United States
Department of
Agriculture

Forest Service

Southern
Research Station

Research Paper
SRS–37

An Outdoor Recreation Use Model with Applications to Evaluating Survey Estimators

Stanley J. Zarnoch, Donald B.K. English, and Susan M. Kocis

The Authors

Stanley J. Zarnoch, Mathematical Statistician, USDA Forest Service, Southern Research Station, Asheville, NC 28802; **Donald B.K. English**, Visitor Use Monitoring Program Manager, USDA Forest Service, WO–RHR, Washington, DC 20250-1125; and **Susan M. Kocis**, Visitor Use Assistant Program Manager, USDA Forest Service, WO–RHR, East Lansing, MI 48823.

Cover photo: Arkansas Buffalo National River in the Ozark Mountains.

Photo by: Bill Lea.

September 2005

Southern Research Station
P.O. Box 2680
Asheville, NC 28802

An Outdoor Recreation Use Model with Applications to Evaluating Survey Estimators

Stanley J. Zarnoch, Donald B.K. English,
and Susan M. Kocis

Contents

An Outdoor Recreation Use Model with Applications to Evaluating Survey Estimators

Stanley J. Zarnoch, Donald B.K. English, and Susan M. Kocis

Abstract

An outdoor recreation use simulator (ORUS) has been developed to simulate recreation survey data currently being obtained by the U.S. Department of Agriculture Forest Service, National Visitor Use Monitoring (NVUM) program's survey of the national forests of the United States. Statistical distributions represent the various behaviors of recreationists during their visit to a recreation site. The beta distribution is used to model arriving times and last-exiting times. The Poisson distribution is used to model the number of intermediate exits from the site, and the times of the exits are selected randomly according to the uniform distribution. Finally, three levels of trap shyness are assigned to the recreationists to quantify the probability that the recreationists will be captured by the interviewer. The beta distributions for arriving and last-exiting are parameterized to the NVUM survey data. The functioning of the simulator is demonstrated with a simple example. The utility of ORUS in evaluating the bias and coefficient of variation of various survey scenario estimators of recreation use is also presented.

Keywords: Last-exiting recreationist, National Visitor Use Monitoring, ORUS, simulation, trap shy.

Introduction

Outdoor recreation has recently become a very important commodity of forests and in many situations it is given higher priority than more traditional forest products. Dependable, accurate estimates of recreation use are increasingly critical in national, regional, and local forest-level decisionmaking and planning. They are necessary in assessing recreation benefits from forests and for quantifying the impacts on other forest resources and local economies.

Although forest inventories have been developed for estimating the traditional timber values of forests, until recently no large scale, statistically based analog has existed for estimating recreation use. However, in 1996 a pilot study was initiated to develop a field survey for estimating recreation use on the national forests throughout the United States (Zarnoch and others 2002). This was modified and expanded substantially to also include characteristics of the visitors, their satisfaction with the recreation resource, and their economic impact on the local community (English and others 2002). This has led to the U.S. Department of Agriculture Forest Service (Forest Service), National Visitor Use Monitoring (NVUM) program, which has now completed a full multiyear cycle of sampling all national forests in the United States. While the NVUM survey was designed for use on the national forests of the United States, the basic methodology is under study by several other Federal and State agencies in the United States that need quality estimates of visitation on recreation lands on which recreation sites are usually dispersed.

One step in the review process of the first NVUM survey cycle is that a critical evaluation of the visitation estimators be performed to determine their potential bias and variance properties and to evaluate methodological changes that can reduce one or the other. To that end, an outdoor recreation use simulator (ORUS) has been developed that can provide data similar to that collected across the national forests during the first NVUM survey cycle.

The simulator benefits efforts to estimate recreation use on forest lands in several ways. First, the model outlines a structure that allows researchers to decompose the complex system of visitor behavior into a set of more easily understood components and to demonstrate how these components are related to the recreation use estimator. Second, the model provides a process that makes possible detailed analysis of the statistical properties of the recreation use estimator. Third, the model enables a researcher to isolate and describe the effects of different assumptions about one or more visitor behaviors on the properties of the recreation use estimator. Finally, the model can provide guidance on the effects of scheduling alternatives for the amount and timing of on-site interviewing on both the quality of the recreation use estimate and the volume of visitor characteristic data collected.

The specific objectives of this paper are to (1) describe a simulation model (ORUS) for outdoor recreation use estimation, (2) parameterize ORUS to the first NVUM survey cycle data, (3) illustrate the functioning of ORUS with a simple example, and (4) demonstrate the evaluation of the recreation use estimator under various theoretical scenarios.

The National Visitor Use Monitoring Sampling Design

The NVUM program performs surveys that collect data about visitors on the U.S. national forests. The survey employs a stratified multistage sampling design based on rotating panels that are spread over a 5-year sampling cycle. All national forests in the United States are sampled once every 5 years, with approximately one-fifth of the forests in each of nine regions sampled each year. This statistical methodology follows conventional sample survey techniques (Cochran 1977) with a few modifications to incorporate specific situations inherent in sampling of national forests for recreation use.

The NVUM sampling design divides each national forest into areas that are called site types, each of which contains a multitude of individual sites exhibiting similar recreational attributes. Four mutually exclusive site types served as stratification variables for reducing variation in the survey's estimates. These sites types are:

- Day-use developed sites (DUDS)—sites intended for day use only, including boating areas, picnic sites, fish-viewing sites, fishing sites, information sites, interpretive sites, observation sites, playground-park sport sites, ski areas (alpine and Nordic), wildlife viewing areas, visitor centers, museums, swimming areas, and winter sports areas; generally, DUDS provide visitor comfort, convenience, and educational opportunities, but they are available only on a day-use basis

- Overnight-use developed sites (OUDS)—include campgrounds, cabins, hotels, lodges, resorts, horse camps, organization sites, and any other overnight facility on national forest lands, whether they are owned or managed by the Forest Service or are private concessions

- Wilderness sites (WILD)—sites that are designated official wilderness areas

- General forest area (GFA)—all national forest sites that are not designated as DUDS, OUDS, or WILD

Since recreationists' behavior could vary by site type and thus exert differential effects on the recreation use estimates, the simulation model was developed to exhibit the unique behavior associated with each of these four site types. For the first survey cycle, there were approximately 2,700 DUDS, 5,000 OUDS, 3,200 WILD, and 16,500 GFA individual sites distributed throughout the national forests of the United States. DUDS were open for recreation use for ~528,000 site days annually, OUDS for ~668,000 site days annually, WILD sites for ~516,000 site days annually, and GFA sites for ~3,368,000 site days annually.

In the NVUM survey the primary measurement variable is the number of recreationists who were completing a visit to a given site on a given day. These are termed last-exiting recreationists, and this term distinguishes them from recreation visitors who are making intermediate (non-final) exits and who will return to the site. An exact value for last-exiting recreationists would be obtained if each site were monitored on an around-the-clock basis and if all visitors exiting each site were required to participate in an on-site interview. Such a protocol is not possible for several obvious reasons. Consequently, NVUM uses methods that estimate the measurement variable indirectly. A 24-hour mechanical count of all traffic is obtained along with 6 hours of vehicle occupant interviewing and exiting vehicle counts at a designated interview point traversed by visitors exiting the site or area of the forest. This process obtains (1) a ratio of observed exiting vehicles to the 6-hour mechanical vehicle count which is then used to calibrate the 24-hour mechanical vehicle count, yielding an estimate of total exiting vehicles for the 24-hour period (*VEHC*); (2) an estimate of the proportion of exiting vehicles that carry last-exiting recreationists (*PBAR*); and (3) the average number of people in a last-exiting recreation vehicle (*PEOPVEH*).[1] These three values are used to estimate recreation use at the site for 24 hours.

The NVUM site visit estimator for recreation use is defined as

$$\widehat{SV} = PBAR*VEHC*PEOPVEH \tag{1}$$

where

$\widehat{SV}$ = the site visit estimator defined as the number of last-exiting recreationists on a given site for the entire 24-hour day

PBAR = the proportion of last-exiting recreationists estimated from a 6-hour interview survey

VEHC = the calibrated number of vehicles exiting the site for the entire 24-hour day

PEOPVEH = the average number of people in a last-exiting recreation vehicle estimated from the 6-hour interview survey

[1] A recreation vehicle refers to any type of vehicle that contains recreationists and does not refer only to the typical outdoor recreation vehicle.

It should be noted that $\widehat{SV}$ is based on the number of last-exiting recreationists but that it could have also been based on first-entering recreationists. In either case, this eliminates multiple counting of recreationists who exit and re-enter the site. When site visits are combined over all sampled days, an annual estimate of site visits is obtained. For more details on this and other aspects of the NVUM methodology, see English and others (2002).

The accuracy of $\widehat{SV}$ depends on how well each of the three components in equation (1) is estimated. *PEOPVEH* is obtained by simply counting occupants in last-exiting recreation vehicles. Obtaining a sufficient sample size for accurate estimation of *PEOPVEH* is usually not problematic, as the range of values observed is relatively small. The accuracy of *VEHC* depends largely on the consistent performance of the mechanical traffic counter over the 24-hour period and a relatively constant ratio of exiting to total traffic. The focus of this paper is on simulating and evaluating the effect of *PBAR* on $\widehat{SV}$. *PBAR* is a complex variable that is highly dependent on several aspects of visitor behavior at the recreation site.

Model Components

Types of Site Visitors

The first distinction in types of visitors at a site is between those who are there for recreation and those who are there for some non-recreation purpose (NREC). The latter may include agency personnel or contractors who are working, and people who have stopped to use restroom facilities, to obtain information, or for other reasons. Although NREC visitors may be important in some situations and ORUS is capable of incorporating them into realistic simulations, they are not discussed further in this paper. The model emphasizes the typical outdoor recreationist, and thus recognizes four distinct types of recreationists who may be at a site on a given site day. The typology is based on their specific behavior patterns of arriving time, last-exiting time, and intermediate (non-final) exits from the site. These types are defined as follows:

- LERB[2] = a recreationist who will be last-exiting the site on that site day and was at the site before the official beginning of the site day at midnight
- LERD = a recreationist who will be last-exiting the site on that site day and arrived on the site during the site day
- NLERB = a recreationist who will not be last-exiting the site on that site day and was on the site before the official beginning of the site day at midnight
- NLERD = a recreationist who will not be last-exiting the site on that site day and arrived on the site during the site day

The distinction between LERBs and LERDs deals mainly with the amount of time on the site. LERBs are generally on site longer and are likely to have more intermediate exits. A similar distinction is made between NLERBs and NLERDs. The four types of recreationists could have similar or different arriving or last-exiting distribution parameters and intermediate exit rates as will be explained in the next sections.

A DUDS site will usually only have recreationists of the LERD type, although it is possible that a few recreationists may stay overnight. On the other hand, an OUDS site typically has LERBs, NLERBs, and NLERDs but few LERDs. However, the simulator is flexible enough to allow occasional day use in an OUDS if this is specified in the simulation. Other special considerations can be accommodated by the simulator. For example, an OUDS may only contain LERBs on a Sunday if it is assumed that all campers will exit and return home on Sunday night to begin the work week on Monday. The GFA and WILD sites may contain all four types if both overnight camping and day use are possible. In addition, for site days where multiple recreation types are permissible, varying proportions of the types may be allocated to mimic a realistic situation and the parameters that control their individual behavior patterns need not be the same.

Arriving and Last-Exiting Times

The fundamental behavior for recreationists involves arriving at the site, engaging in recreation, and then exiting the site. The distributions of these actions relative to interview times are key elements of the simulation model. Arriving and last-exiting times for recreationists are modeled using the beta distribution, which is defined as

$$f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \qquad (2)$$

where

$a > 0$, $b > 0$, and $0 \le p \le 1$.

The mean of this distribution is $\frac{a}{a+b}$ and the variance is $\frac{ab}{(a+b)^2(a+b+1)}$. The beta distribution takes on a wide variety of shapes depending on its parameters a and b. For

[2] The acronyms for the types of site visitors are formed by a combination of (LER or NLER) and (B or D) where LER represents "last-exiting recreationist," NLER represents "nonlast-exiting recreationist," B represents "before," and D represents "during."

instance, the uniform distribution is a special case of the beta when $a = b = 1$ with a mean of 0.50. If $a = 1$ and $b = 5$ then the beta is skewed to the right with a hump in the left of the distribution and, consequently, a mean of 0.17. On the other hand, if $a = 5$ and $b = 1$ then the opposite is true with a mean of 0.83. A symmetric bell-shaped distribution with a mean of 0.5 occurs when $a = b = 5$. If a and b are both < 1 then a u-shaped distribution results. Figure 1 shows the beta distribution for several values of the parameters.

The arriving time (AT) of a recreationist is determined by selecting a random variate p_1 from the specified beta distribution and determining the arriving time as that proportion of the recreation day after the start of the recreation day. Mathematically, for LERD and NLERD this is

$$AT = D_S + p_1(D_E - D_S) \qquad (3)$$

where

D_S = time when the recreation day starts[3]

D_E = time when the recreation day ends

Since LERB and NLERB recreationists are on the site previous to the site day, they have no arriving time for that site day. Thus, if the recreation day is assumed to be from 6.00 to 18.00 (6:00 a.m. to 6:00 p.m.) with a uniform distribution of arriving times ($a = b = 1$) then arrivals occur at random throughout the day and the mean arriving time is

[3] Times used throughout this paper are represented as real numbers for computational purposes. The hour component is analogous to standard military time while the minute component represents the decimal part of the hour. Thus, 6:30 a.m. is represented as 6.50 while 4:15 p.m. is represented as 16.25.

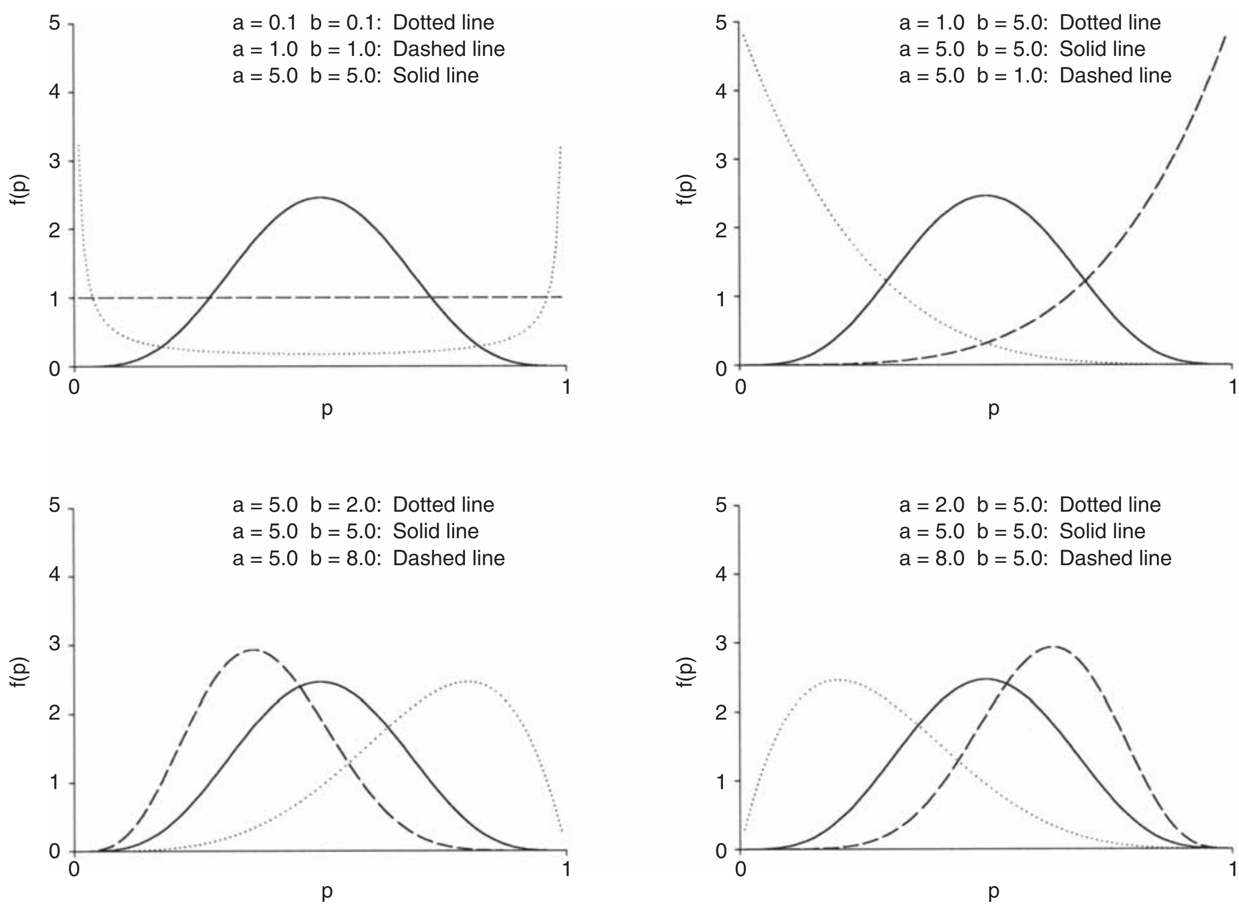

Figure 1—The beta distribution f(p) for various values of the parameters a and b.

12.00 (12:00 noon). Obviously, if morning arrivals or afternoon arrivals are more typical, the beta distribution's parameters could be altered to reflect this situation.

Another variate, p_2, is selected from the beta distribution and is used to calculate the last-exiting time of a recreationist. For LERBs, last-exiting time is defined as

$$LET = D_S + p_2(D_E - D_S) \tag{4}$$

and for LERDs as

$$LET = AT + p_2(D_E - AT) \tag{5}$$

Since D_S and D_E are fixed parameters specified in the simulator depending on the characteristics of the site under study, the arriving time variate p_1 in equation (3) and the last-exiting time variate p_2 in equation (4) are selected from the beta distribution specified in the simulation for the given site. However, for the LERD situation in equation (5), each visitor usually has a unique arriving time and it is conceivable that different beta distributions of last-exiting times may result for individuals. The simulator allows for the beta parameters to be adjusted with a regression model. Since NLERB and NLERD recreationists do not exit the site on the site day, they have no last-exiting time. Thus, NLERB has neither arriving time nor last-exiting time beta distributions.

The mean arriving time, also known as the mathematical expectation E, for LERD and NLERD for a given beta distribution is easily found from equation (3) as

$$E(AT) = D_S + E\{p_1(D_E - D_S)\} = D_S + (D_E - D_S)\left(\frac{a_1}{a_1 + b_1}\right) \tag{6}$$

The mean last-exiting time for LERB from equation (4) is

$$E(LET) = D_S + E\{p_2(D_E - D_S)\} = D_S + (D_E - D_S)\left(\frac{a_2}{a_2 + b_2}\right) \tag{7}$$

The mean last-exiting time for LERD from equation (5) is a little more difficult to obtain. Let

$$E(LET) = E(AT) + D_E E(p_2) - E\{p_2(AT)\} \tag{8}$$

Taking first the component on the far right of equation (8) and substituting AT from equation (3) we have

$$\begin{aligned} E\{p_2(AT)\} &= E\{p_2(D_S + p_1(D_E - D_S))\} = \\ &D_S E(p_2) + (D_E - D_S)E(p_1 p_2) \end{aligned} \tag{9}$$

Note that if p_1 and p_2 are independent betas (which may not be a valid assumption in some instances) then

$$Cov(p_1, p_2) = E(p_1 p_2) - E(p_1)E(p_2) = 0 \tag{10}$$

This implies that

$$E(p_1 p_2) = E(p_1)E(p_2) \tag{11}$$

Thus, substituting equations (6), (9), and (11) into equation (8) we get for LERD

$$\begin{aligned} E(LET) = D_S + (D_E - D_S)\Bigg(&\frac{a_1}{a_1 + b_1} + \frac{a_2}{a_2 + b_2} - \\ &\left(\frac{a_1}{a_1 + b_1}\right)\left(\frac{a_2}{a_2 + b_2}\right)\Bigg) \end{aligned} \tag{12}$$

Equations (6), (7), and (12) estimate the theoretical mean arriving and last-exiting times based on the parameters of the appropriate beta distributions.

Number of Intermediate Exits

Some visitors will make intermediate exits from the site before completing their recreation visit. Intermediate exits are defined as an exit and re-entry into the recreation site on the same day. Intermediate exits affect the accuracy of the *PBAR* estimate through their relation to the probability that an individual will be surveyed and to trap shyness (to be discussed later). The number of intermediate exits a recreationist performs during the site day is modeled with the Poisson distribution which assumes that they occur at random throughout the day. The Poisson distribution is defined as

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{13}$$

where

$x = 0,1,2,3,...$ and $\lambda > 0$.

The mean and variance of the Poisson are both λ. The parameter λ represents the intermediate exit rate of a recreationist for the length of an active recreation day, $D_E - D_S$, during which it is assumed that such exits are possible. This could be assigned to be the total 24-hour day if it is believed that intermediate exits occur throughout this time span. However, more realistically, these exits are usually from around a little before dawn to somewhat after dusk, which would encompass at most 15 hours. The simulator provides for such flexibility by defining λ as the intermediate exit rate only during the assumed active recreation day defined for that specific site day. All individuals within a recreationist type have the same λ, but this λ is adjusted by the length of stay for each recreationist at the site. For instance, an NLERD entering a site at 6.00 has more time that day for intermediate exits than one that enters at 12.00 because neither will complete a visit that day. The rate is adjusted based on the proportion of the active recreation day that is available to a recreationist. Thus, a recreationist who is there only a third of the active recreation day will have the parameter set at $\lambda/3$ and the number of intermediate exits will be selected from a Poisson distribution with this parameter. Specifically, for LERBs this is

$$\lambda_{LERB} = \left(\frac{LET - D_S}{D_E - D_S} \right) \lambda \qquad (14)$$

for LERDs this is

$$\lambda_{LERD} = \left(\frac{LET - AT}{D_E - D_S} \right) \lambda \qquad (15)$$

for NLERBs this is

$$\lambda_{NLERB} = \left(\frac{D_E - D_S}{D_E - D_S} \right) \lambda = \lambda \qquad (16)$$

and for NLERDs this is

$$\lambda_{NLERD} = \left(\frac{D_E - AT}{D_E - D_S} \right) \lambda \qquad (17)$$

It is important to note that λ controls only intermediate exits and does not include exits for the last time. Last exits are controlled by the last-exiting beta distribution. It is assumed that last exits are a special movement and are dependent on the recreationist type. It is possible to assume that the intermediate exit rate is the same for all recreationist types or it may vary depending on the parameters chosen for each.

Time of Intermediate Exits

The specific times of intermediate exits are selected at random from the total length of stay that a recreationist has for the site day. This appears to be a reasonable assumption because each recreationist is unique and has intermediate exit behavior that is nearly impossible to predict. Some may wander off the site as soon as they get there merely to see what's around the next bend. Others may go out to the store only to immediately leave again when they find that they forgot to get an important item. Still others may never leave the site until they depart for home. The total length of stay for the site day is defined by the arriving time and last-exiting time. The number of intermediate exits is then used to randomly select a time for each exit. The uniform distribution is applied to the total length of stay for the site day to generate these variates.

Trap Shyness

The estimation of *PBAR* used for the site visit estimator, $\widehat{SV}$, is based on the assumption that interviewed vehicles are selected at random from those passing over the vehicle counter. Unfortunately, stopping to be interviewed is optional. Forest Service regulations usually prohibit mandatory traffic stops. Thus, some exiting individuals may choose not to be interviewed. The probability that a recreationist stops for an interview may very well depend on the recreationist's previous history of being stopped on that site day. For instance, the probability that a recreationist stops for an initial interview may be 0.9. However, after being interviewed that day on an intermediate exit, the recreationist may not be so eager to be interviewed again and the probability may drop to 0.1. This phenomenon is commonly known as trap shyness, a term that originated in animal studies in which trapped animals learn to avoid traps after they are captured once. Thus, trap shyness by the recreationists will change the probability of being interviewed and invalidate the random sample needed for an unbiased estimate of *PBAR*.

Although an infinite number of trap shyness behaviors could be modeled, only three will be discussed here. First, the not trap-shy situation is defined as

$$P_0 = P_1 = P_2 = \ldots \qquad (18)$$

where

P_i = probability that a recreationist will stop to be interviewed given i previous interviews on that site day.

In this situation, all probabilities are equal. For a mild degree of trap shyness, the probabilities diminish by half after the first interview:

$$P_0 \to P_1 = \frac{P_0}{2} \to P_2 = \frac{P_0}{2} = \ldots \qquad (19)$$

In extreme trap shyness there is zero probability that there will be another interview after the first one:

$$P_0 \to P_1 = 0 \qquad (20)$$

The opposite situation, trap addiction, may also occur when being interviewed is a very good experience to the recreationist. For instance, if a reward of some type is given after the interview process, the recreationist may become more eager to be interviewed again. Any level of trap shyness or addiction can be incorporated into the model.

Methods

The *PBAR* estimator for the proportion of last-exiting recreationists that exit from a site is defined as

$$\widehat{PBAR} = \frac{LC_{11}}{LC_{11} + LC_{01}} \tag{21}$$

where

LC_{11} = number of last-exiting recreationists captured in the interview process

LC_{01} = number of non-last-exiting recreationists that were captured

This could be computed from the data produced by the ORUS model under a specific scenario. The true proportion of last-exiting recreationists could also be computed as

$$PBAR = \frac{LC_{10} + LC_{11}}{LC_{10} + LC_{11} + LC_{00} + LC_{01}} \tag{22}$$

where

LC_{10} = number of last-exiting recreationists that were not captured

LC_{00} = number of non-last-exiting recreationists that were not captured

Comparison of the estimated *PBAR* to the true *PBAR* for a given simulation scenario reveals the quality of the site visit estimator, $\widehat{SV}$. However, since comparisons from only one simulation are difficult to judge because the simulated values are stochastic, 10,000 simulations were performed. An estimate of the bias results from averaging the differences for the 10,000 simulations. Thus, the percent bias is defined as

$$\begin{aligned} Percent\ Bias &= \frac{100}{10000} \sum_{i=1}^{10000} \frac{\widehat{SV_i} - SV_i}{SV_i} = \\ &\frac{100}{10000} \sum_{i=1}^{10000} \frac{\widehat{PBAR_i} - PBAR_i}{PBAR_i} \end{aligned} \tag{23}$$

where

$\widehat{SV_i}$ is defined in equation (1) and

$$VEHC = LC_{10} + LC_{11} + LC_{00} + LC_{01} \tag{24}$$

Note that *VEHC* and *PEOPVEH* are fixed constants in the simulator and cancel out from equation (23).

Although $\widehat{SV}$ could be evaluated under hypothetical beta distributions, it is more realistic and useful to fit the beta distributions to the NVUM sampled survey data. Estimators for the *a* and *b* parameters of the beta distribution were obtained by using the method of moments. This equates to simply setting the sample mean, $\overline{X}$, and variance, S^2, equal to the theoretical beta mean and variance, respectively. This yields two equations and two unknowns, which are easily solved for *b* and then *a*, yielding

$$\hat{b} = \frac{\left(1-\overline{X}\right)\left[\overline{X}\left(1-\overline{X}\right) - S^2\right]}{S^2} \tag{25}$$

and

$$\hat{a} = \frac{\hat{b}\overline{X}}{\left(1-\overline{X}\right)} \tag{26}$$

Arrival times for recreation visitors were obtained from the NVUM survey and beta distributions fitted for LERDs and NLERDs for each of the four site types. Similarly, last-exiting times were used to fit beta distributions to LERBs and LERDs for the site types.

The beta distribution of last-exiting times for LERD recreationists was presumed to be dependent on the arriving time of an individual. Thus, the beta parameters in this situation were not estimated based on the sampled $\overline{X}$ and S^2. Instead, two linear regression models were used to predict $\overline{X}$ and S^2 as functions of arriving time. Thus, each individual recreationist with a unique arriving time had a unique predicted $\overline{X}$ and S^2 which were used in equations (25) and (26) to predict the individual's beta parameters.

Results

Parameterization of ORUS to the NVUM Data

ORUS could be used to evaluate outdoor recreation surveys and their estimators under any hypothetical scenario or assumed distribution of arriving and last-exiting times. However, to illustrate how ORUS could be tailored to fit a specific set of recreation sites, the arriving and last-exiting beta distributions were parameterized to the complete set of NVUM data collected over the first two sampling years. It was assumed that on-site recreation could occur from 6.00 to 21.00 hours, so the beta distributions are based on this recreation day length.

Although there are 4 site types, 4 visitor types, and 2 types of beta distributions (arriving and last-exiting times), there are only 16 beta distributions to parameterize. The LERB recreationists have only a distribution of last-exiting times for a given survey day. The LERDs have both arriving and last-exiting beta distributions. The NLERBs neither enter nor exit during the survey day, so they have no beta distributions to parameterize. Since the NLERDs only enter and do not exit, they have only arriving distributions. The parameter estimates along with the sample size and mean and variance of the distributions are shown in table 1. Distributions for the developed site types (DUDS and OUDS) are graphed in figure 2 and for dispersed site types (WILD and GFA) in figure 3. Generally, the distributions for the DUDS appeared to be quite similar to those for OUDS, and distributions for WILD resembled those for GFA.

Table 1—Parameter estimates for the beta distributions of arriving and last-exiting times pooled over the NVUM survey data from all regions and forests

Visitor type	Movement	DUDS	OUDS	WILD	GFA
LERB	Arriving	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined
LERB	Last-exiting	a = 3.053 b = 3.488 n = 789 Mean = 0.467 Variance = 0.033	a = 3.557 b = 6.112 n = 3727 Mean = 0.368 Variance = 0.022	a = 4.208 b = 4.462 n = 1088 Mean = 0.485 Variance = 0.026	a = 3.694 b = 4.150 c = 1322 Mean = 0.471 Variance = 0.028
LERD	Arriving	a = 2.987 b = 4.119 n = 11365 Mean = 0.420 Variance = 0.030	a = 2.581 b = 3.933 n = 1264 Mean = 0.396 Variance = 0.032	a = 2.196 b = 4.383 n = 3302 Mean = 0.334 Variance = 0.029	a = 1.602 b = 3.422 n = 10822 Mean = 0.319 Variance = 0.036
LERD	Last-exiting[a]	a = 0.903 b = 3.514 n = 10701 Mean = 0.204 Variance = 0.030	a = 1.077 b = 3.262 n = 1184 Mean = 0.249 Variance = 0.035	a = 1.890 b = 3.574 n = 3220 Mean = 0.346 Variance = 0.035	a = 2.022 b = 4.402 n = 10822 Mean = 0.313 Variance = 0.029
NLERB	Arriving	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined
NLERB	Last-exiting	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined
NLERD	Arriving	a = 1.840 b = 1.541 n = 364 Mean = 0.544 Variance = 0.057	a = 2.410 b = 1.803 n = 3618 Mean = 0.572 Variance = 0.047	a = 1.458 b = 1.859 n = 971 Mean = 0.440 Variance = 0.057	a = 1.520 b = 1.687 n = 1240 Mean = 0.474 Variance = 0.059
NLERD	Last-exiting	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined	a = not defined b = not defined

NVUM = National Visitor Use Monitoring; DUDS = day-use developed sites; OUDS = overnight-use developed sites; WILD = wilderness sites; GFA = general forest area; LERB = a recreationist who will be last-exiting the site on that site day and was at the site before the official beginning of the site day at midnight; LERD = a recreationist who will be last-exiting the site on that site day and arrived on the site during the site day; NLERB = a recreationist who will not be last-exiting the site on that site day and was on the site before the official beginning of the site day at midnight; NLERD = a recreationist who will not be last-exiting the site on that site day and arrived on the site during the site day.

[a] These parameters require regression models to predict the mean and variance as a quadratic function of the arriving time of a recreationist. The values for the LERD last-exiting parameters across the site types in this row are averages because they change depending on the specific arriving times that are simulated.

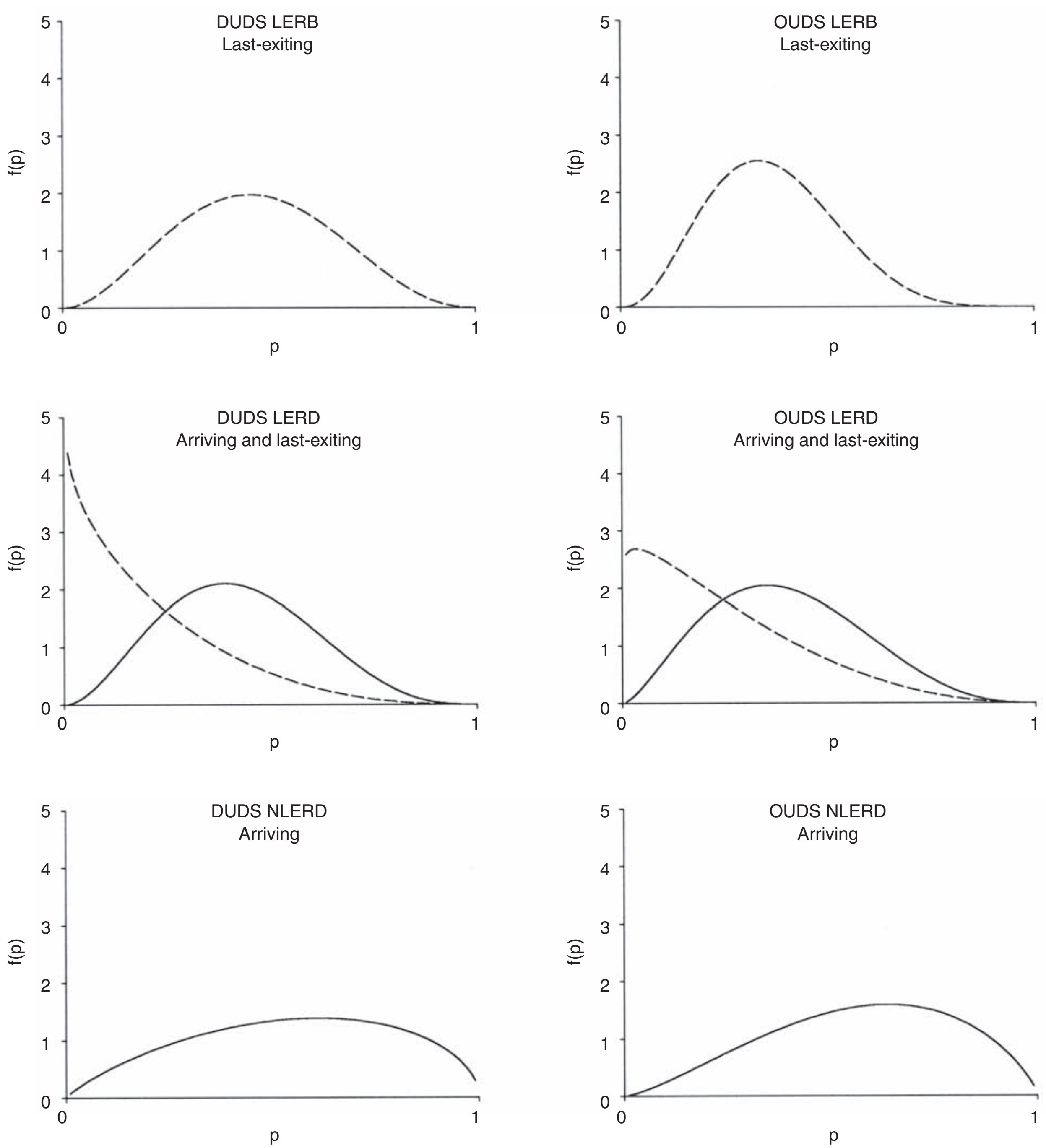

Figure 2—The beta distribution f(p) of arriving and last-exiting times for day-use developed sites (DUDS) and overnight-use developed sites (OUDS) where solid lines are arriving distributions and dashed lines are last-exiting distributions; (LERB = a recreationist who will be last-exiting the site on that site day and was at the site before the official beginning of the site day at midnight, LERD = a recreationist who will be last-exiting the site on that site day and arrived on the site during the site day, NLERD = a recreationist who will not be last-exiting the site on that site day and arrived on the site during the site day).

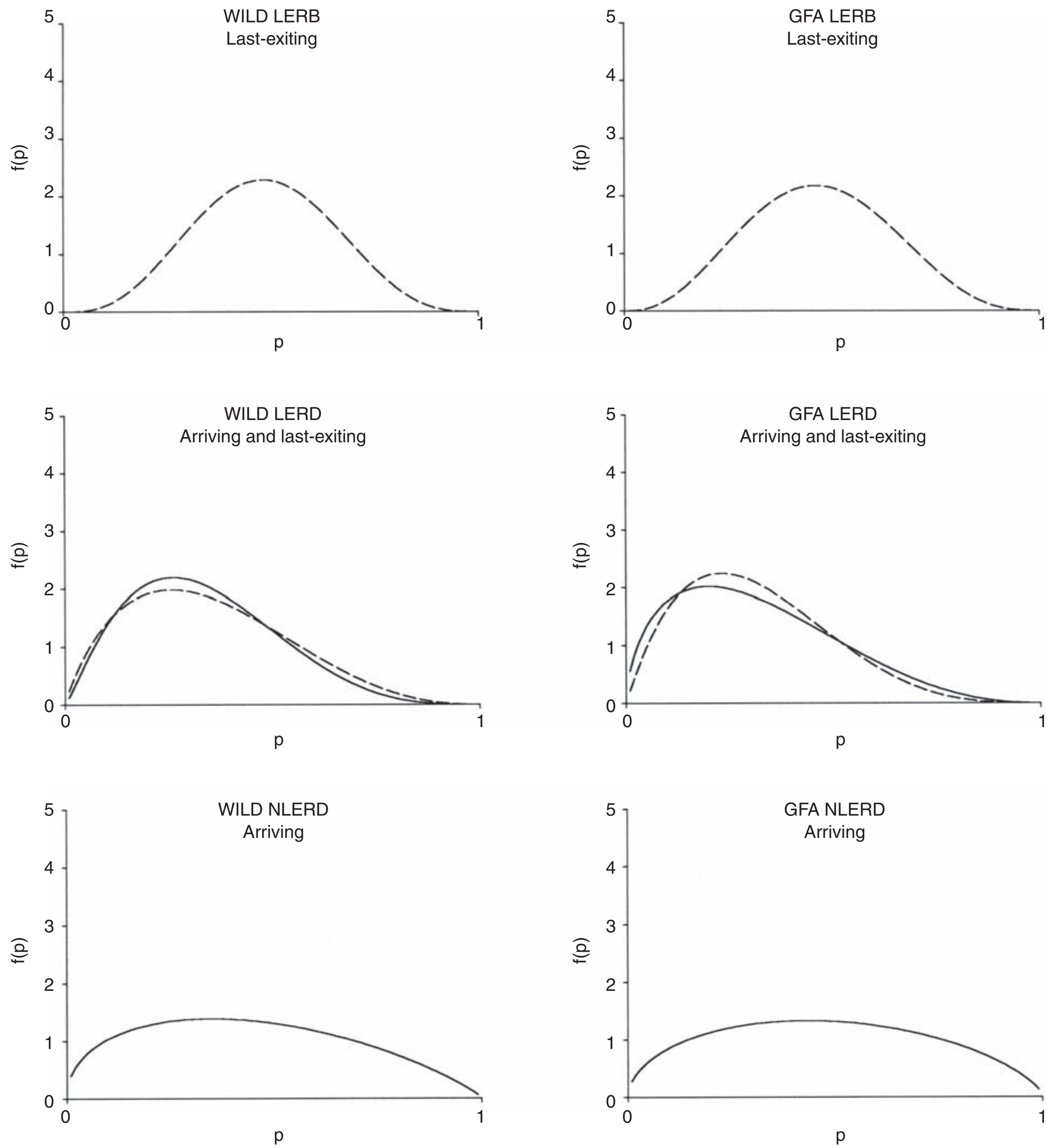

Figure 3—The beta distribution f(p) of arriving and last-exiting times for wilderness sites (WILD) and general forest area sites (GFA) where solid lines are arriving distributions and dashed lines are last-exiting distributions; (LERB = a recreationist who will be last-exiting the site on that site day and was at the site before the official beginning of the site day at midnight, LERD = a recreationist who will be last-exiting the site on that site day and arrived on the site during the site day, NLERD = a recreationist who will not be last-exiting the site on that site day and arrived on the site during the site day).

The LERB recreationists last exited earliest from the OUDS sites, with a mean last-exiting beta variate of 0.368 [11.52 hours computed with equation (7)] and with a distribution slightly skewed to the right. Beta distributions for the other site types were more symmetric throughout the day. WILD had the largest mean last-exiting beta variate of 0.485 (13.28 hours) while values for both DUDS with 0.467 (13.00 hours) and GFA with 0.471 (13.06 hours) were just slightly less. Although logically it would appear that LERB recreationists should not be on DUDS sites because overnight use is not defined on this site type, there were occasional DUDS LERB recreationists in the NVUM database, so this distribution was fitted to accommodate such situations.

LERD recreationists arrived earliest on the GFA sites with a mean beta variate of 0.319 [10.78 hours computed with equation (6)] and latest on the DUDS with a mean beta variate of 0.420 (12.30 hours). Values for the other site types were intermediate with those for WILD being 0.334 (11.01 hours) and OUDS being 0.396 (11.94 hours). For all site types, and especially for WILD and GFA, the distributions were skewed to the right, indicating a tendency for most of these one-day visitors to come early in the day. The last-exiting distributions show two contrasting relationships. Many DUDS and OUDS recreationists tend to depart soon after they arrive, as can be seen by their exponential type patterns. In contrast, most of the WILD and GFA recreationists do not leave as quickly, but tend to exhibit a more symmetric distribution but with a strong skew to the right. The mean last-exiting beta variates were 0.204 [14.07 hours computed with equation (12)] for DUDS, 0.249 (14.20 hours) for OUDS, 0.346 (14.47 hours) for WILD, and 0.313 (13.98 hours) for GFA. Thus, the mean visit duration is simply the difference between $E(LET)$ and $E(AT)$. Performing the calculations reveals that for LERD recreationists the visit duration at DUDS was 1.77 hours, OUDS was 2.26 hours, WILD was 3.46 hours, and GFA was 3.20 hours. It must be emphasized that these visit durations are for LERD recreationists who do not spend the night at the site. Obviously, overnight recreationists such as NLERBs and NLERDs would have a much longer visit duration. It is interesting that for the developed site types (DUDS and OUDS) a LERD recreationist (who by definition makes only a day visit) stays approximately 2 hours while on the dispersed site types (WILD and GFA) the visit duration is over 3 hours. The additional hour may be due to the more remote access to the dispersed sites or to a desire to achieve a longer, more intense outdoor recreation experience or to both of these causes.

The NLERD recreationists at DUDS and OUDS sites arrived an average of 1.51 hours later than did those at the WILD and GFA sites. The DUDS had a mean arriving beta variate of 0.544 [14.16 hours computed with equation (6)] and that for OUDS was 0.572 (14.58 hours) while those for WILD and GFA were 0.440 (12.60 hours) and 0.474 (13.11 hours), respectively. It is interesting that for all site types the NLERD distributions do not display the skewed, asymmetrical bell shape of the LERD distributions.

Simple Simulation Example

A simple example illustrates the ORUS model's capabilities. The assumptions for this scenario site day are:

1. a DUDS site, open from 6.00 until 21.00 hours with LERD = 10 recreationists
2. the interviewer was on site from 8.00 until 14.00 hours
3. the daily rate of intermediate exits was set high at $\lambda = 4$
4. arriving times and last-exiting times were selected from the beta distributions that were parameterized with the NVUM data (table 1)
5. to illustrate the effect of trap shyness, the probability of capture on a recreationist's first exit was set at 1.0 and then set at 0.0 for any subsequent exits, including the last

These assumptions imply that an average recreationist on site for the total 15-hour active recreation day would make four intermediate exits in addition to a last exit. Since most do not stay for 15 hours, the observed number of intermediate exits is considerably less.

Results from this scenario site day are shown in table 2. There were a total of 17 exits from the site during the 15-hour day, 10 of which were obviously last exiting. Only 2 of the 10 last-exiting recreationists were captured. Seven last exited the site after the interviewers left at 14.00 hours and thus could not be captured. The other one was interviewed first during an intermediate exit and trap shyness precluded that individual from being interviewed on the final exit from the site. A total of five recreationists were stopped by the interviewers. Thus, an estimate of $PBAR$ from equation (21) is $\widehat{PBAR} = 2/5 = 0.40$. The true proportion is $PBAR = 10/17 = 0.59$ computed from equation (22). This estimate of $PBAR$ results in poor estimates for recreation use on this site. Assuming that the vehicle counter correctly recorded 17 exiting vehicles for the 24-hour period and that there was an average of 1 person per vehicle (for simplicity), the SV estimate would be $\widehat{SV} = 0.40(17)(1) = 6.8$ while the true SV would be $SV = 0.59(17)(1) = 10.0$. This would represent a negative 32-percent bias.

Although the model has simulated data in accordance with the specified probability distributions, it is interesting to compare the results to what is expected theoretically. From equation (6) the theoretical mean arriving time is calculated as

$$E(AT) = 6 + (21 - 6)(0.420) = 12.30$$

Table 2—Simulation of a DUDS site day where the recreation day was from 6.00 to 21.00 hours, the interview period was from 8.00 to 14.00 hours, λ = 4, and P_0 = 1.0 and P_1 = 0.0; there were 10 LERD recreationists on this site, *PBAR* = 0.59, and $\widehat{PBAR}$ = 0.40

Visitor	Arrived	Last exit	Exited	Captured	Remarks
1	12.23	Yes	14.44	No	Not captured because left after interviewers
2	16.87	Yes	18.37	No	Not captured because left after interviewers
3	9.20	Yes	10.17	Yes	Captured because P_0 = 1.0
4	14.16	Yes	15.11	No	Not captured because left after interviewers
5	9.47	No	9.89	Yes	Captured because P_0 = 1.0
5		No	12.48	No	Not captured because became trap shy
5		Yes	12.60	No	Not captured because became trap shy
6	10.88	Yes	10.91	Yes	Captured because P_0 = 1.0
7	15.27	Yes	16.43	No	Not captured because left after interviewers
8	10.61	No	11.27	Yes	Captured because P_0 = 1.0
8		No	14.69	No	Not captured because left after interviewers and also became trap shy
8		Yes	15.33	No	Not captured because left after interviewers and also became trap shy
9	14.71	Yes	15.19	No	Not captured because left after interviewers
10	11.79	No	12.54	Yes	Captured because P_0 = 1.0
10		No	12.56	No	Not captured because became trap shy
10		No	12.88	No	Not captured because became trap shy
10		Yes	14.50	No	Not captured because left after interviewers and also became trap shy

DUDS = day-use developed sites; LERD = a recreationist who will be last-exiting the site on that site day and arrived on the site during the site day.

which compares quite well to the mean arriving time of 12.52 computed from table 2. Similarly, the theoretical mean last-exiting time is calculated from equation (12) as

$$E(LET) = 6 + (21 - 6)[0.420 + 0.204 - 0.420(0.204)] = 14.07$$

and from table 2 the mean last-exiting time is 14.30. This close agreement between the theoretical and simulated data indicates that the simulator is reflecting the properties that were designed into it. The slight departure is due merely to the simulated sampling error inherent in the statistical models. If this example were replicated many times and the arriving and last-exiting times computed and averaged, they would converge to the theoretical means as the number of simulations increased.

Thus, if a survey crew is sampling from 8.00 to 14.00 hours as in this site, what are the consequences if their sampling stops at or before the mean last-exiting time for the recreationists? In this one simulated example, the data reveal that only 3 of the 10 recreationists last exited before the survey crew was finished interviewing, one of whom was already captured and had become trap shy. Thus there were only five interviews, and only two of those interviewed were last-exiting recreationists. Since very few recreationists were interviewed, it may be more efficient to perform the survey later in the day.

Although mean number of intermediate exits was set at λ = 4 for this simulation example, no recreationist had four intermediate exits. The largest was for recreationist 10, who had three. The observed number of intermediate exits is reasonable because most visitors were on site far less than the total 15 hours. The mean visit duration is computed as the difference between $E(LET)$ and $E(AT)$, which is $14.07 - 12.30 = 1.77$, implying that the average number of intermediate exits is calculated from equation (15) as $(1.77/15)4 = 0.47$. The data from table 2 reveal that there was an average of 0.70 intermediate exits, which is close to the theoretical value of 0.47.

Estimator Evaluation

An evaluation of the site visit estimator $\widehat{SV}$ under several scenarios was performed for DUDS sites. The recreation day was set at 6.00 to 21.00 hours with LERD = 10 recreationists and an interview window of 8.00 to 14.00 hours. The arriving and last-exiting beta distributions parameterized with the NVUM survey data were used. With these as the base parameters, evaluations were performed to determine the effect of trap shyness, intermediate exit frequency, and length of interview period. Each scenario is based on 10,000 replicated simulations and the results are shown in table 3. For simplicity, the results are based on $\widehat{SV}$ assuming that *PEOPVEH* = 1.

Scenarios A and B show the effect of trap shyness with a low level of intermediate exits ($\lambda = 1$). In scenario A

Table 3—Results from several scenarios for a DUDS site where the recreation day was from 6.00 to 21.00 hours with 10 LERD recreationists

Simulation scenario	Interview times	λ	$P_0, P_1, P_2, \ldots$	n[a]	$\widehat{SV}$	Percent bias	CV
	hours						
A	8.00 14.00	1	1, 1, 1, . . .	9,993	9.70	-3.04	11.9
B	8.00 14.00	1	1, 0	9,990	9.57	-4.30	14.6
C	8.00 14.00	4	1, 1, 1, . . .	9,997	9.13	-8.74	19.6
D	8.00 14.00	4	1, 0	9,997	8.79	-12.07	28.1
E	6.00 8.00	1	1, 1, 1, . . .	359	6.99	-30.11	75.4
F	6.00 10.00	1	1, 1, 1, . . .	4,541	8.84	-11.64	45.4
G	6.00 12.00	1	1, 1, 1, . . .	9,370	9.44	-5.60	25.0
H	12.00 18.00	1	1, 1, 1, . . .	10,000	10.17	1.73	7.4
I	14.00 18.00	1	1, 1, 1, . . .	9,978	10.36	3.59	11.7
J	16.00 18.00	1	1, 1, 1, . . .	8,652	10.59	5.94	17.1

DUDS = day-use developed sites; LERD = a recreationist who will be last-exiting the site on that site day and arrived on the site during the site day.

[a] Although the specified target number of simulations for a given scenario was 10,000, not all simulations gave results because some did not capture any recreationists for a given simulation, which resulted in no estimate. Thus, n is the number of simulations that were actually performed for a given scenario that gave valid estimates.

there was no trap shyness and in B there was extreme trap shyness. Both scenarios yielded estimates of *SV* that differed slightly from the true *SV* = 10 with A having a percent bias of -3.04 and B of -4.30. Here, trap shyness negatively affected estimator performance, but only slightly because the bias was small. Trap shyness was also examined in scenarios C (none) and D (extreme) but with an elevated intermediate exit rate ($\lambda = 4$). Results showed a percent bias of -8.74 for C and -12.07 for D, which illustrates that increasing the number of intermediate exits can increase the negative bias almost threefold. Thus, increasing the number of intermediate exits tends to accent the effect of trap shyness.

The effect of interview length on the *SV* estimator was examined under $\lambda = 1$ and non-trap-shy scenarios E to J. These six scenarios altered interviewer time from the first 2 hours in the morning to the first 4 and then first 6, and then to the last 6 in the afternoon, then last 4 and finally last 2 before 18.00 hours. The true *SV* for these situations was 10. Estimates for the morning were all negatively biased, with the earliest 2-hour scenario being the worst at -30.11 percent. In the afternoon a slight positive bias was observed, and this was greatest—only 5.94 percent—in the last-2-hour scenario. The general conclusions from these particular scenarios are that (1) mornings have a large negative bias while afternoons have a small positive bias, (2) longer interview periods are better, and (3) these negative and positive biases will probably not balance out if a 50-to-50 mix of morning and afternoon interview times is selected.

Evaluation of the bias and coefficient of variation of a pre-noon *SV* estimator (8.00 to 14.00 hours) and a post-noon *SV* estimator (12.00 to 18.00 hours) under a range of numbers of intermediate exits was performed on a DUDS site (fig. 4). The site was open from 6.00 to 21.00 hours with LERD = 50 recreationists each with probability of capture of 0.9 for all exits. The results indicate that both estimators are unbiased when $\lambda = 0$. However, as λ increases the pre-noon estimator becomes quite negatively biased, approximately 10 percent when $\lambda = 5$. The post-noon estimator showed the opposite effect, but with a positive bias of only about half the magnitude of that for the pre-noon estimator. The coefficient of variation of both estimators was < 10 percent, which is quite reasonable, with that for the post-noon estimator being somewhat smaller and more desirable.

The previous scenario was performed again but for a GFA site where there was an equal mixture of 10 visitors from each of the 4 recreation types (fig. 5). In this situation, the pre-noon estimator was again negatively biased and the post-noon estimator positively biased, but bias of the post-noon estimator was much larger. In addition, the coefficient of variation was approximately the same. Thus, in this scenario, the pre-noon estimator is preferable. This emphasizes the fact that the effect of recreationists' behavior on the *SV* estimator is complex, that general statements about the quality of an estimator cannot be easily extended to other scenarios, and that the use of a simulation model such as ORUS can help in evaluating alternative estimators.

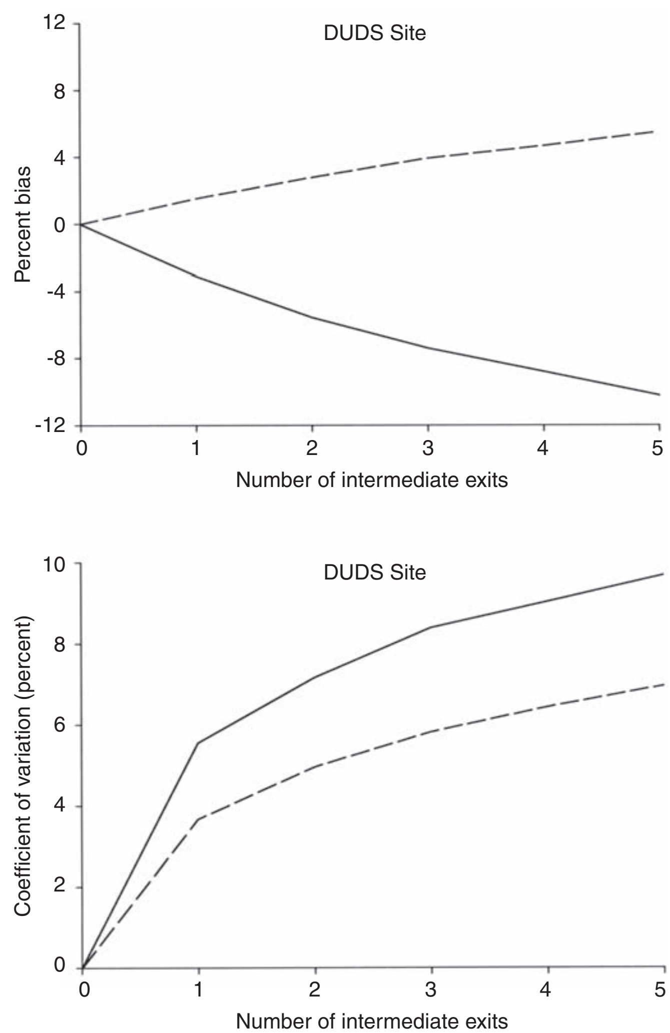

Figure 4—Evaluation of the bias and coefficient of variation of the SV estimator for a day-use developed site (DUDS) that is open from 6.00 to 21.00 hours with 50 recreationists who will be last-exiting the site on that site day and arrived on the site during the site day (LERD) each with probability of capture of 0.9 for all exits. The pre-noon estimator (8.00 to 14.00 hours) (solid line) and post-noon estimator (12.00 to 18.00 hours) (dashed line) are evaluated over a range of numbers of intermediate exits.

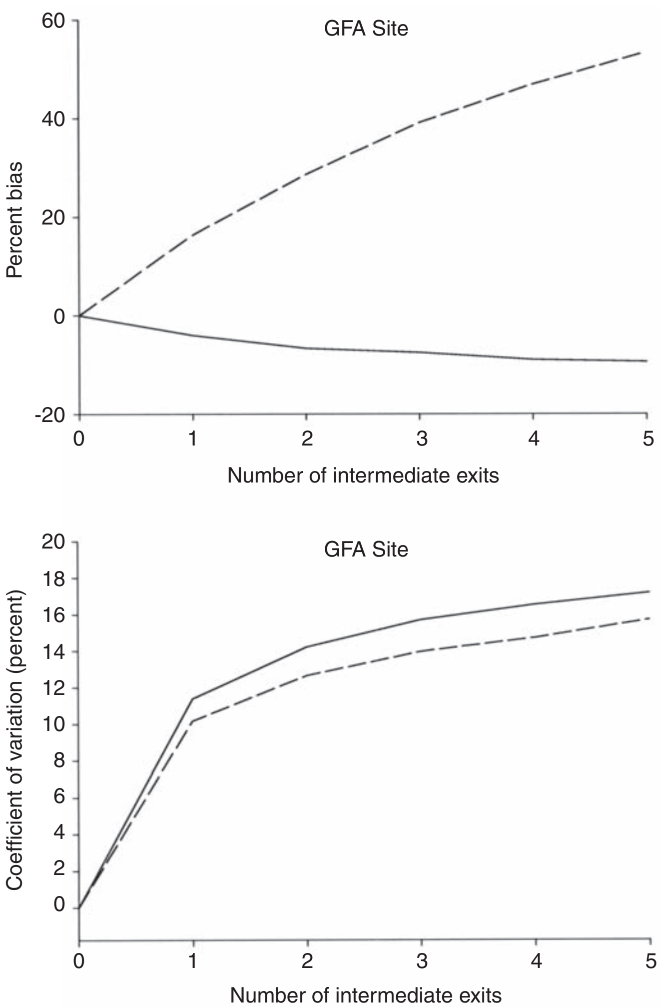

Figure 5—Evaluation of the bias and coefficient of variation of the SV estimator for a general forest area site (GFA) that is open from 6.00 to 21.00 hours with 10 recreationists who will be last-exiting the site on that site day and were at the site before the official beginning of the site day at midnight (LERB), 10 recreationists who will be last-exiting the site on that site day and arrived on the site during the site day (LERD), 10 recreationists who will not be last-exiting the site on that site day and were on the site before the official beginning of the site day at midnight (NLERB), and 10 recreationists who will not be last-exiting the site on that site day and arrived on the site during the site day (NLERD) each with probability of capture of 0.9 for all exits. The pre-noon estimator (8.00 to 14.00 hours) (solid line) and post-noon estimator (12.00 to 18.00 hours) (dashed line) are evaluated over a range of numbers of intermediate exits.

Conclusions

The ORUS model appears to be simulating the behavior incorporated into it by the various statistical distributions that describe the model components. Parameterization of the arriving and last-exiting beta distributions revealed the behavior patterns of the four different types of recreationists. A test simulation of behavior of recreationists at a DUDS site correctly reflected the statistical distributions used for the model's components. Further examination of several scenarios developed by altering the model's parameters showed that a bias of considerable magnitude may be present in many instances. Thus, use of the model to examine various scenarios can help isolate problems and formulate refinements in the survey methodology for future NVUM sampling.

These sample simulation results reveal some difficulties with the NVUM recreation use estimator. However, it should be kept in mind that ORUS is a very simple model at this point and does not yet make provisions for many problems that can occur in field sampling. For instance, the variation in the *SV* estimator does not provide for biases due to commuter traffic or to the "voluntary survey" sign effect that are believed to occur in the field. The effect of these biases on the estimate is unknown. Further refinements in the model are possible to help quantify these sources of bias and to make the recreationist behavior more realistic.

Acknowledgments

The authors acknowledge the help of the many Forest Service personnel who participated in the field sampling for the National Visitor Use Monitoring program.

References

Cochran, W.G. 1977. Sampling techniques. 3d edition. New York: John Wiley. 428 p.

English, D.B.K., Kocis, S.M., Zarnoch, S.J., and Arnold, J.R. 2002. Forest Service national visitor use monitoring process: research method documentation. Res. Pap. SRS-57. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 14 p.

Zarnoch, S.J., Kocis, S.M., Cordell, H.K., and English, D.B.K. 2002. A pilot sampling design for estimating outdoor recreation site visits on the national forests. Res. Pap. SRS-29. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 20 p.

Zarnoch, Stanley J.; English, Donald B.K.; Kocis, Susan M. 2005. An outdoor recreation use model with applications to evaluating survey estimators. Res. Pap. SRS–37. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 15 p.

An outdoor recreation use simulator (ORUS) has been developed to simulate recreation survey data currently being obtained by the U.S. Department of Agriculture Forest Service, National Visitor Use Monitoring (NVUM) program's survey of the national forests of the United States. Statistical distributions represent the various behaviors of recreationists during their visit to a recreation site. The beta distribution is used to model arriving times and last-exiting times. The Poisson distribution is used to model the number of intermediate exits from the site, and the times of the exits are selected randomly according to the uniform distribution. Finally, three levels of trap shyness are assigned to the recreationists to quantify the probability that the recreationists will be captured by the interviewer. The beta distributions for arriving and last-exiting are parameterized to the NVUM survey data. The functioning of the simulator is demonstrated with a simple example. The utility of ORUS in evaluating the bias and coefficient of variation of various survey scenario estimators of recreation use is also presented.

Keywords: Last-exiting recreationist, National Visitor Use Monitoring, ORUS, simulation, trap shy.