ENVIRONMETRICS
Environmetrics 2003; **14**: 559–571 (DOI: 10.1002/env.603)

Stemflow estimation in a redwood forest using model-based stratified random sampling[‡]

Jack Lewis*,†

Redwood Sciences Laboratory, Pacific Southwest Research Station, U.S. Forest Service, 1700 Bayview Drive, Arcata, CA, USA

SUMMARY

Model-based stratified sampling is illustrated by a case study of stemflow volume in a redwood forest. The approach is actually a model-*assisted* sampling design in which auxiliary information (tree diameter) is utilized in the design of stratum boundaries to optimize the efficiency of a regression or ratio estimator. The auxiliary information is utilized in both the design and estimation phases. Stemflow and its variance were modelled as powers of diameter and a generalized non-linear least squares model was used to estimate the exponents and to impute values for missing storm events prior to application of the ratio estimator. The advantage of the ratio estimator over standard stratified sampling formulas is greatest for species in which stemflow is strongly dependent on diameter. With measurements on 24 trees in a 1-hectare stand, annual stemflow was estimated with coefficients of variation of 9 per cent and 10 per cent in 2 years of study. Striking stemflow differences were found between species. Published in 2003 by John Wiley & Sons, Ltd.

KEY WORDS: stemflow; interception; sampling design; model-assisted

1. INTRODUCTION

The portion of rainfall that reaches the forest floor by flowing down the stems of trees is known as stemflow. In order to accurately determine rainfall interception, it is necessary to measure stemflow as well as rainfall and throughfall. Therefore, stemflow is of interest because of its role in the forest water budget. It can also be ecologically important because of its ability to locally concentrate soil water and nutrients.

Sampling error (the error created from observing a sample rather than the entire population) is seldom reported in hydrologic field studies because it is seldom estimated. However, sampling error can be estimated if a well-understood probability sampling design is employed. Studies that include measurements of stemflow in forest stands could easily employ probability sampling of trees to provide estimates of sampling error. However, in a survey of 19 studies that included stemflow measurements (Table 1), only three employed probability sampling. Only one of these (Cape *et al.*, 1991) reported sampling error, and it was calculated without regard to the stratified random sampling

*Correspondence to: Jack Lewis, Redwood Sciences Laboratory, Pacific Southwest Research Station, U.S. Forest Service, 1700 Bayview Drive, Arcata, CA 95521, U.S.A.
†E-mail: jlewis01@fs.fed.us
‡This article is a U.S. Government work and is in the public domain in the U.S.A.

Published online 13 June 2003

Received 29 February 2002
Accepted 28 August 2002

Table 1. Stemflow sampling designs and estimation methods

Study	Species	n	Sampling design	Estimation method
Aboal *et al.*, 1999	49 yr laurel forest, 6 species	30	STRS by DBH for each species, $n_h = 1$, $H = 5$	BA regression
Anderson and Pyatt, 1986	25 yr Sitka spruce 25 yr lodgepole Pine 63 yr Sitka spruce	12 12 20	Judgment within basal area strata	Mean per tree times total number of trees
Asdak *et al.*, 1998	Tropical rainforest Pristine and logged	16 20	Judgment within DBH strata	Not reported
Cape *et al.*, 1991	Scots pine and 5 other species	54	STRS by DBH in each of 9 plots, $n_h = 3$, $H = 2$	STRS for total; SRS for error
Crockford and Richardson, 1990	Eucalypt Pinus radiata	32 21	Judgment	Mean per unit BA times total BA
Ford and Deans, 1978	14 yr Sitka spruce	23	Measured all trees in plot	Not reported
Gash *et al.*, 1995	20 m maritime pine	6	Judgment	Not reported
Gash *et al.*, 1980	30 yr Sitka spruce 41 yr Scots pine 26 yr Sitka spruce	26	Judgment	Not reported
Gash and Morton, 1978	44 yr Scots pine	5	Judgment	Not reported
Hanchi and Rapp, 1997	Pinus pinea	—	Judgment	Stormwise DBH regression
Herwitz and Levia, 1997	Populus grandidentata	5	Judgment	Mean and standard deviation
Johnson, 1990	50 yr Sitka spruce	9	Judgment	Not reported
Hamilton and Rowe, 1949	Chaparral	all	Whole plot	NA
Lawson, 1967	Pine-hardwood	28	SRS	Not reported
Loustau *et al.*, 1992	18 yr maritime pine	12	Judgment	Mean and CI
DeWalle and Paulsell, 1969	Black oak	32	Judgment	By individual tree
Llorens *et al.*, 1997	Scots pine	7	Judgment	Not reported
Spittlehouse, 1998	Five conifer forests	—	Not reported	Mean
Viville *et al.*, 1993	90 yr Norway spruce	4	Judgment	Not reported

STRS, stratified random sampling, SRS, simple random sampling, DBH, diameter at breast height, BA, basal area, CI, confidence interval, n, number of trees sampled, n_h, number of trees per stratum, H, number of strata.

(STRS) design. Aboal *et al.* (1999) also employed STRS, but sampling error could not be estimated because only one tree was selected in each stratum. In most of the remaining studies, the selection method was not reported, or selection was based on the judgment of the investigator. In one study (e.g. Ford and Deans, 1978), a representative plot was selected, and all trees in the plot were measured. That approach is only feasible with very small plots, but then it is difficult to make inferences since the question of selection bias is simply transferred from the individual tree level to the plot level. At some level it is necessary to make a judgment in choosing a representative study area, i.e. a population to sample. However, if a small plot containing only a few trees is chosen, it may be difficult to credibly extend the results to an entire stand or forest.

Several of the studies in Table 1 employed stratification, either by diameter or basal area to improve stemflow estimates. In some studies (Asdak *et al.*, 1998; Ford and Deans, 1978; Lawson, 1967; Hanchi and Rapp, 1997), stemflow was found to be related to physical tree characteristics such as diameter, basal area or crown projection area, suggesting that these relationships could be exploited to estimate stemflow at the stand level. While various studies have employed probability sampling, stratification, or regression, individually, the potential benefits can be maximized by combining the approaches. This article describes a sampling design that combines these approaches in a way that permits nearly unbiased and very efficient estimation of stemflow and its variance. The design is illustrated by a case study in a redwood forest in northern coastal California.

2. METHODOLOGY

Hanchi and Rapp (1997) proposed measuring stemflow on representative trees of each size class in the stand and developing a relationship between stemflow and tree diameter for each storm event. In each storm, the relationship would be applied to all the trees in the stand to compute the total stemflow. The volumes from all storms in any period of interest would then be summed and divided by the area of the stand to obtain the depth of rainfall that was routed to stemflow. The form of the relationship they proposed was a power function, $V = \beta D^{\lambda}$, where V is stemflow volume, D is diameter, and β and λ are constants specific to each storm.

Hanchi and Rapp's proposal is reasonable but statistically inadequate because it does not provide a means of assessing the goodness of the estimate, i.e. a confidence interval. To compute a confidence interval, one needs a reference distribution for the stemflow estimate and an estimate of its variance. The reference distribution is the probability distribution from which inferential statements about the estimate are derived. There are two basic inferential paradigms that may be employed to estimate variance (Gregoire, 1998): (a) design-based inference, and (b) model-based inference. They differ in the way a reference distribution is determined for the statistics being computed.

2.1. Design-based and model-based inference

Under design-based inference, the reference distribution is a consequence of a probability sampling design for a fixed, finite population. No distributional assumptions are made about the population, but the probability of inclusion in the sample is a known positive quantity for each sampling unit in the population. The reference distribution of a statistic derives from its potential outcomes among the population of samples collected according to the specified sampling design. Such designs employ probability sampling and are viewed as objective because they remove bias from the selection process. Sampling error can be reduced by employment of efficient sampling designs that use auxiliary population information in the design phase to restrict randomization in some way.

Under model-based inference, the reference distribution is a consequence of a presumed model of population behavior. For example, a linear regression model typically assumes normally distributed errors with variance independent of x. It is implicit in the model that, for any given observation x, there is an infinite distribution, or *superpopulation*, of possible y values, only one of which is realized in any particular population. The reference distribution of a statistic derives from its possible outcomes among the assumed superpopulation. Random (probability) sampling is not a requirement for model-based inference, but the reliability of results is dependent on the correctness of the assumptions. Information in auxiliary variables (easy-to-measure variables that are closely related to the variable of interest) is utilized in the estimation phase under model-based inference.

2.2. *Model-based stratified sampling*

The efficiency of the design/estimator combination depends on how well the auxiliary information can be utilized. STRS normally utilizes auxiliary information during the design phase only. Application of a regression estimator in simple random sampling utilizes auxiliary information during the estimation phase only. *Model-based stratified sampling* (Wright, 1983) is a hybrid design that utilizes auxiliary information during both the design and estimation phases. The design is described also by Särndal *et al.* (1992), who present an elegant and very general framework for sampling designs that includes all the standard designs as well as some rather complex ones. Many of these are *model-assisted* designs, incorporating elements of both design-based and model-based inference. Model-based stratified sampling is a model-assisted design that optimizes stratum and sample allocation for estimation with a regression or ratio estimator. This study employs a ratio model in which the mean and variance of an observation are given by:

$$\begin{cases} \mathrm{E}(y_k) = \beta x_k^{\lambda} \\ \mathrm{V}(y_k) = \sigma_k^2 = \sigma^2 w_k = \sigma^2 x_k^{\gamma} \end{cases} \tag{1}$$

where y_k, $k = 1, \ldots, N$, are realized values of independent random variables Y_k (the superpopulation), and x_k are the values of an auxiliary variable (e.g. tree diameter, that will be used to predict y_k =stemflow on unmeasured trees). Parameters β and σ^2 are unknown, and λ and γ are known positive constants. This model differs from Hanchi and Rapp's in that, here, the exponent λ is treated as a known constant and a variance σ_k^2 about the regression line is modelled as a power of x_k.

In model-based stratified sampling, the design is nearly optimal in the sense that an approximation of the anticipated variance of the estimator of the total (Equation (2), below) is minimized for a given sample size. The *anticipated variance* is defined as the variance of the difference between the estimated and true total under *both* the sampling design and the regression model. The optimum design is one in which the inclusion probabilities are proportional to σ_k.

If a random sample size is acceptable, then a Poisson sampling design provides a simple solution. A series of N Bernoulli trials would be conducted such that the kth element is given probability of selection $\pi_k = n\sigma_k / \sum_{\mathrm{U}} \sigma_k$, with n being the desired sample size and U denoting the population $\{1, \ldots, N\}$. However, Poisson sampling has the disadvantage of a random sample size. On the other hand, specification of an optimal fixed-size design also meets with difficulties (Särndal *et al.*, 1992). Model-based stratified sampling is a fixed-size design by which one can obtain inclusion probabilities that are close to the optimum.

The procedure for designing strata in model-based stratified sampling is called the *equal aggregate σ rule*. The strata are defined so that σ_k is similar for all k within any given stratum and the stratum sums of σ_k are approximately equal for all strata. In practice, $x_k^{\gamma/2}$ is substituted for the unknown σ_k.

Step 1. Order the values σ_k in increasing magnitude: $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_N$.

Step 2. Specify the number of strata, H, and calculate $T = \sum_{\mathrm{U}} \sigma_k / H$.

Step 3. In the first stratum, U_1, allocate the first N_1 elements ordered as in step 1 up to the point where, as closely as possible $\sum_{U_1} \sigma_k = T$. In the second stratum, U_2, allocate the next N_2 elements ordered as in step 1 up to the point where, as closely as possible, $\sum_{U_2} \sigma_k = T$, and so on. That is, every stratum accounts for very nearly one Hth of the total of the values σ_k.

Step 4. Allocate the sample equally to all the strata. That is, as closely as possible, set the sample size to $n_h = n/H$ in each stratum.
Step 5. Randomly select n_h elements among the N_h elements in each stratum.

When stratifying by a continuous variable, increasing the number of strata, H, generally reduces the variance of the estimated population mean and total up to a point of diminishing returns (Cochran, 1977). If the cost of stratifying is trivial, H may be set to half the desired sample size to minimize variance while still permitting its estimation. Otherwise, a smaller value for H is optimal, as determined by its relationships with cost and variance.

2.3. *Estimation*

The general ratio estimator for the population total $t_y = \sum_U y_k$ is given by

$$\hat{t}_{yr} = \hat{B}t_x + \sum_s \breve{e}_k \tag{2}$$

where $\sum_U$ and $\sum_s$ indicate summation over the population and sample, respectively, $t_x = \sum_U x_k$ and $\breve{e}_k = \left(y_k - \hat{B}x_k\right)/\pi_k$, in which π_k is the inclusion probability for element k and $\hat{B}$ is the estimator for the ratio y_k/x_k:

$$\hat{B} = \left(\sum_s x_k y_k \Big/ w_k \pi_k\right)\left(\sum_s x_k^2 \Big/ w_k \pi_k\right)^{-1} \tag{3}$$

The term $\sum_s \breve{e}_k$ in Equation (2) vanishes for models where $\gamma = 1$ in model (1) and $w_k = x_k$, yielding the most common form of the ratio estimator. The estimator $\hat{t}_{yr}$ for the total is asymptotically design-unbiased. That is, under the sampling design, there is a small bias that approaches zero as the sample size increases. The suggested variance estimator (Särndal *et al.*, 1992) is given by

$$\hat{\mathrm{V}}\left(\hat{t}_{yr}\right) = \sum\sum_s \breve{\Delta}_{kl}(g_k \breve{e}_k)(g_l \breve{e}_l) \tag{4}$$

where

$$\begin{aligned} \breve{\Delta}_{kl} &= (\pi_{kl} - \pi_k \pi_l)/\pi_{kl} \\ g_k &= 1 + \left(\sum_U x_k - \sum_s x_k/\pi_k\right)\left(\sum_s x_k^2/w_k\pi_k\right)^{-1}(x_k/w_k) \end{aligned} \tag{5}$$

in which π_{kl} is the joint inclusion probability of elements k and l in the sample and $w_k = x_k^{\gamma}$ in accordance with the ratio model (1).

In the special case of STRS, the inclusion probabilities for stratum h are $\pi_{kh} = n_h/N_h$ and $\pi_{klh} = n_h(n_h - 1)/(N_h(N_h - 1))$, where n_h and N_h are the stratum sample size and total stratum size, respectively. Substituting these expressions into Equations (2)–(5) yields the following formulas for $\hat{t}_{yr}$ and $\hat{\mathrm{V}}\left(\hat{t}_{yr}\right)$:

$$\hat{t}_{yr} = \hat{B}t_x + \sum_h \frac{N_h}{n_h} \sum_{k \in s_h} y_{kh} - \hat{B}x_{kh} \tag{6}$$

$$\hat{B} = \left(\sum_h \frac{N_h}{n_h} \sum_{k \in s_h} \frac{x_k y_k}{w_k}\right)\left(\sum_h \frac{N_h}{n_h} \sum_{k \in s_h} \frac{x_k^2}{w_k}\right)^{-1} \tag{7}$$

$$\hat{V}(\hat{t}_{yr}) = \sum_h \sum_{k \neq l \in s_h} \sum \breve{\Delta}_{klh}(g_{kh}\breve{e}_{kh})(g_{lh}\breve{e}_{lh}) \tag{8}$$

$$\breve{\Delta}_{klh} = (n_h - N_h)/(N_h(n_h - 1)) \tag{9}$$

$$g_{kh} = 1 + (t_x - \hat{t}_x)\left(\sum_h \frac{N_h}{n_h} \sum_{k \in s_h} \frac{x_k^2}{w_k}\right)^{-1} \frac{x_k}{w_k} \tag{10}$$

$$\hat{t}_x = \sum_U \frac{x_k}{\pi_k} = \sum_h \frac{N_h}{n_h} \sum_{k \in s_h} x_k \tag{11}$$

$$\breve{e}_{kh} = \frac{N_h}{n_h}\left(y_k - \hat{B}x_k\right) \tag{12}$$

where $\sum_{k \in s_h}$ denotes summation over all elements sampled from stratum h and the double-sum in Equation (8) indicates summation over all pairs of sampled elements from stratum h.

3. APPLICATION

3.1. Research context

Runoff peaks and volumes were found to increase during storms after logging in a redwood forest at Caspar Creek Experiment Watershed in northwestern California (Ziemer, 1998; Lewis *et al.*, 2001). During large storm events with recurrence intervals of about 2 years, the increase in storm peak averaged 27 per cent from clearcut areas. Reduced rainfall interception was thought to play a role in these changes, so a study of rainfall interception was undertaken. To measure interception, stemflow as well as rainfall and throughfall measurements were required. These measurements were conducted in a 1-hectare square selected to represent conditions typical of 100 to 135 year-old redwood forest. The plot contained 538 trees and a basal area of 97 m^2 (Table 2).

Measurements in the year 2000 were limited to a period from 17 January to 19 March. This was a wet period in an average year, with measurable rainfall on 46 of 63 days that amounted to 542 mm, about 46 per cent of the annual rainfall. Measurements in water year 2001 were limited to the period from 1 October 2000 to 1 April 2001. This period included 751 mm of rainfall, or 85 per cent of the annual rainfall in what was a relatively dry year. For the purposes of this article these two measurement periods will be discussed as if they represented annual stemflow.

Table 2. Tree species composition of stemflow plot

Species	BA (m^2)	Stems
Redwood	60.6	341
Douglas-fir	30.9	106
Tanoak	5.2	79
Grand-fir	0.0	1
W. Hemlock	0.0	1
Other	0.3	9
Total	97.0	538

3.2. Hardware

From each selected tree, the outer bark was removed from a 5-cm strip near the base of the tree. The notch was then sealed and EPDM sponge rubber collars were wrapped around the notch. The width of the collar varied from 6 to 15 cm, depending on the depth of the notch. The lower edge was sealed with waterproof polyurethane glue or silicone rubber caulking and nailed in the notch. The upper edge was nailed to the bark above the notch, using weather-stripping as a spacer, and with foil tape sandwiched between the weather-stripping and sponge rubber. The edge of the foil tape thus defined the catchment area. The collected water was routed via polyethylene tubing through the collar to collecting barrels. Volumes were usually measured after each storm event.

3.3. Sampling design

Asdak *et al.* (1998), found that stemflow was approximately proportional to basal area. Since the variance of many measurements increases with size, it was assumed for the sampling design that stemflow variance would be proportional to basal area, implying that σ_k would be proportional to x_k, the diameter of the kth tree (measured 1.37 m above the ground). Under that assumption, x was substituted for σ in the equal aggregate σ rule.

The total sample size (24) was determined by budget and logistical factors. The allocation of samples to species was set at 12 redwoods, 8 Douglas-firs, and 4 tanoaks, based on a consideration of the stand composition (Table 2). Because the cost of stratifying was insignificant, the most efficient design was that which maximized the number of strata ($n_h = 2$ for all strata), giving 6 redwood, 4 Douglas-fir, and 2 tanoak strata. Application of the equal aggregate σ rule resulted in the stratum boundaries and sizes shown in Table 3.

The efficiency of the sampling design can be measured by the ratio of the sample size under the optimal design ($\pi_k \propto \sigma_k$) to that required to achieve the same variance under the actual design. In model-based stratified sampling, the efficiency is always at least

$$eff_{\min} = \frac{1}{1 + (\max[cv_{\sigma h}])^2} \tag{13}$$

where $cv_{\sigma h}$ is the coefficient of variation of σ_k in stratum h and max $[cv_{\sigma h}]$ is its maximum among all strata. Under the assumption that σ_k is proportional to x_k, the efficiencies for the strata shown in Table 3 are at least 0.96 for redwood, 0.91 for Douglas-fir and 0.94 for tanoak.

Table 3. Stratum boundaries, sizes, and coefficients of variation of σ_k (assumes σ_k is proportional to x_k)

Redwood			Douglas-fir			Tanoak		
x_h (cm)	n_h	$cv_{\sigma h}$	x_h (cm)	n_h	$cv_{\sigma h}$	x_h (cm)	n_h	$cv_{\sigma h}$
15–26	124	0.20	15–48	48	0.32	15–35	60	0.26
26–39	74	0.12	48–67	27	0.11	35–55	19	0.12
39–50	51	0.07	67–87	18	0.07			
50–66	39	0.10	87–145	13	0.15			
66–88	31	0.07						
88–144	22	0.14						

3.4. Estimation of stemflow by individual event

Rainfall events in this analysis are defined as periods bounded by field measurements of stemflow volume. They do not necessarily coincide with storm events, but the analysis serves to illustrate how storm event stemflow can be estimated. Two preliminary issues needed to be addressed before estimating stemflow. First, since values were not available from preliminary surveys, it was necessary to estimate λ and γ as a preliminary step before calculating the ratio estimator. Second, about 5 per cent of the individual event totals were unreliable because of leaky collars or plumbing. Missing values are an important issue when stratum sample sizes are equal to two, because then variance estimation is impossible if any data are missing. Both of these issues were addressed by fitting a model for event stemflow volumes. The model is identical to model (1), but the parameter β is indexed by event. The mean and variance of an observation from event *j* on tree *k* are given by:

$$\begin{cases} \mathrm{E}(y_{jk}) = \beta_j x_k^{\lambda} \\ \mathrm{V}(y_{jk}) = \sigma^2 x_k^{\gamma} \end{cases} \tag{14}$$

The parameters β, λ, γ and σ were estimated using the *gnls* (generalized non-linear least squares) function in the mixed-effects modelling package of the S language. The solution, which employs maximum-likelihood estimation despite the name of the function, assumes a normally distributed response for any given *j* and *k*. The estimated values of γ were 0.64, 1.36 and 4.36, and the estimated values of λ were 0.85, 1.08 and 1.19, respectively, for redwood, Douglas-fir and tanoak. The solution was used to impute missing values in order to proceed with estimation using the design-based formulas (6)–(12). Table 4 shows the stemflow estimates by date and species, with totals for each water year. To facilitate comparisons, variances in Table 4 are expressed in terms of the coefficient of variation, i.e.

$$CV = \sqrt{\hat{V}(\hat{t}_{yr})} \Big/ \hat{t}_{yr} \tag{15}$$

The *CV* in the 'Total' column assumes independent estimates for the three species and averages 0.14. Thus, the half-width of a 95 per cent confidence interval, on average, would be 28 per cent of the stemflow estimate. The *CV* columns in Table 4 are blank in the HY2000 and HY2001 sum-of-event rows because it is not possible to compute variance estimates without assuming a covariance structure among individual event estimates. The estimates for individual events are almost surely positively correlated because the particular sample of trees is the same for all events. Thus, for example, if the sample represents above-average stemflow for one storm, it is likely to do so for all storms. But the degree of correlation is unknown. Therefore, to estimate annual stemflow, the procedure described in the following section is preferred.

3.5. Estimation of annual stemflow

Annual stemflow estimation is analogous to event stemflow estimation, except the basic data are annual stemflow for each collared tree. The imputed missing values from model (14) were summed into the annual stemflow to ensure that the entire period was represented.

If model (14) is modified so that *j* indexes years instead of individual events, the solution is unstable (probably because of the small sample size). However, if we assume that γ is independent of the aggregation period, we can use the event-based estimates of γ in a model for annual stemflow, and re-estimate only the exponent λ. That is, the mean and variance of an observation from water year *j* on

Table 4. Estimates of storm stemflow by storm and species

Event	Redwood		Douglas-fir		Tanoak		Total	
yr/mo/dy	$\hat{t}_{yr}$	*CV*	$\hat{t}_{yr}$	*CV*	$\hat{t}_{yr}$	*CV*	$\hat{t}_{yr}$	*CV*
00/01/24	0.07	0.18	0.48	0.15	0.41	0.22	0.96	0.12
00/01/28	0.05	0.32	0.21	0.30	0.20	0.25	0.46	0.18
00/01/31	0.05	0.17	0.29	0.24	0.37	0.18	0.71	0.14
00/02/14	0.98	0.16	1.89	0.17	1.49	0.12	4.36	0.09
00/02/22	0.15	0.25	0.62	0.21	0.45	0.20	1.22	0.13
00/02/25	0.33	0.21	0.44	0.20	0.36	0.22	1.12	0.12
00/02/28	1.06	0.15	1.56	0.18	0.84	0.21	3.45	0.11
00/03/09	0.32	0.18	1.04	0.23	0.67	0.41	2.03	0.18
HY2000	3.00		6.53		4.79		14.31	
00/10/23	0.00	0.63	0.00	0.31	0.03	0.08	0.03	0.08
00/10/30	0.19	0.46	0.63	0.24	0.62	0.19	1.44	0.15
00/11/30	0.04	0.22	0.21	0.38	0.22	0.18	0.48	0.19
00/12/15	0.02	0.30	0.34	0.20	0.35	0.08	0.71	0.11
00/12/28	0.00	0.43	0.00	0.27	0.00	0.31	0.01	0.20
01/01/09	0.01	0.30	0.10	0.31	0.17	0.15	0.28	0.14
01/01/11	0.12	0.28	0.44	0.17	0.31	0.17	0.88	0.11
01/01/25	0.08	0.27	0.36	0.19	0.30	0.16	0.73	0.12
01/01/30	0.12	0.40	0.48	0.20	0.32	0.31	0.92	0.16
01/02/12	0.04	0.30	0.51	0.26	0.42	0.23	0.97	0.17
01/02/20	0.47	0.20	1.50	0.20	0.91	0.32	2.88	0.15
01/02/22	0.64	0.20	1.16	0.16	0.53	0.21	2.33	0.11
01/02/25	0.40	0.26	0.79	0.18	0.49	0.20	1.68	0.12
01/03/05	0.39	0.21	1.03	0.20	0.66	0.20	2.08	0.12
01/04/02	0.05	0.27	0.26	0.30	0.33	0.18	0.64	0.16
HY2001	2.57		7.81		5.67		16.05	

tree k are given by:

$$\begin{cases} \mathrm{E}(y_{jk}) = \beta_j x_k^{\lambda} \\ \mathrm{V}(y_{jk}) = \sigma^2 x_k^{\gamma} \end{cases} \tag{16}$$

The parameters β, λ and σ were estimated using *gnls* and γ was assumed known from the event-based analysis. The estimated values of λ were 0.74, 1.09 and 1.31, respectively, for redwood, Douglas-fir and tanoak.

Figure 1 shows the relation of annual stemflow to stem diameter for each species. The curves shown are those estimated by model (16). In addition, a quadratic fit ($\lambda = 2$) has been added for tanoak, since that also appeared as a reasonable fit to the data in both years. The quadratic fit effectively places more weight on the highest point. Table 5 shows the annual stemflow estimates with λ estimated from model (16) and with an alternate set of λ selected by casual observation of the data ($\lambda = 1{,}1{,}2$ and $\gamma = 0{,}1{,}2$ for redwood and Douglas-fir, and tanoak). In addition, Table 5 shows estimates from standard STRS formulas that do not utilize auxiliary (diameter) information. The standard formulas are equivalent to ratio estimates with $\lambda = \gamma = 0$.

The ratio estimates based on the casually selected λ have lower *CV* than those based on model (16) (Table 5). The difference is almost entirely due to the tanoak variance estimates. The standard STRS

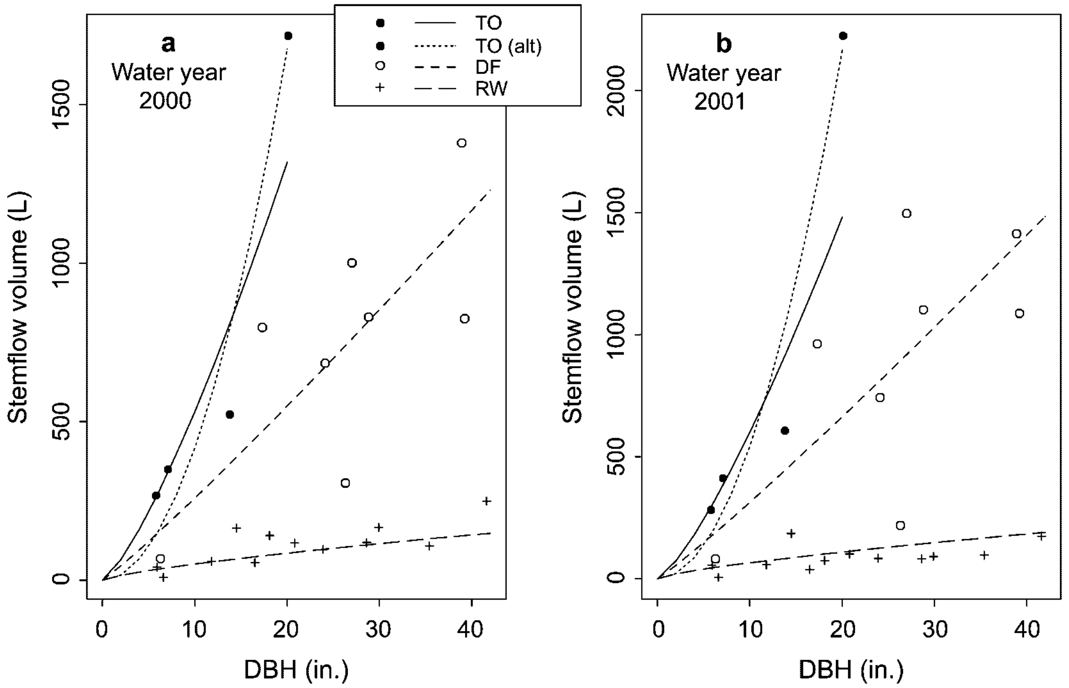

Figure 1. Relation between stemflow and diameter at breast hight (DBH) during the two measurement periods for tanoak (TO), Douglas-fir (DF), and redwood (RW). Curves are maximum likelihood fits by *gnls* to model (16) except TO (alt), in which $\lambda = \gamma = 2$ are fixed. Measurement periods are: a) 17 Jan 2000–19 Mar, 2001, b) Oct, 2001–01 Apr, 2001.

formulas gave lower estimates of total stemflow and higher variance estimates than the ratio estimators. The differences between the standard and ratio estimates is most pronounced for tanoak and least pronounced for redwood. For redwood, in which the dependency of stemflow on diameter is relatively weak, there is very little advantage of using the ratio estimate. This indicates that, after stratifying redwoods by diameter, there is little or no remaining (within-stratum) dependency of stemflow on diameter.

The sums of event stemflow estimates (Table 4) agree closely with the annual stemflow estimates from Table 5 (top tier) for each species as well as for the total. However, the estimates from Table 5 are preferred because, as mentioned before, the uncertainty of the sum-of-event estimates cannot be reliably estimated.

Figure 2 shows 95 per cent confidence intervals for stemflow by species, as a percentage of rainfall. Total stemflow was about 2.7 per cent for the monitoring period in water year 2000 and 2.2 per cent for water year 2001. Stemflow was higher in water year 2000, because the monitoring period that year was a period of relatively continuous rainfall. In 2001, the entire wet season was measured. In both years, the half-width of the confidence interval for total stemflow (2 standard errors) amounted to about 0.5 per cent of the rainfall. Although redwood comprises 62 per cent of the basal area in the stand, it accounted for 18 per cent of the total stemflow. While tanoak is a minor component of the stand (less than 6 per cent of the basal area), it accounts for about 35 per cent of the stemflow. Douglas-fir, comprising 32 per cent of the basal area, accounted for 47 per cent of the stemflow.

Table 5. Estimates of annual stemflow and its coefficient of variation

			2000		2001	
	λ	γ	$\hat{t}_{yr}$ (mm)	CV	$\hat{t}_{yr}$ (mm)	CV
Ratio estimates						
Redwood	0.745	0.642	2.98	0.165	2.56	0.223
Douglas-fir	1.086	1.364	6.53	0.183	7.81	0.190
Tanoak	1.309	4.356	4.88	0.127	5.78	0.167
Total			14.40	0.099	16.15	0.115
Redwood	1.000	0.000	3.00	0.160	2.57	0.219
Douglas-fir	1.000	1.000	6.53	0.183	7.80	0.191
Tanoak	2.000	2.000	4.99	0.035	5.98	0.073
Total			14.52	0.090	16.35	0.101
STRS estimates						
Redwood	0.000	0.000	2.89	0.177	2.48	0.232
Douglas-fir	0.000	0.000	6.50	0.280	7.77	0.285
Tanoak	0.000	0.000	3.98	0.277	4.77	0.315
Total			13.37	0.164	15.02	0.182

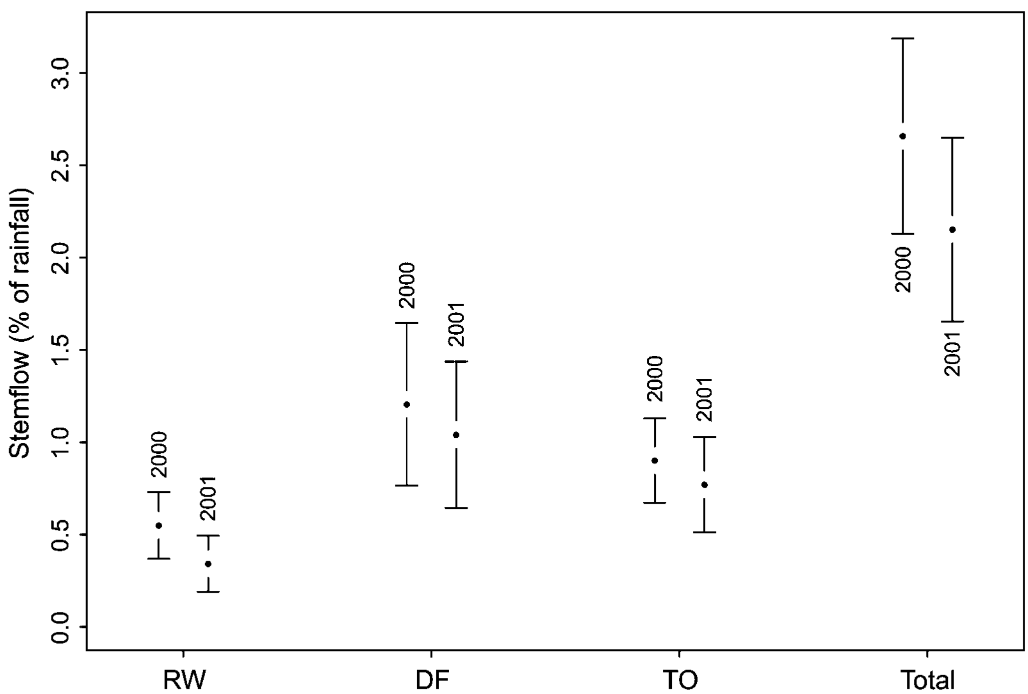

Figure 2. 95% confidence intervals for stemflow percentage of rainfall by measurement period and species: Redwood (RW), Douglas-fir (DF), and tanoak (TO). Water year 2000 (17 Jan, 2000–19 Mar, 2000), and water year 2001 (01 Aug, 2001–01 Apr, 2001).

4. DISCUSSION

For many populations, the equal aggregate σ rule produces strata in which $cv_{\sigma h}$ are approximately equal. However, it is the structure of a particular population that determines how well the rule achieves optimality, so it is a good idea in practice to examine the series of $cv_{\sigma h}$ values before finalizing the stratum boundaries. For all species, the maximum of $cv_{\sigma h}$ occurred in stratum 1 (Table 3). For Douglas-fir and tanoak, $\max[cv_{\sigma h}]$ was more than twice the value of the next largest $cv_{\sigma h}$. Therefore, efficiency (Equation (13)) could have been improved somewhat by lowering the boundaries between strata 1 and 2 for each species to reduce $\max[cv_{\sigma h}]$. This could have been done in the design stage since the diameter information was available. Changing the boundaries to 23, 30 and 26 cm would have raised the efficiencies to 0.97, 0.95 and 0.96, respectively, for redwood, Douglas-fir and tanoak. Further marginal improvement might have been obtained by adjusting other stratum boundaries using a trial-and-error search process. However, any improvement based on this approach would be contingent on a correct model for σ_k, which is difficult to identify with much certainty.

Table 5 suggests that the stemflow estimates are not highly sensitive to the chosen values of λ and γ. It also raises the question of whether a casual inspection of the data might have provided values that result in better estimates of stemflow than those corresponding to the λ and γ from models such as (14) and (16). The annual tanoak stemflow estimate based on $\lambda = \gamma = 2$ had $CV = 0.035$ compared to $CV = 0.127$ when λ and γ were estimated by model (16). However, the former *CV* is very likely be *under*estimated. It is a result of relatively low residual variance that occurred because emphasis was placed on an extreme value from a small data set. Estimating γ from the event-based data obtains a more realistic representation of the dependency of variance on diameter. Incorporating that estimate of γ into the model for annual stemflow provides an appropriate weighting for the data that recognizes increasing variability with diameter and does not place undue emphasis on large trees that can exert strong leverage in the fitting process.

The ratio estimation approach is increasingly advantageous as the dependency on diameter becomes more pronounced. This is evidenced by the increasing disparity between the *CV*s of the ratio and STRS estimates in Table 5 as one moves from redwood to Douglas-fir to tanoak. Fitting a model such as Equation (14) or (16) can help choose appropriate values of λ and γ, but the ratio estimates are not very sensitive to the exact values selected except in the case of tanoak, which had a very small sample size of four trees, one of which was an outlier in the sample of measured trees.

It would have been possible to apply models (14) and (16) to each tree in the stand to estimate stemflow. Such an entirely model-based approach ignores the sampling design. In fact, such an approach does not even require a probability sampling design. However, as mentioned before, the results would be entirely dependent on how well the models actually describe the data. Model-based stratified sampling is really a misnomer; it is actually a *model-assisted* approach that employs design-based formulas, (6)–(12), to achieve results that are valid regardless of the truth of the models that they employ. The validity of the model only improves the efficiency of the method, i.e. a good model will result in low variance estimates.

ACKNOWLEDGEMENTS

The success of the field implementation was largely due to the diligent effort of Elias Steinbuck in the course of his Master's research. Financial support was provided by the California Department of Forestry and Fire Protection and the U.S.D.A. Forest Service, Pacific Southwest Research Station.

REFERENCES

Aboal JR, Morales D, Henrandez M, Jimenez MS. 1999. The measurement and modelling of the variation of stemflow in a laurel forest in Tenerife, Canary Islands. *Journal of Hydrology* **21**: 161–175.

Anderson AR, Pyatt DG. 1986. Interception of precipitation by pole-stage Sitka spruce and lodgepole pine and mature Sitka spruce at Kielder Forest, Northumberland. *Forestry* **59**(1): 29–38.

Asdak C, Jarvis PG, van Gardingen P, Fraser A. 1998. Rainfall interception loss in unlogged and logged forest areas of central Kalimantan, Indonesia. *Journal of Hydrology* **206**: 237–244.

Cape JN, Brown AHF, Robertson SMC, Howson G, Paterson IS. 1991. Interspecies comparisons of throughfall and stemflow at three sites in northern Britain. *Forest Ecology and Management* **46**: 165–177.

Cochran WG. 1977. *Sampling Techniques*, 3rd edn. Wiley & Sons: New York; 132–134.

Crockford RH, Richardson DP. 1990. Partitioning of rainfall in a eucalypt forest and pine plantation of southeastern Australia, II: stemflow and factors affecting stemflow in a dry aclerophyll eucalypt forest and *Pinus radiata* plantation. *Hydrological Processes* **4**: 145–155.

DeWalle DR, Paulsell LK. 1969. Canapy interception, stemflow and streamflow on a small drainage in the Missouri Ozarks. *Forestry Abstracts* **13**(2): 1854.

Ford ED, Deans JD. 1978. The effects of canopy structure on stemflow, throughfall and interception loss in a young Sitka spruce plantation. *Journal of Applied Ecology* **15**: 905–917.

Gash JHC, Morton AJ. 1978. An application of the Rutter model to the estimation of the interception loss from Thetford Forest. *Journal of Hydrology* **38**: 49–58.

Gash JHC, Wright IR, Lloyd CR. 1980. Comparative estimates of interception loss from three coniferous forests in Great Britain. *Journal of Hydrology* **48**: 89–105.

Gash JHC, LLoyd CR, Lachaud G. 1995. Estimating sparse forest rainfall interception with an analytical model. *Journal of Hydrology* **170**: 79–86.

Gregoire TG. 1998. Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forestry Research* **28**: 1429–1447.

Hamilton EL, Rowe PB. 1949. Rainfall interception by chaparral in California published by the State of California under terms of a cooperative agreement with the California Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture: Riverside, California.

Hanchi A, Rapp M. 1997. Stemflow determination in forest stands. *Forest Ecology and Management* **97**: 231–235.

Herwitz SR, Levia DF. 1997. Mid-winter stemflow drainage from bigtooth aspen (*Populus grandidentata* Michx.) in central Massachusetts. *Hydrological Processes* **11**: 169–175.

Johnson RC. 1990. The interception, throughfall, and stemflow in a forest in Highland Scotland and the comparison with other upland forest in the U.K. *Journal of Hydrology* **118**: 281–287.

Lawson ER. 1967. Throughfall and stemflow in a pine-hardwood stand in the Ouachita Mountains of Arkansas. *Water Resources Research* **3**(3): 731–735.

Lewis J, Mori SR, Keppeler ET, Ziemer RR. 2001. Impacts of logging on storm peak flows, flow volumes and suspended sediment loads in Caspar Creek, California. In *Land Use and Watersheds: Human Influence on Hydrology and Geomorphology in Urban and Forest Areas: Water Science and Application*, Vol. 2, Wigmosta MS, Burges SJ (eds). American Geophysical Union: Washington, DC; 85–125.

Llorens P, Poch R, Latron J, Gallart F. 1997. Rainfall interception by a Pinus sylvestris forest patch overgrown in a Mediterranean mountainous abandoned area, I: monitoring design and results down to the event scale. *Journal of Hydrology* **199**: 331–345.

Loustau D, Berbigier P, Granier A, El Hadj Moussa F. 1992. Interception loss, throughfall and stemflow in a maritime pine stand, I: variability of throughfall and stemflow beneath the pine canopy. *Journal of Hydrology* **138**: 449–467.

Särndal C, Swensson B, Wretman J. 1992. *Model Assisted Survey Sampling*. Springer-Verlag: New York.

Spittlehouse DL. 1998. Rainfall interception in young and mature conifer forests in British Columbia. In *Proceedings of 23rd Conference on Agricultural and Forest Meteorology.* Albuquerque: New Mexico; 171–174.

Viville D, Biron P, Granier A, Dambrine E, Probst A. 1993. Interception in a mountainous declining spruce stand in the Strengbach catchment (Vosges, France). *Journal of Hydrology* **144**: 273–282.

Wright RL. 1983. Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association* **78**(384): 879–884.

Ziemer RR. 1998. Flooding and stormflows. In *Proceedings of the Conference on Coastal Watersheds: The Caspar Creek Story*, Ukiah, California, 6 May 1998. *General Tech. Rep. PSW GTR-168*; Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture: Albany, California; 15–24.

