# The Poor Man's Geographic Information System: Plot Expansion Factors

**Paul C. Van Deusen**[1]

**Abstract.—**Plot expansion factors can serve as a crude Geographic Information System for users of Forest Inventory and Analysis (FIA) data. Each FIA plot has an associated expansion factor that is often interpreted as the number of forested acres that the plot represents. The derivation of expansion factors is discussed and it is shown that the mapped plot design requires a different expansion factor than the old periodic plot design.

## Introduction

Plot expansion factors are important to most users of U.S. Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) data. Expansion factors are typically viewed as the number of forested acres that a plot represents. In fact, they are constructed to ensure that the sum of all expansion factors for plots in a particular county is equal to FIA's current estimate of the forested acres in the county. Therefore, there is some basis in reality for this view of expansion factors. A potential downside to the standard view is the implication that somehow the plot with the largest expansion factor is most important, i.e., it should be given the most weight in an analysis. This is not generally true because all plots were initially selected with equal probability. Plots may receive different weights due to a poststratification process that is performed in some regions, but this has nothing to do with initial selection probability. The purpose here is to show how expansion factors are derived and to point out the implications of using them in several common analyses.

## Expansion Factor Computation

The rule of thumb that leads to the formula for plot expansion factors is that expansion factors within a county must sum to the forest area in the county. This leads immediately to

$$w_i = A_c(i) / n_c(i) \qquad (1)$$

where $A_c(i)$ and $n_c(i)$ are the forest area and number of plots in county c, which contains plot i. The same rule applies for poststratification with the additional proviso that the sum of expansion factors in stratum h and county c must sum to the stratum area in the county. Poststratified weights are then computed as

$$w_i = A_{c,h}(i) / n_{c,h}(i) \qquad (2)$$

where $A_{c,h}(i)$ and $n_{c,h}(i)$ are the stratum h forest area and number of plots in county c which contains plot i.

Expansion factors computed with equation (1) should not be used in an analysis of the data to give varying weights to the plots. They have no relationship to the plot selection probability and do not indicate the importance of the plot in any way. This is not necessarily true for the poststratified weights (equation 2). Some FIA regions use these weights to compute stratified means and variances (Bechtold and Patterson 2005, Van Deusen 2005).

## Computations With Poststratified Weights

Plot expansion factors can provide a clever mechanism for users to compute stratified means and variances using standard formulas for weighted means. The generic formula for the stratified sample mean (Thompson 2002, Cochran 1977) is

$$\bar{y}_{st} = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{y}_h \qquad (3a)$$

where:
N = total population size.
$N_h$ = the number of population elements in stratum h.

[1] Principal Research Scientist, National Council for Air and Stream Improvement, Inc., 600 Suffolk Street, Fifth Floor, Lowell, MA 01854. E-mail: PVanDeusen@ncasi.org.

$\bar{y}_h$ = the estimated mean for stratum h.

An unbiased estimator of the stratified variance is

$$v\left(\bar{y}_{st}\right) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{s_h^2}{n_h} \tag{3b}$$

Equations (3a) and (3b) are formulated in terms of total population size (N) for the area of interest, but N is unknown in most forest inventory settings. This equation can be converted to a useful form for FIA data by ignoring the finite population correction factor and replacing Ns with As to get

$$\bar{y}_{st} = \frac{1}{A} \sum_{h=1}^{L} A_h \bar{y}_h , \tag{4a}$$

and

$$v\left(\bar{y}_{st}\right) = \sum_{h=1}^{L} \left(\frac{A_h}{A}\right)^2 \frac{s_h^2}{n_h} \tag{4b}$$

where the variance estimate for plots in the same stratum is

$$s_h^2 = \sum_{j \varepsilon h} \left(y_j - \bar{y}_h\right)^2 / \left(n_h - 1\right) \tag{4c}$$

Equations (4a) and (4b) are based on the assumption that the stratum areas are known. In practice, these values might come from counting pixels on a classified raster image that covers the region of interest. For the purposes here, assume that the variance of area estimates is small enough to ignore. A further complication is that equation (4c) must be recomputed for each county and stratum and there may not always be sufficient observations in all county strata.

It is easy to show (Van Deusen 2005) that equations (4a) and (4b) are implemented for unmapped plots with the following weighted mean and variance equations:

$$\bar{y}_{st} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} , \tag{5a}$$

and

$$V\left(\bar{y}_{st}\right) = \frac{\sum_{i=1}^{n} w_i^2 \sigma_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2} \tag{5b}$$

where $\sigma_i^2$ is estimated from equation (4c). The county and stratum of the plots can be ignored for equation (5a) because this data is incorporated into the expansion factors. The variance ($\sigma_i^2$) in equation (5b), however, changes for each county and stratum. This variance could be problematic if sample size in a particular stratum is small. Regardless, expansion factors are a very useful component of the FIA database (FIADB) for the old periodic data.

It can be shown (Van Deusen 2005) that when working with mapped plots the weights must be changed to $w_i = A_{c,h}\left(i\right) a_i / \sum_{j \varepsilon c,h} a_j$ where $a_i$ is the proportion of plot i that falls into the condition of interest. This means that the weights must be modified for each mapped condition. This eliminates the simplicity of working with poststratified data in the FIADB and only a very committed user should attempt to use the weights to compute stratified estimates with mapped plot data. The weighted estimation formulas (5a) and (5b) also change for mapped plots (Van Deusen 2005).

## Conclusions

Plot expansion factors provide users with information on forested area by county. In this sense, they serve the same purpose as a Geographic Information System. Users need to understand, however, that these weights were derived to ensure that the expansion factors for all plots in a county add to the forest area in the county. Trying to extract additional information by using expansion factors as weights in a statistical analysis of the plot data is not justified. Some FIA regions incorporate poststratification information into plot expansion factors that enable one to extract the forest area by strata for a county. This information is difficult to properly use for mapped plot data and most users of the FIADB should probably ignore it.

## Literature Cited


Bechtold, W.A.; Patterson, P.L. 2005. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. New York: John Wiley.

Thompson, S.K. 2002. Sampling. 2nd ed. New York: John Wiley.

Van Deusen, P.C. 2005. Stratified forest inventory estimation with mapped plots. Canadian Journal of Forest Research. 35: 2382-2386.