**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 are removed. 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 require processing for comparisons with other spatial data. The anonymized mobility data came as 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 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.

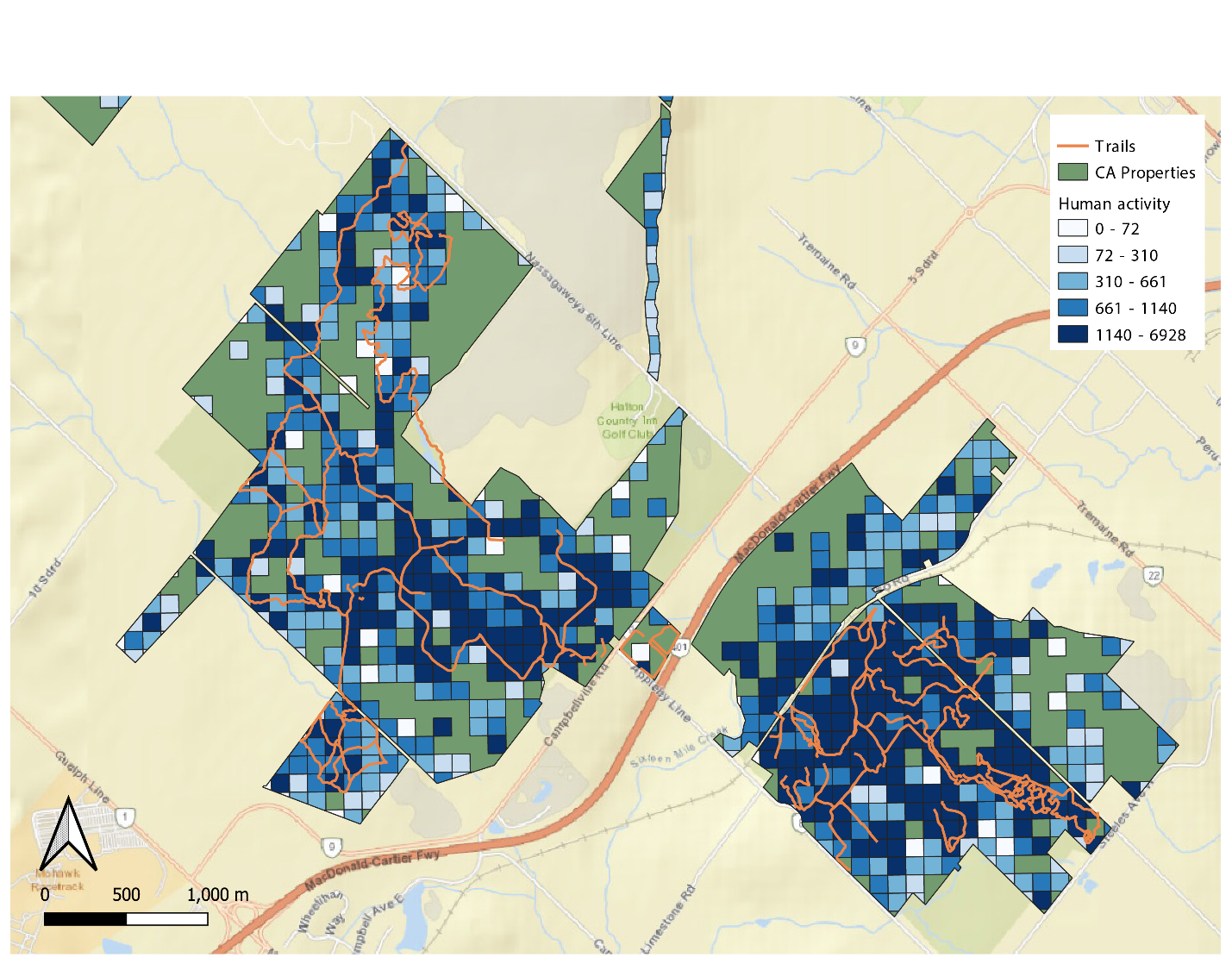
A significant challenge with using the mobility data for urban greenspaces was the accidental inclusion of activity outside of the park. Roads and highways were especially challenging with boundaries of greenspace properties have high activity patterns that are likely not reflective of the activity within the property. Removing 100 x 100 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 properties are closed to access overnight and the remaining properties likely experience substantially lower traffic. 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 activity during overnight hours. We excluded any grid cell with activity during these select hours to remove activity outside of the greenspaces from being reported (Appendix A).

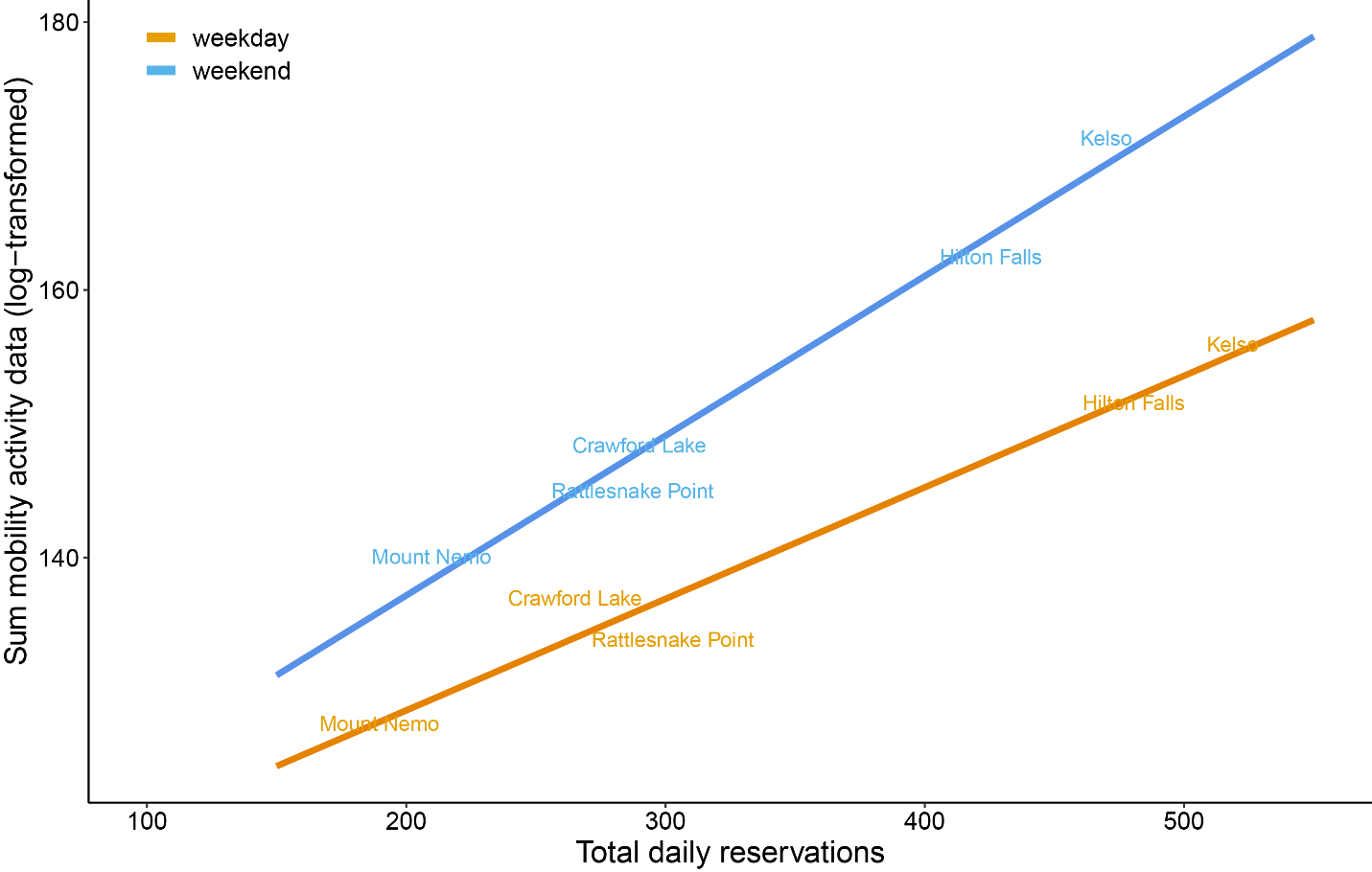
**3.0 Results**

3.1

**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)