren, A. (2005). The role of nectar sources for oviposition decisions of the common blue butterfly *Polyommatus icarus* . *Oikos*, *109*(3), 535–538. https://doi.org/10.1111/j.0030-1299.2005.13817.x

Kallioniemi, E., Zannese, A., Tinker, J. E., & Franco, A. M. A. (2014). Inter- and intra-specific differences in butterfly behaviour at boundaries. *Insect Conservation and Diversity*, *7*(3), 232–240. https://doi.org/10.1111/icad.12046

Klappstein, N., Michelot, T., Fieberg, J., Pedersen, E., & Flemming, J. M. (2024). Step selection analysis with non-linear and random effects. *Methods in Ecology and Evolution*, 2024–01. https://doi.org/10.1111/2041-210X.14367

Kuussaari, M., Saarinen, M., Korpela, E.-L., Pöyry, J., & Hyvönen, T. (2014). Higher mobility of butterflies than moths connected to habitat suitability and body size in a release experiment. *Ecology and Evolution*, *4*(19), 3800–3811. https://doi.org/10.1002/ece3.1187

León-Cortés, J. L., Cowley, M. J. R., & Thomas, C. D. (1999). Detecting decline in a formerly widespread species: How common is the common blue butterfly *Polyommatus icarus* ? *Ecography*, *22*(6), 643–650. http://www.jstor.org/stable/3683318.

McCaslin, H. M., Caughlin, T. T., & Heath, J. A. (2020). Long-distance natal dispersal is relatively frequent and correlated with environmental factors in a widespread raptor. *Journal of Animal Ecology*, *89*(9), 2077–2088. https://doi.org/10.1111/1365-2656.13272

Morton, A. C. (1982). The effects of marking and capture on recapture frequencies of butterflies. *Oecologia*, *53*(1), 105–110. https://doi.org/10.1007/BF00377143

Munguira, M. L., & Thomas, J. A. (1992). Use of road verges by butterfly and burnet populations, and the effect of roads on adult dispersal and mortality. *Journal of Applied Ecology*, *29*(2), 316–329. https://doi.org/10.2307/2404501

Nathan, R. (2006). Long-distance dispersal of plants. *Science*, *313*(5788), 786–788. https://doi.org/10.1126/science.1124975

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), Article 49. https://doi.org/10.1073/pnas.0800375105

Nathan, R., Perry, G., Cronin, J. T., Strand, A. E., & Cain, M. L. (2003). Methods for estimating long-distance dispersal. *Oikos*, *103*(2), Article 2. https://doi.org/10.1034/j.1600-0706.2003.12146.x

Noé, F., & Clementi, C. (2015). Kinetic distance and kinetic maps from molecular dynamics simulation. *Journal of Chemical Theory and Computation*, *11*(10), 5002–5011. https://doi.org/10.1021/acs.jctc.5b00553

Nowicki, P., Vrabec, V., Binzenhöfer, B., Feil, J., Zakšek, B., Hovestadt, T., & Settele, J. (2014). Butterfly dispersal in inhospitable matrix: Rare, risky, but long-distance. *Landscape Ecology*, *29*(3), 401–412. https://doi.org/10.1007/s10980-013-9971-0

Otto, S. P., & Day, T. (2007). *A biologist's guide to mathematical modeling in ecology and evolution.* Princeton University Press.

Panzacchi, M., Van Moorter, B., Strand, O., Saerens, M., Kivimäki, I., St. Clair, C. C., Herfindal, I., & Boitani, L. (2016). Predicting the continuum between corridors and barriers to animal movements using Step Selection Functions and Randomized Shortest Paths. *Journal of Animal Ecology*, *85*(1), 32–42. https://doi.org/10.1111/1365-2656.12386

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ 7*, *e6876*. https://doi.org/10.7717/peerj.6876

Pellissier, V., Muratet, A., Verfaillie, F., & Machon, N. (2012). Pollination success of *Lotus corniculatus* (L.) in an urban context. *Acta Oecologica*, *39*, 94–100. https://doi.org/10.1016/j.actao.2012.01.008

Pyšek, P., & Richardson, D. M. (2006). The biogeography of naturalization in alien plants. *Journal of Biogeography*, *33*(12), 2040–2050. https://doi.org/10.1111/j.1365-2699.2006.01578.x

R Core Team. (2021). *R* (Version 4.2.1) [R]. R Foundation for Statistical Computing. ttps://www.R-project.org/

Reim, E., Arnstedt, I., Barwisch, I., Baumgarten, M., Bock, S., Eberspach, J., Ellerbrok, J., Gebremeskel, M., Küpper, S., Guth, L., Lassen, A., Letro, L., Meth, R., Möller, M., Närmann, F., Neunaber, I., Seliger, A., Stein, W. V., Vallinga, C., … Fischer, K. (2018). Movement patterns differ between sexes and depend on weather conditions in the butterfly *Lycaena tityrus*. *Journal of Insect Behavior*, *31*(3), Article 3. https://doi.org/10.1007/s10905-018-9679-8

Reim, E., Widderich, F., & Fischer, K. (2019). Sexual differences in the morphology and movement of a butterfly: Good shape does not make good dispersers. *European Journal of Entomology*, *116*, 468–476. https://doi.org/10.14411/eje.2019.048

Revillaa, E., & Wiegand, T. (2008). Individual movement behavior, matrix heterogeneity, and the dynamics of spatially structured populations. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19120–19125. https://doi.org/10.1073/pnas.0801725105

Ricciardi, A., & Cohen, J. (2007). The invasiveness of an introduced species does not predict its impact. *Biological Invasions*, *9*(3), 309–315. https://doi.org/10.1007/s10530-006-9034-4

Richardson, D. M., & Pyšek, P. (2008). Fifty years of invasion ecology – the legacy of Charles Elton. *Diversity and Distributions*, *14*(2), 161–168. https://doi.org/10.1111/j.1472-4642.2007.00464.x

Rivest, S. A., & Kharouba, H. M. (2021). Anthropogenic disturbance promotes the abundance of a newly introduced butterfly, the European common blue (*Polyommatus icarus*; Lepidoptera: Lycaenidae), in Canada. *Canadian Journal of Zoology*, *99*(8), Article 8. https://doi.org/10.1139/cjz-2021-0009

Rivest, S. A., & Kharouba, H. M. (2024). Taxonomic and functional homogenization of butterfly communities along an urban gradient. *Insect Conservation and Diversity*, *17*, 273–286. https://doi.org/10.1111/icad.12729

Sen, A. (1973). Behaviour and the Concept of Preference. *Economica,* 40, 241-259. 10.2307/2552796

Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, *2*(1), Article 1. https://doi.org/10.1186/2051-3933-2-4

Turchin, P. (1998). *Quantitative analysis of movement: Measuring and modeling population redistribution in animals and plants*. Sinauer Associates.

Turkington, R. A., & Burdon, J. J. (1983). The biology of Canadian weeds.: 57. *Trifolium repens* L. *Canadian Journal of Plant Science*, *63*(1), 243–266. https://doi.org/10.4141/cjps83-024

Turkington, R. A., & Cavers, P. B. (1979). The biology of Canadian weeds.: 33. *Medicago lupulina* L. *Canadian Journal of Plant Science*, *59*(1), 99–110. https://doi.org/10.4141/cjps79-015

Turkington, R. A., Cavers, P. B., & Rempel, E. (1978). The biology of Canadian weeds.: 29. *Melilotus alba* Desr. and *M . officinalis* (L.) Lam. *Canadian Journal of Plant Science*, *58*(2), 523–537. https://doi.org/10.4141/cjps78-078

Turkington, R. A., & Franko, G. D. (1980). The biology of Canadian weeds.: 41. *Lotus corniculatus* L. *Canadian Journal of Plant Science*, *60*(3), 965–979. https://doi.org/10.4141/cjps80-140

USDA NRCS Idaho Plant Materials. (n.d.). *Red Clover* (United States Department of Agriculture Naturall Resources Conservation Service) [Plant Guide]. United States Department of Agriculture. https://plants.usda.gov/DocumentLibrary/plantguide/pdf/pg_trpr2.pdf

Van Dyck, H., & Baguette, M. (2005). Dispersal behaviour in fragmented landscapes: Routine or special movements? *Basic and Applied Ecology*, *6*(6), 535–545. https://doi.org/10.1016/j.baae.2005.03.005

Wang, Z., & Sakiroglu, M. (2021). The origin, evolution, and genetic diversity of alfalfa. In *The Alfalfa Genome* (pp. 29–42). Springer International Publishing.

Wilson, K., Hanks, E., & Johnson, D. (2018). Estimating animal utilization densities using continuous-time Markov chain models. *Methods in Ecology and Evolution*, *9*(5), 1232–1240. https://doi.org/10.1111/2041-210X.12967

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2023). *mgcv: Mixed gam computation vehicle with automatic smoothness estimation* (Version 1.9-1) [Computer software]. https://cran.r-project.org/web/packages/mgcv/index.html

Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, *7*(6), 636–645. https://doi.org/10.1111/2041-210X.12577

Supplementary Material

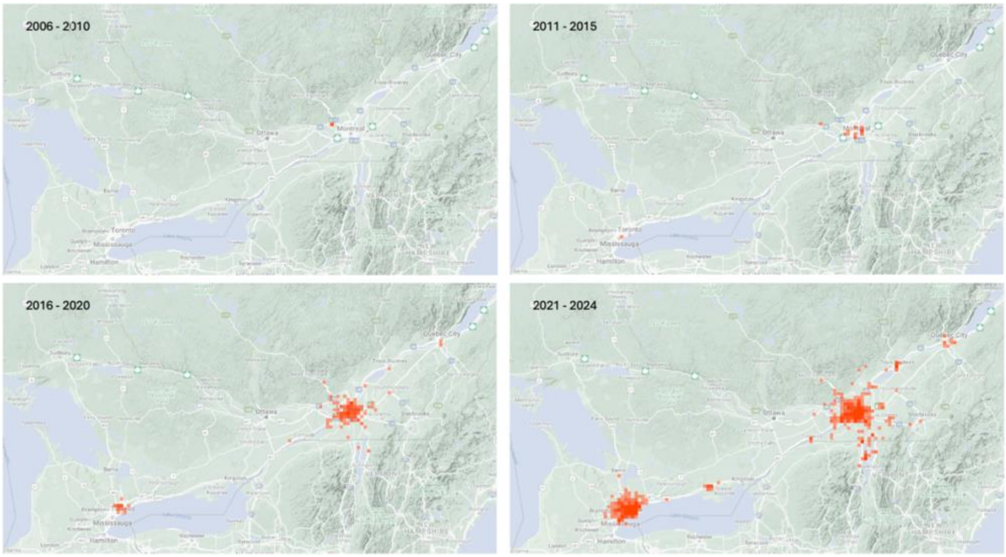
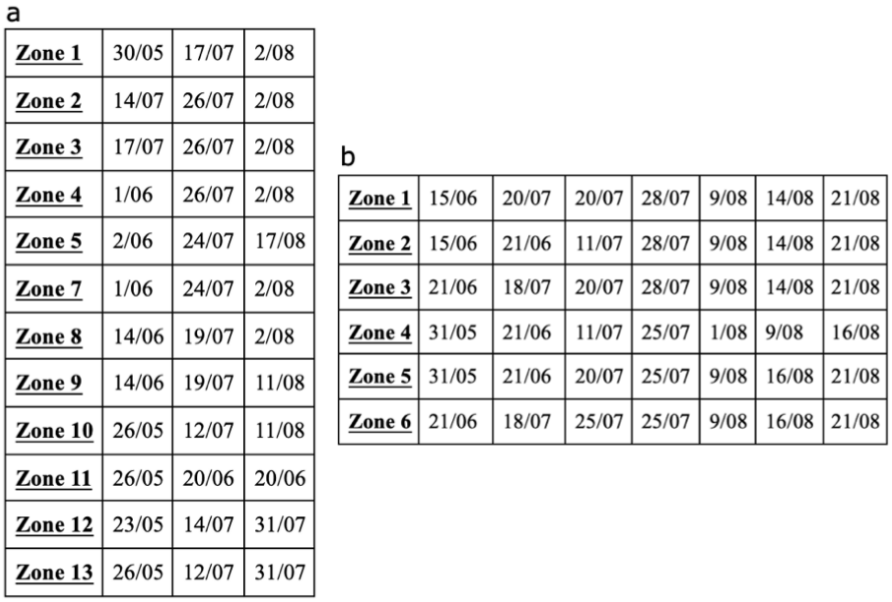


Figure S1: Observations of *P. icarus* over time in North America as reported on the community science platform “iNaturalist” (Quebec, Ontario and Vermont). Red points indicate where a *P. icarus* individual was reported being seen. Images are grouped by observations that occurred within five years of each other, other than the most recent image which only contains four years.



a

Zone 1	30/05	17/07	2/08
Zone 2	14/07	26/07	2/08
Zone 3	17/07	26/07	2/08
Zone 4	1/06	26/07	2/08
Zone 5	2/06	24/07	17/08
Zone 7	1/06	24/07	2/08
Zone 8	14/06	19/07	2/08
Zone 9	14/06	19/07	11/08
Zone 10	26/05	12/07	11/08
Zone 11	26/05	20/06	20/06
Zone 12	23/05	14/07	31/07
Zone 13	26/05	12/07	31/07

b

Zone 1	15/06	20/07	20/07	28/07	9/08	14/08	21/08
Zone 2	15/06	21/06	11/07	28/07	9/08	14/08	21/08
Zone 3	21/06	18/07	20/07	28/07	9/08	14/08	21/08
Zone 4	31/05	21/06	11/07	25/07	1/08	9/08	16/08
Zone 5	31/05	21/06	20/07	25/07	9/08	16/08	21/08
Zone 6	21/06	18/07	25/07	25/07	9/08	16/08	21/08

Figure S2: Dates of short-term data collection at the MUHC (a) and aqueduct (b) sites during the 2023 summer field season. Dates are formatted as day/month. Note that Zone 6 at the MUHC does not exist, as it was planned in a location that was not accessible for sampling, and Zone 14 at the MUHC site does not exist as it was absorbed into Zone 13. Short-term movement sampling did not occur in Zones 15 or 16 at the MUHC site as there were too many hazards to safely do so.

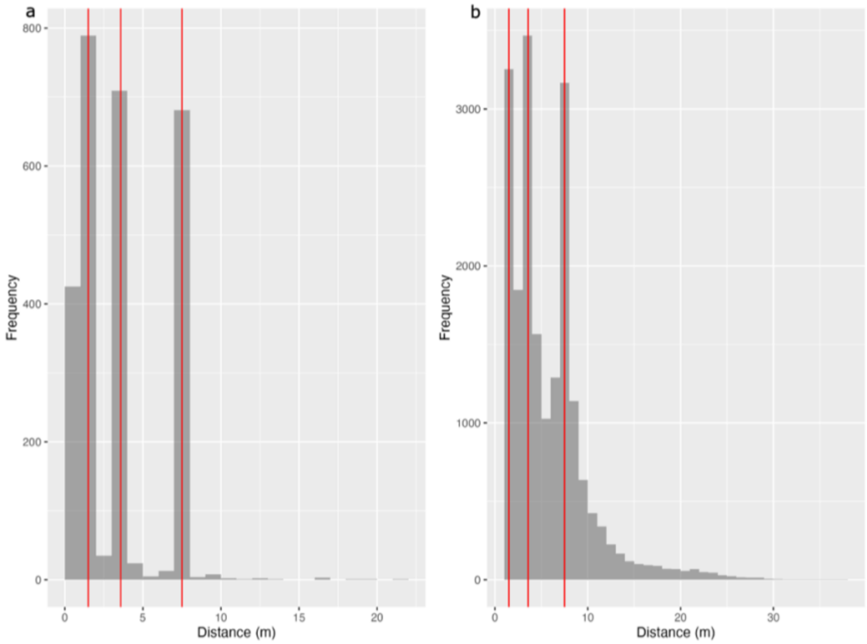


Figure S3: Histogram of all step lengths (observed and alternate steps) recorded during short-term movement data collection from all paths. Panel (a) shows the distribution of distances between steps based on original measurements. Panel (b) shows the distribution of distances between steps after aggregating sequential observed steps shorter than 1 m from the same path into one step and calculating the distance of all observed steps to all alternate steps within a path. Red vertical lines represent the lengths of the three pre-measured alternate step lengths, 1.5 m, 3.5 m, and 7.5 m. Note the difference in scales on the x and y axes.

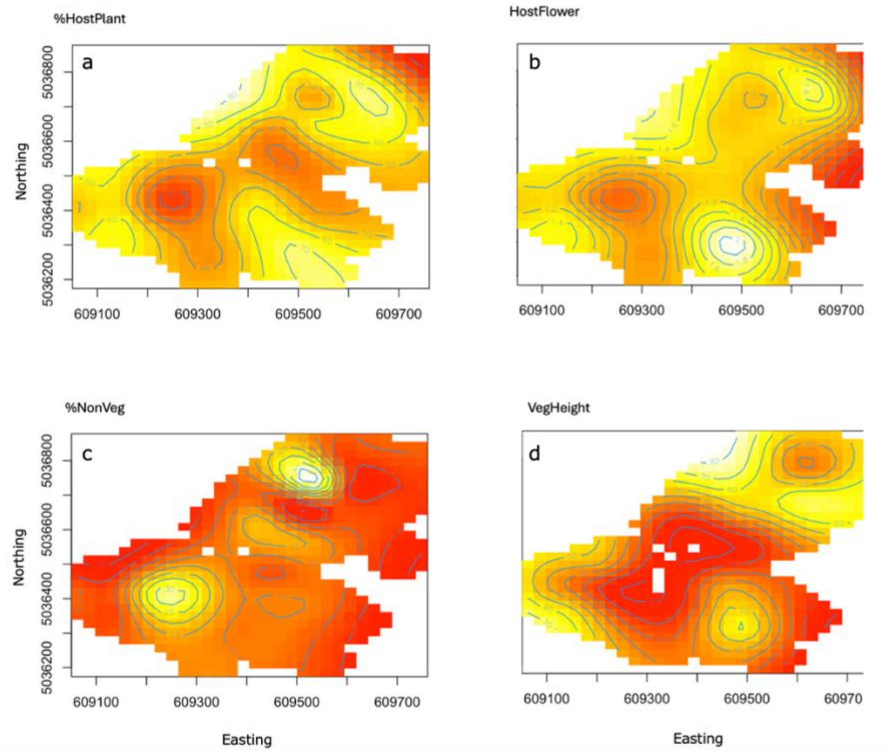


Figure S4: Visual outputs of spatial GAMs generated by equation (4) for the MUHC site. Panels represent the predicted values of (a) percent host-plant coverage, (b) number of host-plant species in bloom, (c) percent of non-vegetative coverage, and (d) maximum vegetation height. Darker colouring represents smaller predictions, while lighter colouring represents larger predictions. Empty pixels represent areas with insufficient information to extrapolate a prediction, due to insufficient sampling in that area. Note that roads and buildings were not removed from this image but were in the matrix created from it.

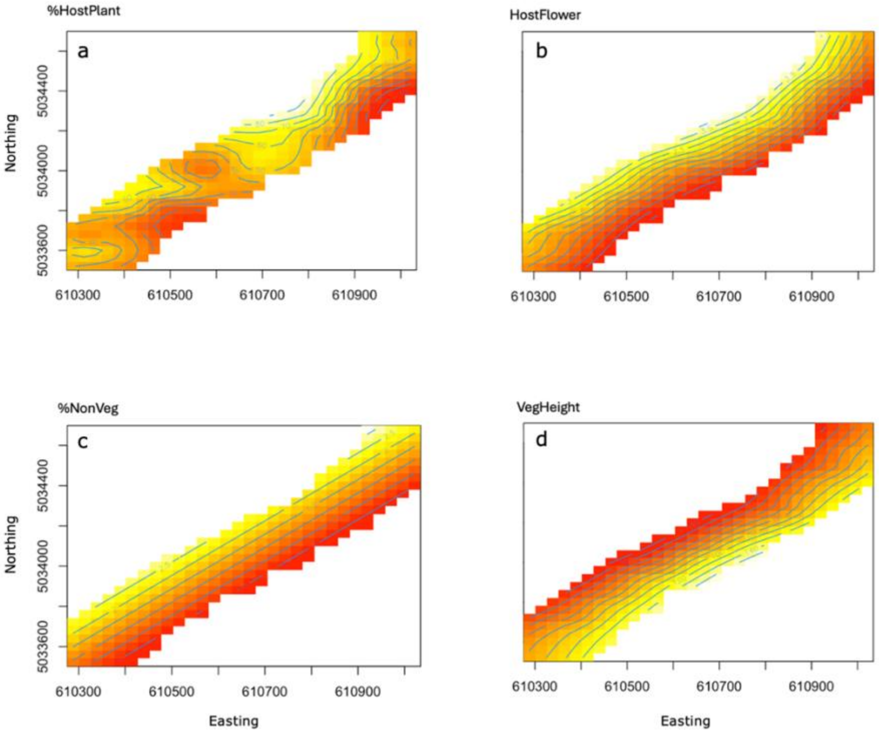


Figure S5: Visual outputs of spatial GAMs generated by equation (4) for the aqueduct site. Panels represent the predicted values of (a) percent host-plant coverage, (b) number of host-plant species in bloom, (c) percent of non-vegetative coverage, and (d) maximum vegetation height. Darker colouring represents smaller predictions, while lighter colouring represents larger predictions. Empty pixels represent areas with insufficient information to extrapolate a prediction, due to insufficient sampling in that area. Note that roads and buildings were not removed from this image but were in the matrix created from it.

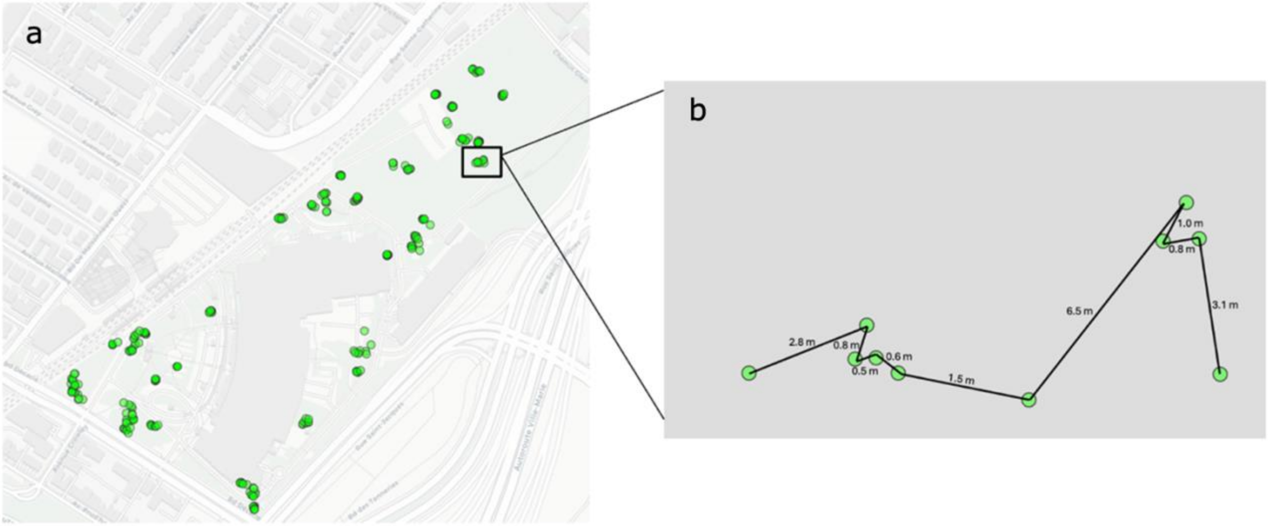


Figure S6: Example of a short-term movement path of a *P. icarus* female. Green points represent landing locations. Depicted on the left (a) are all the short-term movement paths recorded at the MUHC study site, and on the right (b) is a zoomed in path with the sequence of steps connected to each other with the measured distances reported. Note that distances may not appear to be to scale as the GPS used was only accurate up to three meters.

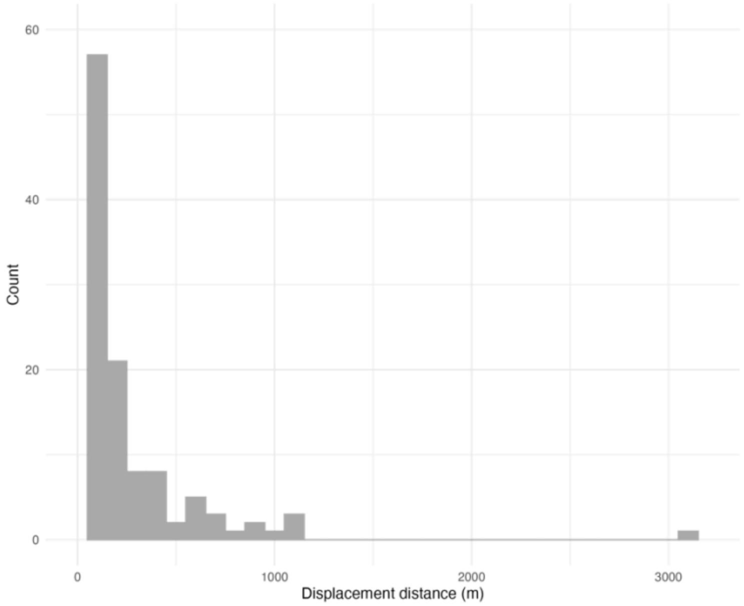


Figure S7: Histogram of distribution of displacements of recaptured *P. icarus* individuals collected via a mark-release-recapture study.

Figure S8: Maps of both study sites, with MUHC on the left (a) and aqueduct on the right (b). Points represent all capture locations of *P. icarus* individuals, with males represented in green and females represented in blue. Locations shown outside of study sites are due to GPS error.

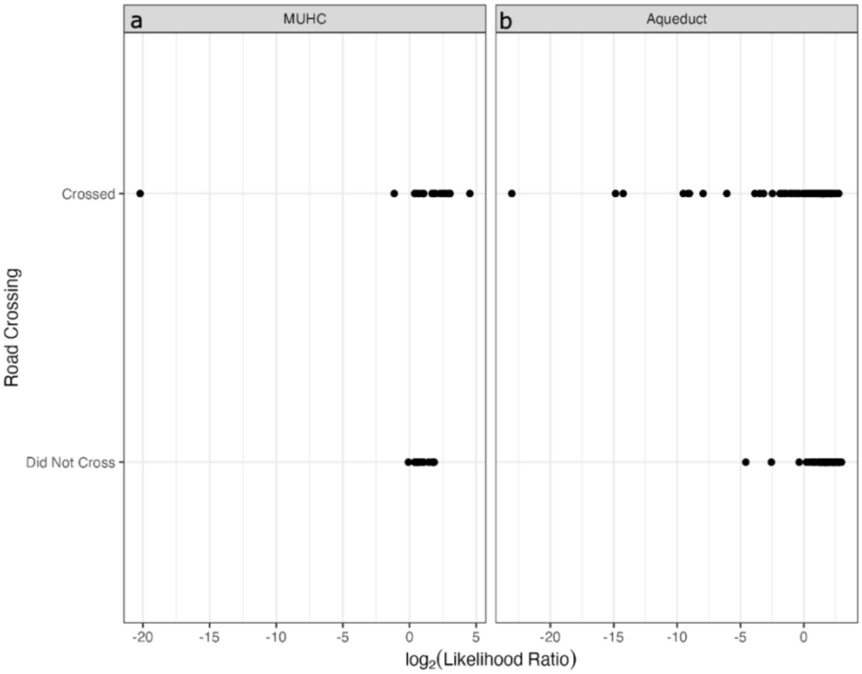


Figure S9: Scatterplot comparing the log-2 transformed likelihood ratios *P. icarus* individuals who crossed roads and those that did not. The left panel (a) depicts the MUHC site, and the right panel (b) depicts the aqueduct site. Note the difference in x-axis scale between sites.

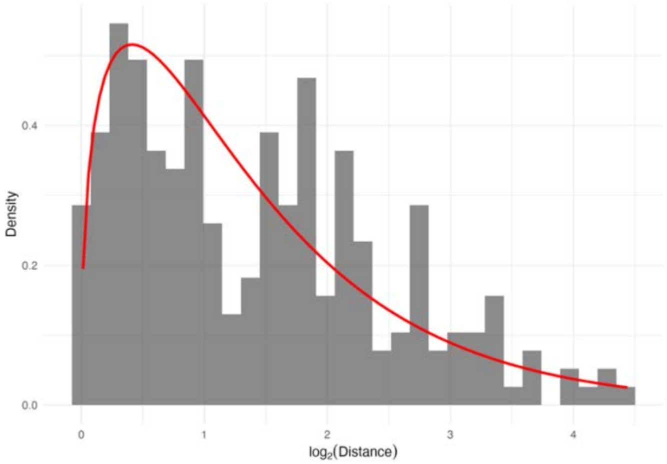


Figure S10: Histogram of observed step lengths of *P. icarus* females collected during short-term movement data collection. The red distribution curve represents the Gamma distribution used to calculate the distance parameter in $Pref(x, y)$. The Gamma distribution has a shape parameter of 1.41 and a scale parameter of 1.0.

Table S1: Information about distance moved and time passed between initial capture and recapture of *P. icarus* individuals at two field sites, collected via a MRR study.

	MUHC	**Aqueduct**
Max hours	456	1027
Mean hours	84	121
Third quartile hours	95	168
Max distance (m)	500	1133
Mean distance (m)	60	178
Third quartile distance (m)	67	207