**Using anonymized mobile data to quantify areas across urban green spaces**

Alessandro Filazzola1,2 Garland Xie3, Kimberly Barrett4, Andrea Dunn4, Marc T.J. Johnson1,5, & J. Scott MacIvor1,2,3

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

**Keywords:** parks, urban ecology, mobile cell data, human-wildlife, urban environments, human activity

**Abstract**

1. Cities are increasing in density and coverage globally, increasing the value of remaining urban green spaces. Understanding the interactions between city residents and green spaces is critical for human health, biological conservation, and sustainable development. However, quantifying green space use is particularly challenging.
2. We used anonymized mobile cell data to characterize human activity in green spaces for the Greater Toronto Area, Canada. We describe some of the main challenges associated with using anonymized mobility data, especially in the context of green spaces and biodiversity monitoring. Importantly, we provide solutions including *R* scripts that can assist in analyzing mobility data.
3. We found anonymized mobility data is strongly correlated with visitation records and that this data may be used to identify hotspot or coldspot areas. Parks with a more elaborate trail network typically experienced higher visitation rates and a substantial proportion of activity remain on-trails. We identified certain land covers that were more popular, such as rock formations, and found a relationship between human activity and tree composition.
4. For the first time, our study demonstrates anonymized mobility data as a powerful tool for quantifying human activity in green spaces. Management of green spaces both for human use and biological conservation, will be a significant challenge over the coming decades because of continuing urbanization. We have provided a framework to allow usage of mobility data that overcomes some common limitations and provide details for connecting with this information with biodiversity data. Importantly, we include a series of recommendations for green spaces that may assist in biomonitoring and supporting sustainable human use.

**1. Introduction**

Cities are rapidly expanding, creating new challenges for managing urban green spaces. 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 (Nations, 2015; Riahi et al., 2017). As cities increase in size and area, green spaces including remnant natural areas, protected reserves, and urban parks, face growing stressors from human activity. Direct human use of green spaces can negatively impact urban wildlife including trampling, introduction of non-native species, and pollution (McDonnell et al., 2008; Shochat et al., 2010; Mason, Newsome, Moore, & Admiraal, 2015; Cadotte, Yasui, Livingstone, & MacIvor, 2017). However, both managed and unmanaged green spaces are important for city residents as a place for exercise, recreation, socialization, and supporting mental well-being (Jim & Chen, 2006; Lee & Maheswaran, 2011; Nutsford, Pearson, & Kingham, 2013; Grzyb, Kulczyk, Derek, & Woźniak, 2021). This need has been amplified during the Covid-19 pandemic (Rice et al., 2020; Volenec, Abraham, Becker, & Dobson, 2021). Thus, managing green spaces is a delicate balancing act between utility for people and conservation of biodiversity.

One of the main limitations in effectively managing green spaces is the uncertainty around how and when people use these areas. Some managed parks use a reservation-based system with controlled points of entry, whereas other green spaces have unrestricted access points. Trails are created to facilitate human movement and reduce disturbance to biodiversity, but residents will still venture off-trail or erode new paths of easily navigable terrain, e.g., desire lines (Lynn & Brown, 2003). 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 green spaces. However, capturing human mobility at a resolution fine enough for management, such as less than 100 x 100 m, is challenging. Typical methods for quantifying human activity include record keeping visitors at entrance points, video monitoring, or *post-hoc* assessment of visitor impacts, such as campsite use (D’Antonio, Monz, Newman, Lawson, & Taff, 2013; Ward et al., 2014; Marion, 2016). However, this data often neglects any spatial component of what visitors do outside of control points. Using social media can be effective to track actions and activity from geotags of publicly shared images, posts, or tweets (Tenkanen et al., 2017; Donahue et al., 2018), but this data can be biased towards individual behaviours and points of interest (Wilkins, Wood, & Smith, 2021). With the widespread adoption of mobile smart phones, using anonymized mobility data can be an effective tool in determining patterns in use of green spaces

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

Using location data from mobile devices that identify user location (hereafter mobility data) is not without limitations. The challenges associated with using mobility data for estimating human activity can be generalized into social, technical, and data issues (Table 1). Understanding these challenges is crucial when deciding to infer human activity, but there are specific considerations with respect to green spaces. For example, mobility data is often anonymized by aggregating activity patterns to coarse resolutions to prevent harassment, crime, or injustice (de Montjoye, Hidalgo, Verleysen, & Blondel, 2013; UnitedNations, 2015). This prevents tracking individual behaviours, activity by demographics, or fine resolution of activity patterns (e.g., < 100 m). Often green spaces found in urbanizing areas are not very large, so discerning activity within the green space relative to nearby city development can be difficult. This becomes particularly problematic on green space boundaries that are delineated by private residential properties or high-traffic roads. Mobility data are rarely separated by mode of transportation (e.g., pedestrian, cyclist, motorist) and thus differentiating between cars driving along the boundaries and hikers within the adjacent green space can be difficult at coarse scales. Similarly, determining behaviour of individual from mobility data requires some assumptions (Table 1). One can infer activity at a beach or picnic area could represent swimming and socializing respectively, but neither is definitive and requires knowledge of how land cover matches activity type. 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. Green spaces often have relatively less activity compared to adjacent paved city spaces (e.g., roads, buildings), causing some areas to report no activity when activity is low. One approach to resolving the above challenges is to examine every green space 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 and reproducible methodology that can synthesize accurate patterns of mobility data across green spaces.

Connecting biodiversity observations to mobility data can pose a unique set of challenges 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 Jr, 2021). Unlike mobility data, biodiversity data are rarely collected hourly or covering a broad spatial area, presenting a challenge when trying to connect these two disparate types of data (e.g., evaluating effects of human activity on biodiversity). Additionally, biodiversity surveys are often conducted away from areas with high human activity (e.g., trails, playgrounds, picnic areas) reducing the chance that any activity would be recorded. Using community science can be an effective tool at obtaining surveys with broad spatial and temporal coverage of green spaces (Callaghan, Ozeroff, Hitchcock, & Chandler, 2020; Jimenez, Pejchar, & Reed, 2021) 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 green spaces to determine if certain areas, particularly where there is sensitive habitat, receive disproportionate levels of human activity.

* 1. *Objectives*

Anonymized mobility data can be a powerful tool in managing green spaces, but methods are needed that can properly assess patterns of human activity. Using Mapbox Movement ([www.mapbox.com/movement-data](http://www.mapbox.com/movement-data)) we obtained anonymized mobile data representing human activity aggregated to 100 x 100 m grid cells and two-hour windows. We partnered with Conservation Halton ([www.conservationhalton.ca](http://www.conservationhalton.ca)), a local conservation authority responsible for natural areas, protected reserves, and urban parks in the regional municipality of Halton. In the following study, we develop methods for the synthesis, management, and analysis of anonymized mobility data in green spaces by answering the following three questions:

1. How does anonymized mobility data compare to traditional measures of human activity in urban green spaces, 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 the environment?

To our knowledge, this is the first-time Mapbox Movement Data has been used to explore questions relating to human impacts on the environment. Thus, we needed to develop tools for management, validation, and comparison, especially when evaluating biodiversity patterns to connect of people to nature and support biological conservation in growing urban areas. We share the related methods and scripts to facilitate future users of mobility data for monitoring human activity in green spaces.

**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) which includes the regional municipality of Halton. Mapbox is a private company that specializes in location data with products for application development. The data were provided in aggregated, 100 x 100 m grid cells, for June, July, and August 2020 by Mapbox. Each grid cell has a monthly average value for 2-hour time windows throughout a complete 24-hour 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 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 ([www.computecanada.ca/](http://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 green spaces.

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 green spaces was the accidental inclusion of activity outside of the green space. Roads and highways were especially challenging when adjacent to property boundaries, causing 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 green spaces 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 green spaces from being reported (Appendix S1).

*2.2 Green space data*

As a case study for using anonymized mobility data with green spaces, we selected 53 green spaces managed by Conservation Halton in Ontario, Canada. Conservation Halton is a conservation authority empowered by the provincial government to manage green spaces for biological conservation, the preservation of ecosystem services, and for human recreation. The 53 green spaces 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 2).

During the summer of 2020, visitation to seven of the green spaces managed as conservation areas was controlled through reservations because of the Covid-19 pandemic. Individuals with reservations were allowed to visit the green space between 9 am and 6 pm for a maximum of 2 hours. These seven green spaces are among the most popular within Conservation Halton with features including waterbodies, rock formations, look-out sand well-developed trail networks. We obtained the reservation data for visitors that attended these seven properties for June, July, and August 2020. Each reservation included the number of individuals, the time of check-in, and the park visited. Additionally, we obtained information about the property boundaries, ecological land classification, and officially managed trail network from green space Conservation Halton’s open data portal ([www.conservationhalton.ca/mapping-and-data](http://www.conservationhalton.ca/mapping-and-data)). 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 grouped into different classifications such as marsh, forest, dune, and swamp.

Tree surveys were conducted at eleven unique sites in Conservation Halton between 2006 and 2019. Although the mobility data was acquired for 2020, the composition of tree species at these sites remains relatively constant between years. To confirm composition did not change significantly between years, we conducted a constrained correspondence analysis with year as a predictor and found it did not significant effect tree species (F26 = 1.26, p = 0.23).

*2.3 Data analysis*

For every green space, we calculated the total adjusted mobility data and divided it by the area of the respective property (Eq. 1).To quantify total mobile activity, we used sum of an unbiased estimate, rather than median or mean, because the number of grid cells varies over time due to 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 the area of the green space. 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 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 be the percent area of the property where human activity is at a non-detectable level 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 June, July, and August 2020 with the sum adjusted mobility data and fit a linear model. The number of visitors and day-of-week were fitted as interacting predictors.

We compared the density of trails among all properties with official trails (number of green spaces = 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 greater visitation to a green space relates to activity on trails, we fit a linear model comparing the percent of human activity on trails to the percent of human activity in the property.

Next, we determined how the anonymized mobility data 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 established by dividing the area of human activity in each ELC class by the total area of that ELC class in the respective green space.

Lastly, we tested if mobility data had any relationship with biodiversity patterns in these green spaces. We compared tree composition part of a long-term monitoring project at 11 sites using a partially constrained correspondence analysis (pCCA). We fit as predictors the percent area of human activity, the sum adjusted mobility data (averaged across weekday and weekend), and the proportion of weekend activity (i.e., weekend activity / weekday activity). Year was fit as a conditioning matrix to partial out any interannual differences in tree composition. However, many of these trees long-lived individuals without substantial differences in species among years. We conducted a permutation test of pCCA (function *anova.cca*, package *vegan*) to determine model significant and percent of variation explained (Legendre, Oksanen, & ter Braak, 2011).

All analyses were conducted in *R* version 4.1.2 (R Core Team, 2019). All scripts and source codes are available on a public repository that can be found at <https://github.com/afilazzola/CUERecreationEcology> including a *functions.r* file that contains tools for processing mobility data. The function *spatial\_Mapbox\_Find* is useful for finding Mapbox activity patterns that intersect a defined spatial object of SF class but is designed to be conducted in parallel (see repository for more documentation).

**3.0 Results**

*3.1 Patterns of mobility data*

With adjustments, the anonymized mobility data was found to accurately capture human visitation within green spaces. All properties were found to have at least on grid-cell with activity levels above the threshold used to anonymize the data. Most properties had between 30 and 50% of the total area with a detectable observation of mobility data (Table 2). Reserve areas had the lowest percentage of mobility data (i.e., cool spots) (Table 2) 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 green spaces 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 green space 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 the environment*

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 (i.e., recreational) were used relatively infrequently in proportion to their abundance among green spaces.

We found that our measures of human activity were significantly related to patterns of tree composition across the eleven green space properties (F15 = 2.74, p = 0.001; Figure 5), explaining 25% of the variation in species. Although few species were unique identified to correlate with human activity, there were some correlations observed. For instance, Sassafras species (Lauraceae: *Sassafras spp.*) were found to closely associated with green spaces that have a larger human footprint, i.e., areas with greater human coverage (Figure 5). Yellow birch (Betulaceae: *Betula alleghaniensis*) and black ash (Oleaceae: *Fraxinus nigra*) were both found to correlate with total human activity and proportion of weekend activity (Figure 5). Some species appear to be negatively associated with human activity including bitternut hickory (Juglandaceae: *Carya cordiformis*), ironwood (Betulaceae: *Ostrya virginiana*), and American elm (Ulmaceae: *Ulmus americana*).

**4.0 Discussion**

For the first time, we examine relationships between human activity levels and green spaces environments using anonymized mobility data. We found a significant correlation between number of visitors and total mobility data (R2 =0.97; Figure 2) demonstrating that anonymized mobility data effectively captures human activity in green spaces. However, it is important to note that pre-processing is required to reflect the activity more accurately in green spaces. The mobility data proved an effective tool at capturing human activity patterns in green spaces including patterns of trail and land use category (Figures 3, 4). For land managers looking to balance human use with biological conservation in green spaces, we illustrate that anonymized mobility data is a powerful and accessible tool for pinpointing hot spots of human activity, prospective ecological refugia (i.e., cold spots where human activity is low), and encroachment of activity on restricted areas. This information can be used to *a priori* plan biomonitoring to capture impacts along a gradient of human activity level.

*4.1 Description of observed patterns and interpretation*

Activity patterns varied considerably, but predictably among green spaces. Off-trail use is a significant problem in conserving biodiversity, causing disturbance, trampling, and introduction of non-native species (Nepal & Way, 2007; Mason et al., 2015; Barros, Aschero, Mazzolari, Cavieres, & Pickering, 2020). In the green spaces evaluated, we found that most visitors appear to remain within the designated use areas, and highest activity observed along trails (Figure 1) or in recreational spaces (e.g., picnicking areas) (Figure 4). Still, the activity outside of designated use areas persisted across green spaces with the percent of activity on-trails dropping below 5% of total activity (e.g., Kilbride). Some of these green spaces have unofficial trails that are managed by non-profits or local communities. For example, the Bruce Trail Conservancy manages a 904 km trail that intersects some of these green spaces but that is independent of Conservation Halton ([www.brucetrail.org/](http://www.brucetrail.org/)). Future land managers interested in relating trail networks to human activity may need to aggregate trail locations from multiple data sources, such as AllTrails ([www.alltrails.com/](http://www.alltrails.com/)). The mobility data used here can also guide managers to areas of frequent or abundant human activity, but where no trails exist, to determine where off-trail incursions are most common. For green spaces in Conservation Halton, highest activity of off-trail use appeared near the entrances or in between adjacent trails (e.g., Figure 1). These perceived negative impacts to green spaces could be flipped to a positive, if human behaviour (i.e., where people go off-trail the most) is useful for guiding future trails and accessibility. However, biodiversity in these areas will be more susceptible to disturbance and if off-trail activity occurs in sensitive habitats, this human activity should be monitored and modified. Additionally, while trails are typically managed mitigate impacts to the landscape, many are still constructed in ecologically sensitive areas (e.g., Tomczyk, 2011). Thus, high activity on-trails are not enough information alone to determine effective management.

*4.2 Human activity and the environment*

Certain land covers had higher activity relative to others. Forests were the most visited land cover, but also represented the largest area of the green spaces, particularly surrounding the trail network (Figure 4). Cliff and rock formations were visited disproportionately high relative to their cover, likely because these areas are scenic points of interest. Rock formations are consistently popular attractions in green spaces (e.g., Sinclair, Mayer, Woltering, & Ghermandi, 2020), but can pose a hazard both to human health from climbing accidents and to local biodiversity from trampling (Newsome, Dowling, & Leung, 2012; Gstaettner, Kobryn, Rodger, Phillips, & Lee, 2019). There is certainly a delicate balancing act between restricting access to preserve biodiversity and facilitating recreation by the public. Many green spaces visitors are often seeking quiet, uncongested areas (Frick, Degenhardt, Buchecker, & others, 2007; Home, Hunziker, & Bauer, 2012) and overcrowding can cause negative interactions among users, which decreases support for conservation (Rossi, Byrne, Pickering, & Reser, 2015). Yet, some areas in green spaces will inherently by more sensitive than others. Knowledge about preferred land cover types can inform land managers to either prioritize conservation potential in areas with low human activity (e.g., swamps, marshes, or grasslands) or implement measures (e.g., fencing) to mitigate impacts to high activity areas.

*4.3 Limitations of location data*

Anonymized mobility data is a powerful tool with broad spatial and temporal resolution, but there are biases and considerations in its use (Table 1). Although mobile device use has expanded rapidly across the globe (Obile, 2016), there remain large differences among regions and demographics (e.g., van Biljon & Kotzé, 2007; Baishya & Samalia, 2019). In areas where mobile device adoption is high, mobility data may more accurately reflect human activity relative to others location-based data (e.g., social media, geotagged photos, iNaturalist). Our study took place in Canada where LTE mobile networks cover 99% of the population (CRTR, 2018) and 80% of Canadians report having a mobile data plan for personal use ([www150.statcan.gc.ca](file:///D:\RStudio\CUERecreationEcology\manuscript\www150.statcan.gc.ca)). 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) and in Canada 81% living within a city metropolitan area (CMA) had mobile data plans compared to 73% in non-CMA areas (<www150.statcan.gc.ca>). The devices and the software applications used will also be prone to biases (Table 1). There can be noise in quantifying activity caused different accuracies among devices, operating system, software, and location (Lin, Kansal, Lymberopoulos, & Zhao, 2010). 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 a green space may not have any application open. As green spaces 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 demographics or regions (Table 1).

Associating human activity with patterns of biodiversity is inherently challenging. We demonstrated preliminary evidence that human activity interpreted from mobility data is related to patterns of tree composition in green spaces (Figure 5). There may be a few reasons for this observed relationship. Human visitors may prefer a certain assemblage of tree species or land cover that supports certain species. For instance, we saw a negative relationship between human activity and ironwood, but ironwood is more commonly found in swamps or high-moisture areas (Metzger, 1990), which we saw markedly lower human activity. Conversely, yellow birch has visually appealing bark and thus may attract greater human activity. In addition to biodiversity patterns driving human activity, there may be an affect of human activity on tree composition. Higher human activity leads to greater disturbance, potentially limiting recruitment or survivability of some tree species. Eastern Hemlock is a relatively tolerant to disturbance compared to other species, and thus may be more likely to associated with higher intensity of human activity (Frelich & Lorimer, 1991). There may also be indirect factors driving the observed relationships, such as high deer densities, which have a significant effect on tree composition in Southern Ontario (Tanentzap et al., 2011; Filazzola, Tanentzap, & Bazely, 2014) but may also be appealing to visitors. Other types of biodiversity data may be useful for disentangling some of these patterns. We used long-term monitoring data of trees that is relatively robust to survey biases but is restricted to select survey locations and certain years. Community science databases (e.g., iNaturalist, eBird) would have larger spatial and temporal coverage allowing for greater overlap with the mobility data. However, many of these data sources will be inherently correlated to mobility data since they have smartphone applications. Relating patterns of biodiversity to mobility data is thus a promising opportunity that requires careful consideration of data source or experimental design.

*4.4 Management implications*

The prospects of anonymized mobility data for managers of green spaces and more broadly, to determine human impacts on the environment, is exciting. We show it surpasses information available from traditional methods for monitoring visitation by including human activity hot spots and cool spots, off-trail movement and human behaviour that could inform future planning for accessibility and development of more parsimonious trail routes that mitigate off-trail activity. With improved understanding of human activity across green spaces, biomonitoring inventories can be coordinated in ways that capture biodiversity information across gradients of human activity levels. This could be critical to disentangle local disturbance and modification to community composition, loss of sensitive species, and management of at-risk species. Since human activity in green spaces is correlated with invasive species propagules (Cadotte et al., 2017), biomonitoring inventories designed to include sites along human activity gradients may be critical to proactively manage invasive species and mitigate economic costs associated with eliminating established species.

Anonymized mobility data collected by Mapbox is recorded hourly and so continuous monitoring of green spaces for changes in activity patterns are possible. With the Covid-19 pandemic, many local parks have seen a significant increase in the number of people visiting, and with this come a wider range of behaviours (Rice et al., 2020; Volenec et al., 2021). Knowledge of when green spaces are visited and where in the park visitation is highest, will permit significant refinement of both biomonitoring, management, and engagement to more effectively (and economically) link the needs of humans to access nature and recreation opportunities while also conserving biodiversity in green spaces.

**Acknowledgments**

We thank Mapbox for providing the anonymized mobility data. This research was funded by a post-doctoral fellowships awarded to AF by the Center for Urban Environments and School of Cities at the University of Toronto, Canada.

**Author’s Contributions**

All authors participated in the conceptualization of the project, interpretation of the results, and editing of the manuscript. AF and GX developed the methodology and software including all validation and formal analysis. AF was responsible for data curation, presentation of results, and writing the original draft. JSM managed and supervised the project. Funding for the post-doctoral fellowship was acquired by MJ.

**References**

Baishya, K., & Samalia, H. V. (2019). Factors Influencing Smartphone Adoption: A Study in the Indian Bottom of the Pyramid Context. *Global Business Review*, *21*(6), 1387–1405. doi:10.1177/0972150919856961

Barros, A., Aschero, V., Mazzolari, A., Cavieres, L. A., & Pickering, C. M. (2020). Going off trails: How dispersed visitor use affects alpine vegetation. *Journal of Environmental Management*, *267*, 110546. doi:https://doi.org/10.1016/j.jenvman.2020.110546

Cadotte, M. W., Yasui, S. L. E., Livingstone, S., & MacIvor, J. S. (2017). Are urban systems beneficial, detrimental, or indifferent for biological invasion? *Biological Invasions*, *19*(12), 3489–3503. doi:10.1007/s10530-017-1586-y

Callaghan, C. T., Ozeroff, I., Hitchcock, C., & Chandler, M. (2020). Capitalizing on opportunistic citizen science data to monitor urban biodiversity: A multi-taxa framework. *Biological Conservation*, *251*, 108753. doi:https://doi.org/10.1016/j.biocon.2020.108753

CRTR, C. R. and T. C. (2018). *Communications Monitoring Report*. Ottawa, Canada. Retrieved from https://crtc.gc.ca/pubs/cmr2018-en.pdf

D’Antonio, A., Monz, C., Newman, P., Lawson, S., & Taff, D. (2013). Enhancing the utility of visitor impact assessment in parks and protected areas: A combined social–ecological approach. *Journal of Environmental Management*, *124*, 72–81. doi:https://doi.org/10.1016/j.jenvman.2013.03.036

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*(1), 1376. doi:10.1038/srep01376

Donahue, M. L., Keeler, B. L., Wood, S. A., Fisher, D. M., Hamstead, Z. A., & McPhearson, T. (2018). Using social media to understand drivers of urban park visitation in the Twin Cities, MN. *Landscape and Urban Planning*, *175*, 1–10. doi:https://doi.org/10.1016/j.landurbplan.2018.02.006

Filazzola, A., & Cahill Jr, J. F. (2021). Replication in field ecology: Identifying challenges and proposing solutions. *Methods in Ecology and Evolution*, *12*(10), 1780–1792. doi:https://doi.org/10.1111/2041-210X.13657

Filazzola, A., Tanentzap, A. J., & Bazely, D. R. (2014). Estimating the impacts of browsers on forest understories using a modified index of community composition. *Forest Ecology and Management*, *313*, 10–16.

Frelich, L. E., & Lorimer, C. G. (1991). Natural disturbance regimes in hemlock-hardwood forests of the upper Great Lakes region. *Ecological Monographs*, *61*(2), 145–164.

Frick, J., Degenhardt, B., Buchecker, M., & others. (2007). Predicting local residents’ use of nearby outdoor recreation areas through quality perceptions and recreational expectations. *Forest Snow and Landscape Research*, *81*(1–2), 31–41.

Grzyb, T., Kulczyk, S., Derek, M., & Woźniak, E. (2021). Using social media to assess recreation across urban green spaces in times of abrupt change. *Ecosystem Services*, *49*, 101297. doi:https://doi.org/10.1016/j.ecoser.2021.101297

Gstaettner, A. M., Kobryn, H. T., Rodger, K., Phillips, M., & Lee, D. (2019). Monitoring visitor injury in protected areas - analysis of incident reporting in two Western Australian parks. *Journal of Outdoor Recreation and Tourism*, *25*, 143–157. doi:https://doi.org/10.1016/j.jort.2018.04.002

Home, R., Hunziker, M., & Bauer, N. (2012). Psychosocial outcomes as motivations for visiting nearby urban green spaces. *Leisure Sciences*, *34*(4), 350–365.

Jim, C. Y., & Chen, W. Y. (2006). Recreation–amenity use and contingent valuation of urban greenspaces in Guangzhou, China. *Landscape and Urban Planning*, *75*(1), 81–96. doi:https://doi.org/10.1016/j.landurbplan.2004.08.008

Jimenez, M. F., Pejchar, L., & Reed, S. E. (2021). Tradeoffs of using place-based community science for urban biodiversity monitoring. *Conservation Science and Practice*, *3*(2), e338.

Lee, A. C. K., & Maheswaran, R. (2011). The health benefits of urban green spaces: a review of the evidence. *Journal of Public Health*, *33*(2), 212–222. doi:10.1093/pubmed/fdq068

Legendre, P., Oksanen, J., & ter Braak, C. J. F. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, *2*(3), 269–277. doi:https://doi.org/10.1111/j.2041-210X.2010.00078.x

Lin, K., Kansal, A., Lymberopoulos, D., & Zhao, F. (2010). Energy-accuracy aware localization for mobile devices. In *Proceedings of 8th International Conference on Mobile Systems, Applications, and Services (MobiSys’ 10)*.

Lynn, N. A., & Brown, R. D. (2003). Effects of recreational use impacts on hiking experiences in natural areas. *Landscape and Urban Planning*, *64*(1), 77–87. doi:https://doi.org/10.1016/S0169-2046(02)00202-5

Marion, J. L. (2016). A Review and Synthesis of Recreation Ecology Research Supporting Carrying Capacity and Visitor Use Management Decisionmaking. *Journal of Forestry*, *114*(3), 339–351. doi:10.5849/jof.15-062

Mason, S., Newsome, D., Moore, S., & Admiraal, R. (2015). Recreational trampling negatively impacts vegetation structure of an Australian biodiversity hotspot. *Biodiversity and Conservation*, *24*(11), 2685–2707. doi:10.1007/s10531-015-0957-x

McDonnell, M. J., Pickett, S. T. A., Groffman, P., Bohlen, P., Pouyat, R. V, Zipperer, W. C., … Medley, K. (2008). Ecosystem Processes Along an Urban-to-Rural Gradient BT - Urban Ecology: An International Perspective on the Interaction Between Humans and Nature. In J. M. Marzluff, E. Shulenberger, W. Endlicher, M. Alberti, G. Bradley, C. Ryan, … C. ZumBrunnen (Eds.) (pp. 299–313). Boston, MA: Springer US. doi:10.1007/978-0-387-73412-5\_18

Metzger, F. T. (1990). Ostrya virginiana (Mill.) K. Koch: eastern hophornbeam. *RM Burns and B. H. Honkala, Technical Coordinators. Silvics of North America. USDA Forest Service, Washington, DC, USA*, 490–496.

Nations, U. (2015). Transforming our world: the 2030 Agenda for Sustainable Development.

Nepal, S. K., & Way, P. (2007). Comparison of vegetation conditions along two backcountry trails in Mount Robson Provincial Park, British Columbia (Canada). *Journal of Environmental Management*, *82*(2), 240–249. doi:https://doi.org/10.1016/j.jenvman.2005.12.016

Newsome, D., Dowling, R., & Leung, Y.-F. (2012). The nature and management of geotourism: A case study of two established iconic geotourism destinations. *Tourism Management Perspectives*, *2*–*3*, 19–27. doi:https://doi.org/10.1016/j.tmp.2011.12.009

Nutsford, D., Pearson, A. L., & Kingham, S. (2013). An ecological study investigating the association between access to urban green space and mental health. *Public Health*, *127*(11), 1005–1011. doi:https://doi.org/10.1016/j.puhe.2013.08.016

Obile, W. (2016). Ericsson mobility report. *Nov*.

R Core Team. (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*, URL https://www.R-project.org/.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O’Neill, B. C., Fujimori, S., … Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. doi:https://doi.org/10.1016/j.gloenvcha.2016.05.009

Rice, W. L., Mateer, T. J., Reigner, N., Newman, P., Lawhon, B., & Taff, B. D. (2020). Changes in recreational behaviors of outdoor enthusiasts during the COVID-19 pandemic: analysis across urban and rural communities. *Journal of Urban Ecology*, *6*(1), juaa020. doi:10.1093/jue/juaa020

Rossi, S. D., Byrne, J. A., Pickering, C. M., & Reser, J. (2015). ‘Seeing red’ in national parks: How visitors’ values affect perceptions and park experiences. *Geoforum*, *66*, 41–52. doi:https://doi.org/10.1016/j.geoforum.2015.09.009

Shochat, E., Lerman, S. B., Anderies, J. M., Warren, P. S., Faeth, S. H., & Nilon, C. H. (2010). Invasion, Competition, and Biodiversity Loss in Urban Ecosystems. *BioScience*, *60*(3), 199–208. doi:10.1525/bio.2010.60.3.6

Sinclair, M., Mayer, M., Woltering, M., & Ghermandi, A. (2020). Using social media to estimate visitor provenance and patterns of recreation in Germany’s national parks. *Journal of Environmental Management*, *263*, 110418. doi:https://doi.org/10.1016/j.jenvman.2020.110418

Tanentzap, A. J., Bazely, D. R., Koh, S., Timciska, M., Haggith, E. G., Carleton, T. J., & Coomes, D. A. (2011). Seeing the forest for the deer: Do reductions in deer-disturbance lead to forest recovery? *Biological Conservation*, *144*(1), 376–382. doi:https://doi.org/10.1016/j.biocon.2010.09.015

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*, *7*(1), 17615. doi:10.1038/s41598-017-18007-4

Tomczyk, A. M. (2011). A GIS assessment and modelling of environmental sensitivity of recreational trails: The case of Gorce National Park, Poland. *Applied Geography*, *31*(1), 339–351. doi:https://doi.org/10.1016/j.apgeog.2010.07.006

UnitedNations. (2015). Mapping the risk-utility landscape: mobile data for sustainable development and humanitarian action. *Global Pulse Project Series No18*.

van Biljon, J., & Kotzé, P. (2007). Modelling the Factors That Influence Mobile Phone Adoption. In *Proceedings of the 2007 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries* (pp. 152–161). New York, NY, USA: Association for Computing Machinery. doi:10.1145/1292491.1292509

Vogels, E. A. (2021). Some digital divides persist between rural, urban and suburban America. *Pew Research Center, August*, *19*.

Volenec, Z. M., Abraham, J. O., Becker, A. D., & Dobson, A. P. (2021). Public parks and the pandemic: How park usage has been affected by COVID-19 policies. *PLOS ONE*, *16*(5), e0251799. Retrieved from https://doi.org/10.1371/journal.pone.0251799

Ward, P., McKenzie, T. L., Cohen, D., Evenson, K. R., Golinelli, D., Hillier, A., … Williamson, S. (2014). Physical activity surveillance in parks using direct observation. *Preventing Chronic Disease*, *11*, 130147. doi:10.5888/pcd11.130147

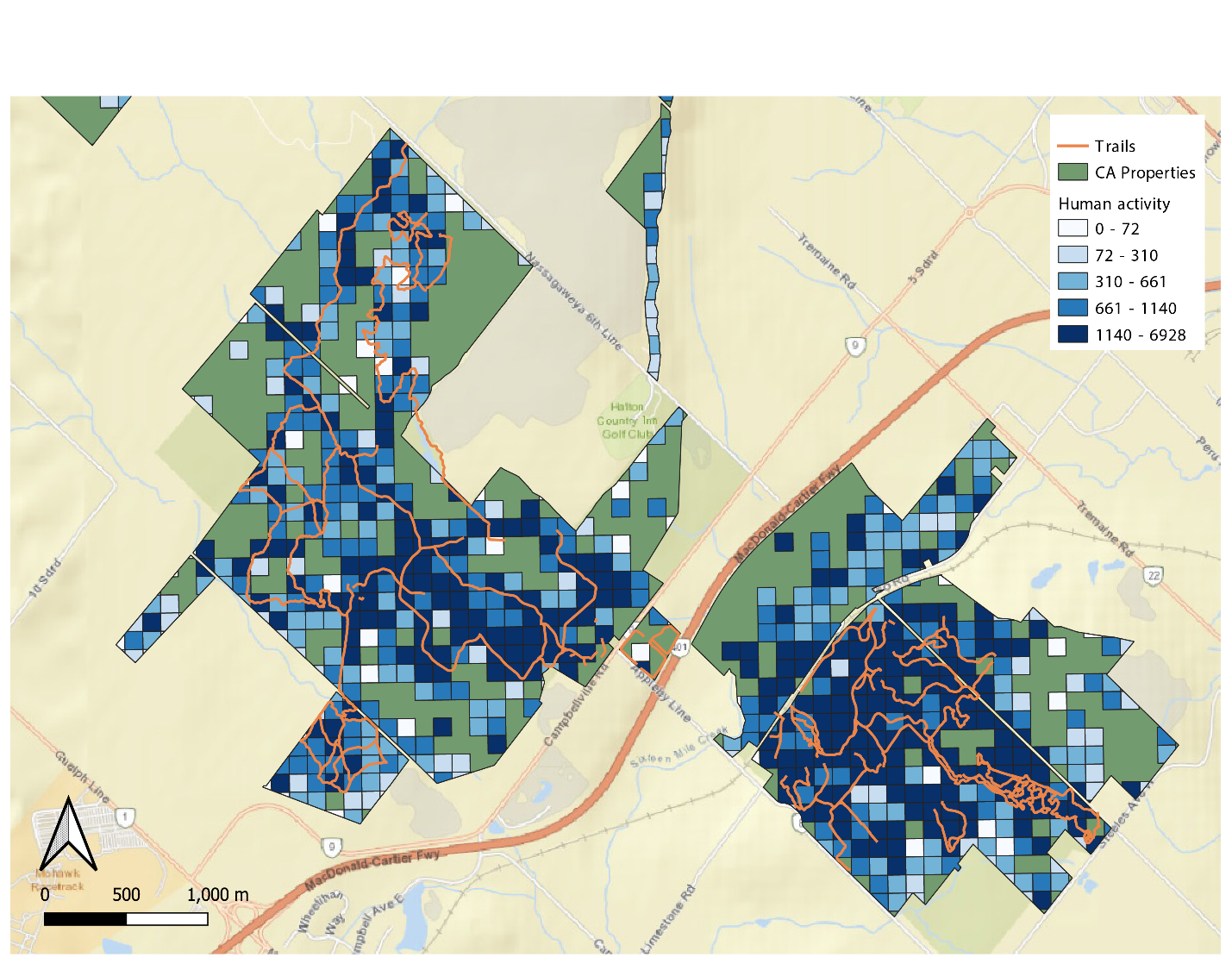
Wilkins, E. J., Wood, S. A., & Smith, J. W. (2021). Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. *Environmental Management*, *67*(1), 120–132. doi:10.1007/s00267-020-01373-7

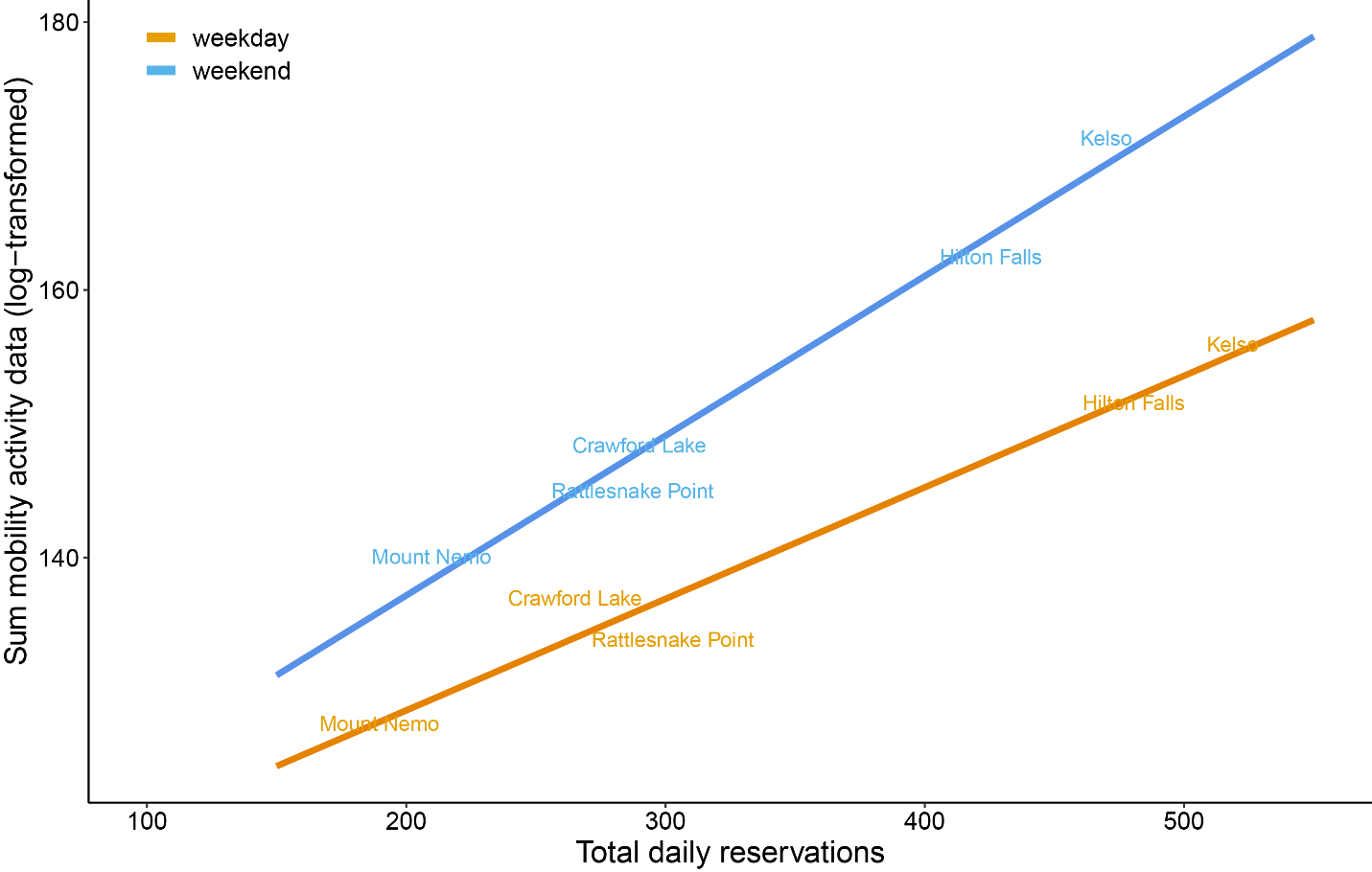
**Table 1:** Considerations for using anonymized mobility data for estimating human activity.

|  |  |  |  |
| --- | --- | --- | --- |
| **Issue** | **Challenge** | **Details** | **Examples** |
| *Social* | Privacy | Data should not be too specific because it can infringe on people’s personal privacy or used for nefarious purposes. | Data is often anonymized and aggregated to a coarse spatial resolution. |
|  | Ownership differences | The number of mobile devices differs among demographics of people. | More affluent communities will likely possess more devices and thus have higher activity patterns. |
|  | Safety | Use of devices may vary depending on location or security. | Out of concerns for safety, individuals may avoid using their device. |
|  | Behaviour of individual | Use of device may vary depending on activity of individual | When exercising someone may not use their device. |
|  | Behaviour on phone | Location tracking services may vary depending on application use on mobile device. | Ride-sharing applications will more frequently report location relative to a messaging application. |
|  | Land type | Certain land types are inherently more likely to have higher activity. | Roads are vectors for transportation and thus will have more activity relative to private land. |
| *Technical* | Accuracy | Not all devices or networks equally estimate the location of a device. | Older cell phones may have a larger radius of location accuracy. |
|  | Coverage | Some areas have limited or no coverage of a cellular network because of geographic barriers or placement of towers. | Alpine areas may have low coverage because of mountains preventing coverage and infrequent towers. |
|  | Noise | To minimize privacy concerns, activity data is often masked below a certain threshold or has a random amount of activity added. | Areas with low human activity will have no values reported. |
|  | Time | Over time, the number or usage of devices can increase. | Devices with location services were less common 10 years ago than today. |
| *Data* | Training | Requires expertise in spatial analysis and big data. | Mobility data is often vectored tiles based spatial extents. |
|  | Cost | Data is not publicly available and requires purchase from companies. | Mobility data is costly even when purchased as snapshots based on certain timeframes or areas. |
|  | Management | The size and type of the data often requires high-performance computing resources. | Regional datasets often exceed the storage and memory capabilities of personal computers. |

**Table 2:** 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 determined from Mapbox anonymized mobility data.

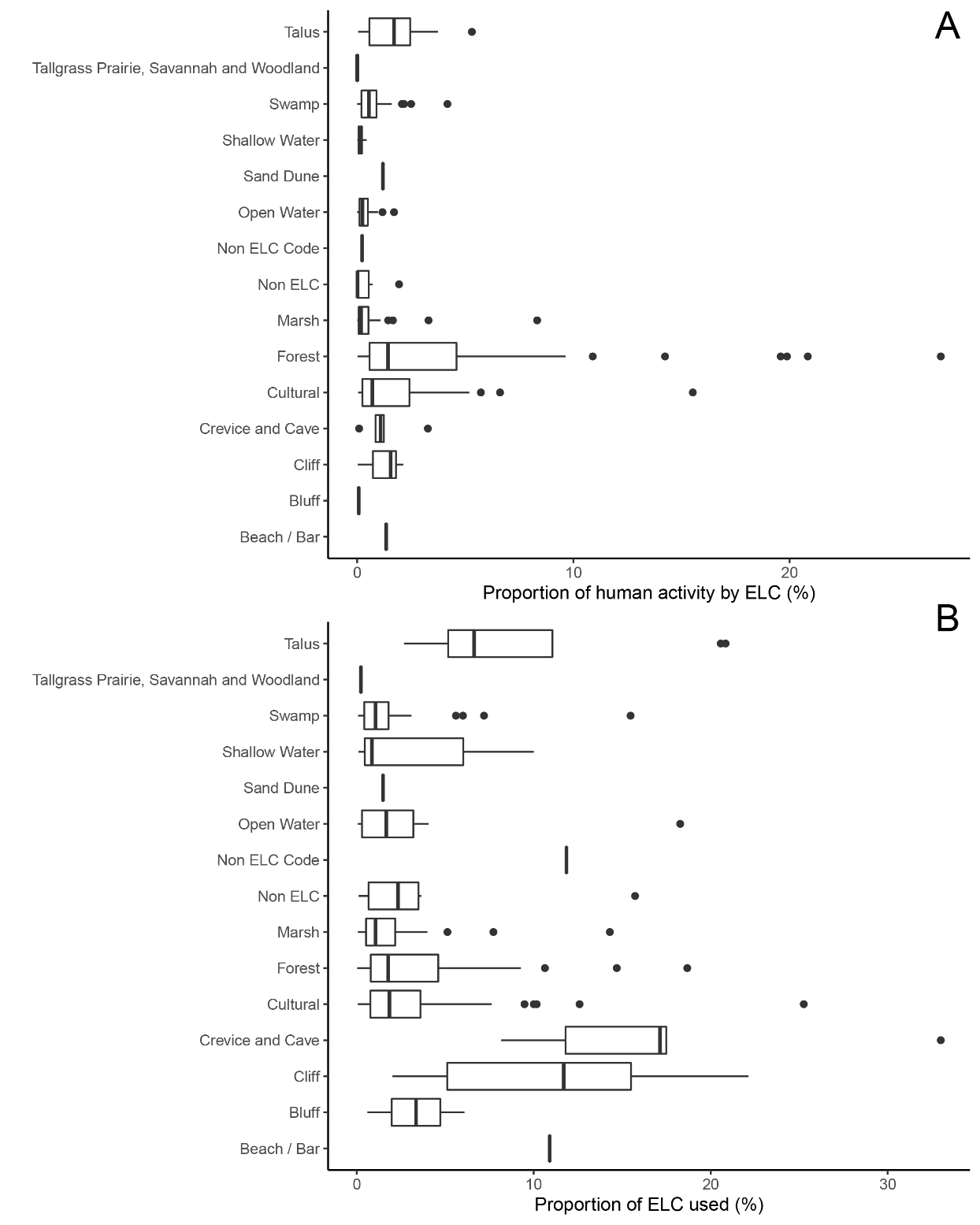
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Land Type** | **Managed** | **Properties** | **Property Size (km2)** | **Activity coverage (%)** |
| Conservation Area | Managed Land | 7 | 3.68 | 59.8 |
| Conservation Area | Non-Managed Land | 8 | 0.88 | 48.4 |
| Natural Area | Non-Managed Land | 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 anonymized 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 green space. 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.

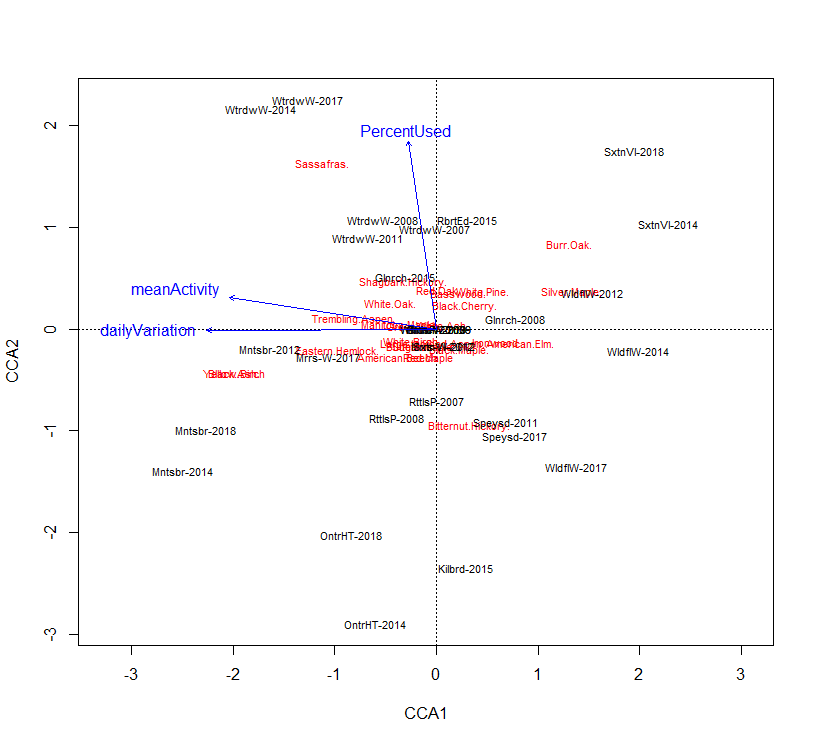


**Figure 2:** The total daily reservations for Conservation Halton properties that had reservation-only entrance were strongly positively associated with mobility data. On average, mobility data was much higher on weekends compared to weekdays. Reservation and mobility data are the total of activity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



**Figure 5:** Patterns of human activity relate to tree composition in green spaces (F15 = 2.74, p = 0.001). The partially constrained correspondence analysis (pCCA) explained 25% of the variation in tree species among properties. Year was included as a conditioning matrix to partial out any interannual variability.

**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. | |