Visitor Use Management Fact Sheet

Mobile Device & Big Data in Visitor Use Management (VUM)

Ashley D'Antonio, Troy E. Hall, Conner Wanless, & Madeline Aberg
Oregon State University, Department of Forest Ecosystems & Society

Big Data in a VUM Context

Definition

- Large datasets, often from secondary data sources
- Used to understand use levels, patterns of use, and, in some cases, visitor experience

Examples of Big Data in VUM

- Social Media
- Crowdsourced and/or Volunteered Geographic Information (VGI)
- Mobile Device (Cellphone or Human Mobility) Data
- Reservation system (i.e., rec.gov)
- Active participation via apps
- Connected vehicles (GPS-enabled)
- Community science apps (e.g., eBird, iNaturalist)

Why Use Big Data in VUM?

- **Volume** – There is large amounts of data available.
- **Velocity** – data is generated quickly
- **Variety** – many components and types available
- **Passive participation** – Data often comes from secondary sources that collect or generate the data automatically

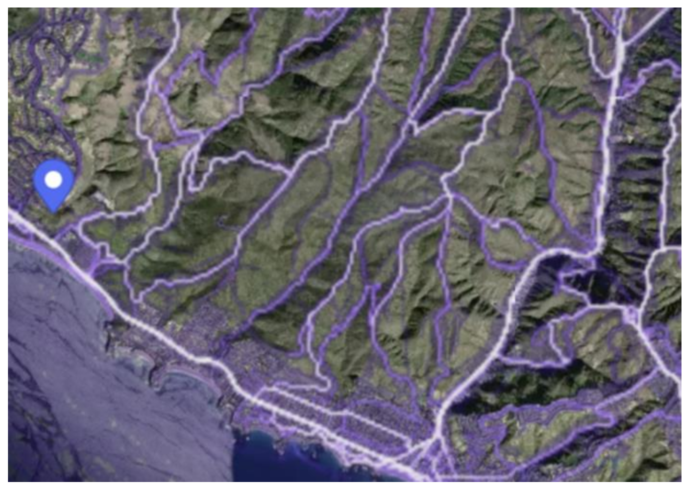

Figure 1. A hotspot map of trail use near Crystal Cove State Park generated by Strava Global Heatmap, one source of big data used in VUM. Photo credit: Ashley D'Antonio, 2023.

Ethical & Privacy Considerations

- Data are passively generated by users who may not realize their data will be used in this way
- Some big data sources (e.g., social media) will need to be deidentified to protect anonymity

Questions to consider:

- What are the **best data** for the question or management need?
- Can you **validate** the data?
- What does the data **represent** and what **biases** are present?
- How will you use the data **ethically** and protect the **privacy** of users?

Social Media Data

- User-generated content that is publicly shared on digital platforms
- Downloaded with an application programming interface (API) that "talks" to the platform
- Examples:
    - Flickr
    - X (formally named Twitter)
    - Instagram

Opportunities

- Reduced time & resources in the field
    - However, there are significant time and resources required to process and analyze the data
- May contain geospatial information
- Contextual information from captions or dialogue
- Engagement between users
- Publicly available (using API) & low cost

Best Use Case Scenarios

- There is a validated correlation between the number of posts on social media and visitor use
    - Relationships are variable and are generally weaker in less visited or less aesthetic parks
- Broad temporal scale predictions are needed
    - i.e., Use estimations for an entire week
- Text and/or photo content is needed
- Less precise geospatial information is needed

Limitations

- User bias
    - Social media users are not representative of park users' demographics, age, etc.
    - The number of posts may vary
- Privacy regulations and data availability may change
- Choosing a relevant platform for your user group of interest
- Best if validated or calibrated with on-the-ground counts, such as automatic trail counters.

Popularity of Australian Beaches

Mota & Pickering (2022).

Ocean & Coastal Management

Researchers used Flickr photos, which contain spatial information, to estimate relative levels of use across beaches. They also examined the captions included with posts and used the emotion of the post as an indicator of visitor experience.

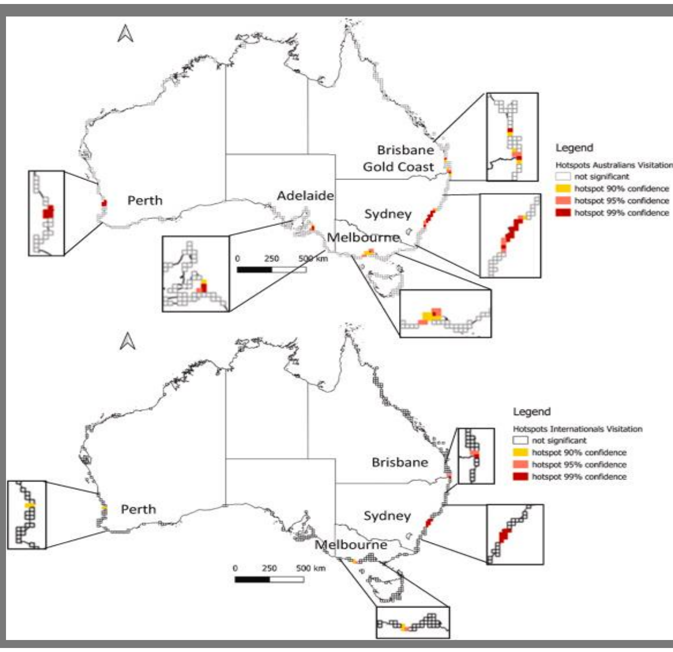

Figure 2. Relative popularity of Australian beaches estimated from Flickr data. Hotspots are shown with yellow (90% confidence), orange (95% confidence), and red (99% confidence). From Mota & Pickering 2022.

Volunteered Geographic Information & Crowdsourced Data

- User-generated content with geospatial information
- Used to understand behaviors (e.g., visitor movement through a site, the amount of time visitors spend at an attraction)
- Examples:
  - Strava
  - AllTrails
  - Trailforks
  - Most fitness-tracking apps

Opportunities

- Reduces time and resources in the field
- Publicly available and low cost
- Precise at broad spatial and temporal scales
- Provides metric of relative use levels and behavior

Best Use Case Scenarios

- Destinations with a fitness or activity focused visitors (i.e., mountain biking optimized trails)
- Works well for comparing relative use levels or patterns at broad temporal and spatial scales
- Can be used to monitor use off of designated trails or where visitor use is more dispersed

Limitations

- User bias
  - For example, Strava tends to underestimate use by women, older visitors, and visitors under 18 years old
- Best if validated or calibrated with on-the-ground counts, such as automatic trail counters.
- Less precise at smaller spatial and temporal scales
- Apps used may change in popularity over time
- Requires an understanding of who uses your park

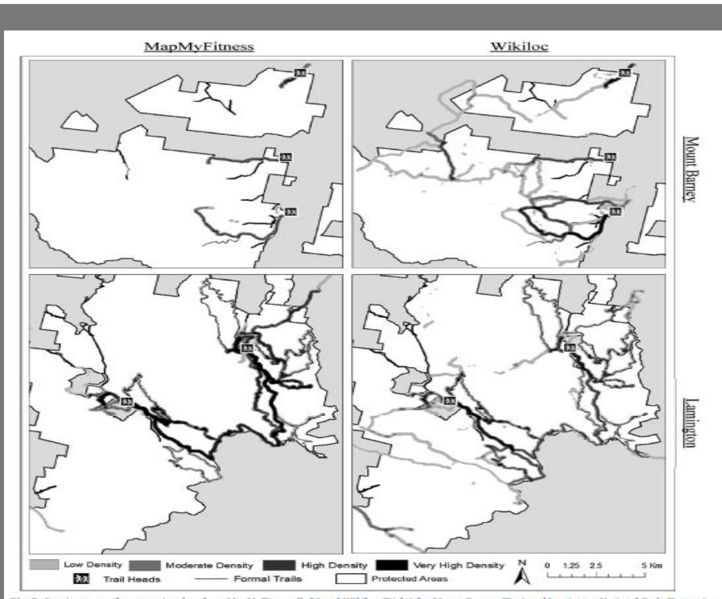

Figure 3. Estimations of use patterns from two VGI apps (columns). Trail use is shown through two sites (rows) with darker colors indicating higher relative use. From Norman & Pickering 2017.

VGI to Estimate Park Visitation

Norman & Pickering (2017). *Applied Geography*

Researchers used multiple sources of volunteered geographic information (VGI) to assess park visitation. They found that data differed depending on which VGI source they used, which highlights the importance of using the most relevant apps for your purpose.

Mobile Device Data

- Geospatial data associated with the movement of mobile device
- Where is the data coming from?
  - A device's internal GPS
  - Location-based Services (LBS)
- Often acquired through a vendor
  - Many different vendors
  - Rapidly changing landscape
  - Available through a fee or contract

Opportunities

- Relatively high geospatial precision
- Variety of temporal and spatial scales available
- Reduces time & resources in the field
- Can include additional information
  - Demographics (some limitations)
  - Estimate of home location

Mobile Device Data in Orange County, California

Creany et al. (2021). *Environmental Challenges*

Researchers compared estimates of mean daily visitor use produced with mobile data from Streetlight to estimates from TRAFx (automatic trail counters). They found that the median estimate matched closely. They also compared on-trail use and found that Streetlight data overestimated use in the border of the park, near urban areas.

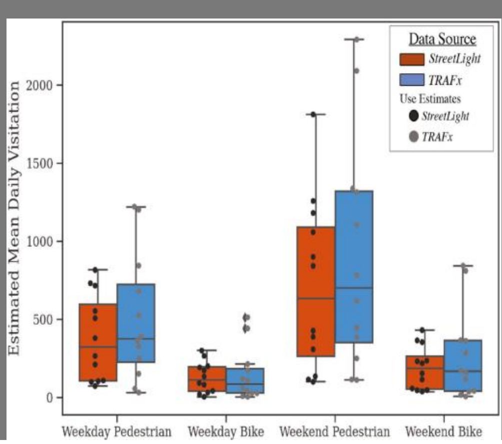

Figure 4. Comparison of Streetlight (red) and TrafX (blue) estimates of mean daily visitation. From Creany et al. 2021.

Best Use Case Scenarios

- Estimating visitation at broad scales (i.e., the entire park)
- Parks with porous boundaries where it is difficult to accurately estimate visitation
- Broad spatial patterns and network-level VUM questions
- Parks with decent cellphone coverage
- In cases where data can be calibrated or validated

Limitations

- Not suitable for all VUM questions
- Can be expensive to purchase data
- Transparency: Data are typically processed by vendors who may not provide their data processing steps and algorithms
- Demographic data is often based on Census blocks
  - Park visitors do not match census blocks
  - Disparities in smartphone use and ownership
- Staff and volunteers may be overrepresented in use estimates
- Spatial accuracy with mobile devices varies by device, locations, and environment
- Behavioral aspect
  - For some activities (e.g., going to the beach) visitors may be more likely to leave their phone behind
  - Local and non-local users may use phone or GPS-related apps differentially

Challenging Settings

- Parks with close proximity to urban settings
- Parks with limited signal
- Small spatial scales
- Where calibration or validation data are not available

RESOURCES

Baird, T., Stinger, P., Cole, E., & Collins, R. (2022). Mobile Device Data for Parks and Public Lands Transportation Planning: A Framework for Evaluation and Applications. Transportation Research Record, 2676(8), 490–500. https://doi.org/10.1177/03611981221083911

Creany, N., Monz, C., D'Antonio, A., Sisneros-Kidd, A., Wilkins, E., Nesbitt, J., & Mitrovich, M. (2021). Estimating trail use and visitor spatial distribution using mobile device data: An example from the nature reserve of orange county, California USA. Environmental Challenges, 4, 100171. https://doi.org/10.1016/j.envc.2021.100171

Dagan, D. T., & Wilkins, E. J. (2023). What is "big data" and how should we use it? The role of large datasets, secondary data, and associated analysis techniques in outdoor recreation research. Journal of Outdoor Recreation and Tourism, 100668. https://doi.org/10.1016/j.jort.2023.100668

Liang, Y., Yin, J., Pan, B., Lin, M. S., Miller, L., Taff, B. D., & Chi, G. (2022). Assessing the validity of mobile device data for estimating visitor demographics and visitation patterns in Yellowstone National Park. Journal of Environmental Management, 317, 115410. https://doi.org/10.1016/j.jenvman.2022.115410

McKitrick, M. K., Schuurman, N., & Crooks, V. A. (2022). Collecting, analyzing, and visualizing location-based social media data: Review of methods in GIS-social media analysis. GeoJournal, 88(1), 1035–1057. https://doi.org/10.1007/s10708-022-10584-w

Rice, W., Mueller, J. T., Graefe, A., & DerrickTaff. (2019). Detailing an Approach for Cost-Effective Visitor-Use Monitoring Using Crowdsourced Activity Data. Journal of Park and Recreation Administration, 37(2), 144–155. https://doi.org/10.18666/JPRA-2019-8998

Schirck-Matthews, A., Hochmair, H., & Paulus, G. (2023). Comparison of reported outdoor activities in Florida State Parks among three fitness tracker apps. Journal of Leisure Research, 54(1), 46–71. https://doi.org/10.1080/00222216.2022.2153097

Whitney, P., Rice, W. L., Sage, J., Thomsen, J. M., Wheeler, I., Freimund, W., & Bigart, E. (2023). Developments in big data for park management: A review of mobile phone location data for visitor use management. Landscape Research, 1–19. https://doi.org/10.1080/01426397.2023.2198762

Venter, Z. S., Gundersen, V., Scott, S. L., & Barton, D. N. (2023). Bias and precision of crowdsourced recreational activity data from Strava. Landscape and Urban Planning, 232, 104686. https://doi.org/10.1016/j.landurbplan.2023.104686

RESOURCES FOR DATA ANALYSIS

- R Statistical Software R: The R Project for Statistical Computing (r-project.org)
- Georeferencing Strava Global Heatmap methods-estimating-trail-use.pdf (headwaterseconomics.org)

SUGGESTED FACT SHEET CITATION

D'Antonio, A., Hall, T. E., Wanless, C., & Aberg, M. (2023). Mobile Device & Big Data in Visitor Use Management (VUM). Fact Sheet prepared for the Visitor Use Management Toolkit. Corvallis, OR: Oregon State University, Department of Forest Ecosystems & Society.