Project No. 08-132

ACCESSING AMERICA'S GREAT OUTDOORS: FORECASTING RECREATIONAL TRAVEL DEMAND

TECHNICAL MEMORANDUM

MODEL IMPLEMENTATION HANDBOOK

Prepared for
NCHRP
Transportation Research Board

of

The National Academies of Sciences, Engineering, and Medicine

RSG
White River Junction, Vermont

May 2023

CONTENTS

CHAPTER 1

Introduction

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 state and 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 "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, on the other hand, 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 passive "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 socio-economics, land use data, road and air network data, and climate and topology data, each of which 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.

Overview of the Remaining Chapters

This document provides a guide to the software implementation of the models documented in the project Final Report. Chapter 2 of this document provides a brief overview of the model system and how the implementation software works.

Chapter 3 provides a summary of all the model parameters used in the software implementation. The model results are described in more detail in the Final Report, but all model coefficients are repeated here for reference purposes.

Chapter 4 documents the input data needed for the model system, distinguishing between those that are provided and are not park-specific and those that are park-specific and may need to be provided by the user to run the model for a park that is not already included in the park database.

Chapter 5 documents the various output files produced by the model implementation and how they can be used in practice.

CHAPTER 2

Overview of the Model System and Software Implementation

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. For example, a person who lives in New Jersey, flies to Las Vegas, spends some time there, rents a car or RV, drives to the Grand Canyon, drives to Zion, drives to Arches, drives to Lake Mead, then drives back to Las Vegas and flies back to New Jersey, can be treated as making four separate visits to different destination recreational sites – Grand Canyon, Zion, Arches, and Lake Mead/Hoover Dam. For our modeling purposes, we treat these as four separate visits from New Jersey, without trying to model the combination of different destination recreational sites visited (which would require a much more complex type of simulation model that is not computationally tractable for a national-scale model). Rather than focusing on the itinerary across destinations, our model can better focus on the attributes of the visits to each destination—not just where the visitors come from, but also things that mainly determine the patterns of visitation to the destination recreational site, including seasonality of visits, duration of stay in the destination recreational site, and the local trip-making and VMT generated by visitors (by day of week/period of the day, and inside vs. outside the boundaries of designated recreational sites.) As part of the "impedance" measurement, and to provide more accurate origin-destination information for the long-distance approach and departure trips, we also model the distribution across "gateway airports" for visiting the area. For Zion, for example, gateways could include major airports within a few hundred miles of the area for people who use air (Salt Lake City, Phoenix, Las Vegas).

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.

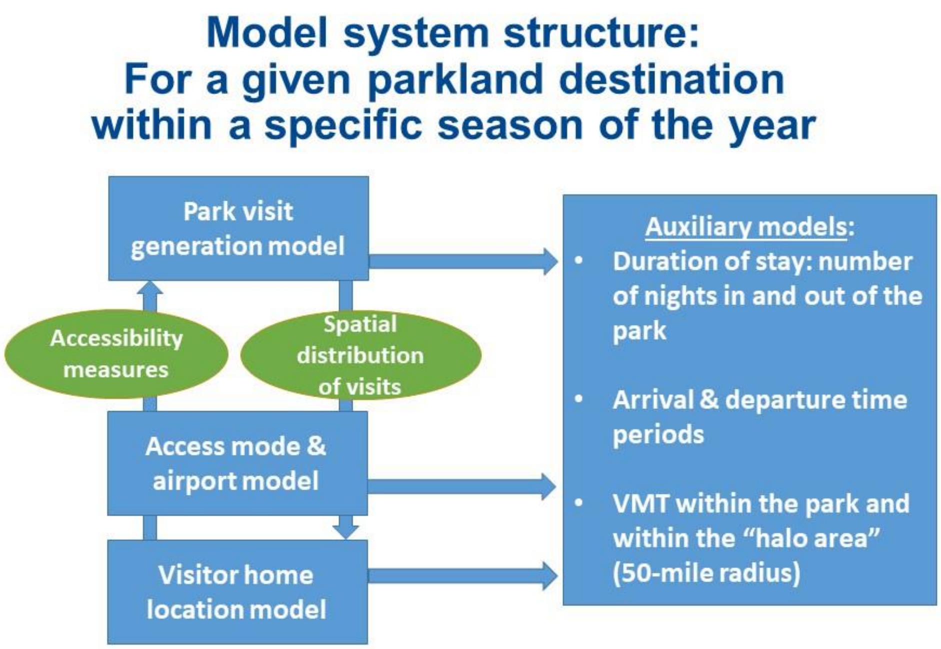

Figure 1. Model system structure

Note that the visitor home location model and the access mode and airport model interact with the park visit generation model in two directions. These models predict where the homes (and airports in some cases) are that the trips to the park location originate from and return to. At the same time, the expected utility from these two models (the "logsum" across all alternatives) provide very important accessibility measures that help explain the generation of park visits. The more accessible the recreation destination to areas that are more likely to generate visitor trips (e.g., due to socio-demographics and relative climate patterns), and the more accessible to well-served airports, the more visits the destination is likely to attract, all also equal. Note that the terms "accessible" and "accessibility" are used in this report (and generally in travel modeling) to apply to all travelers, and not only to people with disabilities.

A key decision for estimating the model system was which recreation sites to use in model estimation and in the initial model application. Because the big data includes thousands of visitation records for each site, it was necessary to select a subset of possible sites to keep the amount of data manageable. The final selection included:

- 89 Federal sites, including nearly all national parks plus selected national monuments
- 47 California State Parks
- 29 Pennsylvania State Parks

Sites with major highways running through or adjacent to the park boundaries were excluded, as it can be difficult using passive big data to identify park visits for recreation as opposed to through traffic that does not stop within the park. The states of California and Pennsylvania were selected for state park sites because both states have fairly comprehensive and accessible on-line databases of park amenities that are somewhat compatible with the databases available from the National Park Service. Other types of park sites such as National Forests, National Wildlife Refuges, US. Army Core of Engineers (USACE) sites, Bureau of Reclamation Sites, and Bureau of Land Management (BLM) lands were not included in this project, as they generally do not have the same types of data available from published sources, and the decisions to visit such types of sites may be different than the decisions to visit national and state parks, possibly requiring a different model design. Although it would be ideal to expand this type of research at some point to include such areas, the changes to the data collection and modeling approaches would first need to be thought through in detail.

The data files provided with the code include all input data for the Federal sites, although a discussion is provided for how the necessary input variables could be created for other park sites.

An Overview of the Software Implementation

The code to apply the model system was first written in the Pascal programming language. This language was selected for coding efficiency and accuracy, as it is the language most familiar to the project PI who designed and estimated the models and wrote the initial application code. The software was then translated into the Python language, which is a more popular coding language that is familiar to many travel modelers and can be run interactively from several software packages, including network software packages such as Cube and Visum. The Python language typically involves longer computing times than the older compiled languages such as C++ and Pascal, but computing time is not a bottleneck for this particular model implementation.

In broad "pseudo-code" terms, the software performs the following steps:

1. Get the user control options, including which park and which season(s) of the year the current run is being done for. These options can be entered as “keys” on a remote run command or can be input interactively after the run is started.
2. Initialize all the model parameters. These are documented in Chapter 3 of this document and interpreted in more detail in the project Final Report.
3. Read in all input data files. These are documented in Chapter 4 of this document.
4. Apply the various models using the following logic:
    a. Loop across all possible residence zones in the model system. For each zone:
        i. Apply the Visitor Home Location Model to calculate the model utility for that residence zone alternative,
        ii. Apply the Access Mode & Airport Model to calculate the model probabilities of using air and road and the probability of using each relevant access airport, conditional on the residence zone alternative.
    b. After all possible residence zones have been considered, the probabilities of visitors residing in each residence zone are calculated, as well as the accessibility “logsum” measures from both models that are later used as inputs to the Park Visit Generation Model.
    c. Loop across all of the possible residence zones in the model system again to apply the Auxiliary models, as follows:
        i. Apply the model to calculate the fraction of visitors from the residence zone that live inside the park “halo area.” (For zones with a zone-to-zone distance of over 150 miles to the park zone, all visitors are designated as living outside the halo area.)
        ii. For both halo area and non-halo residents, apply the Duration of Stay models to get the probability of stays of each possible duration (0 nights up to 15+ nights) inside the park and outside the park within the halo area. (Halo residents are assumed to spend 0 nights away from home outside the park—i.e., any nights they spend away from home are inside the park.)
        iii. For each possible combination of residence type (halo area or non-halo area), number of nights stayed in the park (0 to 15+) and nights stayed outside the park (0 to 15+), do the following:
            1. Apply the regression model of the Number of Park Person-Entries and update the average number of person-entries per adult visitor for halo area residents and non-halo area residents.
            2. Apply the Vehicle-Miles Traveled (VMT) models to calculate the number of miles traveled inside the park and (for non-halo residents) outside the park in the halo area)
            3. Apply the tables to calculate the distribution of arrival trips into the park and halo area and departure trips from the park and halo area by day of the week/period of the day combination.
    d. Apply the Park Visit Generation model to calculate the predicted number of person-entries to the park, using the accessibility logsums from above. (It would also be possible to apply this model in the code before the Auxiliary models are applied, as it would only affect the order of calculations but not the model outcomes.)

   e. For each possible residence zone and resident type (halo area vs. non-halo-area) apply the results from the above models to get the number of the person-entries and divide by the average number of person-entries per adult visitor to calculate the number of adult visits to the park from each zone (and from each access airport). Divide by the average of 2.25 adults per travel party (based on NPS survey data) to translate from adult-visits to vehicle-visits.

5. Using the stored results from the above steps, write output records of various types, as documented in Chapter 5.

The code will be available via the TRB website (where this guidebook will also be made available.)

CHAPTER 3

Documentation of the Model Parameters

This section documents the final parameters for each model component. A much more detailed explanation and interpretation of the model estimation results is provided in Chapter 3 of the project Final Report.

The Visitor Home Location Model

This is a multinomial logit (MNL) model, with the choice alternatives being 4,486 zones that encompass the 50 United States (plus D. C) and are the smaller of Counties or PUMAs. The final estimated model coefficients and t-statistics are shown in Table 1. The model was estimated on 1,193,571 observations. The fit of the model in terms of rho-squared (the fraction of log-likelihood explained by the model) is 0.303, which is statistically robust for a disaggregate location choice model. The following is an initial interpretation of the results:

Table 1. Visitor home location model estimation results

VARIABLE	COEFFICIENT	T-STATISTIC
Log (adults residing in zone)	1.0	constrained
Fraction of HH that are families with no children	0.0	constrained
Fraction of HH that are families with children under age 6	-2.860	-29.5
Fraction of HH that are families with all children aged 6 or older	-0.656	-8.3
Fraction of HH that are non-family	-1.292	-31.2
Fraction of persons aged 0-17	-4.269	-40.3
Fraction of persons aged 18-34	-3.342	-50.4
Fraction of persons aged 35-49	0.0	constrained
Fraction of persons aged 50-64	-3.369	-31.8
Fraction of persons aged 65 or older	-4.257	-55.2
Fraction of HH with income under $25,000	0.672	14.5
Fraction of HH with income $25,000 to $49,999	1.105	15.8
Fraction of HH with income $50,000 to $99,999	0.0	constrained
Fraction of HH with income $100,000 to $149,999	-2.478	-31.1
Fraction of HH with income $150,000 or more	0.057	1.7
Fraction of HH with HH head is Hispanic	-0.571	-55.1
Fraction of HH with HH head is white (non-Hispanic)	0.0	constrained
Fraction of HH with HH head is Black (non-Hispanic)	-1.340	-125.1
Fraction of HH with HH head is Asian/Pacific Isld (non-Hispanic)	-0.748	-41.4
Fraction of HH with HH head other/multiple race (non-Hispanic)	-1.036	-61.9
Fraction of adults that are employed	2.691	80.6
Fraction of workers who work from home	0.710	13.1
Log of zone population density (persons per square mile)	-0.115	-128.1
Home is in a zone with coastal access	0.881	69.3
Park winter sport index minus home winter sport index	0.0050	4.7
Park mountain hiking index minus home mountain hiking index	0.0613	63.3
Park coastal recreation index minus home coastal recreation index	0.1785	48.8
Park extreme heat index minus home extreme heat index	-0.0703	-39.3
Miles of road distance in range 0-100	-0.03088	-409.5

VARIABLE	COEFFICIENT	T-STATISTIC
Miles of road distance in range 100-200	-0.01418	-260.6
Miles of road distance in range 200-400	-0.00543	-191.2
Miles of road distance in range 400-800	-0.00340	-232.6
Miles of road distance in range 800-1,600	-0.00083	-101.7
Miles of road distance over 1,600	-0.00022	-31.4
Miles of road distance in Winter (Jan-Mar)	-0.00038	-79.2
Miles of road distance to Summer (Jul-Sep)	0.000086	26.2
There is no US land route between home and recreational site	-7.633	-511.2
Home is in the same zone as the recreational site	0.426	66.7
Miles of road distance to Pennsylvania state parks	-0.00314	-15.0
Miles of road distance to California state parks	-0.00015	-16.5
Home is in Pennsylvania, to Pennsylvania state parks	1.462	64.5
Home is in California, to California state parks	-0.582	-34.7

The Long-Distance Mode and Airport Choice Model

To model trips to the park destination area, it is important to account for cases where the visitor flies to an airport and goes by road from the airport to the park, rather than traveling by road all the way from home to the parkland destination. The same passive data that was used to estimate the Visitor Home Location Model was used to estimate a Mode and Airport Choice model. The two mode alternatives are (1) road for the entire distance between home and the park versus (2) using air for at least part of the journey.

The airport choice set for each park was limited to a maximum of 26 possible airports, each within a road distance of 1,000 miles from the park. This size of choice set was set to include over 99% of the observed airports used in the data set. The model was only estimated for cases where the home zone is 100 miles or more from the park zone, as trips of less than 100 miles using air are very rare. The form of the model is nested logit (NL), with the airport choice alternatives nested under the air mode. The model results are shown in Table 2. The model is based on almost 1.2 million observations, approximately 10% of which used air. The final rho-squared (pseudo-R-squared) values are in the range that would be expected for disaggregate mode choice models.

Table 2. The long-distance mode and airport choice model

Observations	**1,194,712**	
Rho-squared versus 0 coefficients	0.6981	
Rho-squared versus constants only	0.1713	
Variable	**Coefficient**	**T-ratio**
Mode nesting parameter	0.4305	120.5
Airport distance (miles)	-1.219E-02	-267.2
Airport distance squared	6.638E-06	118.2
Log of 2019 airport passengers	0.8308	405.9
Road distance (miles)	-1.004E-02	-110.5
Road distance squared	2.192E-06	103.3
Road mode constant	23.41	215.1
Road-California State Park	-0.9985	-27.9
Road-Pennsylvania State Park	0.6562	5.6

The Park Visit Generation Model

The Park Visit Generation Model is based on data published by the National Park Service for monthly visitation to National Parks and Monuments. 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. The results of the model estimation are shown in Table 3. There are 329 observations (park/season combinations for which parks were open and data available). The R-squared value adjusted for the number of parameters is 0.777. Because the number of observations is so much smaller than the number used for the two preceding models based on passive data, the t-statistics for the coefficients are generally much smaller than in those models. Nevertheless, the model includes a large number of significant coefficients with t-statistics of 1.9 or higher.

Table 3. Park visit generation model

Independent variables	Coefficient	T-ratio
Constant (intercept)	0.295	0.2
Home location model utility logsum	0.400	2.2
Airport choice model utility logsum	0.006	0.1
Winter (base = Autumn)	-1.022	-3.3
Spring (base = Autumn)	0.425	0.8
Summer (base = Autumn)	1.852	3.4
Winter sport index- Winter	0.126	2.6
Winter sport index- Spring	0.074	1.2
Winter sport index- Autumn	0.029	0.6
Mountain hiking index- Winter	0.277	5.1
Mountain hiking index- Spring	0.177	1.7
Mountain hiking index- Summer	0.029	0.3
Mountain hiking index- Autumn	0.161	2.6
Coastal recreation index- Winter	0.725	6.5
Coastal recreation index- Spring	0.498	4.6
Coastal recreation index- Summer	0.318	3.1
Coastal recreation index- Autumn	0.509	4.8
Extreme heat index	-0.454	-4.3
Site has RV camping	0.389	2.5
Site has indoor lodging	0.659	4.4
Fraction of trails w/ wheelchair access	0.555	2.6
Log of land area	0.203	4.6
Log of years since site established	1.146	5.8
Lodging/food employment in halo area	0.077	1.6
Lodging/food emp. In 100–200-mile band	0.114	1.8
Lodging/food emp. In 200–300-mile band	-0.217	-2.8
Site is in Hawaii	-1.450	-3.2
Site is in Alaska	-2.331	-3.8
Site is only accessible by water	-2.620	-7.5
Site is a Washington/Oregon national park	-1.349	-4.7
Site is in a Utah national park	1.287	5.0

Independent variables	Coefficient	T-ratio
Site is an Arizona national park	0.883	2.6
Site is a Colorado national park	0.163	0.6
Site is a New Mexico national park	0.418	1.1
Site is a California national park	0.506	2.1
Site is a Wyoming/Montana national park	0.491	1.5
Site is in Great Smoky Mountains national park	1.989	4.0
National Monuments in 4-corners states	-0.306	-1.4
National Monuments in other states	-0.097	-0.5

Auxiliary Models of Duration of Stay, Day and Time of Arrival and Departure, and Vehicle Miles Traveled

The auxiliary models include:

1. A model of whether or not the visitor lives within the halo area around the parkland site.
2. A model to predict the number of nights the visitor stays inside the park.
3. For non-halo residents, a model to predict the number of nights the visitor stays outside the park but inside the halo area.
4. A model to predict the number of vehicle-miles traveled inside the park.
5. For non-halo residents, a model to predict the number of vehicle-miles traveled outside the park but inside the halo area.
6. A classification-table to indicate what day of week the visitor first arrives at the park.
7. A classification table to indicate what periods of the day the visitor first arrives at the park and last departs from the park.

A model of whether or not the visitor lives within the halo area

The Visitor Home Location Model predicts which zone the person resides in, but the zone system is based on County and PUMA boundaries that do not conform to the halo area boundaries that are 50 miles from the nearest parkland site boundary. It is important to predict if the person lives within the halo area as the passive data processing used the halo area boundaries to generate many of the variables, and because the other auxiliary models are different for halo area residents versus non-residents.

The passive LBS data uses the exact home location to determine if the person lives within the halo area. In the estimation data, there are almost no cases for which the zone-to-zone road distance is more than 150 miles, but the person lives in the halo area. So, the model excludes all cases where the zone-to-zone road distance is more than 150 miles, as those visitors are assumed in model application to live outside the halo area. This left 416,194 cases for model estimation, of which 85% lived within the halo area. The model was estimated as a binary logit model, with the coefficients shown in Table 4, and a rho-squared (pseudo-R-squared) of 0.592.

Table 4. Model of whether or not the visitor lives within the halo area

Variable	Coefficient	T-statistic
constant (lives in halo area)	5.408	73.6
car distance miles in range 0.50	-0.043	-26.5
car distance miles in range 50-100	-0.062	-116.4
car distance miles in range 100-150	-0.079	-92.7
lives in park zone	1.752	16.5
park zone approximate radius	-0.031	-30.9
park zone approximate radius * lives in park zone	-0.073	-17.8
park approximate radius	0.039	31.1
park approximate radius * lives in park zone	-0.031	-10.4

Models to predict the number of nights staying Inside the park and halo area

The models to predict the number of nights stayed in the park and outside the park in the halo area were estimated using ordered logit (OL) models, which are appropriate when the dependent variable is a count variable taking on only integer values. The estimation results for three models are shown in Table 5. The models are:

- Halo residents- nights stayed in the park, 364,367 observations, rho-squared = 0.022
- Non-residents- nights stayed in the park, 647,758 observations, rho-squared = 0.031
- Non-residents- nights stayed outside the park, 647,758 observations, rho-squared = 0.031

Table 5. Models of number of nights stay in the park and the halo area

Model	Halo resident nights in park		Non-residents nights in park		Non-residents nights outside park	
Thresholds	**Coeff.**	**T-stat.**	**Coeff.**	**T-stat.**	**Coeff.**	**T-stat.**
0 nights	0.737	10.8	1.285	23.3	1.608	29.4
1 night	2.285	33.6	2.599	47.0	2.580	47.1
2 nights	3.004	44.0	3.355	60.6	3.330	60.8
3 nights	3.521	51.3	3.961	71.4	3.928	71.6
4 nights	3.925	57.0	4.448	80.0	4.421	80.5
5 nights	4.247	61.3	4.850	86.9	4.841	87.9
6 nights	4.530	65.0	5.190	92.6	5.220	94.6
7 nights	4.789	68.2	5.510	97.8	5.584	100.9
8 nights	5.027	70.9	5.727	101.3	5.856	105.4
9 nights	5.261	73.4	5.920	104.2	6.085	109.2
10 nights	5.464	75.4	6.097	106.7	6.287	112.4
11 nights	5.664	77.2	6.258	108.9	6.476	115.2
12 nights	5.850	78.5	6.412	110.9	6.649	117.8
13 nights	6.000	79.5	6.564	112.8	6.801	119.9
14 nights	6.185	80.4	6.717	114.5	6.955	121.9
Explanatory variables	**Coeff.**	**T-stat.**	**Coeff.**	**T-stat.**	**Coeff.**	**T-stat.**

Model	Halo resident nights in park		Non-residents nights in park		Non-residents nights outside park	
Winter (Jan-Mar)	0.263	21.8	-0.089	-10.1	-0.306	-34.1
Spring (Apr-Jun)	0.009	0.7	-0.296	-34.2	-0.158	-18.5
Summer (Jul-Aug)	-0.278	-19.7	-0.050	-5.5	0.162	18.1
Site has indoor lodging	-0.301	-26.1	-0.064	-9.1	0.289	9.0
Winter sport index	-0.00020	-3.9	0.00135	43.0	0.00014	20.2
Mountain hiking index	0.00024	6.6	0.00049	26.9	0.00034	19.3
Coastal recreation index	0.0060	13.1	0.0036	9.8	-0.0038	-10.8
Extreme heat index	0.013	7.2	-0.047	-48.1	-0.015	-17.3
Log of land area	0.114	32.3	0.126	55.0	-0.116	-54.1
Log of age of park	-0.096	-7.0	0.166	14.1	0.350	30.3
National monument	0.117	7.9	-0.326	-27.3	-0.116	-10.3
Road distance/1000			0.156	12.1	0.988	78.5
(Road distance/1000) squared			-0.066	-13.4	-0.307	-64.5
Traveled by air			-0.111	-14.9	0.473	67.9
Live in same zone as park	-0.009	-1.0	-0.191	-3.0	-0.349	-5.0
Park in Hawaii	0.049	1.6	-0.267	-11.5	2.011	91.7
Park in Alaska	0.059	1.2	0.237	10.5	0.595	25.8
Halo area food lodging employment	-2.3E-06	-41.0	3.7E-07	7.7	5.4E-06	116.8
Halo area fraction water	-1.746	-17.7	-0.785	-9.4	0.475	5.8
Halo area fraction protected land	0.107	4.7	-0.877	-74.7	0.020	1.7
Nights spent in the park					-0.071	48.7

Models to predict vehicle-miles traveled Inside the park and halo area

Given the nights spent inside and outside the park, the next set of models predicts the number of vehicle-miles traveled (VMT) within the park and within the halo area outside the park. They were estimated using log-log regression models, where the log of the vehicle-miles traveled during the visit is the dependent variable, and most of the explanatory variables also use the log form. These models are based on passive LBS data, after screening the data for outliers. Note that although we use the term "vehicle-miles traveled" because very slow trips imputed to be walking/hiking trips were excluded from the total, we do not know if the device from which the passive data was recorded is from the vehicle driver or a passenger. A more accurate term might be "person-miles traveled in vehicles." To translate that outcome to actual VMT, it needs to be divided by the average number of adults per travel party, which was estimated as 2.25 using NPS survey data.

The estimation results for three models are shown in Table 6. The models are:

- Halo residents- VMT within the park, 356,902 observations, r-squared = 0.196
- Non-residents- VMT within the park, 604,181 observations, rho-squared = 0.197
- Non-residents- VMT outside the park within the halo area, 504,789 observations, rho-squared = 0.307

Table 6. Models of vehicle miles traveled in the park and the halo area

Model	Halo resident VMT in park		Non-residents VMT in park		Non-residents VMT outside park	
	coef	t-stat	Coef	t-stat	coef	t-stat
Constant (intercept)	4.123	101.0	2.975	78.9	5.143	102.2
Log of (nights in park + 1)	0.085	5.3	0.832	65.2	-0.777	-45.9
Log of (nights in park squared + 1)	0.353	35.2	-0.266	-37.1	0.807	84.6
Log of (nights in halo + 1)			-0.437	-34.6	3.050	176.8
Log of (nights in halo squared + 1)			0.218	31.7	-0.995	-105.7
Log of (nights in park * nights in halo + 1)			0.028	7.9	-0.123	-25.2
Winter (base is Autumn)	0.546	78.2	0.649	103.2	-0.305	-36.1
Spring (base is Autumn)	0.436	63.2	0.507	87.8	0.373	49.2
Summer (base is Autumn)	0.101	14.1	0.154	27.2	0.191	26.2
Extreme heat	-0.004	-3.9	-0.018	-28.8	-0.0004	-0.5
Log of land area	0.128	80.9	0.225	166.7	-0.116	-63.8
Log of age of park	-0.447	-62.9	0.001	0.2	-0.066	-6.4
National monument	0.315	40.8	0.253	32.9	-0.269	-25.6
Site in Hawaii	0.668	51.1	0.264	21.3	0.068	3.3
Site in Alaska	-0.452	-14.3	-1.452	-88.8	0.633	30.2
Halo area food lodging employment	-0.068	-42.5	-0.153	-129.0	0.024	14.6
Halo area fraction water	0.548	14.4	1.880	43.4	-1.236	-20.9
Halo area fraction protected land	0.900	88.8	0.808	102.3	-0.710	-65.8
Log of home-to-park road distance	1.648	31.3	0.090	17.7	0.280	41.3

The day of first arriving at the park and In the halo area

For gauging local traffic effects, it may be important to know on what days the visitors first enter and last leave the area. Table 7 shows the distribution of arrival day of the week, separately for halo residents and non-residents, and varying by how many nights are spent in the park (for day arriving at the park) or how many nights are spent in the park and/or the surrounding halo area (for day arriving in the halo area).

Table 7. Table of the day of the week visitors first arrive in the park or halo area

Halo resident- day first arriving at the park	SUN	MON	TUE	WED	THU	FRI	SAT	
0 nights	15.3%	13.3%	12.8%	13.2%	13.3%	14.3%	17.8%	100.0%
1 night	13.0%	13.5%	13.0%	13.4%	13.6%	16.3%	17.1%	100.0%
2 nights	11.2%	12.3%	13.0%	13.1%	13.2%	23.9%	13.3%	100.0%
3-4 nights	12.6%	15.6%	13.8%	14.0%	16.4%	16.3%	11.2%	100.0%
5+ nights	13.9%	16.3%	14.7%	14.0%	13.9%	13.8%	13.4%	100.0%
Non-resident- day first arriving at the park	**SUN**	**MON**	**TUE**	**WED**	**THU**	**FRI**	**SAT**	**Total**
0 nights	15.7%	13.1%	12.0%	11.9%	12.8%	15.0%	19.6%	100.0%
1 night	13.8%	12.9%	12.0%	12.3%	13.2%	17.3%	18.4%	100.0%
2 nights	12.4%	12.8%	11.8%	11.8%	14.4%	23.0%	13.8%	100.0%
3-4 nights	15.1%	15.9%	12.1%	13.0%	16.6%	14.4%	12.8%	100.0%

5+ nights	19.3%	13.4%	11.9%	11.8%	11.6%	12.9%	19.2%	100.0%
Non-resident- day first arriving in the halo area	**SUN**	**MON**	**TUE**	**WED**	**THU**	**FRI**	**SAT**	
0 nights	15.7%	12.7%	10.9%	11.3%	13.0%	16.8%	19.6%	100.0%
1 night	13.7%	12.4%	10.8%	11.9%	13.3%	19.1%	18.7%	100.0%
2 nights	11.9%	11.3%	9.9%	11.3%	14.1%	27.2%	14.2%	100.0%
3-4 nights	13.4%	13.2%	10.4%	14.4%	18.5%	16.6%	13.4%	100.0%
5+ nights	18.1%	13.0%	11.4%	12.1%	11.8%	13.6%	20.0%	100.0%

The periods of the day first arriving and last departing the park or the halo area

Besides the day of the week visitors first arrive at and last leave from the park (for halo residents) or the halo area (for non-residents), traffic effects may also vary by the period of the day that the visitors enter and leave the relevant area. Table 8 shows the combination of arrival and departure periods at the park boundaries for halo residents. The day is split up into six periods—two 6-hour periods from 6 pm to 6 am, and then four 3-hour periods between 6 am and 6 pm. The 6-9 am and 3-6 pm periods roughly correspond to peak traffic periods on weekdays, although the heaviest traffic periods on weekend days may differ.

Table 8 for halo residents is split into two different sub-tables, one for day trips with no overnights stays in the park (the majority of visits by halo residents), and one for trips with overnights stays of one or more nights in the park. Table 9 for non-halo residents is analogous but is the for the arrival in the halo area rather than the park.

Table 8. Table of park arrival and departure period of the day combinations for halo residents

halo resident **day trip**	**depart before 6 am**	depart 6-9 am	depart 9 am-noon	depart noon-3 pm	depart 3-6 pm	depart after 6 pm	*total*
arrive before 6 am	0.2%	0.4%	0.6%	0.5%	0.5%	0.2%	*2.5%*
arrive 6-9 am		1.4%	6.8%	9.0%	10.1%	3.2%	*30.5%*
arrive 9 am-noon			2.5%	13.6%	11.5%	3.7%	*31.3%*
arrive noon-3 pm				2.8%	12.6%	5.8%	*21.2%*
arrive 3-6 pm					2.4%	8.0%	*10.4%*
arrive after 6 pm						4.1%	*4.1%*
total	*0.2%*	*1.8%*	*9.8%*	*26.0%*	*37.1%*	*25.0%*	*100.0%*
halo resident overnight	**depart before 6 am**	**depart 6-9 am**	**depart 9 am-noon**	**depart noon-3 pm**	**depart 3-6 pm**	**depart after 6 pm**	***Total***
arrive before 6 am	4.8%	4.8%	3.6%	2.5%	2.2%	1.8%	*19.6%*
arrive 6-9 am	0.7%	2.7%	1.8%	1.7%	2.0%	1.7%	*10.6%*
arrive 9 am-noon	0.6%	1.9%	2.9%	2.6%	2.4%	2.0%	*12.4%*
arrive noon-3 pm	0.8%	1.9%	3.0%	3.1%	2.6%	2.1%	*13.5%*
arrive 3-6 pm	1.0%	2.9%	3.2%	3.1%	3.4%	2.9%	*16.5%*
arrive after 6 pm	2.7%	5.1%	4.2%	3.6%	3.6%	8.2%	*27.4%*
total	*10.6%*	*19.1%*	*18.6%*	*16.7%*	*16.2%*	*18.7%*	*100.0%*

Table 9. Table of halo arrival and departure period of the day combinations for non-residents

non-resident day trip	**depart before 6 am**	**depart 6-9 am**	**depart 9 am-noon**	**depart noon-3 pm**	**depart 3-6 pm**	**depart after 6 pm**	***Total***
arrive before 6 am	0.1%	0.2%	0.5%	0.8%	0.4%	0.1%	*2.1%*
arrive 6-9 am		1.0%	4.3%	7.7%	5.2%	1.5%	*19.8%*
arrive 9 am-noon			3.5%	17.0%	14.7%	4.0%	*39.1%*
arrive noon-3 pm				4.5%	17.0%	5.6%	*27.0%*
arrive 3-6 pm					3.3%	6.5%	*9.8%*
arrive after 6 pm						2.2%	*2.2%*
Total	*0.1%*	*1.2%*	*8.4%*	*30.0%*	*40.6%*	*19.8%*	*100.0%*
non-resident overnight	depart before 6 am	depart 6-9 am	depart 9 am-noon	depart noon-3 pm	depart 3-6 pm	depart after 6 pm	*Total*
arrive before 6 am	1.7%	1.7%	3.2%	2.3%	1.4%	0.8%	*11.1%*
arrive 6-9 am	0.3%	1.4%	2.2%	1.9%	1.6%	1.0%	*8.3%*
arrive 9 am-noon	0.4%	2.2%	5.0%	4.5%	3.9%	2.1%	*18.1%*
arrive noon-3 pm	0.4%	2.5%	5.8%	5.0%	4.1%	2.4%	*20.2%*
arrive 3-6 pm	0.6%	3.0%	6.3%	4.9%	4.2%	2.6%	*21.6%*
arrive after 6 pm	1.2%	3.1%	5.0%	3.8%	3.2%	4.4%	*20.7%*
Total	*4.5%*	*14.0%*	*27.4%*	*22.4%*	*18.4%*	*13.3%*	*100.0%*

Translating between person-entries and adult visits and vehicle visits

Additional parameters used in the model implementation are needed because the NPS visitation data used to estimate the Park Visit Generation Model includes both adults and children and would count all entries for people who entered the park more than once during their visit to the area. The other models, however, use adult visits to the area as the unit of observation, as that is the most reliable way of using the passive data. Based on NPS survey data, Table 10 shows factors used to translate between person-entries and adult entries, with between 0.211 and 0.311 children (age 0-17) per adult (age 18+), increasing somewhat with longer stays in the park and halo area.

Table 10. Average number of children per adult visitor, from NPS survey data analysis

Nights in the area	**Average children per adult**
0	0.211
1	0.225
2	0.234
3	0.259
4	0.266
5	0.268
6 or more	0.311
Total	0.250

Table 11 shows the results of a simple regression model based on NPS survey data of the number of park entries per visit as a function of the number of nights staying in the park and the number of nights staying outside the park but in the area.

Table 11. Regression model of number of park entries, from

Variable	**Coeff**	T-statistic
Constant	1.146	25.8
Nights in the park	0.454	37.0
Nights near the park	0.303	36.5
Day trip	0.440	7.4

A final parameter based on the NPS survey data is the average number of adults per travel party. This factor of 2.25 is used to translate from person-trips and person-miles traveled (PMT) to vehicle-trips and vehicle-miles traveled (VMT).

CHAPTER 4

Documentation of the Input Data Files

The Zonal Data File

As described in the project Final Report, the zone system is the “NUMA” zone system of 4,486 zones that is also used for the National Long-Distance Passenger Travel Demand Model, created by RSG for the Federal Highway Administration (FHWA). The zones are the intersection of counties and Census Bureau Public Use Microdata Areas (PUMAs) used for the American Communities Survey (ACS). In urban areas where there are typically multiple PUMAs per county, the zones are PUMAs, while in more rural areas where there are often multiple counties per PUMA, the zones are counties.

Table 12 shows the data item labels and descriptions for the input data file *zone_acs_19.dat.* The file has 4,486 records, with spaces as delimiters between fields. The file contains a mix of ACS socio-demographic data, land use density data, and climate and topography data. The ACS data was aggregated up from Census tract level data for the 2015-2019 5-year ACS tables. Chapter 3 of the project Final Report has further information about how these data items were prepared.

It is not expected that the user will have to change this data in any way, although, as stated in the Recommendations section of the Final Report, the data could be updated sometime in 2024 or beyond when new PUMA other Census boundaries have been specified and new ACS 5-year tables from (predominantly) post-2020 data are available.

Table 12. Zonal data items (provided in Zone_acs_19.dat)

Data label	Description
zone	The zone index number, from 1 to 4486
zone_sqm	The area of the zone, in square miles
hholds	The number of households (HH) in the zone, from ACS
persons	The number of persons in the zone, from ACS
adults	The number of adults age 18+ in the zone, from ACS
ffamcu6	The fraction of family HH in the zone that have any children under age 6, from ACS
ffamc6p	The fraction of family HH in the zone that have children, all ages 6 or older, from ACS
famnoc	The fraction of family HH in the zone that have no children, from ACS
fnonfam	The fraction of non-family (single-person or non-related) HH in the zone, from ACS
fage017	The fraction of persons in the zone aged 0-17, from ACS
fage1834	The fraction of persons in the zone aged 18-34, from ACS
fage3549	The fraction of persons in the zone aged 35-49, from ACS
fage5064	The fraction of persons in the zone aged 50-64, from ACS
fage6579	The fraction of persons in the zone aged 65-79, from ACS
fage80up	The fraction of persons in the zone aged 80 or older, from ACS
fwhite	The fraction of HH in the zone with a White, non-Hispanic householder, from ACS
fblack	The fraction of HH in the zone with a Black, non-Hispanic householder, from ACS

fhisp	The fraction of HH in the zone with a Hispanic householder, from ACS
asian	The fraction of HH in the zone with an Asian, non-Hispanic householder, from ACS
fothrc	The fraction of HH in the zone with other/multiple race, non-Hispanic householder, ACS
f0veh	The fraction of HH in the zone with no vehicles available, from ACS
fwathome	The fraction of workers in the zone who usually work from home, from ACS
femployed	The fraction of adults in the zone who are employed, from ACS
fincu25	The fraction of HH in the zone with gross income of $0-$24,999, from ACS
finc2550	The fraction of HH in the zone with gross income of $25,000-$49,999, from ACS
finc5075	The fraction of HH in the zone with gross income of $50,000-$74,999, from ACS
finc75100	The fraction of HH in the zone with gross income of $75,000-$99,999, from ACS
finc100150	The fraction of HH in the zone with gross income of $100,000-$149,999, from ACS
finc150200	The fraction of HH in the zone with gross income of $150,000-$199,999, from ACS
finco200	The fraction of HH in the zone with gross income of $200,000 or more, from ACS
state	The state FIPS code for the zone
popden	The population density in the zone, persons per square mile, from EPA SLD
empden	The employment density in the zone, jobs per square mile, from EPA SLD
enteden	The food and accommodation job density in the zone per square mile, from EPA SLD
fracwater	The fraction of the area in the zone covered by water, from EPA SLD
fracprotl	The fraction of the land area in the zone that is public protected land, from EPA SLD
coast	1 if the zone has Pacific or Atlantic coastal access, 0 otherwise (based on maps)
elevation	The average elevation of weather stations in the zone from 2006-2020
avgtemp1	The average JAN-MAR temperature at weather stations in the zone from 2006-2020
maxtemp1	The average JAN-MAR maximum daily temperature at weather stations from 2006-2020
mintemp1	The average JAN-MAR minimum daily temperature at weather stations from 2006-2020
avgprec1	The average JAN-MAR daily precipitation at weather stations in the zone from 2006-2020
avgtemp2	The average APR-JUN temperature at weather stations in the zone from 2006-2020
maxtemp2	The average APR-JUN maximum daily temperature at weather stations from 2006-2020
mintemp2	The average APR-JUN minimum daily temperature at weather stations from 2006-2020
avgprec2	The average APR-JUN daily precipitation at weather stations in the zone from 2006-2020
avgtemp3	The average JUL-SEP temperature at weather stations in the zone from 2006-2020
maxtemp3	The average JUL-SEP maximum daily temperature at weather stations from 2006-2020
mintemp3	The average JUL-SEP minimum daily temperature at weather stations from 2006-2020
avgprec3	The average JUL-SEP daily precipitation at weather stations in the zone from 2006-2020
avgtemp4	The average OCT-DEC temperature at weather stations in the zone from 2006-2020
maxtemp4	The average OCT-DEC maximum daily temperature at weather stations from 2006-2020
mintemp4	The average OCT-DEC minimum daily temperature at weather stations from 2006-2020
avgprec4	The average OCT-DEC daily precipitation at weather stations in the zone from 2006-2020

The other pre-prepared zonal data file is a file with zone-to-zone road network car distances, using the network for the FHWA National Long-Distance Passenger Travel Demand Model. Although there are only 3 data items on each record, there are 4,486 x 4,486 records (with spaces as delimiters between fields, making it the largest data file used. Table 13 shows the data items in input file ODCarDist.dat. Note that for zones with no road access through the U.S, such as zone pairs with one zone in Hawaii and Alaska and the other zone outside that state, the file has a value of -1.

Table 13. Zone to zone road distance data items (provided in ODCarDist.dat)

Data label	Description
oZone	The origin zone index number (1-4486)
dZone	The destination zone index number (1-4486)
odRoadDist	The origin-destination road network distance, in miles

The model also uses a fixed set of airports, each with a number and familiar three letter code, along with a correspondence to a particular zone in the zone system, and the number of trips originating from the airport (not including transfers) based on the DB1B Market data for 2019. The data items are listed in Table 14.

Table 14. Airport data items (provided in APdat_19.dat)

Data label	Description
apCode	The airport three letter code
apZone	The model zone that the airport is located in
apPAX2019	The number of air trips originating from the airport in 2019 (from DB1B data)
apNum	The airport index number

The next input file is the file with park attributes, including attributes of the park halo area. The fields are shown in Table 15. Records in this file have been pre-prepared for nearly all National Parks, and for several National Monuments. Much of this data has also been compiled for several California and Pennsylvania state parks, although a few of the variables, such as the year established and the fraction of trails with wheelchair access have not yet been compiled for those parks but could be via internet searches.

Filling in many of the variables for a park that is not currently in the database involve some GIS work, including identifying the Census tract (and zone) with the most land area in common with the park, identifying all Census tracts that lie at least partially withing the park and within the park halo area and how much spatial area in each tract is within the park and the halo area, This information can then be joined with tract-level data on population, employment, water area, protected land area, etc. to calculate several of the variables used in the models. RSG has example GIS scripts to create such variables.

The variables for the number of visitors during each of the four seasons are optional, since that is one of the things that the model system is designed to predict. The model user can leave those “blank” by using a value of -1 (or any value less than 0). If the user does provide a value greater than 0, then they are given the option of using that value instead of applying the Park Visit Generation Model. In that case, the model system will still apply all of the other models to predict the long-distance and local travel generated by those visitors but will not attempt to predict the actual number of visitors. (Note that a value of exactly 0 for the number of visitors is interpreted to mean that the park is not open to visitors during that season, so all outputs will be 0.)

The final numeric value on the file is the assumed fraction of foreign visitors that reside outside the U.S. This is an optional input that can be supplied by the user, and a discussion of how international visitors are modeled is provided below.

Table 15. Park specific data items in (provided in NPS_ParkData.dat)

Data label	Description
parkCode	A unique 4-letter park identification code
parkType	1=National Park, 2=National Monument, 3=CA State Pk, 4=PA St.Pk 5=Other
park_tractgeoid	The 2019 Census tract with the most land area in common with the park
park_numaindex	The NUMA zone with the most land area in common with the park
park_sqm	The park area, in square miles
year_founded	The year the park was established
Frwcaccess	The fraction of park trails with that are wheelchair accessible
Hascamping	1 if the park has any camping sites, 0 otherwise
Hasrvcamping	1 if the park has any recreational vehicle (RV) camping sites, 0 otherwise
Haslodging	1 if the park has any indoor (lodge, cabin, etc.) lodging, 0 otherwise
Halopop	The population of the 50-mile halo area around the park
Halohh	The number of households in the halo area around the park
Haloemp	The number of jobs in the halo area around the park
Haloentemp	The number of lodging and food service jobs in the halo area around the park
Halowater	The area in the halo area covered by water, in square meters
Haloprotl	The area of public predicted land in the halo area, in square meters
Haloarea	The total area of the halo area, in square meters
lentemp1	The log of the number of lodging and food jobs within 100 miles of the park
lentemp2	The log of the number of lodging and food jobs 100-200 miles from the park
lentemp3	The log of the number of lodging and food jobs 200-300 miles from the park
Wateraccessonly	1 if the park is only accessible by water, 0 otherwise
win_visits	The number of visitors during JAN-MAR, if known, -1 otherwise
spr_visits	The number of visitors during APR-JUN, if known, -1 otherwise
sum_visits	The number of visitors during JUL-SEP, if known, -1 otherwise
aut_visits	The number of visitors during OCT-DEC, if known, -1 otherwise
fr_foreign	The fraction of visitors that are from other countries, if known, -1 otherwise
parkName	The full name of the park (used for labeling purposes only)

The final input file defines the choice set of access airports for each park, with the record content shown in Table 16. This file has already been generated for all of the National Parks and Monuments and California and Pennsylvania state parks used for model estimation in this project. To add a new park to this database, the user can define a set of up to 26 access airports. Typically, this set should include all local commercial airports within 100 miles or so, all regional airports within 250 miles or so, and any major airports within 500 miles or so. The distance between the park and each airport could be estimated using apps such as Google Maps or Google Earth. If the user leaves the distance "blank" by entering -1 (or any value less than 0), the model implementation code will use the zone-to-zone road distance between the park zone and the airport zone, which may not be very accurate for airports that are very nearby.

Table 16. Park specific access airport choice set data (provided in ParkAPorts.dat)

Data label	Description
parkCode	A unique 4-letter park identification code
apCode1	The 3-leter code for the 1st access airport (XXX if no more airports in set)
apDist1	The distance from the 1st airport to the park, miles (-1 to use zone-to-zone)
apCode2	The 3-leter code for the 2nd access airport (XXX if no more airports in set)
apDist2	The distance from the 2nd airport to the park, miles (-1 to use zone-to-zone)
apCode3	The 3-leter code for the 3rd access airport (XXX if no more airports in set)
apDist3	The distance from the 3rd airport to the park, miles (-1 to use zone-to-zone)
etc.	The same for up to 26 airports
apCode26	The 3-leter code for the 26th access airport (XXX if no more airports in set)
apDist26	The distance from the 26th airport to the park, miles (-1 to use zone-to-zone)

Some specific aspects of generating inputs for the implementation are discussed in more detail in the following paragraphs.

Input Variables for Parkland Destinations

RSG has already prepared all of the necessary model inputs for 89 Federal sites, including nearly all national parks plus selected national monuments, and we have prepared most of the necessary inputs for 47 California state parks and 29 Pennsylvania state parks. To make predictions for a new parkland destination, the first step is mainly a GIS exercise overlay the park boundaries on a map of Census tract boundaries, and determine:

- Which tracts fall within the park boundaries, and how much land area in each tract falls within the park boundaries.
- Which tracts fall within a “halo area” which has boundaries extending 50 miles around the park nearest park boundary, and how much land area in each tract falls within the park boundaries.

RSG has example GIS scripts to carry out this first step, as well as other scripts to use the use the results to calculate specific variables used in the model such as:

- The “NUMA” zone that has the most land area in common with the park.
- The total land area in the park
- The number of lodging and food service employees in the halo zone
- The fraction of the halo area covered by water.
- The fraction of the halo area that is protected public land.

Other model input variable must be supplied by the user from local knowledge, including:

- The designation of park type (national park, national monument, state park, or other)
- The year the park was first established.
- Whether or not the park has RV camping
- Whether or not the park has indoor lodging
- The fraction of trails in the park that are accessible by wheelchair

Note that the Census tract boundaries used in this project are from 2010, while the PUMA and County boundaries used to define the "NUMA" zone system are from 2012. At some point, these geographic boundaries will need to be updated, and other data used by the model system from 2019 and before should be updated as well. This point is discussed further in the Recommendations chapter.

Travel Made by International Visitors

Passive data sets do not contain trace data for international visitors. So, international visitors were not included in the models based on the passive data. The Park Visit Generation Model, on the other hand, is based on data on total visits published by the National Park Service, and those data presumably do include international visitors. Our approach to dealing with international visitors is as follows:

1. The user can provide an estimate of the fraction of visits to the site that are by foreign visitors who do not reside in the US. Table 17 provides estimates of the percent of foreign visitors at fourteen park sites, based on National Park Service Visitor Survey reports from 2015-2018. (The first five shaded rows are parks that were also used for this study, so the passive data and model results can be compared to the survey data results.) The percentage of foreign visitors is clearly higher for the more famous parklands, particularly those in the desert Southwest, and also for remote parklands in Alaska.
2. After the Park Visit Generation Model is used to predict the total visitation (or estimates of actual visitation provided by the user are used instead of applying the Park Visit Generation Model), the total visitation is multiplied by the input fraction of international visitors, and those visitors are treated in the remaining models as if their residence zone is in Honolulu if the park is in the continental U.S., or as if their residence zone is an Atlanta if the park is in Hawaii or Alaska. Assuming these residence zones ensures that the visitors will be modeled as arriving by air, with no possible road trip between home and the park. In essence, including Alaskan and Hawaiian zones and parks in the model and including the Mode and Airport Choice model allows us to use travelers between Hawaii and the mainland as a proxy for international visitors. In reality, international visitors may have different patterns of stay and travel around the park and the halo area compared to domestic visitors, but with no data to model such differences, including international visitors in a proxy manner seems more accurate than ignoring them altogether.

Table 17. Percent of visitors who live outside of the US, based on NPS visitor surveys

National Site	% Foreign Visitors
ACAD: Acadia National Park	9%
BISC: Biscayne National Park	4%
GLAC: Glacier National Park	13%
GLBA: Glacier Bay National Park and Preserve	20%
GRCA: Grand Canyon National Park	31%
BLRI: Blue Ridge Parkway	3%
CACO: Cape Cod National Seashore	7%
COLO: Colonial National Historical Park	3%
GLCA: Glen Canyon National Recreation Area	20%
GOGA: Golden Gate National Recreation Area	13%
KLGO: Klondike Gold Rush National Historical Park	64%
NIOB: Niobrara National Scenic River	0.5%
SAAN: San Antonio Missions National Historical Park	5%
VALR: World War II Valor in the Pacific National Monument	5%

CHAPTER 5

Documentation of the Output Data Files

The output files provide various types of model outputs that may have different users. This chapter discusses what information is provided in each file and how it might be used in practice.

Output from the Visitor Home Location Model and Mode & Airport Choice Models

The file as outlined in Table 18 has an output record for each of the 4486 possible residence zones in the U.S. There is an additional record for international visitors, for which the residence zone, state, county, and PUMA are all coded as 0. The results include:

- The total number of adult visitors from the residence zone
- The number of adult visitors split out by the two main modes (road only and air & road)
- The number of visitors arriving via each of the airports that are in the choice set for the park.

This file provides the outputs in their most disaggregate form. The file can be analyzed and aggregated in a number of ways. For example, one can calculate the percentage of visitors residing in each state. One can also calculate the overall air mode share and airport shares summing across the visitors from all residence zones. In fact, those particular aggregated outputs were used for the validation comparisons presented in the project Final Report.

Table 18. Output Records from the Visitor Home Location and Mode & Airport Choice Models

Data label	Description
parkCode	A unique 4-letter park identification code
quarter	1 = JAN-MAR, 2 = APR-JUN, 3 = JUL-AUG, 4 = OCT-DEC
resZone	Residence zone number
resState	Residence state FIPS code
resCounty	Residence county FIPS code
resPUMA	Residence PUMA FIPS code
roadDistance	Road distance from the residence zone to the park (-1 if no land route)
adultVisits	Number of adults from the residence zone visiting the park
adultVisitsViaRoad	Number of adult visits using road all the way from residence zone to park
adultVisitsViaAir	Number of adult visits using air partway from residence zone to park
aoCode1	3-leter code for the first airport in the choice set for the park (XXX=none)
adultVisitsAirport1	Number of adult visits from residence zone via airport 1
etc.	Same two fields for airports 2-25
aoCode26	3-leter code for the 26th airport in the choice set for the park (XXX=none)
adultVisitsAirport26	Number of adult visits from residence zone via airport 26

Detailed Outputs for the Trips Arriving in and Leaving from the Park/Halo Area

The trips in this file 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 model predicts the number of vehicle-trips arriving from each zone during each combination of day of week and period of the day. This file is somewhat similar to the previous one in that it contains records for each zone. It is different from that file, however, in some important ways:

- Instead of focusing on the residence zone, it focuses on the "origin zone," which is the residence zone if the person travels all the way by road but is the access airport zone if the person travels partway by air. As the model does not predict itineraries for multi-destination tours, this provides the best indication of where vehicle trips come from when first arriving in the park area and where they go to when last leaving the park area.
- There is no separate record for international visitors, since all international visitors are modeled as arriving by air.
- Instead of using units of adult visits, it uses units of vehicle visits, by dividing adult visits by an average factor of 2.25 adult occupants per vehicle.
- Any potential origin zones that have no land route to the park (roadDistance=-1) have no vehicle trips, as those visitors must use an intervening access airport to visit the park.
- The results of the auxiliary models are used to split out the trips by day of week and period of the day.

If a user wishes to factor the results to average daily vehicle trips, then there are approximately 13 weeks during each of the four quarters, so the simplest method would be to divide the values in the file by 13. However, users may have more information on local month-to-month or even week-to-week variations in traffic, so could use a different factor to go from the JUL-SEP results to a peak weekend in July, for example.

Table 19. Output Records for Vehicle Trips Arriving in and Leaving the Park Area

Data label	Description
parkCode	A unique 4-letter park identification code
Quarter	1 = JAN-MAR, 2 = APR-JUN, 3 = JUL-AUG, 4 = OCT-DEC
originZone	origin zone number
originState	origin state FIPS code
originCounty	origin county FIPS code
originPUMA	origin PUMA FIPS code
roadDistance	Road distance from the origin zone to the park
vehicleVisits	Number of vehicles from the origin zone visiting the park area.
arriveSundayPer1	Number of vehicles arriving in the area on Sunday in Period 1 (before 6 am)
arriveSundayPer2	Number of vehicles arriving in the area on Sunday in Period 2 (6-9 am)
arriveSundayPer3	Number of vehicles arriving in the area on Sunday in Period 1 (9 am-noon)
arriveSundayPer4	Number of vehicles arriving in the area on Sunday in Period 4 (noon-3 pm)
arriveSundayPer5	Number of vehicles arriving in the area on Sunday in Period 5 (3-6 pm)
arriveSundayPer6	Number of vehicles arriving in the area on Sunday in Period 6 (after 6 pm)
departSundayPer1	Number of vehicles departing the area on Sunday in Period 1 (before 6 am)

Data label	Description
departSundayPer2	Number of vehicles departing the area on Sunday in Period 2 (6-9 am)
In	Number of vehicles departing the area on Sunday in Period 1 (9 am-noon)
departSundayPer4	Number of vehicles departing the area on Sunday in Period 4 (noon-3 pm)
departSundayPer5	Number of vehicles departing the area on Sunday in Period 5 (3-6 pm)
departSundayPer6	Number of vehicles departing the area on Sunday in Period 6 (after 6 pm)
arriveMondayPer1	Number of vehicles arriving in the area on Monday in Period 1 (before 6 am)
etc.	In total 7 days x 2 directions x 6 periods = 84 fields
departSaturdayPer6	Number of vehicles departing the area on Saturday in Period 6 (after 6 pm)

The way that each agency uses this information for their model can depend on the size of the area modeled and the location of the park relative to the modeled area. For an MPO regional model for which the parkland is inside the modeled area, the parkland halo area may encompass most or all of the modeled area. In that case, the longer distance arriving non-resident trips can be treated as external-internal (X-I) trips and the longer-distance departing non-resident trips can be treated as internal-external (I-X) trips that can be fed into the regional model and assigned to the road network. If the parkland is somewhat outside the MPO modeled area, then some of the trips will be external-external (X-X) trips that pass through the modeled area on the way to or from the parkland site.

For DOT statewide models, the usage of the outputs may be similar to that described above for MPO models, but in that case (depending on the size of the state), many trips from outside the halo area may still be made by residents of the same state, so some of the trips will be internal-internal (I-I) trips totally within the state. For those trips, the user will need to provide a translation between their own model's internal zones and the "NUMA" PUMA/county zones used in this model system. When the NUMA zones are larger than the statewide model zones, the trips can be apportioned to those zones proportionally to the number of households in each zone. Alternatively, if the NUMA zones are much larger than the statewide model zones, the trips could be apportioned using a simple gravity model approach that takes into account both the number of households in each statewide model zone as well as the distance that each zone is from the parkland destination.

For trips that have at least one end in an external zone outside the MPO or DOT modeled area, it will be useful to aggregate the zones to make it easier to assign those trips to specific "external stations"—key access roads by which travelers enter or leave the modeled area. Users can designate one or more thresholds to use to indicate that zones should be aggregated outside that threshold, and the level of aggregation. Thresholds can be region or state boundaries or distance thresholds from the parkland site. Zones can be aggregated up to the state level outside of those boundaries, or even up to the 9 Census Divisions.

As discussed in the project Final Report, this way of using the model implicitly assumes that the visitors travel directly from the residence zone or access airport to the parkland destination. In actual cases, travelers may be on multi-destination tours and may approach the park from another destination on the tour or may depart towards another destination on the tour rather than towards home. In the project Final Report, we suggest some contexts in which this limitation could be addressed for a more limited geography with extensive data, such as for a single statewide model.

Detailed Outputs for Local Trips that Visitors Make Within the Parkland and Halo area

The auxiliary models provide estimates of person-nights spent, vehicle-nights spent, person-miles traveled, and vehicle-miles traveled, both within the parkland and within the halo area around the park. These can also be used to provide an estimate of vehicle-miles traveled for each day of the week. What the model does not predict is the exact number of trips and the location of each trip-end in the halo area or within the park—e.g., at the Census tract level. Detailed local trip generation and destinations were not modeled because that level of local travel detail cannot be reliably generated from LBS big data, and, even if it could be, it would be very difficult to model such data in a way that could be generalized to any model region in the US.

The most detailed outputs for local travel are written to a file with the record structure shown in Table 20. The file contains 16*16=256 records for non-halo residents, for each combination of nights spent inside the park and in the halo area, plus another 16 records for halo residents for nights spent in the park. The fields are total VMT in the park, total VMT in the halo area (for non-halo residents), and then the same split out by day of the week. The allocation to days of the week splits the VMT evenly for all days that the visitor is in the area, as the models (and the passive data that the models were based on) do not provide that information.

In most cases, model users will want to aggregate the records in this file across the nights in the park and the nights in the halo area, as most users are more interested in the VMT generated by all visitors rather than how it is distributed by the duration of stay in the region.

Table 20. Output Records for Local Travel Made in the Area During the Visit

Data label	Description
parkCode	A unique 4-letter park identification code
Quarter	1 = JAN-MAR, 2 = APR-JUN, 3 = JUL-AUG, 4 = OCT-DEC
resType	1 = residence within the halo area, 2 = residence outside the halo area
nightsInPark	Nights stayed in the park (0-15+)
nightsInHalo	Nights stayed in halo area (0-15+, always 0 for halo residents)
VMTinPark	Vehicle miles traveled in the park
VMTinHalo	Vehicle miles traveled in the halo area (always 0 for halo residents)
parkVMTSunday	Vehicle miles traveled in the park on Sundays
parkVMTMonday	Vehicle miles traveled in the park on Mondays
parkVMTTuesday	Vehicle miles traveled in the park on Tuesdays
parkVMTWenesday	Vehicle miles traveled in the park on Wednesdays
parkVMTThursday	Vehicle miles traveled in the park on Thursdays
parkVMTFriday	Vehicle miles traveled in the park on Fridays
parkVMTSaturday	Vehicle miles traveled in the park on Saturdays
haloVMTSunday	Vehicle miles traveled in the halo on Sundays (always 0 for halo residents)
haloVMTMonday	Vehicle miles traveled in the halo on Mondays (always 0 for halo residents)
haloVMTTuesday	Vehicle miles traveled in the halo on Tuesdays (always 0 for halo residents)
haloVMTWenesday	Vehicle miles traveled in the halo on Wednesdays (always 0 for halo residents)
haloVMTThursday	Vehicle miles traveled in the halo on Thursdays (always 0 for halo residents)
haloVMTFriday	Vehicle miles traveled in the halo on Fridays (always 0 for halo residents)
haloVMTSaturday	Vehicle miles traveled in the halo on Saturdays (always 0 for halo residents)

Note that the VMT for travel in the halo area is not modeled for halo residents because the previous file already includes model outputs for halo residents' first trip to the park and last trip leaving the park. For non-halo residents, however, there will be some double counting of miles, unless their trips arriving in the halo area and leaving the halo area are modeled as stopping at the edge of the halo area, with all other VMT part of local travel during the visit.

Reconciling the Local Travel Outputs with the Local Model System

The zone system and regional boundaries for the local DOT or MPO model will generally be different than the zone system and halo area boundaries used for this model system. The issue then becomes how to generate reasonable trips within and between the parkland site and the halo area that can be calibrated to the model outputs on daily VMT and then assigned to the traffic network. We recommend using a gravity model formulation, where the distance from the parkland zone to each other zone in the local model zone system is used together with the lodging and food service employment in each local model zone to get a distribution of local non-resident trips between the park and other locations. (Retail employment could also be used as a proxy.)   The distribution of trips will give an average trip distance, which can be divided into the total VMT to get a number of vehicle trips. The model predicts the distribution of trips across days of the week but not across periods of the day, so a knowledge of local park visitor tourist patterns trip for local trips in and out of the park could be used to assume typical time of day factors for weekdays and weekends.

For both halo residents and non-residents, the model system provides estimates of person-nights spent in the park, and vehicle-miles traveled within the parkland. If the park area is more than one zone in the local model, and the user wants to predict trips between zones within the parkland, the model system does not provide that information. That could also be done using a gravity model formulation using distance between the zones and a variable representing attractions within the parkland zones, such as the number of camping and lodging sites in each zone. If the local model includes the parkland site as (part of) a single zone, then all travel within the park will be intrazonal-travel in the local model. In that case, it is more important to focus on travel on the roads entering and surrounding the park that can be used by park visitors.

In a more general sense, when the NUMA zones used in the model system are larger than the zones in the local model system (which will generally be the case), then the numbers of trips can be allocated to smaller zones based on the number of residents (as production variables) or the number of lodging and food service (or retail jobs) as production variables.

Some regions or states may have extensive local household travel survey data that could be used in validating and calibrating the outputs from this general model system. This possibility is discussed further in the final chapter of the project Final Report.

CHAPTER 6

Running the Stand-alone Model

A stand-alone version of the model has been provided as an executable file NCHRP_08_132_Model_1.exe. This version was compiled from Pascal code in the file NCHRP_08_132_Model_1.pas, using the Lazerus Pascal IDE and compiler. (The Lazerus project files are provided in NCHRP_08_132_Model_1.lpi and NCHRP_08_132_Model_1.lps.

The stand-alone version reads in a configuration file NCHRP_08_132_Model_1_Config.txt. The contents of the configuration file provided is listed below. Each line has a label and an argument, separated by one or more spaces. All text is converted into lowercase letters upon input, so the labels and arguments are not case-sensitive.

The first label ParkCode tells the model which park in the data file to run—in this case, it is Zion. It must match one of the 4-letter park codes in the park data file.

The second label, Season, tells the model which season(s) to run. Only the first three letters of the input are used, so valid labels are Winter, Spring, Summer, or Autumn (or just the first three letters) to run one of the four seasons alone, or All to run all four seasons in succession. (The default is to run all four seasons.)

The third level RunVisitModel tells the program whether it should run the ParkVisitGenerationModel or to use provided visitation data instead. If the argument begins with letter N for No, and if non-negative values for visits in each quarter are provided in the input park data file, then the ParkVisitGenerationModel is not run, and the input data is used instead to provide the number of visits to the park during the quarter.

The remaining lines give the file names for the 5 input files and 3 output files. If no directory name is given, the files must be in the same directory as the .EXE file. Otherwise, directory paths can be given using standard syntax—e.g., c:\parkmodel\input_data\custom_parkdata.dat.

All input files for the standalone model must currently be in space-delimited format, except for the park airport file, which is in comma-delimited CSV format. All output files are in CSV format.

```
parkcode        zion
season          sum
runvisitmodel   y
parkdatafile        nps_parkdata.dat
airportdatafile     apDat_19.dat
parkapsdatafile     parkAPs_19.csv
zonaldatafile       zone_acs_19.dat
roaddistdatafile    odcardist.dat
homelocandmodeoutfile     vishomelocandmode_out.csv
tripstoandfromparkoutfile tripstoandfrompark_out.csv
localvmtoutfile           localvmt_out.csv
```