**Using anonymized mobile data to quantify human-wildlife interactions in protected areas**

Alessandro Filazzola1,2 Garland Xie3, Kimberly Barrett4, Andrea Dunn4, Marc T.J. Johnson1,5, & J. Scott MacIvor1,2

1. Centre for Urban Environments, University of Toronto Mississauga, Mississauga, Ontario, Canada.

2. School of Cities, University of Toronto, Toronto, Ontario, Canada.

3. Department of Biological Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario Canada.

4. Conservation Halton, Burlington, Ontario, Canada.

5. Department of Biology, University of Toronto Mississauga, Mississauga, Ontario, Canada.

\* Corresponding author E-mail address: alex.filazzola@utoronto.ca

**Abstract**

**1. Introduction**

Cities are rapidly expanding, creating new challenges for managing urban greenspaces. More than half of the global population currently live in cities and that number is projected to increase to almost 90% by the end of the century (citations). As these cities increase in size and area, greenspaces, such as parks, remnant natural areas, and protected reserves, face new stressors from human activity. Direct human use of green spaces can negatively impact urban wildlife including trampling, introduction of exotics, and pollution (citations). However, urban green spaces are important for city residents as a place for exercise, recreation, socialization, and supporting mental well-being (citations). Thus, managing these spaces is a delicate balancing act between utility for people and conservation of biodiversity.

One of the main limitations in effectively managing urban greenspace is the uncertainty around how and when people use these areas. Although some parks use a reservation-based system with controlled points of entry, other urban greenspaces are more open with many access points. Trails are meant to facilitate human movement and reduce disturbance to biodiversity, but residents will still venture off-trail or erode new paths of easily navigable terrain, i.e., desire lines (citations). Determining areas of high disturbance (i.e., high traffic), potential off-trail use, and overlap with sensitive species, can all be achieved through understanding human mobility in greenspaces. However, capturing human mobility at a resolution fine enough for management, such as less than 100 x 100 m, is challenging. Previous methods for quantifying human activity include record keeping visitors at entrance points or camera traps to track the number of visitors. However, this data neglects any spatial component of what visitors do past the control point. Using social media can be effective to track actions and activity from geotags of images, but this data can be biased towards individual behaviours and points of interest (Wilkins et al. 2021). With the widespread adoption of mobile smart phones, using anonymized mobility data can be an effective tool in determining use of urban green spaces.

* 1. *Challenges with anonymized mobile data for conservation*

Using location data from mobile smart phones (hereafter mobility data) is not without limitations. Rightly so, mobility data is often anonymized by aggregating activity patterns to coarse resolutions to prevent harassment, crime, or injustice (de Montjoye et al. 2013; UN Global Pulse). This prevents tracking individual behaviours, activity by demographics, or fine resolution of activity patterns (e.g., < 10 m). Often urban greenspaces are not very large, so discerning activity within the space relative to nearby city development can be difficult. This becomes particularly problematic on greenspace boundaries that are often residential or high-traffic roads. Mobility data is rarely separated by mode of transportation (e.g., pedestrian, cyclist, motorist) and thus differentiating between cars driving along the boundaries and hikers within the greenspace can be difficult at too coarse scales. Similarly, determining type of activity from mobility data requires some assumptions. One can infer activity at a beach or picnic area could represent swimming and socializing respectively, but neither is definitive and requires knowledge about the land cover. Another common challenge with anonymized mobility data is thresholding to remove activity patterns below a certain level to prevent tracking select individuals on private properties. Urban greenspaces often have relatively less activity compared to adjacent city spaces, causing some areas to report no activity when activity is low. One approach to resolving the above challenges is to examine every greenspace case-by-case to validate activity patterns. However, for municipal land managers responsible for many properties, this approach is laborious and subjective. Clearly, there is a need for a systematic methodology that can synthesize accurate patterns of mobility data in urban greenspaces.

Connecting biodiversity observations to mobility data can pose a unique set of problems beyond validating human activity patterns. While mobility data has broad spatial and temporal coverage across a region, biodiversity data is often restricted to long-term monitoring plots static in location or multiple experimental sites that are short-term (Filazzola & Cahill 2021). Rarely is biodiversity data collected hourly or more than a few grid-cells of mobility data, presenting a challenge trying to connect these two disparate types of data. Additionally, biodiversity surveys are often conducted away from human activity, such as away from trails, playgrounds, or picnic areas. Using community science can be an effective tool at obtaining surveys with broad spatial and temporal coverage of urban greenspaces (Jimenez et al. 2020; Callaghan et al. 2020), but this type of data is inherently correlated with mobility data. A preliminary exploration of biodiversity and mobility data would include examining the relative use of land cover types in urban greenspaces to determine if certain areas, particularly sensitive ecozones, receive disproportionate levels of human activity.

* 1. *Objectives*

Anonymized mobility data can be a powerful tool in managing urban greenspaces, but methods are needed that can properly assess patterns of human activity. Using Mapbox Movement (<https://www.mapbox.com/movement-data>) we obtained anonymized mobile data representing human activity for the Greater Toronto Area in Canada. Mapbox presents mobile activity data aggregated to 100 x 100 m grid cells and to two-hour windows. We partnered with a local conservation authority, Conservation Halton, responsible for the management of urban greenspaces including nature reserves, parks, and unmanaged lands. Using Conservation Halton properties as a case study, we developed methods for the synthesis, management, and analysis of mobility data in urban greenspaces. We answered the following three questions:

1. How does anonymized mobility data compare to traditional measures of human activity in urban greenspaces, such as reservation data or trail density?
2. What information does mobility data capture that is different from traditional measures of greenspace use that is beneficial for land management?
3. Can mobility data be used to correlate patterns of human activity to biodiversity?

To our knowledge, this is the first-time Mapbox Movement Data has been used to explore questions in ecology and evolution. Thus, we needed to develop tools for management, validation, and comparison, especially when comparing to biodiversity patterns. We share the related methods and scripts to facilitate future users of this type of data.

**2.0 Data and Methods**

*2.1 Mapbox Movement Data*

We obtained anonymized mobility data from Mapbox for the Greater Toronto Region (43.23° N – 44.35° N, 78.83° W - 80.26° W). Mapbox is a private company that specializes in location data with products for application development. The data was provided aggregated to 100 x 100 m grid cells for June, July, and August 2020. Each grid cell has a monthly average value for 2-hour time windows throughout the day. Monthly averages are also separated into weekdays (Monday-Friday) and weekends (Saturday and Sunday). To anonymize the mobility data, grid cells with activity levels below a certain threshold were removed by Mapbox. Additionally, all activity patterns are normalized (i.e., scaled between 0 and 1) across the entire dataset. The activity pattern found within a grid-cell therefore represents the relative human activity between areas and not the raw total.

The data provided by Mapbox requires processing for comparisons with other spatial data. The anonymized mobility data is a text delimited file with a column specifying the boundaries of the grid cell. We found the intersection of each grid cell with the properties managed by Conservation Halton through an iterative loop. The file size for this dataset was large (> 4 million observations) and difficult to manage by personal computers. We conduct this intersection process in parallel for efficiency in runtime on the Compute Canada high-performance computing cluster (<https://www.computecanada.ca/>). We provide a function for matching spatial files in R (e.g., sf, sfc) to the grid cells from the mobility data for future users (*functions.r*). The output produced was a spatial data file (SF class; package sf) that had grid cells masked to the Conservation Halton properties.

Grid cells that were found to intersect on Conservation Halton properties often ended up masked to an area smaller than the full 100 x 100 m bounding box (for an example, see Figure 1). Many grids were reduced to areas that were only a fraction of the full size. However, the activity value for that grid cell remained unadjusted. To adjust the activity patterns to more accurately reflect the activity registered, we multiplied the mobility data for every grid cell by the polygon area of the respective cell. We also log-transformed the adjusted mobility data to account for wide right-skew in the data.

A significant challenge with using the mobility data for urban greenspaces was the accidental inclusion of activity outside of the greenspace. Roads and highways were especially challenging with boundaries of greenspace properties have high activity patterns that are likely not reflective of the activity within the property. Removing grid cells individually based on proximity to road is labourious, requires spatial information about roads, and can be subjective. For a more systematic approach, we identified any grid cell with human activity between 12 – 6 am. Many of the greenspaces are closed to access overnight and the remaining properties likely experience substantially lower traffic compared to daytime hours. The activity in these areas between 12 – 6 am are likely below the threshold identified by Mapbox for human activity. Conversely, roads and adjacent commercial operations remain active during overnight hours. Therefore, we excluded any grid cell with activity during these select hours to remove activity outside of the greenspaces from being reported (Appendix S1).

*2.2 Urban greenspace data*

As a case study for using anonymized mobility data with urban greenspaces, we selected 53 greenspace properties managed by Conservation Halton in Ontario, Canada. Conservation Halton is a conservation authority empowered by the provincial government to manage urban greenspaces for biological conservation, the preservation of ecosystem services, and for human recreation. The 53 properties include a range of management types including conservation areas used for recreation and conservation, natural areas where human visitation is not facilitated (i.e., no parking lots or trails), reserve areas where human activity is limited (e.g., fencing), and other areas that include stormwater management spaces and city parks (Table 1).

During the summer of 2020 visitation to seven of the conservation areas was controlled through reservations because of the Covid-19 pandemic. Individuals with reservations were allowed to visit the greenspace between 9 am and 6 pm for a maximum of 2 hours. These seven properties are among the most popular areas within Conservation Halton with popular features including waterbodies, rock formations, and a well-developed trail network. We obtained the reservation data for visitors that attended these seven properties for June, July, and August 2020. The reservation data included the number of individuals, the time of check-in, and park visited. Additionally, we obtained information about the greenspace property boundaries, ecological land classification, and officially managed trail network from Conservation Halton’s open data portal (<https://conservationhalton-camaps.opendata.arcgis.com/>). The ecological land classification categorizes land formations and vegetation communities to assist in the characterization of the landscape (Lee et al., 1998). Through ground surveys, lands are classified into different classifications such as marsh, forest, dune, swamp.

*2.3 Data analysis*

For every greenspace property, we calculated the total adjusted mobility data and divided it by the area of the respective property (Eq. 1). We used sum rather than median or mean because the number of grid cells varies over time because of the thresholding of activity that is applied to anonymize the data (i.e., no grid cells have zero values, they are simply absent). Thus, to obtain density of activity for comparisons among properties, we divided the sum adjusted activity by area of the property. We calculated the density of adjusted mobility data across all months, but separated for weekdays and weekends. For every property, we also calculated the percentage of area with any human activity. If a grid-cell had a mobility data value for any of the time periods within our dataset, that grid-cell was treated as having human activity. We totalled the area with identified with human activity and divided it by the total area of the property to determine the percent area (Eq 2.). The inverse of this value would the percent area of the property where human activity has been essentially non-existent throughout the timeframe.

Eq.1

Eq. 2

To compare mobility data to traditional estimates of visitation and to validate the estimated activity patterns through our adjusted metric, we examined the seven Conservation Halton properties that had reservation-only access. We summarized total number of visitors on weekends and weekdays for summer of 2020 with the sum adjusted mobility data and fit a linear model. The number of visitors and day-of-week were fitted as crossed-predictors.

We compared the density of trails among all properties with official trails (n = 16) to the sum adjusted mobility data and percent area with human activity using linear models. We identified any grid cell from the mobility data that intersected with the trail network (function st\_intersection; package sf) and summed the area of activity on trails divided by area with human activity. The resulting percentage represents the amount percent of human activity that is spent on trails. To determine if increasing greenspace use relates to activity on trails, we fit a linear model comparing the percent of activity on trails to the percent of human activity in the property.

We determined the mobility data that intersected with the ecological land classifications (ELC). The proportion of human activity was determined by dividing the area of human activity in each ELC class by the total area of human activity. The proportion of ELC used was determined by dividing the area of human activity in each ELC class by the total area of that ELC class in the respective greenspace property.

All analyses were conducted in R 4.1.2

**3.0 Results**

*3.1 Patterns of mobility data*

Mobility data was found to capture human activity within urban greenspaces. Almost all properties were found to have activity levels above the threshold used to anonymize the data. Most properties had between 30 and 50% of the total area with some observation of mobility data (Table 1). Reserve areas had the lowest percentage of mobility data (Table 1) as would be expected for lands where access is limited. For the properties where reservations were required, we found a strong positive relationship between the total number of reservations and total mobility data (F6 = 86.6, p < 0.0001, R2 = 0.97; Figure 2). The relationship between number of reservations and mobility data was mediated by day-of-week (p = 0.047), where mobile activity was much higher on weekends. This pattern suggests that for the same number of reservations, people often spend more time at these properties on the weekend relative to weekdays.

*3.2 Unique metrics from mobility data*

Mobility data provides greater spatial and temporal resolution on human activity relative to tracking visitation patterns. Many parks had hot spots of mobility data where activity was substantially higher than adjacent areas. For example, two properties (Hilton Falls and Kelso Conservation Area) had high activity patterns within their trail network (Figure 1). On average, parks with higher trail densities were found to have higher amounts of mobile activity (F14 = 18.2, p < 0.0001, R2 = 0.75; Figure 3A). Similarly, properties with high densities of trails also correlated with more area of the property containing some human activity (F9 = 6.57, p = 0.035, R2 = 0.36; Figure 3B). Properties with waterbodies tended to have higher percentages of coverage by human activity. The percent of activity on-trail relative to off-trail was significantly correlated with the percent area with human activity (F9 = 18.0, p = 0.002, R2 = 0.63; Figure 3C), suggesting increased use typically occurs on trails.

*3.3 Mobility patterns and biodiversity*

Human activity varied considerably by ELC class. Forest and cultural were the land classes where most of the mobility data occurred, followed by talus and cliff (Figure 4A). However, relative to the abundance of the land classes in each property, rock formations were disproportionately visited relative to other land types, including talus, cliff, crevice and cave, and bluff (Figure 4B). By contrast, forests and cultural were used relatively infrequently in proportion to their abundance among properties.

**4.0 Discussion**

We examined the relationship with mobility data and urban greenspace use for 56 properties in the Greater Toronto Area. We found that anonymized mobility data effectively captures activity in urban greenspaces, with a significant correlation between number of visitors and total mobility data (R2 =0.97; Figure 2). However, it is important to note that pre-processing is required to reflect the activity more accurately in greenspaces. The mobility data proved an effective tool at capturing human activity patterns in urban greenspaces including patterns of trail and land use (Figures 3, 4). For land managers looking to balance human use with biological conservation, the mobility data appears a powerful tool in determining hot spots of activity, ecological refugia, and encroaching activity on restricted areas.

*4.1 Description of observed patterns and interpretation*

* Points of interest and recreation areas with typically have higher activity patterns
* Small properties are prone to noise
* Other data may be required to fill in the blank. E.g., managed trails vs. all trails.
* Biodiversity surveys often do not overlap

*4.2 Limitations of location data*

Location data is a powerful tool with broad spatial and temporal resolution, but there are biases in use. Although mobile device use has expanded rapidly across the globe (citations), there remain large differences among countries (citations). In countries where mobile device adoption is high, such as Canada where this study took place, mobility data may more accurately reflect human activity relative to others. However, even within countries there are differences in mobile device use between rural, sub-urban, and urban communities. In the United States, rural Americans have consistently few mobile devices relative to urban or sub-urban areas (Vogels 2021). The devices and the software they use are also prone to biases. There can be noise in quantifying activity caused different accuracies among devices and operating system (citations). The choice of software application by the device user can also determine activity patterns. For example, a person using a ride sharing application is more likely to have location services turned on, whereas a person in an urban greenspace may not have any application open. As greenspaces are often viewed as a place to “disconnect” or be engaged in activities that discourage mobile device use (e.g., swimming, jogging), activity patterns may be less accurate than when compared to roads. These biases are important when considering expanding the applicability of mobility data to other areas or regions.

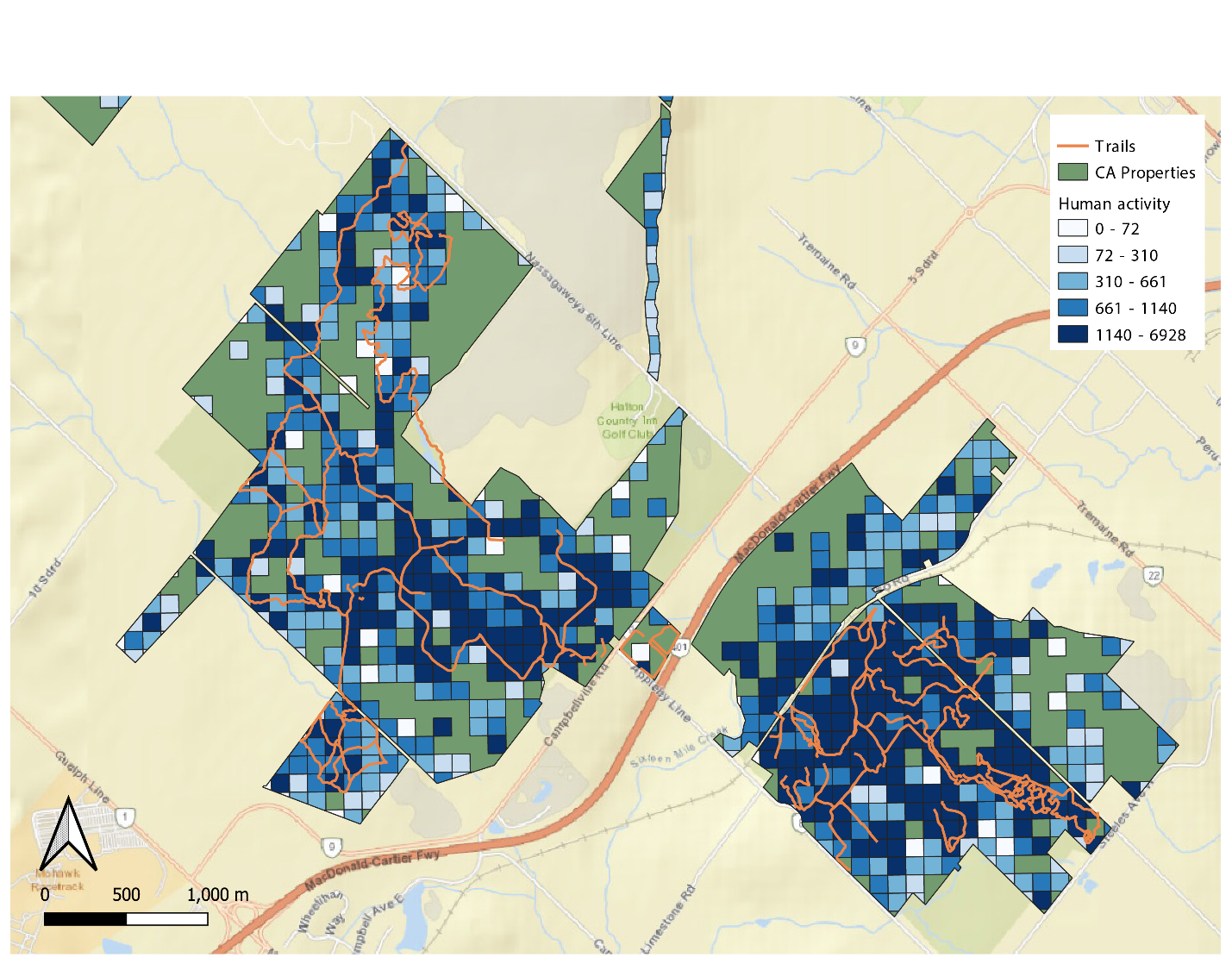
For future users of mobility data

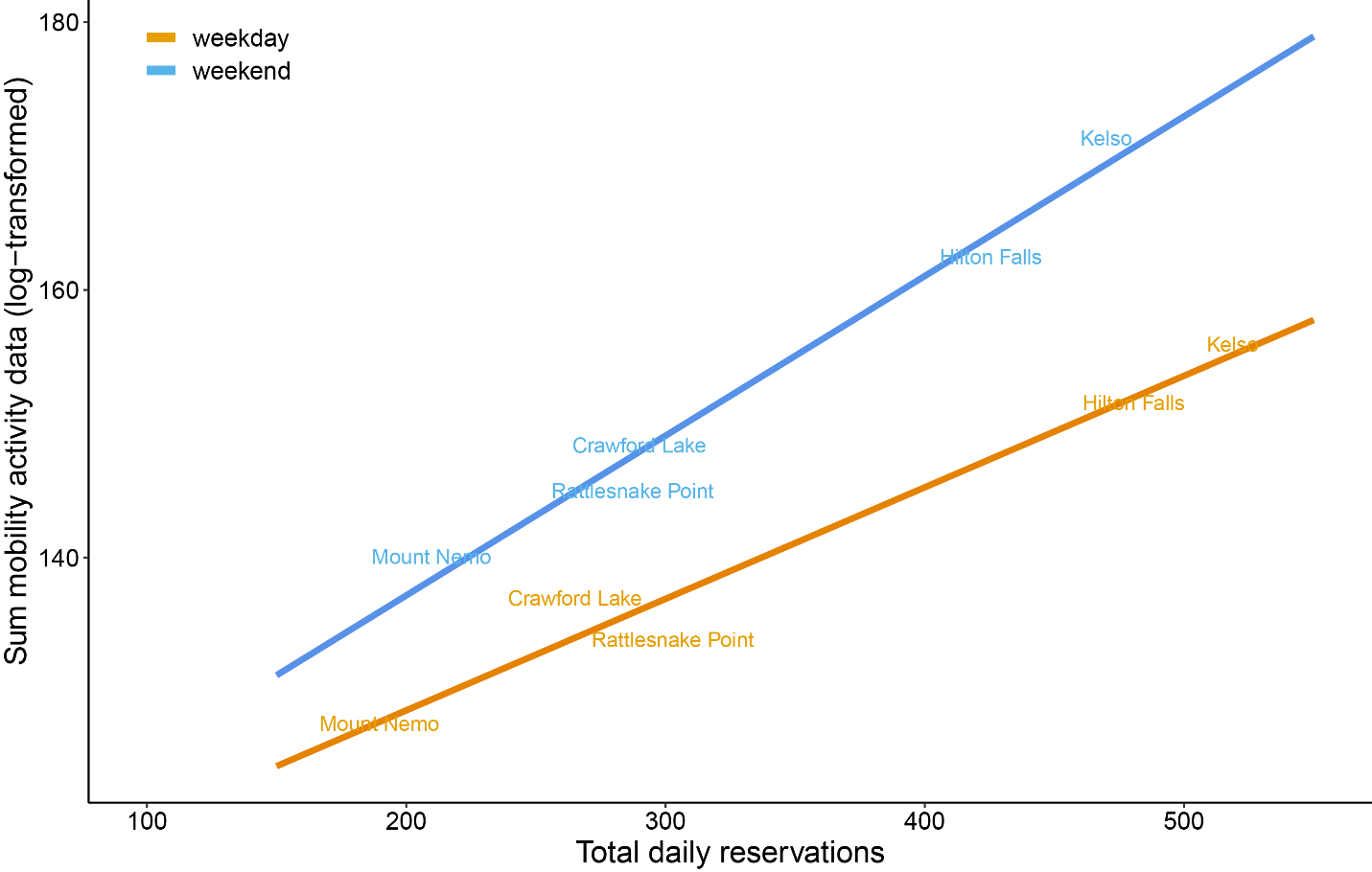
* Table for meta-data: <https://www.mapbox.com/blog/how-to-utilize-mapbox-movement-data-for-mobility-insights-a-guide-for-analysts-data-scientists-and-developers>
* Using sum rather than median/median, where are based on the denominator which varies based on quadkeys. Divide by area.
* Available scripts for processing the data

Management implications

**Table 1:** General characteristics of the properties within Conservation Halton’s jurisdiction including type of land, whether actively managed, number of properties, and average property size. Activity coverage represents the percentage of area within the property that has any human activity data from Mapbox.

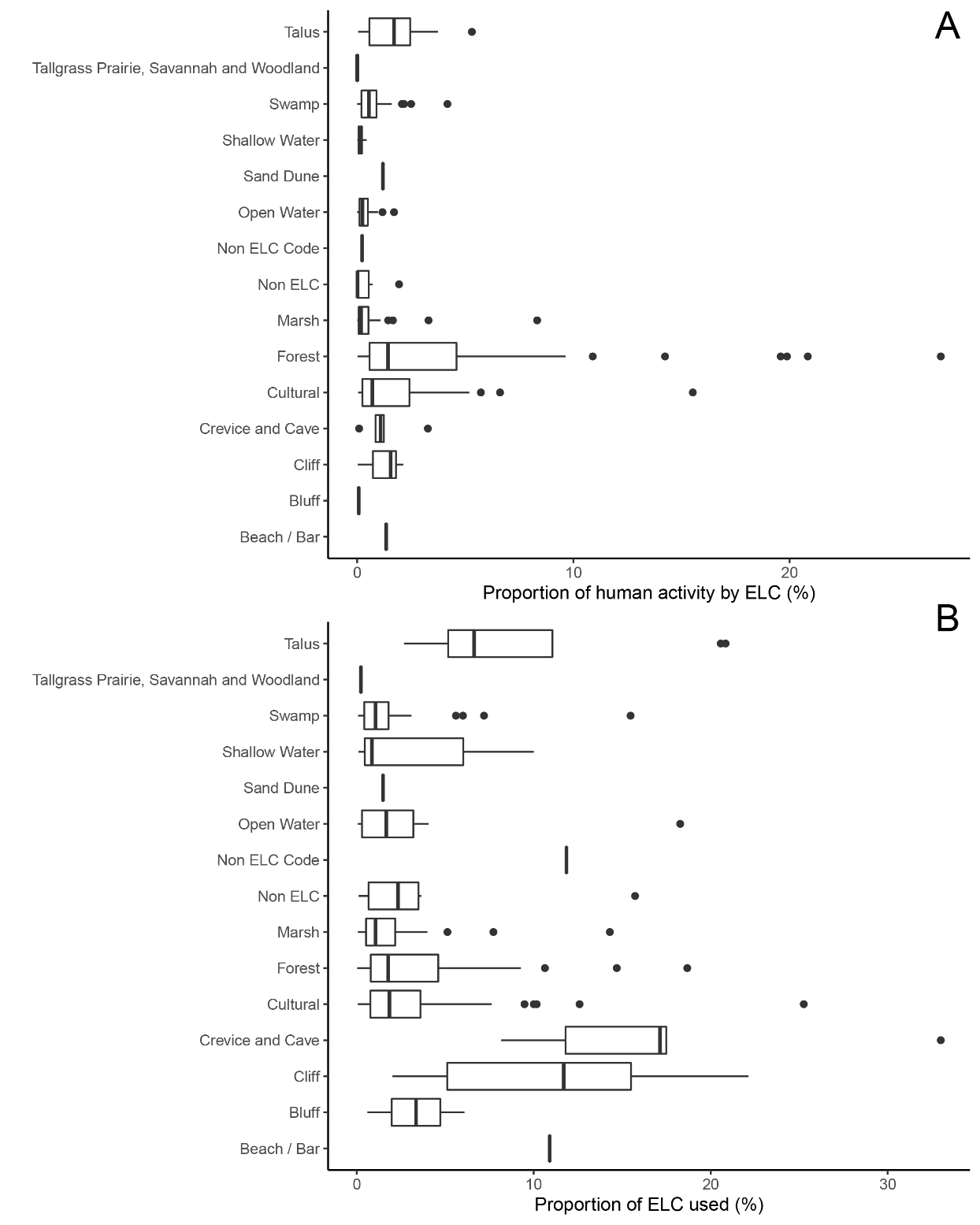
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Land Type** | **Managed** | **Properties** | **Property Size (km2)** | **Activity coverage (%)** |
| Conservation Area | Managed Land | 7 | 3.68 | 59.8 |
| Conservation Area | Non-Managed Land | 8 | 0.88 | 48.4 |
| Natural Area | Non-Managed Land | 16 | 0.5 | 44.5 |
| Other | Non-Managed Land | 9 | 0.1 | 34.3 |
| Reserve Area | Non-Managed Land | 13 | 0.24 | 26.3 |

**Figure 1**: A representation of mobility data in two Conservation Halton properties Hilton Falls and Kelso Conservation Area. Each blue grid-cell represents a 100 x 100 m pixel of anonymized mobility data with darker colours representing higher densities of human activity. Orange lines within property boundaries are official trails managed within the greenspaces. Green areas within property boundaries have activity levels too low to be available in the mobility dataset, potentially representing refugia in the park where human activity is negligible.



**Figure 2:** The total daily reservations for Conservation Halton properties that had reservation-only entrance were strongly positively associated with mobility data. On average, mobility data was much higher on weekends compared to weekdays. Reservation and mobility data are the total of activity within each conservation authority property between May and August 2020.

**Figure 3:** Patterns of trail use with mobility data. Properties with higher trail densities were found to have significantly high activity patterns once controlling for the two smallest properties (Shanahan and Robert Edmondson). There was also a positive relationship with area of human activity with trail densities, especially properties with water bodies that had high percentages of human interaction. The percentage of activity on trails (i.e., on-trail vs. off-trail use).



**Figure 4**: Patterns of human activity in each of the ecological land classifications (ELC) for Conservation Halton properties. Proportion of human activity by ELC represents the percent of human activity within the property separated by each ELC class (Panel A). Proportion of ELC used represents the percent of human activity in that ELC class relative to the total coverage of that ELC class within the property (Panel B).

**Appendix S1**

|  |  |
| --- | --- |
|  |  |
|  | |
| **Figure S1:** Exclusion of grid cells with high activity patterns overnight (12 – 6 am) maximizes activity found within greenspace poperties. Each grid cell represents a 100 x 100 m of mobility data. Purple grid cells are those identified to have high activity overnight. These identified grid-cells were commonly along highways, major roads, or residental areas. | |