A Model for Evaluating Dispersed Outdoor Recreation Use Estimation

Stanley J. Zarnoch[1], Donald B. K. English[2] & Susan M. Kocis[3]

[1]US.D.A. Forest Service, Asheville, NC, USA
szarnoch@fs.fed.us
[2]U.S.D.A. Forest Service, Athens, GA, USA
denglish@fs.fed.us
[3]U.S.D.A. Forest Service, Bridgeville, CA, USA
skocis@fs.fed.us

Abstract: An outdoor recreation use simulator (ORUS) has been developed to simulate dispersed recreation survey data similar to that collected by the National Visitor Use Monitoring (NVUM) Project's survey of the national forests of the U.S.A. Statistical distributions are used to represent the various behaviors of recreationists during their visit to a dispersed area. The beta distribution is used to model arriving times and last exiting times. The number of intermediate exits from a site is determined by the Poisson distribution while their times are selected randomly according to the uniform distribution. Finally, three trap shy behaviors are assigned to the recreationists to quantify their probability of capture by the interviewer. The arriving and last exiting beta distributions are fitted to the NVUM data. The functioning of the simulator is demonstrated with a simple example with explanations of each recreationist's actions with respect to the sampling methodology. The utility of ORUS in evaluating the bias and coefficient of variability of various estimating scenarios is also presented.

Introduction

Since outdoor recreation has become an important valued component of forests, accurate recreation use estimates have become critical necessities in forest level planning. In 1996 a pilot study was performed to develop a field survey for estimating recreation use on the national forests throughout the United States (Zarnoch et al. 2002). This was later modified and expanded to include characteristics of the visitors, their satisfaction with the recreation resource and their economic impact on the local community (English et al. 2002). This has lead to the National Visitor Use Monitoring Project (NVUM) that currently surveys recreation use across the national forests of the U.S.A.

To validate the NVUM survey, a critical evaluation of the visitation estimators must be performed to determine the potential bias and variance properties under realistic recreation site scenarios. Thus, an outdoor recreation use simulator (ORUS) has been developed that has the capabilities of providing typical data similar to what has been collected by NVUM sampling.

There were several purposes for the creation of ORUS. First, the model outlines a structure that decomposes the complex system of visitor behavior into a set of more easily understood components and demonstrates their relationship to the visitation estimator. Second, the model provides the ability to evaluate the statistical properties (bias and coefficient of variation) of the visitation estimator. Third, the model enables a researcher to evaluate the effects of different assumptions about one or more visitor behaviors on the properties of the visitation estimator.

The objectives of this paper are to (1) describe the ORUS simulation model for outdoor recreation use estimation in dispersed areas and (2) demonstrate the evaluation of the NVUM visitation estimator under various site day scenarios.

The NVUM Sampling Design

The NVUM survey consists of a stratified multistage sampling design based on rotating panels that are spread over a five year sampling cycle. All national forests in the U.S. are sampled once every five years, with approximately one-fifth of the forests in each of 9 regions sampled each year. The statistical methodology follows conventional sample survey techniques with a few modifications to incorporate specific situations inherent in sampling national forests for recreation use.

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

– Day-Use Developed Sites (DUDS) – thoses sites intended for day use only.

– Overnight-Use Developed Sites (OUDS) – include campgrounds, cabins, hotels and any other overnight facility.
– Wilderness Sites (WILD) – sites that are designated official wilderness areas.
– General Forest Area (GFA) – all other areas in the national forest that are not DUDS, OUDS or WILD.

In this paper, only dispersed area recreation sites that are defined as GFA's will be considered.

As in all sample surveys, it is important to accurately determine the measurement variable on each sampling unit selected for the survey. In most natural resource monitoring and sampling situations, this issue is of little concern because a standard measuring device is used. For instance, in forest inventory a standard diameter tape is used to measure tree diameter. 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, called last exiting recreationists. The term distinguishes these individuals from recreation visitors who are making intermediate (non-final) exits and then returning to the site. An exact value for the measurement variable would be obtained under a 24-hour monitoring on-site interview protocol wherein all people exiting the site were required to participate in the survey process. Such a protocol is not possible for several reasons. Consequently, the NVUM project uses a methodology that estimates 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. This is performed at a designated interview point traversed by visitors exiting the site. This process obtains (1) a calibrated estimate of total exiting vehicles for the 24-hour period (VEHC), (2) an estimate of the proportion of exiting vehicles that are last exiting (PBAR), and (3) average number of occupants in a last exiting vehicle (PEOPVEH). These three quantities are used to estimate recreation site visits at the site for 24 hours.

The site visit estimator used by NVUM is defined as

$$\widehat{SV} = PBAR * VEHC * PEOPVEH \qquad (1)$$

For more details on the NVUM methodology, see English et al. (2002).

The accuracy of the site visit estimator depends on how well each of the three components in (1) is estimated. PEOPVEH is an easily observed quantity because it is obtained by simply counting occupants in vehicles determined to be last exiting recreation vehicles. The accuracy of VEHC depends largely on the consistent performance of the mechanical traffic counter over the 24-hour period. PBAR is a complex variable that is highly dependent on several aspects of visitor behavior at the recreation site. Thus, the focus of this paper is on simulating and evaluating the effect of PBAR on the site visit estimator.

Model Components

Types of Site Visitors

The model recognizes five distinct types of visitors who may be at a site. The typology is based on their specific behavior patterns of arriving time and last exiting time. These types are defined as follows:

– LERB = a recreationist that will be last exiting the site on the survey day but was at the site before the official beginning of the survey day at midnight
– LERD = a recreationist that will be last exiting the site and arrived on the site during the survey day
– NLERB = a recreationist that will not be last exiting but was on the site before the official beginning of the survey day at midnight
– NLERD = a recreationist that will not be last exiting the site and arrived on the site during the survey day
– NREC = a visitor who is on the site for non-recreation purposes (agency personnel, contractors etc.)

The four types of recreationists could have similar or different arriving or last exiting distributions and intermediate exit rates as will be explained in the next sections.

Arriving and Last Exiting Times

The fundamental behavior for visitors involves arriving at the site, engagement in recreation, and then leaving the site. The distributions of these actions relative to the six hour interview times are key elements of the simulation model. Arriving and last exiting times 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 a/(a+b) and the variance is 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 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 occurs when a=b=5 with a mean of 0.50. If a and b are both less than 1 then a u- shaped distribution results. Figure 1 shows the beta distribution for some 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

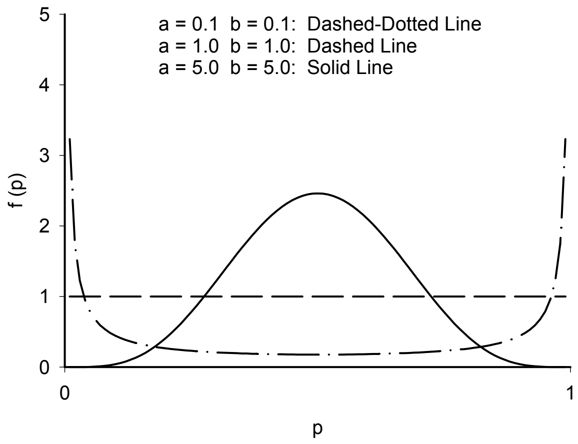

Figure 1. The beta distribution f(p) for some values of the parameters a and b.

arriving time as that proportion of the recreation day after the start of the recreation day. Mathematically, for LERD and NLERD this is

$$AT = D_S + p_1(D_E - D_S) \tag{3}$$

where
D_S = time when the recreation day[1] starts and
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.

The last exiting time (*LET*)of a recreationist also uses a variate, p_2, selected from the beta distribution and is defined for LERB as

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

and for LERD as

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

Since NLERB and NLERD recreationists do not exit the site on the site day, they have no last exiting time.

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. The number of intermediate exits a recreationist performs for the site day is modeled with the Poisson distribution which assumes that they are at random. The Poisson distribution is defined as

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{6}$$

where x = 0, 1, 2, … 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 the recreation day, $D_E - D_S$. Although this could be the total 24 hour day, 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 intermediate exiting period defined for that specific survey day. All individuals within a recreationist type have the same λ but adjusted by the length of the individual's time on the site. Thus, a recreationist that 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. It is possible to assume that the intermediate exit rate is the same for all recreationist types or it may vary depending on the parameter 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 survey day. This appears to be a reasonable assumption because each recreationist is unique and its intermediate exiting behavior is nearly impossible to predict. Some may wander off the site as soon as they get there just to merely see what's around the next bend. Others may go out to the store only to immediately leave again when they find out they forgot to get an important item. Still others may never leave the site until they depart for home. The total length of stay interval is defined by the arriving times and last exiting times. Then the number of intermediate exits is used to select a time at random from the interval for each exit. The uniform distribution defined on the interval length is used to generate these variates.

Trap Shy Behavior

The estimation of PBAR used for the site visit estimator 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. Thus, some exiting individuals may choose not to be interviewed. In particular, the probability that a recreationist stops for an interview may very well depend on the previous history of being stopped on that survey 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 where trapped animals learn to avoid traps after they are captured once. Thus, trap shy behavior 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 shy behaviors could be modeled, only three will be discussed here. First, the not trap shy situation is defined as

$$P_0 = P_1 = P_2 \tag{7}$$

where P_i is the 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 in half after being interviewed once, specifically

$$P_0 \to P_1 = \frac{P_0}{2} \to P_2 = \frac{P_0}{2} \tag{8}$$

The extreme case of trap shyness results in zero probability of an interview after the first, yielding

$$P_0 \to P_1 = 0 \to P_2 = 0 \tag{9}$$

Here it is assumed that the probability of an interview after the second is equivalent to P_2, although this assumption could be easily modified.

Methodology

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

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

where LC_{11} is the number of last exiting vehicles that were stopped for an interview and LC_{01} is the number of non-last exiting vehicles that were stopped for an interview. These could be computed from the data produced by ORUS under a specific scenario. The true proportion of last exiting recreationists could be computed as

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

where LC_{10} is the number of last exiting vehicles that were not stopped for an interview and LC_{00} is the number of non-last exiting vehicles that were not stopped for an interview.

Comparison of the estimated $\widehat{PBAR}$ to the true $PBAR$ for a given simulation scenario reveals the quality of the site visit estimator. However, since comparisons from only one simulation are difficult to judge because the simulated values are stochastic, 10,000 simulations were performed. The percent bias is used as a criterion for the quality of the site visit estimator and is defined as

$$\%\,Bias = \frac{100}{10000} \sum_{i=1}^{10000} \frac{\widehat{PBAR}_i - PBAR_i}{PBAR_i} \tag{12}$$

To judge the variability of the site visit estimator, the typical coefficient of variation is used. Although the site visit estimator could be evaluated under hypothetical beta distributions, it is more realistic 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 methods of moments and are defined as

$$\hat{b} = \frac{\left(1-\overline{X}\right)\left[\overline{X}\left(1-\overline{X}\right) - S^2\right]}{S^2} \tag{13}$$

and

$$\hat{a} = \frac{\hat{b}\overline{X}}{\left(1-\overline{X}\right)} \tag{14}$$

Recreation visitor arriving times were obtained from the NVUM survey to fit arriving beta distributions for LERD and NLERD. Last exiting times were used to fit beta last exiting distributions to the LERB and LERD. The last exiting beta distribution for LERD recreationists was assumed to be dependent on the arriving time of an individual. Thus, two linear regression models were used to predict $\overline{X}$ (mean) and S^2 (variance) for each individual as functions of arriving time and then used in (13) and (14) to estimate the individual's beta parameters.

Results

Parameterization of Dispersed GFA Sites

The arriving and last exiting beta distributions for the dispersed GFA sites were parameterized to the NVUM data collected over the first two sampling years. It was assumed that on-site recreation could occur only from D_S=6.00 to D_E=21.00, so the beta distributions are based on this recreation day length. There were only four distributions to parameterize. The LERB recreationists have only a last exiting distribution for a given survey day. The LERD type has both arriving and last exiting beta distributions. The NLERB neither enter nor exit during the survey day, so they have no beta distributions to parameterize. Since the NLERD only enter and do not exit, they have only arriving distributions. These fitted beta distributions are shown in Figure 2. The LERB recreationists (n=1,322) were fitted to the last exiting beta distribution, yielding a=3.694 and b=4.150. The distribution was approximately symmetric with a mean last exiting beta variate of 0.471, which when equated to last exiting time with equation (4) represents the time 13.06. The LERD recreationists (n=10,822) had an arriving beta distribution with a=1.602 and b=3.422 which was highly skewed to the right, indicating a tendency for most of these one day visitors to come early in the day. Their average beta variate was 0.319 which represents a time of 10.78 based on equation (3). These same LERD recreationists had a last exiting beta distribution with a=2.022 and b=4.402 which gave a mean beta variate of 0.313. Using the mean arriving time and equation (5), this represents an average last exiting time of 13.98. The NLERD recreationists (n=1,240) had an arriving beta distribution with a=1.520 and b=1.687 and a mean beta variate of 0.474, which yields an average arriving time of 13.11 using equation (3). This was not a skewed, asymmetrical bell shape distribution like the LERD. This is probably because these recreationists arrived on the site more uniformly throughout the day.

Simple Simulation Example

A simple example illustrates the ORUS model's capabilities. Assume for simplicity that a dispersed GFA site that is open for recreation from 6.00 until 21.00 will be surveyed from time 8.00 to 14.00. In addition, let the site have LERD=10 recreationists each with a high daily rate of intermediate exits set at λ=4. Their arriving and last exiting distributions were both selected from the NVUM fitted GFA beta distributions. To illustrate the effect of trap shyness, the probability of capture on the visitor's first exit was set at 1.0, and set at 0.0 for any subsequent exits, including the last.

Results from this scenario site day are shown in Table 1. There were a total of 18 exits from the site during the 15 hour day, 10 of which were obviously last exiting. Only 4 of the 10 last exiting recreationists were captured. Four visitors last exited the site after the interviewers left at 14.00 and, thus, could not be captured. The other two were interviewed first during an intermediate exit and trap shyness precluded these individuals from being interviewed on their final exit from the site. A total of 10 recreationists were stopped by the interviewers. Thus, an estimate of *PBAR* from equation (10) is $\widehat{PBAR}$ = 4/10=0.40. The true proportion is *PBAR*=10/18=0.56 computed from equation (11). This is a considerably poor estimate and could result in poor estimates for visitation on this site. Assuming that the vehicle counter recorded correctly 18 exiting vehicles for the 24 hour period and there was an average of one person per vehicle (for simplicity), the SV estimate would be $\widehat{SV}$ =0.40(18)(1)=7.2 while the true would be *SV* =0.56(18)(1)=10.0. This represents a negative 28 % bias.

Estimator Evaluation

Evaluation of the bias and coefficient of variation of an AM estimator (8.00 to 14.00 survey window) and a PM estimator (12.00 to 18.00 survey window) under a range of number of intermediate exits was performed on a dispersed GFA site (Figure 3). Specifically the site was open from 6.00 to 21.00 with an equal mixture of 10 visitors from each of the four recreation types each with probability of capture of 0.9 for all exits. The results indicate that both estimators are unbiased when λ=0. However, as λ increase the AM estimator becomes negatively biased, approximately 10 percent when λ=5. The PM estimator showed the opposite effect with a large positive bias of approximately 50 percent when λ=5. The coefficient of variation of both estimators average about 10 percent, which is quite reasonable, with the PM being somewhat smaller.

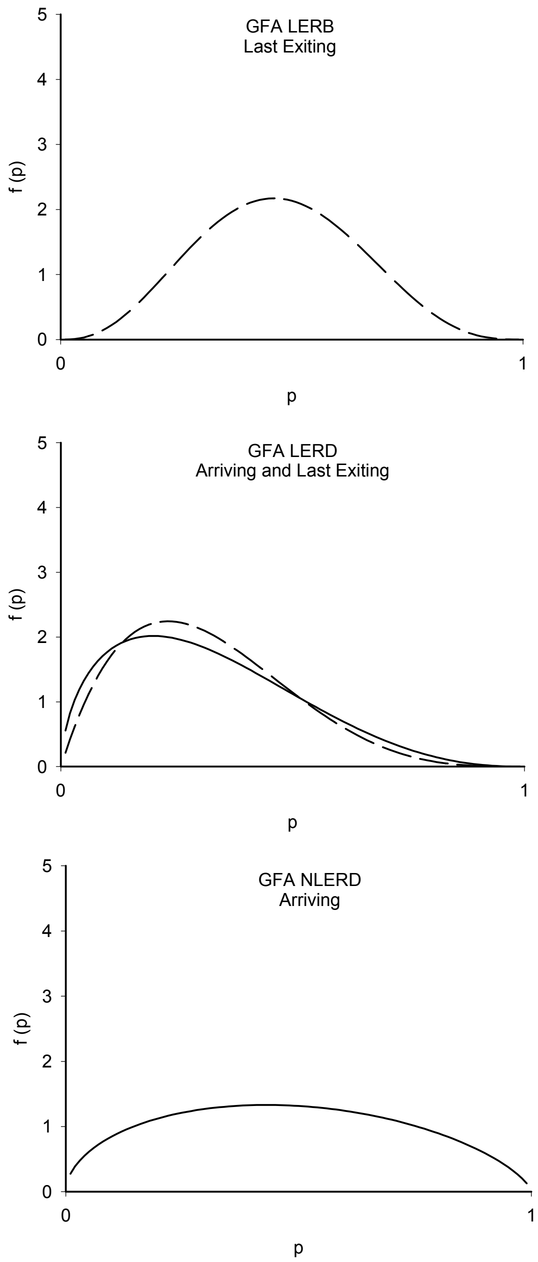

Figure 2. The beta distribution f(p) for dispersed GFA recreation visitors where solid lines are arriving distributions and dashed lines are last exiting distributions.

Table 1. Simulation of a dispersed GFA survey day.

Visitor	Last Exit	Time	Captured	Remarks
1	Yes	12.10	Yes	Captured because P_0=1.0.
2	Yes	12.75	Yes	Captured because P_0=1.0.
3	No	11.62	Yes	Captured because P_0=1.0.
3	No	14.77	No	Not captured because left after interviewers and became trap shy.
3	Yes	16.74	No	Not captured because left after interviewers and became trap shy.
4	No	10.80	Yes	Captured because P_0=1.0.
4	No	14.47	No	Not captured because left after interviewers and became trap shy.
4	Yes	14.63	No	Not captured because left after interviewers and became trap shy.
5	No	11.01	Yes	Captured because P_0=1.0.
5	Yes	11.22	No	Not captured because became trap shy.
6	No	11.34	Yes	Captured because P_0=1.0.
6	Yes	14.17	No	Not captured because left after interviewers and became trap shy.
7	Yes	13.50	Yes	Captured because P_0=1.0.
8	No	10.73	Yes	Captured because P_0=1.0.
8	Yes	12.91	No	Not captured because became trap shy.
9	No	12.99	Yes	Captured because P_0=1.0.
9	Yes	17.72	No	Not captured because left after interviewers and became trap shy.
10	Yes	9.83	Yes	Captured because P_0=1.0.

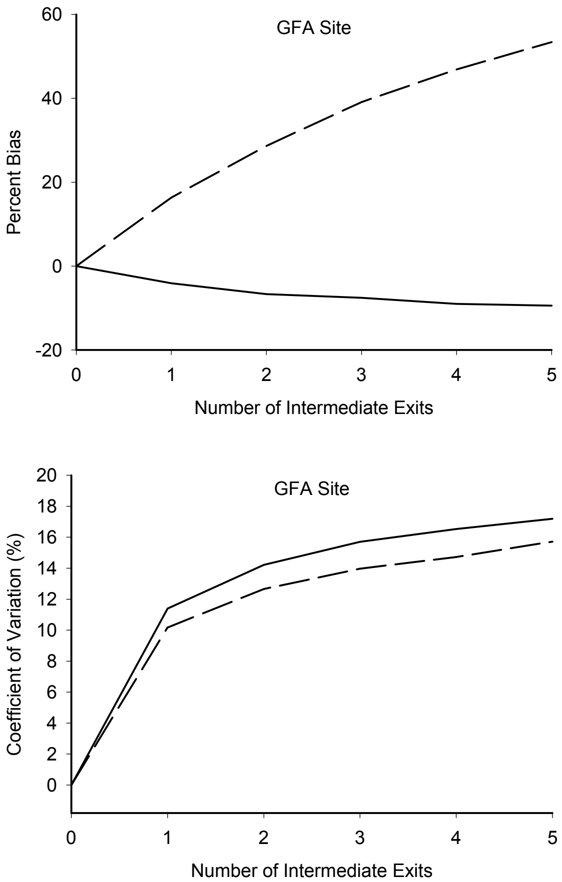

Figure 3. Evaluation of the bias and coefficient of variation of a GFA site that is open from 6.00 to 21.00 with LERB=10, LERD=10, NLERB=10 and NLERD=10 recreationists each with probability of capture of 0.9 for all exits. The AM (8.00 to 14.00) (solid line) and PM (12.00 to 18.00) (dashed line) estimators are evaluated over a range of number of intermediate exits.

Conclusion

The ORUS model appears to be simulating the behavior incorporated into it by the various statistical distributions that describe the model components. Examination of several survey site scenarios demonstrated the evaluation of the bias and coefficient of variation. Similar analyses should isolate problems and help formulate refinements in future survey methodology. It should be kept in mind that ORUS is a very simple model at this point and does not include many other problems that can occur in field sampling. For instance, the variation in the SV estimator does not incorporate any biases due to commuter traffic or to the "voluntary survey" sign effect that are believe to occur in the field. The effect of these on the estimate is unknown. Future refinements in the model are possible to help quantify these sources of bias or to make the recreationist behavior more realistic.

References

English, D.B. K., Kocis, S.M., Zarnoch, S.J. & 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. & 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.

[1] Times used throughout this manuscript are represented as real numbers for computation purposes. The hour component is analogous to standard military time while the minute component represents the decimal part of the hour. Thus, 6:30 am is represented as 6.50 while 4:15 pm is represented as 16.25.