USDA United States Department of Agriculture

Forest Service

Rocky Mountain Research Station

Research Note
RMRS-RN-8

October 2000

Estimating the Number of Tree Species in Forest Populations Using Current Vegetation Survey and Forest Inventory and Analysis Approximation Plots and Grid Intensities

Hans T. Schreuder
Jin-Mann S. Lin
John Teply

Abstract—We estimate number of tree species in National Forest populations, using the nonparametric estimator, $CM2f = c_s + fc_1^2/(2c_2)$, where c_s is the number of species observed, c_1 and c_2 are the numbers of species occurring once and twice, respectively, in the sample plots, and f is the finite population correction for sampling. Data from the Current Vegetation Survey (CVS) of Region 6 of the USDA Forest Service were used to estimate the number of tree species with a plot close in size to the Forest Inventory and Analysis (FIA) plot and the actual CVS plot for the 5.5 km FIA grid and the 2.7 km CVS grid.

Ignoring seedlings, FIA missed seven out of 55 species estimated by the CVS plot on the 2.7 km grid, five with the 5.5 km grid. For observed tree species, FIA missed three species observed on the CVS plot. For the 2.7 km grid including seedlings, FIA missed six out of 53 species, four tallied on the CVS plot were not tallied on the FIA plot. For the 5.5 km grid, FIA missed six estimated to be there with CVS, CVS observed five species not observed by FIA. Ignoring seedlings, the estimate indicates that five species were not observed on the CVS plots, with seedlings two species were not observed, but number of species estimated was less than if seedlings were not considered in estimation. Similar results were obtained by zones and forests except that the results were sometimes erratic because of the small sample sizes involved. Even for the large sample of over 10,000 plots over the whole region, the number of species estimated if seedlings were included was two less (53 versus 55) than if they were ignored in estimation. One more species was actually observed when seedlings were included than when they were ignored (51 versus 50).

There is a definite dilemma. Clearly a first step in protecting rare and endangered species is to determine if they are there. For this purpose the CVS plot and grid intensity are a must at a minimum but still not adequate. To ensure all species are observed requires an even more intensive grid. The problem may be more acute though in Oregon and Washington with the large number of tree species occurring there than in other areas that are more homogeneous in tree species such as the Intermountain West.

Keywords: number of tree species, plot design, plot intensity, nonparametric estimation

Introduction

The USDA Forest Service is charged with inventorying all forest lands in the United States through its Forest Inventory and Analysis (FIA) units. Traditionally interest has been primarily in timber-oriented variables and the focus has been on timberland. However, the information collected can also be used for other ecological variables of interest. One parameter that has been ignored is estimating the number of tree species in a population. The objective of this paper is to apply the estimator CM2f recommended by Schreuder and others (1999) to all National Forests in Region 6 of the Forest Service combined, in zones, and by National Forest using the Current Vegetation Survey (CVS) plot used by Region 6 and a plot more or less equivalent to the current FIA plot used for both the 2.7 km (1.7 mile) grid of the Region and the 5.5 km (3.4 mile) national grid used by FIA to determine what is gained by an increase in plot sample size and grid intensity.

Review of Literature

Considerable literature is available on estimating the number of species in a population. It is usually a difficult prediction problem[1] but may be less so for FIA

Hans T. Schreuder is Mathematical Statistician, Rocky Mountain Research Station, Fort Collins, CO. Jin-Mann S. Lin is a Graduate Research Assistant, Department of Statistics, Colorado State University, Fort Collins, CO. John Teply is a former Forester (retired) Region 6, Portland, OR.

[1]Bunge and Fitzpatrick (1993) note that I. J. Good told them in 1991 that it is usually not possible to estimate the number of unseen species but only a lower bound to that number because there are often many extremely rare species.

and similar surveys because with FIA a probabilistic sample is used to estimate the number of species in a finite rather than a hypothetical population (or situation). Other data sources may be available to support or contradict the findings obtained and lead to improved estimation techniques. In some other disciplines this is not possible, for example, how many words did Shakespeare know or how many different coins by year of issue are still in existence (Efron and Thisted 1976).

Bunge and Fitzpatrick (1993) note that if n sample units are drawn from a population consisting of C classes (species), where C is unknown, then the outcome can be represented by $N^* = (n_1, ..., n_C)$, where n_i is the number of sample units in class i. The observed sample is $c = (c_1, ..., c_n)$ where c_j is the number of classes represented j times in the sample $\{n_i: n_i = j\}$, j = 1, ..., n; and $c_s = \sum_{j=1}^{n} c_j$ is the number of classes in the sample. C is to be estimated based on the sample *c*.

Bunge and Fitzpatrick (1993) summarized the methods available for estimating C. They are not optimistic about using data-analytic methods. They argue that deriving a relationship between an expected number of observed classes and a function of the size of the sample from a sampling model is likely to be less efficient than direct estimation of C by survey sampling methods. Based on a simulation study, Bunge and others (1995) recommend a nonparametric estimator CM3, discussed later, for estimating number of species in a population.

We believe that nonparametric estimators will have to be used to estimate number of tree species because forest populations are frequently disturbed.

For situations where no assumptions are made about the probabilities, π_i, of selecting species, no unbiased estimators exist. And the bias of any estimator of C based on the sample *c* has no bound over all possible populations because there is no way to disprove the existence of an arbitrarily large number of classes with very small probabilities. A nonparametric lower bound estimator (CM2) of C is

$$CM2 = c_s + c_1^2 / (2c_2) \qquad (1)$$

Another nonparametric estimator (CM3) based on the idea of sample coverage is (Chao and Lee 1992):

$$CM3 = \frac{c_s}{1 - c_1/n} + \frac{c_1}{(1 - c_1/n)} \hat{\gamma}^2$$

$$= \frac{n}{n - c_1}\left[c_s + c_1 \hat{\gamma}^2\right] \qquad (2)$$

where $\hat{\gamma}$ is an estimator of γ, the coefficient of variation of the π_i's. Chao and Lee propose CM3 because the coverage of a sample is $\mu = \sum_{i=1}^{c} \pi_i I_{(x_i > 0)}$ where μ = the sum of the probabilities π_i for the c observed classes in the sample and I = 1 if class i is in the sample and = 0 otherwise. Sample coverage (CCOV) can be estimated by $\hat{\mu} = 1 - c_1/n$ under the hypothesis $\mu = c_s/C$. Thus:

$$CCOV = c_s/(1 - c_1/n) \qquad (3)$$

and $CM3 = CCOV + \frac{c_1}{\hat{\mu}} \hat{\gamma}^2$.

Bunge and Fitzpatrick (1993) note that there is some empirical and theoretical evidence that CM3 is preferable to CM2 when used in actual applications.

Schreuder and others (1999) recommend the use of CM2f for use with FIA data where $CM2f = f * CM2 + (1 - f) * c_s$. If $c_2 = 0$, CM2f is undefined. In this case, we set it equal to c_s.

They note that a large sample is necessary for reliable estimation, something on the order of at least 500 to 700 sample units.

Criteria for Selecting a Preferred Estimator

Given the heavy emphasis on protecting rare species, it is desirable to not overestimate the number of tree species in a National Forest. In this way we have some assurance that if our estimate is larger than the observed number of species, c_s, that at least some species were missed in our sample. In addition, although CVS has a large sample of ground plots in each forest, the sample sizes are still small for several forests. Hence, good performance with a (relatively) small sample size is desirable as is consistency over populations. Convergence of estimation to C as $n \rightarrow N$ is guaranteed by adding the finite population correction $f = (N - n)/N$ to each estimator in an appropriate manner. Hence, the criteria recommended for selecting an estimator are:

1. It should not overestimate C.
2. Given several estimators, select the one consistently closest to C for small sample sizes over a wide range of populations.

Given the criteria used and this caveat, Schreuder and others (1999) recommended CM2f in estimating C for a given population, and it was used here. They also note that large sample sizes of the order of 500 to 700 units are desirable and probably essential for reliable estimation of number of species. This is to a considerable degree due to the fact that the nonparametric estimators such as CM2f rely on the number of times a given species occurs only once or twice in the sample to improve the estimation beyond the actual number of species, c_s, observed. For small sample sizes anomalous results can be expected because of that, and when this appears to be the case we will compare the actual number of species observed instead.

FIA and CVS Methods for Estimating C

In FIA we select a cluster of four circular 0.017 ha (1/24 acre) subplots on a 5.5 km (3.4 mile) grid for forest surveys (Birdsey and Schreuder 1992). The four subplots can be considered a sample of a 1 ha primary sample unit. The CVS plot design used by Region 6 of the Forest Service uses a 1 ha primary sample unit subsampled by five locations. At each of the latter, three subplots are located and trees are measured on 0.017 ha (1/24 acre) subplots for 7.6 to 32.8 cm (3.0 to 12.9 inch) trees, 1/5.3 acre (0.076 ha) for trees 32.9 to 81.0 cm (13.0 to 31.9 inch) trees on the east side and 32.9 to 121.9 cm (13.0 to 47.9 inches) on the west side, and trees larger than 81.0 cm (32.0 inches) on the east side or larger than 121.9 cm (48.0 inches) on the west side (Max and others 1996). Seedlings—trees less than 7.4 cm (3.0 inches)—are tallied on 0.004 ha (1/100 acre) subplots. Every tree selected is identified by species, and n plots are selected with each plot (cluster) of trees containing various tree species.

The FIA plot was not implemented on the 1 ha CVS plots, but we approximated it by considering it equivalent to sampling a 1/5 ha subplot of the CVS plot, representing roughly 1/5 of the CVS plot.

All CVS plots available, 10,567 plots, were used. In addition this was broken down by individual National Forests, which in turn were combined into six ecological zones recognized as such by the Region.

Results

Assuming the subsampled 1/5 ha plot selected at random from the five 1/5 ha subplots in the CVS plot to be equivalent to an FIA plot and the subsampled 1 ha plot as the CVS plot, we conclude the following using the 2.7 km grid ignoring seedlings (table 1). Clearly we will have more confidence in the results for the larger sample sizes and we start with those:

1. All forests combined for a sample size of 10,567 plots.
   - Missed 7 out of 55 species using FIA plot relative to the CVS plot using CM2f.
   - 5 of the species were missed with the CVS plot at this intensity of sampling.
   - Three of the species tallied on the CVS plot were not tallied on the FIA plot.

Table 1—Estimating number of species with various subplot sizes (with observed number in parentheses) using CM2f on a 2.7 km grid intensity.

Forest	1/5 ha	2/5 ha	3/5 ha	4/5 ha	1 ha	Missed	Sample
Deschutes	24(19)	19(19)	20(20)	23(21)	31(23)	7(4)	717
Fremont	13(12)	13(13)	13(13)	13(13)	14(13)	1(1)	570
Gifford Pinchot	23(23)	24(24)	25(24)	25(24)	27(25)	4(2)	602
Malheur	11(11)	11(11)	12(12)	12(12)	12(12)	1(1)	720
Mt. Baker-Snoqualmie	17(17)	17(17)	21(20)	21(21)	22(21)	5(4)	580
Mt. Hood	22(22)	24(23)	23(23)	26(25)	25(25)	3(3)	477
Ochoco	9(9)	9(9)	9(9)	9(9)	9(9)	0(0)	443
Okanogan	19(19)	19(19)	19(19)	19(19)	21(20)	2(1)	581
Olympic	18(18)	20(19)	20(20)	20(20)	23(21)	5(3)	287
Rogue River	25(25)	32(27)	29(27)	28(28)	30(30)	5(5)	299
Siskiyou	35(31)	32(31)	32(31)	34(32)	32(32)	–3(1)	482
Siuslaw	13(13)	28(15)	15(15)	20(16)	21(17)	8(4)	315
Umatilla	11(11)	12(11)	11(11)	11(11)	11(11)	0(0)	609
Umpqua	33(29)	31(30)	36(31)	35(33)	34(33)	1(4)	481
Wallowa-Whitman	13(12)	12(12)	14(13)	13(13)	13(13)	0(1)	887
Wenatchee	28(24)	25(24)	26(25)	28(26)	36(30)	8(6)	707
Willamette	28(27)	27(27)	29(28)	28(28)	28(28)	0(1)	734
Winema	16(14)	14(14)	14(14)	15(15)	15(15)	–1(1)	510
Colville	14(14)	14(14)	16(15)	15(15)	15(15)	1(1)	566
Ecological Zone 1	25(25)	26(26)	26(26)	26(26)	28(28)	3(3)	1,469
Ecological Zone 2	32(30)	31(31)	32(32)	38(33)	35(33)	3(3)	1,526
Ecological Zone 3	36(36)	36(36)	37(37)	38(38)	41(39)	5(3)	1,262
Ecological Zone 4	21(21)	22(21)	22(22)	32(24)	38(25)	17(4)	2,240
Ecological Zone 5	14(14)	14(14)	14(14)	14(14)	14(14)	0(0)	2,216
Ecological Zone 6	27(25)	25(25)	27(26)	29(27)	37(31)	10(6)	1,854
All forests	48(47)	49(49)	49(49)	50(50)	55(50)	7(3)	10,567

Note: The c_s values are listed in the parentheses. FIA samples at about 1/5 the rate of R6 CVS, about 1/5 ha versus 1 ha.

2. For zones 1 through 6, with sample sizes ranging from 1,262 to 2,240:
   - 0-17 species were missed with FIA plot relative to the CVS plot using CM2f.
   - 0-6 were missed in actual observations.
   - CVS missed from 0-6 species estimated to be there with CM2f.
3. For individual forests with sample sizes ranging from 287 to 887:
   - The FIA plot missed from 0-8 species relative to CVS with CM2f with the exception of the Siskiyou and Winema forests, where the FIA plot estimates, respectively, 3 and 1 more species than CVS. This is likely due to the small sample size yielding odd results at times, depending on the number of times 1 species is observed only once or twice.
   - When looking at the actual observed number of species, certainly a safer statistic in this case of smaller sample sizes, 0-6 species are missed, 1 for the Siskiyou and 1 for the Winema with the FIA plot relative to the CVS plot.
   - Estimation with the CVS plot using CM2f indicates that up to 8 species were not observed in the field.

When we use the 5.5 km FIA grid, we get the following equivalent results to the above (table 2):

1. For all forests, with a sample size of 3,045 plots:
   - FIA missed 5 species using both CM2f and the actual number of species observed relative to the CVS plot at this grid intensity.
   - CVS observes all 47 species estimated to be there with CM2f.
2. For ecological zones 1 through 6, with sample sizes of 361 to 637:
   - From 0 to 16 species were missed with FIA relative to CVS with CM2f estimation except that 17 extra species were estimated with FIA for zone 1.
   - With actual number of observed species, FIA missed 0 to 4 species observed by the CVS plot.
3. For individual forests, with sample sizes of 79-266:
   - From 0 to 22 species were missed by FIA CM2f estimation relative to CVS CM2f estimation except for -2 for the Deschutes, Olympic, and Siskiyou.
   - With these small sample sizes it is clearly better to rely on only the observed species,

Table 2—Estimating number of species with various subplot sizes using CM2f (with observed number in parentheses) on a 5.5 km grid intensity.

Forest	⅕ ha	⅖ ha	⅗ ha	⅘ ha	1 ha	Missed	Sample
Deschutes	19(17)	17(17)	17(17)	17(17)	17(17)	–2(0)	187
Fremont	11(11)	11(11)	11(11)	11(11)	11(11)	0(0)	143
Gifford Pinchot	21(21)	22(23)	22(24)	22(22)	22(22)	1(1)	168
Malheur	9(9)	9(9)	10(10)	10(10)	12(11)	3(2)	188
Mt. Baker-Snoqualmie	17(15)	18(16)	21(19)	19(19)	20(19)	3(4)	197
Mt. Hood	19(18)	20(20)	21(21)	22(22)	22(22)	3(4)	140
Ochoco	7(7)	7(7)	7(7)	7(7)	7(7)	0(0)	114
Okanogan	18(18)	19(19)	20(19)	19(19)	19(19)	1(1)	207
Olympic	19(16)	16(16)	18(17)	17(17)	17(17)	–2(1)	80
Rogue River	33(25)	30(27)	35(27)	32(27)	33(28)	0(3)	87
Siskiyou	28(26)	27(26)	27(26)	26(26)	26(26)	–2(0)	143
Siuslaw	9(9)	10(10)	10(10)	12(12)	12(12)	3(3)	79
Umatilla	9(9)	9(9)	9(9)	9(9)	10(10)	1(1)	183
Umpqua	21(21)	24(24)	33(25)	33(27)	28(28)	7(7)	131
Wallowa-Whitman	10(10)	10(10)	11(11)	11(11)	11(11)	1(1)	266
Wenatchee	21(21)	22(21)	27(22)	27(22)	43(25)	22(4)	236
Willamette	25(23)	26(25)	27(25)	25(25)	25(25)	0(2)	213
Winema	14(13)	14(14)	14(14)	14(14)	14(14)	0(1)	137
Colville	13(13)	13(13)	13(13)	13(13)	13(13)	0(0)	146
Ecological Zone 1	42(24)	24(24)	26(25)	25(25)	25(25)	–17(1)	445
Ecological Zone 2	26(26)	28(28)	29(29)	32(30)	31(30)	5(4)	432
Ecological Zone 3	33(32)	36(34)	34(34)	35(35)	35(35)	2(3)	361
Ecological Zone 4	18(18)	18(18)	18(18)	18(18)	18(18)	0(0)	581
Ecological Zone 5	11(11)	11(11)	12(12)	12(12)	13(13)	2(2)	637
Ecological Zone 6	23(23)	24(23)	26(24)	26(24)	39(26)	16(3)	589
All forests	44(43)	54(46)	46(46)	47(47)	47(47)	5(4)	3,045

Note: The c_s values are listed in the parentheses. FIA samples at about ⅕ the rate of R6 CVS, about ⅕ ha versus 1 ha.

and there the FIA plot misses from 0 to 7 species.

When we also include seedlings, we obtain the following results for the 2.7 km grid (table 3).

1. All forests combined. For this sample of size 10,720 plots:
   - We miss 6 out of 53 species using the FIA plot relative to the CVS plot using CM2f.
   - 2 of the species were missed with the CVS plot at this intensity of sampling.
   - 4 of the species tallied on the CVS plot were not tallied on the FIA plot.
2. For zones 1 through 6, with sample sizes ranging from 1,275 to 2,277:
   - 1 to 5 species were missed with FIA plot relative to the CVS plot using CM2f; 1 to 5 were missed in actual observations.
   - CVS missed from 0 to 2 species estimated to be there with CM2f.
3. For individual forests with sample sizes ranging from 289 to 910:
   - The FIA plot misses from 0 to 21 species relative to CVS with CM2f with the exception of the Deschutes, Siskiyou, and Umpqua Forests, where the FIA plot estimates, respectively 10, 1, and 17 more species than CVS. This is likely due to the small sample size as indicated earlier.
   - When looking at the actual observed number of species 0-6 species are missed, 3 for the Deschutes, 1 for the Siskiyou, and 5 for the Umpqua.
   - Estimation with the CVS plot using CM2f indicates that up to 18 species may not have been observed in the field.

When we use the 5.5 km FIA grid, we get the following equivalent results to the above (Table 4):

1. For all forests, with a sample size of 3083 plots:
   - FIA respectively misses 6 and 5 species using CM2f by the actual number of species observed relative to the CVS plot at this grid intensity.
   - CVS observes 49 out of 53 species estimated to be there with CM2f.
2. For ecological zones 1 through 6, with sample sizes of 365 to 649:
   - From 0 to 7 species were missed with FIA relative to CVS with CM2f estimation except

Table 3—Estimating number of species with various subplot sizes using CM2f (with observed number in parentheses) on a 2.7 km grid intensity—includes seedlings.

Forest	1/5 ha	2/5 ha	3/5 ha	4/5 ha	1 ha	Missed	Sample
Deschutes	34(21)	24(22)	32(24)	29(24)	24(24)	–10(3)	724
Fremont	13(12)	13(13)	13(13)	13(13)	14(13)	1(1)	578
Gifford Pinchot	24(24)	27(26)	27(27)	27(27)	27(27)	3(3)	616
Malheur	12(12)	12(12)	12(12)	12(12)	12(12)	0(0)	738
Mt. Baker-Snoqualmie	25(20)	22(20)	24(22)	25(23)	25(23)	0(3)	587
Mt. Hood	27(27)	32(28)	35(30)	32(30)	34(30)	7(3)	483
Ochoco	9(9)	9(9)	9(9)	9(9)	9(9)	0(0)	459
Okanogan	22(22)	24(22)	22(22)	22(22)	23(23)	1(1)	583
Olympic	18(18)	22(20)	20(20)	20(20)	22(21)	4(3)	289
Rogue River	29(27)	29(29)	29(29)	30(30)	50(32)	21(5)	303
Siskiyou	34(32)	32(32)	32(32)	33(33)	33(33)	–1(1)	486
Siuslaw	13(13)	23(15)	15(15)	16(16)	25(17)	12(4)	320
Umatilla	11(11)	12(12)	12(12)	12(12)	12(12)	1(1)	613
Umpqua	57(31)	34(32)	41(33)	33(33)	40(36)	–17(5)	486
Wallowa-Whitman	12(12)	14(13)	15(14)	14(14)	16(15)	4(3)	910
Wenatchee	29(25)	25(25)	27(26)	30(28)	35(31)	6(6)	715
Willamette	27(27)	27(27)	28(28)	29(29)	30(30)	3(3)	742
Winema	14(14)	14(14)	14(14)	15(15)	15(15)	1(1)	516
Colville	16(15)	15(15)	16(16)	16(16)	17(17)	1(2)	572
Ecological Zone 1	27(27)	30(29)	31(29)	29(29)	30(30)	3(3)	1,492
Ecological Zone 2	33(33)	42(34)	42(36)	43(37)	38(37)	5(4)	1,545
Ecological Zone 3	40(38)	38(38)	39(39)	40(39)	42(42)	2(4)	1,275
Ecological Zone 4	23(23)	27(24)	33(25)	34(26)	27(26)	4(3)	2,277
Ecological Zone 5	14(14)	14(14)	14(14)	14(14)	15(15)	1(1)	2,261
Ecological Zone 6	31(28)	29(28)	31(29)	35(31)	35(33)	4(5)	1,870
All forests	47(47)	49(49)	51(49)	56(51)	53(51)	6(4)	10,720

Note: The c_s values are listed in the parentheses. FIA samples at about 1/5 the rate of R6 CVS, about 1/5 ha versus 1 ha.

Table 4—Estimating number of species with various subplot sizes using CM2f (with observed number in parentheses) on a 5.5 km grid intensity—includes seedlings.

Forest	1/5 ha	2/5 ha	3/5 ha	4/5 ha	1 ha	Missed	Sample
Deschutes	20(18)	20(19)	19(19)	19(19)	19(19)	–1(1)	189
Fremont	12(11)	11(11)	11(11)	11(11)	11(11)	–1(0)	146
Gifford Pinchot	28(22)	22(22)	23(23)	24(23)	23(23)	–5(1)	171
Malheur	12(11)	11(11)	11(11)	11(11)	11(11)	–1(0)	193
Mt. Baker-Snoqualmie	16(16)	17(17)	20(19)	19(19)	19(19)	3(3)	198
Mt. Hood	22(22)	28(24)	38(25)	25(25)	30(25)	8(3)	142
Ochoco	7(7)	7(7)	7(7)	7(7)	7(7)	0(0)	117
Okanogan	23(21)	22(22)	22(22)	24(22)	24(22)	1(1)	207
Olympic	16(16)	16(16)	18(17)	18(18)	19(18)	3(2)	80
Rogue River	28(25)	29(27)	27(27)	29(27)	28(28)	0(3)	88
Siskiyou	28(26)	28(26)	29(27)	28(27)	28(28)	0(2)	145
Siuslaw	10(10)	11(11)	11(11)	12(12)	12(12)	2(2)	81
Umatilla	9(9)	11(11)	11(11)	12(11)	11(11)	2(2)	183
Umpqua	31(23)	30(25)	27(25)	28(28)	28(28)	–3(5)	132
Wallowa-Whitman	11(11)	11(11)	12(12)	12(12)	12(12)	1(1)	273
Wenatchee	41(23)	23(23)	25(24)	26(25)	28(26)	–13(3)	239
Willamette	24(23)	25(25)	25(25)	26(26)	26(26)	2(3)	215
Winema	13(13)	14(14)	14(14)	14(14)	14(14)	1(1)	137
Colville	13(13)	14(14)	14(14)	16(15)	18(16)	5(3)	147
Ecological Zone 1	24(24)	24(24)	26(26)	26(26)	26(26)	2(2)	449
Ecological Zone 2	30(28)	30(30)	31(31)	32(32)	37(32)	7(4)	438
Ecological Zone 3	34(32)	36(34)	35(35)	37(35)	35(35)	1(3)	365
Ecological Zone 4	19(19)	22(20)	20(20)	20(20)	20(20)	1(1)	589
Ecological Zone 5	13(13)	13(13)	13(13)	13(13)	13(13)	0(0)	649
Ecological Zone 6	39(26)	27(26)	29(27)	30(28)	29(28)	–10(2)	593
All forests	46(44)	60(47)	55(47)	62(49)	53(49)	6(5)	3,083

Note: The c_s values are listed in the parentheses. FIA samples at about 1/5 the rate of R6 CVS, about 1/5 ha versus 1 ha.

that 10 extra species were estimated with FIA for zone 6.

- With actual number of observed species, FIA missed 0-4 species observed by the CVS plot.

3. For individual forests, with sample sizes of 80 to 273:
   - From 0 to 8 species were missed by FIA CM2f estimation relative to CVS CM2f estimation except for -1 for the Deschutes, Fremont, and Malheur forests, 5 for the Gifford-Pinchot, 3 for the Umpqua, and 13 for the Wenatchee.
   - Considering only observed species, the FIA plot missed from 0 to 5 species.
   - Note that the number of species estimated by CM2f including seedlings is smaller than ignoring seedlings even though 1 more species is actually observed when seedlings are included. This too illustrates the problem of small sample sizes. Even though a sample of over 10,000 plots is usually considered very large, it can still cause problems for rare events such as estimating the presence of a rare tree species.

Discussion

The FIA plot size and grid intensity is clearly not as good in Region 6, Oregon and Washington, to estimate the number of tree species relative to the CVS plot and grid intensity. Three species tallied on the CVS plot were not observed on the FIA plot sampling both on the 2.7 km grid. FIA misses seven out of 55 species estimated to be there with the CVS CM2f estimator but the CVS plot only observed 50 of those species. With the 5.5 km grid, FIA missed five of the species observed on the CVS plot whereas at this intensity the CVS plots observe all 47 species estimated to be there with the CM2f estimator. Region 6 is rich in tree species, so the problem may not be as acute in several other regions such as the Intermountain West, but a similar problem with missed species may well occur in the South. Clearly there is a dilemma from an ecological point of view in several parts of the country. A first step in protecting rare and endangered species is to determine if they are there, and even the CVS plot with its 2.7 km intensity grid does not accomplish that. To ensure that all species are observed requires an even more intensive grid or perhaps a complete 1 ha tally on the CVS

plots, neither one of which appears practical with current funding.

The results show that a large sample size is critical. With both plots and grids, the CM2f estimator in several cases estimated a substantial number of species more than observed for small sample sizes. Ideally one would not estimate number of tree species for small sample sizes such as for the individual Forests even with the CVS plot and grid. But if that is not possible, it is clearly preferred to use the observed number of species over CM2f because the latter can be quite misleading for such sample sizes.

References

Birdsey, R. A.; Schreuder, H. T. 1992. An overview of forest inventory and analysis estimation procedures in the eastern U.S.—with an emphasis on components of change. USDA Forest Service, Rocky Mountain Research Station. Gen. Tech. Rep. RM-214. 11 p.

Bunge, J.; Fitzpatrick, M. 1993. Estimating the number of species: a review. J. Amer. Stat. Assoc. 88: 364-373.

Bunge, J.; Fitzpatrick, M.; Handley, J. 1995. Comparison of three estimators of the number of species. J. Appl Stat. 22:45-59.

Chao, A.; Lee, S-M. 1992. Estimating the number of classes via sample coverage. J. Amer. Stat. Assoc. 87: 210-217.

Efron, B.; Thisted, R. 1976. Estimating the number of unseen species: how many words did Shakespeare know? Biometrika 63: 435-447.

Max. T.A.; Schreuder, H.T.; Hazard, J.W.; Teply, J.; Alegria, J. 1996. The Region 6 Vegetation Inventory and Monitoring System. USDA Forest Service, Pacific Northwest Station. Res. Pap. PNW-RP-493. 22 p.

Schreuder, H.T.; Williams, M.S.; Reich, R. 1999. Estimating the number of tree species in a forest population using survey data. Env. Mon. and Asst. 56:293-303.

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please specify the publication title and Research Note number.

Fort Collins Service Center

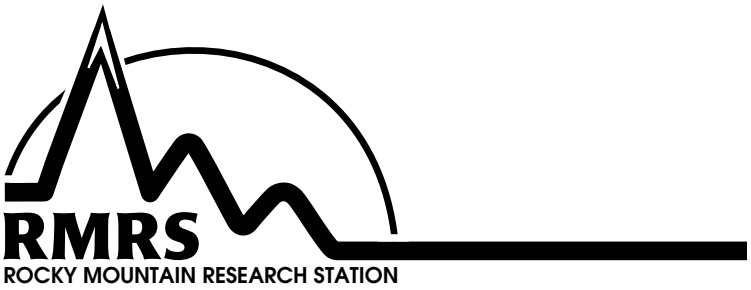
Telephone (970) 498-1392

FAX (970) 498-1396

E-mail rschneider@fs.fed.us

Web site http://www.fs.fed.us/rm

Mailing Address Publications Distribution
Rocky Mountain Research Station
240 West Prospect Road
Fort Collins, CO 80526



The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of the forests and rangelands. Research is designed to meet the needs of National Forest managers, Federal and State agencies, public and private organizations, academic institutions, industry, and individuals.

Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications may be found worldwide.

Research Locations

Flagstaff, Arizona
Fort Collins, Colorado*
Boise, Idaho
Moscow, Idaho
Bozeman, Montana
Missoula, Montana
Lincoln, Nebraska
Reno, Nevada
Albuquerque, New Mexico
Rapid City, South Dakota
Logan, Utah
Ogden, Utah
Provo, Utah
Laramie, Wyoming

*Station Headquarters, Natural Resources Research Center,
2150 Centre Avenue, Building A, Fort Collins, CO 80526