Project No. 08-132

ACCESSING AMERICA'S GREAT OUTDOORS: FORECASTING RECREATIONAL TRAVEL DEMAND

EXECUTIVE SUMMARY

Prepared for
NCHRP
Transportation Research Board

of

The National Academies of Sciences, Engineering, and Medicine

RSG
White River Junction, Vermont
June 2023

Executive Summary

Purpose of the Project

Recreational travel is a major and growing activity in the United States. Recreation trips to outdoor parklands tend to have different characteristics than trips for other types of recreation (entertainment, sporting events, regular exercise, etc.). Despite the significance of recreational travel to outdoor parklands and its impacts on the transportation system in many regions, regional and statewide travel demand models do not explicitly represent the factors that determine demand for this specific type of travel.

The purpose of this project is to build a set of models and open-source tools to predict recreational travel demand to "America's Great Outdoors." The demand models and tools created in this project will allow states, regional agencies, and public land managers to predict visitor demand and vehicle travel to their recreational sites under a variety of future scenarios. The models use passively collected data ("big data") in conjunction with data on recreational-site visitation and amenities, Census data, and data on climate and topography to replicate visitation patterns to various federal and state parklands in the US. These models are appropriate for major recreational sites located in natural settings across the country which attract significant local and out-of-town visitation. Local parks in urban areas (where recreational visits are already largely accounted for in regional models) are excluded from this study.

The large majority of currently collected recreational travel data and modeling efforts are both site- and time-specific, meaning that they were not intended to be generalized across various locations nor to be used in forecasting future demand. The models developed in this project were created to be generalized to any major parkland site. This is a unique and challenging study, for a variety of reasons:

- It is the first attempt at creating predictive models of parkland visitation that can be used by various state and regional agencies in the US.
- The modeling uses big data as a primary data source, supplementing it with various types of data, such as National Park Service data on visitation and amenities; Census data on socioeconomics, land-use data, road and air network data, and climate and topology data adds explanatory power to the models.
- The models are implemented into software tools and supporting data inputs that are available to be used by a wide variety of agencies—including state DOTs, regional MPOs, and parkland management agencies.

Key Research Findings

The research team began this work by reviewing existing literature, including academic publications, reports from federal and state public lands, and federal documentation. This review revealed several limitations specific to current models of recreational travel demand and demonstrated that there is a strong need to improve visitor demand forecasting and planning.

Current State of the Practice

This literature review highlighted four salient takeaways. First, there are many sources of visitor data; these sources have varying levels of detail and consistency across recreational sites. Second, data collection

and modeling efforts are most often driven by local recreational site decision-makers who wish to understand specific issues. Third, these data collection/modeling efforts are often neither transferrable between recreational sites nor are they designed to be applied over long periods of time. Finally, there are many recreational site characteristics that can influence demand that need to be thoughtfully considered in developing transferrable and generalizable forecasting models. The takeaway is that this project is both timely and necessary. Many opportunities exist for improving existing recreational travel demand modeling efforts, especially by designing them so that they can be included into existing statewide and regional travel demand models.

The research team also conducted in-depth interviews with project stakeholders, including members of the NCHRP project panel. These interviews confirmed interest in, and the need for, developing improved methods for estimating outdoor recreation travel demand. They also highlighted the wide variety of approaches and data that are currently used to represent this type of travel in statewide and metropolitan travel modeling practice, as well as the fact that many such models do not include any direct representation of this type of travel.

Data Available for Developing Recreational Travel Models

There is a significant amount of data available on recreational sites, particularly national parks. These data sources include much, but not all, of the information needed to develop models of outdoor recreational travel. Much of the currently available data is collected infrequently and outdated. Further, some of the data does not cover all national parks (nor other recreational sites). Some of the best data describing demand has been collected under the National Park Service's Social Science Program under a pilot implementation of an in-park Socioeconomic Monitoring (SEM) visitor survey. To date this program has covered only 14 NPS sites and data for recreational sites that are not national parks is not nearly as robust. There are many other data sources that could provide relevant information on travel to these and other sites, including other data collected by the National Park Service, U.S. Forest Service, U.S. Bureau of Land Management, and some state park departments. However, none of these data sources provides comprehensive, consistent data across all U.S. outdoor recreation sites sufficient to support development of recreational travel demand forecasting models. Upon investigation, only two states appeared to have data on state park visitation and amenities approaching that available for the National Park Service. Those two states are California and Pennsylvania. Selected state park sites in those two states were also used in the project, along with relevant data available for those sites. We found that it was feasible to include those state park sites for some of the spatial models estimated, but that the park visitation and amenities data was not sufficiently detailed to include the state parks in the model of park site visitation levels. This does not mean that the models developed in this study could not be applied to state park sites or other types of national parkland sites. If the model user can provide the data needed by the models for the local park site(s) of interest—either from published sources or by obtaining estimates from park management—then the models can be applied for the site(s). For this national-level study dealing with dozens of sites, however, it was only feasible to use readily available published data.

Big data, which includes location-based services (LBS) data collected from mobile devices, is potentially quite helpful in filling existing data gaps. The key benefits of this type of data are that they cover travel to and from all of the U.S. and thus all outdoor recreation sites and are collected continuously. However, there are also significant challenges to using and interpreting these data. These challenges and limitations were addressed directly in this research and are the subject of one set of recommendations for further work.

The "Great Outdoors" Recreational Travel Demand Model

The most challenging task of the project was to move beyond the existing literature and data sources and to design and implement a new modeling framework to meet the ambitious objectives of the project. Adding to the challenge is the goal of going beyond modeling the demand to a single recreational site, or even to a single type of recreational site, and creating a modeling framework that can be applied to a wide variety of recreational sites across the U.S., each site differing in the quantity and quality of available data that is specific to that recreation destination. Given these challenges, it seemed preferable to design a framework that can successfully satisfy the most important project objectives for a reasonable variety of types of recreational sites, rather than trying to satisfy all possible objectives for all types of recreational sites and, as a result, not being very successful in meeting most of those objectives. This is the first time that such an ambitious scope has been attempted in practice, so the project team was conservative in limiting the scope to provide a successful and useful framework that can be built upon in the future.

The Overall Model Framework and Approach

To meet the objectives of not only predicting parkland visits, but also predicting, to the extent possible, the spatial and temporal patterns of auto traffic generated by those visits, the model system requires several interrelated components. The concepts behind these components are similar to the concepts used in trip-based or tour-based models for generation, attraction, and distribution, but here the unit of demand is a "visit" rather than a single trip or tour.

Figure *1* provides a schematic of the model system components and structure. The component models are:

1. The **park visit generation model**: This model predicts the number of people visiting a recreation park site during a specific three-month period of the year: January-March, April-June, July-September, and October-December. (These periods are often referred to as the seasons Winter, Spring, Summer, and Autumn in this report, even though they run until the end of March, June, September, and December rather than changing at the solstice or equinox.) A "visit" is defined as entering the park one or more times between the time a person leaves their home location and the time that they return to their home location. Thus, a "visit" can involve entering and leaving the park boundaries more than once, a fact that must be accounted for in the model system.

2. The **visitor home location model**: This model determines the home location zone of each park visitor, thus providing an indication of the trip origin and destination for the visitors' first trip arriving at the park area and last trip leaving the park area during their visit. (In reality, some visitors make multi-destination tours so do not travel directly between home and the park location, but as mentioned above, trying to predict multi-destination tour patterns at the national level would be too complex for any practical model.)

3. The **access mode and airport model**: A fraction of parkland visitors do not drive all the way between home and the destination, but travel by air to an airport nearer the destination and then (typically) rent a car or recreational vehicle (RV) to drive the rest of the way to the destination. Depending on a visitor's home location and the recreation site location, this model predicts whether the person travels part of the way by air or not, and, if so, which airport they land at before traveling to the destination site. This model thus further determines the origin and destination for visitors' road trips to and from the destination area, substituting the airport location for the home location for the visitors predicted to go partway by air.

4. The **auxiliary models**: These models provide further details of the visit and local trips made during the visit for each visitor including:

a. Whether or not the person lives within the park "halo area"—defined as an area within 50 miles of the park boundaries. The remaining models below are estimated and applied separately for those who are residents of the halo area versus those who are not, as halo area residents are much more likely to make day trips rather than stay overnight, and very rarely lodge at non-home accommodations in the halo area outside the park.
b. The number of nights staying in the park
c. For those who live outside the halo area, the number of nights staying outside the park but inside the halo area.
d. The day of week and time of day first arriving and last leaving the park, and (for those who live outside the halo area) the day of week and time of day first arriving and least leaving the halo area.
e. The vehicle-miles traveled (VMT) driven within the park and within the halo area during the entire visit.

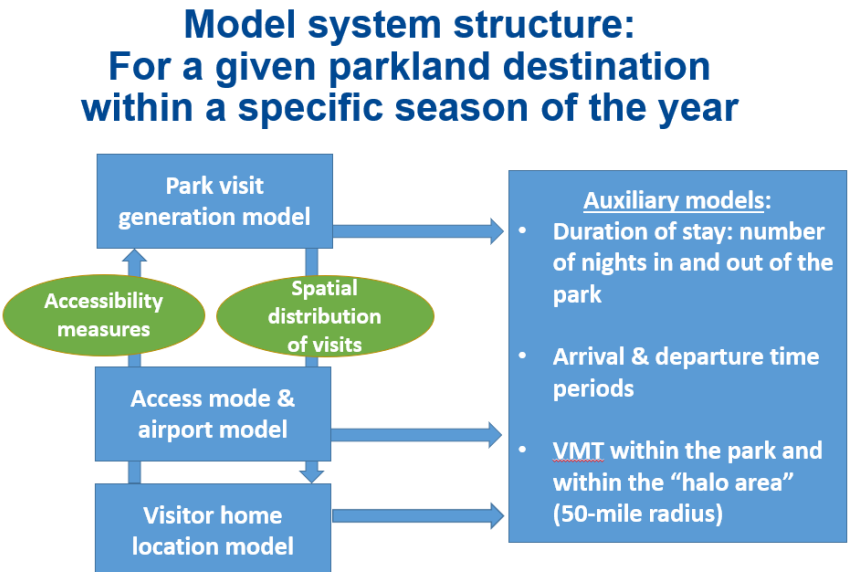

Figure 1. Model system structure

Data Used to Develop the Models

A number of data sources were used to develop these models.

- **Socio-demographic data** were compiled from the 2015-2019 American Communities Survey for residents of the approximately 73,00 U.S. Census tracts.
- **Land use data** was compiled primarily from the Environmental Protection Agency's (EPA) Smart Location Database (SLD)
- **Climate and elevation data** were used to develop three indices that were then used as indicators of attractiveness of recreation sites in different seasons, and the relative attractiveness of recreation sites compared to home locations. These included a Winter sport index, a Coastal recreation index: and an Extreme heat index.
- **Travel times and costs** were taken from the Federal Highway Administration's National Long-Distance Passenger Travel Demand Model, Recreation Destination

- **Recreational site characteristics** were compiled for 89 federal sites and 76 state sites. These included information about the site derived from GIS analysis as well as detailed data on the amenities available at each site.
- **Location-based services (LBS) data** from location-aware apps on smartphones describing travel to and from all of the outdoor recreation sites. The dataset used for this study includes data from nearly 10% of all smartphones in the US. For the entire year of 2019, each smartphone device entering any of the 165 national or state parklands used for the study was traced both backwards and forwards in time to create a full "tour" from the traveler's home location and back, compiling information about the trips to and from the destination recreational site, as well as local trips made in the halo area. In total, the passive data processing yielded well over 3 million records of individual park visits. There were a number of issues to be dealt with in processing the passive data, however, that resulted in reducing the number of visits used in the modeling to approximately 1 million records. This screening process is typical with LBS data, as there are a number of ways in which the data may prove to be inadequate for modeling.

Statistical Estimation of Model Parameters

The parameters of the outdoor recreation travel models were estimated using the types of advanced statistical methods that are commonly used in developing other types of travel demand models. The parameters of each model were estimated several times, each time addressing issues discovered in the data or improving the model specification to give a better model fit or to make the definition of the variables as consistent as possible across the models.

The Visitor Home Location Model relied on LBS data that described more than 1 million visits to one of the 165 national and state parks was used as an observation, and the model predicts the residence zone from the system of roughly 4,500 "NUMA" zones covering all 50 states plus D. C .The model is of multinomial logit form (MNL). This model component takes advantage of the strongest feature of "big data"—that it can provide the most comprehensive data on spatial origin-destination patterns.

The Long-Distance Mode and Airport Choice Model is based on almost 1.2 million observations, approximately 10% of which used air. The choice of airport is a non-liner function of the road distance between the park and airport (based on the 4,500-zone system), and the number of passenger trips originating from the airport in 2019, derived from the BTS DB1B Air Market Data, which is publicly available.

The Park Visit Generation Model uses National Park and Monument visitation data. The form of the model is a log-log regression model, with the logarithm of the number of visitors to a given park/monument during a specific season as the dependent variable, and any continuous independent variables also using a logarithmic form.

Auxiliary Models of Duration of Stay, Day and Time of Arrival and Departure, and Vehicle Miles Traveled were developed to estimate further details of the travel to, from and in the park and the surrounding area—additional model outcomes that agencies can use to augment information from their existing travel models. These models are relatively simple in form, based on the LBS passive data set and validated against the NPS survey data to the extent possible.

Translating The Model Results to Traffic on the Local Road Network

One of the most challenging aspects of this project was to implement the models in a software platform that agencies such as DOTs, MPOs and other potential users can set up and customize for their use. The goal is that users can use this model as a "special generator" to add a new sector of demand to their existing

model system. In that regard, it is important that users be able to adapt the outputs to be used with their existing model structure, zone system and zonal data,

Open-source model application software was developed to produce three main types of outputs:

1) **Non-resident trips to and from the parkland halo area** - These trips are (generally) long-distance road trips from the visitors' residence zones or an intervening arrival airport to the parkland area, and then the return trip upon departing. For a given parkland destination and season, the models predict (a) the number of person-trips and vehicle-trips arriving from each zone and airport during each combination of day of week and period of the day, and (b) the number of person-trips and vehicles trips departing toward each zone and airport during each combination of day of week and period of the day.
2) **Non-resident travel within the parkland and halo area -** The auxiliary models provide estimates of person-nights spent, person-miles traveled, and vehicle-miles traveled, both within the within the parkland and within the halo area around the park. These can also be used to provide an estimate of vehicle miles per day for each day of the week.
3) **Halo resident travel within the parkland area and to and from the parkland area -** For halo residents, the model system provides estimates of person-nights spent, person-miles traveled, and vehicle-miles traveled, within the within the parkland.

Example Model Applications and Validation Against Survey Data

A key source of data for checking model results (and adjusting the models where needed) was visitor surveys. Several national parks administered visitor surveys not long before 2019, and there are survey data for five sites that were also used in model estimation and test model application. These are Acadia (ME), Biscayne (FL), Glacier (MT), Glacier Bay (AK) and Grand Canyon (AZ) national parks—a selection that is small in number but quite varied in terms of the size and accessibility. The data from these five visitor surveys mentioned above was coded and compared to similar analyses done on the passive LBS data used for model estimation, as well as the results from applying the models to those same five national parks.

The comparisons between the model estimates and survey data for these five parks illustrate strengths and limitations of both the models and the survey data. In general, the survey data show lower visitation estimates from areas proximate to the parks than do the models, but this may be due to the survey samples' over-representation of longer park stays. Model estimates of total visitation levels compared to actual visitation show considerable heterogeneity across park sites with the models performing reasonably well for some sites but with notable outliers for other sites. In general, the example applications indicate both the potential usefulness of these models in planning applications as well as the need for further research.

Tools for Model Implementation

Open-source code was developed to allow the models developed as part of this project to be adapted into existing state-level, metropolitan or regional travel demand model as a "special generator" component for outdoor recreation sites within the areas covered by those models. A technical guidebook for travel demand modeling professionals was also created to describe the technical details of the models and their implementation.

Recommendations from the Research

The recreational travel demand models that were developed as part of this research provide a new set of capabilities and tools that have not previously been incorporated into travel demand forecasting practice.

Because these models are entirely new, the research identified a number of improvements that could be made in the data and model structure in any future research efforts. The modeling effort was based on 2019 data and as, with all travel demand models, updates would be needed in the future. However, the future uses of these models and tools in practice would benefit from several types of updates related to the specific data that were used in their development. The project team recommends the following types of updates be considered to support future implementations of the models and tools.

- **Update the model to augment or replace data from mobile device-based passive data with other data sources.** This project relied heavily on mobile-device based passive data from 2019 but the data sources used for those data are likely to continue to be diminished in quality and coverage while vehicle-based passive data will likely increase in coverage. In addition, it is possible that expansion of the NPS survey program and surveys at other outdoor recreation sites could at least in part replace the need for passive data. This would not be a simple swap of like data types as the vehicle-based datasets and NPS survey data have a different set of strengths and limitations that must be accounted for. But any future updates of model coefficients beyond the set estimated in this research using 2019 mobile device data will likely require transition to these new data sources.

- **Deal with mutual causality between visitor demand and the supply of park amenities.** Using cross-sectional data from just one or two years, it is not possible determine the direction of causality between the level of visitation and the provision of park amenities such as parking, lodging, camping, concessions, and trails. Future work should consider expanding the cross-sectional Park Visit Generation Model developed using only the 2019 NPS visitation data into a multi-year dynamic model. Such a model could also include trend effects and effects of external variables such as inflation rates, fuel prices, unemployment rates, demographic shifts, and even transient effects such as the COVID-19 pandemic. Such model enhancements may be of most interest to parkland management agencies, as they would allow the model to include a wider variety of actionable variables that could be used to influence visitation.

- **Update Census boundaries, Census data, and other data sources.** Since the model is based on 2019 demand data, the tract boundaries are the 2010-19 Census boundaries. The Census Bureau changed the block and tract boundaries in 2020, and gradually agencies will base the zones for their regional or statewide models on 2020 Census geography as well. PUMA boundaries will change in 2023 and many county boundaries have changed as well, so the National Long-Distance Model "NUMA" zone system used for this project should also be updated at some point. When the various types of data inputs that rely on Census boundaries have all been updated, create a set of inputs for the model to update to a new base year. For example, this work could be done in 2025 or 2026 to update the base year to 2024. This change could be made at the same time as one or more of the other updates or enhancements recommended above.

- **Represent the multi-destination aspect of outdoor recreation travel.** The model outputs generated as part of this study use the simplifying assumption that trips to a parkland destination area travel directly from the direction of the home location or an intervening airport, and that trips leaving the area return in that same direction. In reality, many visitors come from or go to other non-home destinations on a multi-destination tour. As we have learned in practice, developing tour-based models that predict the entire itinerary for multi-destination tours is much more complex than developing simpler models that do not attempt to predict multi-step itineraries. Such models require more complex estimation methods and software implementation, and also more complete, high-quality data (such as vehicle-based passive data). It would be more feasible to attempt modeling multi-destination park visit tours at the level of a single state, rather than attempting to tackle the entire US in the first attempt at such a model. When this type of model is completed successfully

for one state, it would be more feasible to replicate it for other states and/or expand it to cover multiple states.

- **Customize the model specification and calibration more closely to local conditions.** One of the several challenging aspects of this project is how to translate the results of model that uses a coarse national-level zone system and road network into results that can be assigned to the local zone system and network for any MPO or statewide model. A first step would be to identify one or two agencies with adequate high-quality multi-day travel data and/or connected vehicle passive data to estimate models of local trips that parkland visitors make while staying in the park and the surrounding area. As part of the project, they could apply the model system from this NCHRP project to that same parkland and use the outputs from this model as inputs to the estimation of models of the local trip pattern of park visitors. This would be the most rigorous way of translating the results of the auxiliary models to use in local models and may even make it possible to replace some or all of the auxiliary models with new models based on more complete and detailed data.

- **Apply the models for other types of outdoor recreational sites.** Most of the models in the model system were estimated only on data for visitors to national parks and monuments, while some included data for visitors to selected state parks in two states. We hypothesize that the models are transferable to similar types of parks, such as national recreation areas and national seashores, and to major state parks in other states. There are also other types of outdoor recreation sites for which these models are not likely to be transferable. One such type is scenic byways. Travel data is usually analyzed in terms of trips. Both travel survey data and passive data typically require a stop of at least a minimum threshold duration to be mode to register a trip being made to a particular destination. Scenic byways have many through-travelers who do not stop, so modeling use of these facilities would require a much different approach to data process and analysis. Another type of site for which the models may not be transferable are large national forests which have little in the way of amenities except for limited campsites. If there is interest in modeling demand and related traffic effects for these other types of outdoor recreation sites, that should involve a different data collection and model design strategy than the ones used for this project.

It is normally expected that any set of travel demand models used in practice should be updated regularly to reflect current behavior, to incorporate new and improved methods and to use new sources of data. This certainly applies to these outdoor recreation travel demand models and, since these are entirely new types of models, the above recommendations reflect the opportunities for further evolution of these particular models as they are incorporated into practice.