

Standardization, Residualization, and Adjusted Statistical Thresholds for Forest Health Indicators

William A. Bechtold ¹

Stanley J. Zarnoch ¹

Michael E. Schomaker ²

¹ USDA Forest Service, Southern Research Station, Asheville, NC
² Colorado State Forest Service, Ft. Collins, CO

Corresponding author:

W.A. Bechtold, USDA Forest Service, Southern Research Station,
 P.O. Box 2680, Asheville, NC 28802
 Phone: 828-257-4357 e-mail: wabechtold@fs.fed.us

Example

Composite Crown Volume (CCV), a composite indicator derived from multiple tree crown parameters, was computed for 6,179 individual trees located on 250 FHM plots distributed across Virginia, Georgia, and Alabama:

$$CCV = 0.5 \pi R^2 CL CD \quad (4)$$

where

$$\pi = 3.14159,$$

$$R = \text{crown diameter} / 2,$$

$$H = \text{total tree length},$$

$$CL = H (\text{crown ratio}) / 100, \text{ and}$$

$$CD = (\text{crown density} / 100)(1 - \text{dieback} / 100)(1 - \text{transparency} / 100)$$

Tree-level CCV Indicators

Raw and Standardized Values

Trees were then grouped by species. Mean CCV differed substantially by species, ranging from 697 cu ft (slash pine) to 6,422 cu ft (American beech), with coefficients of variation often exceeding 100%, thus demonstrating the need to standardize across species. Equation (1) was used to obtain adjusted tree-level CCV values that were standardized, by species, to a mean of zero and a standard deviation of one. Indicators standardized in this manner reflect the deviation of all trees from their species mean (in terms of standard deviation units) and are invariant to the species distribution on the plot. At this point, it is valid to combine or compare data across species, since all tree-level values are now on the same basis with respect to species.

Residualized and Standardized-Residualized Values

The distributional properties of CCV in its raw (equation 4, figure 1a) and standardized (equation 1, figure 1b) forms are shown in figure 1. Although standardization results in a mean of zero and standard deviation of one, note that standardization alone has little effect on skewness. Skewness is often improved by residualization with models, so the following model designed to adjust CCV for tree size and stand density was employed:

$$CCV = b_0 + b_1 (dbh) + b_2 (BA) \quad (5)$$

where

dbh = tree diameter (inches) at breast height,

BA = stand-level basal area per acre, and

b_i = regression parameters estimated from the data.

Note that the residuals (equation 2, figure 1c) and standardized residuals (equation 3, figure 1d) associated with the model (equation 5) have substantially smaller skewness coefficients and are more normally distributed than standardization alone (figure 1b). Reduction in skewness is advantageous when performing statistical tests based on assumptions of normality. It also diminishes reliance on transformations and nonparametric statistical methods.

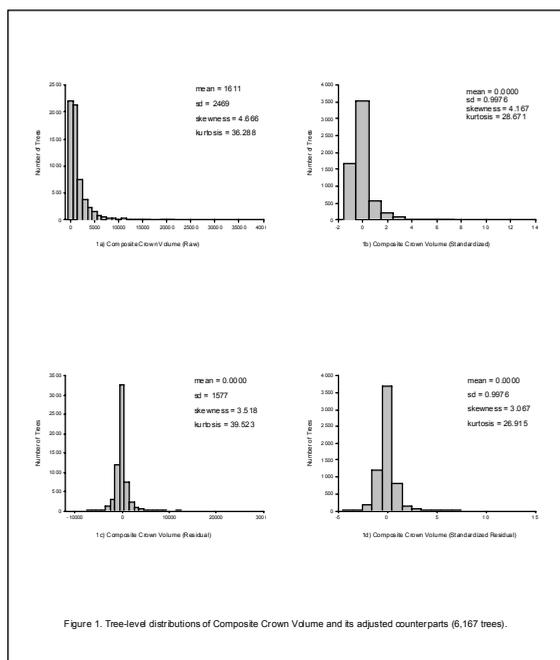


Figure 1. Tree-level distributions of Composite Crown Volume and its adjusted counterparts (6,167 trees).

Abstract. Recent forest health reports have focused on arbitrary thresholds for indicator values that are not adjusted for differences among species. To circumvent this difficulty, we show how indicators can be standardized, by species, to a mean of zero and variance of one. This enables direct comparison among species and permits tree-level indicators to be summed across species to yield plot-level (i.e., stand-level) indicators. We also demonstrate how to produce residualized indicators (which can also be standardized) that adjust for tree size and local stand conditions. Indicators based on crown volume are given as examples.

Thresholds and Statistical Distributions

Thresholds are important when assessing forest health because they separate the sampled population into categories of good and poor. Ideally, thresholds should be developed on a biological basis. Biological thresholds, the point at which a tree becomes noticeably stressed and begins to decline, are difficult to pinpoint. This requires extensive research designed to establish correlations between indicator variables and other signs of tree stress such as damage symptoms, reduced growth, and prospective mortality. Threshold establishment is further complicated because thresholds are often species dependent, and because the effect of normal stand dynamics must be partitioned from the analysis. The ultimate goal in establishing any threshold is to identify signals that appear to be beyond the range of what is expected.

Statistical thresholds have advantages and disadvantages when compared to biological thresholds. They are disadvantaged because they are somewhat arbitrary, and always result in a set of observations designated as poor, even in the absence of a problem. On the other hand, they are easily established by isolating observations at the tails of statistical distributions, and they are quite useful for detecting spatial patterns, establishing empirical correlations, and measuring change over time.

Since biological thresholds have yet to be developed, arbitrary thresholds for some indicators (especially visual crown ratings) have been used in a variety of recent forest health reports. Such analyses have the virtue of simplicity, but contain no adjustment for species differences, other tree attributes, or stand conditions. These types of adjustments and related analyses can be accomplished with statistical distributions and thresholds. This poster utilizes data from the Crown Indicator to demonstrate how such adjustments are accomplished by standardization and residualization of indicator values. These techniques work for individual indicator variables, as well as composite values that may be derived from multiple indicator variables.

Standardized and Residualized Indicator Values

Standardized Indicators

An indicator can be adjusted for different statistical distributions among species by standardizing to a mean of zero and a standard deviation of one. Values are thus expressed in terms of standard deviation units from the mean for a given species. This adjustment results in more meaningful interpretations when comparisons are made among species, and enables summations across species to yield stand-level indicators. The standardization of an indicator is defined as

$$I'_{ij} = \frac{I_{ij} - \bar{I}_j}{s_j} \quad (1)$$

where

I'_{ij} = the standardized indicator for tree i within species j ,

I_{ij} = the non-standardized indicator for tree i within species j ,

\bar{I}_j = the average for the non-standardized indicator for species j , and

s_j = the standard deviation for the non-standardized indicator for species j .

Residualized Indicators

Another method to adjust an indicator is to define the indicator as its residual from a model based on tree and/or stand conditions. Each tree is thus adjusted for its specific competitive situation, resulting in an indicator more suitable for detecting abnormalities because such adjustment identifies individuals that do not conform to model predictions. Specifically, let

Y_{ij} = an indicator for tree i within species j , and

\hat{Y}_{ij} = the predicted value of the indicator for tree i within species j based on the appropriate model.

The associated residualized indicator is then defined as

$$R_{ij} = Y_{ij} - \hat{Y}_{ij} \quad (2)$$

Standardized-Residualized Indicators

Like raw indicator values, residualized indicators can be standardized for comparisons across species:

$$R'_{ij} = \frac{R_{ij} - \bar{R}_j}{s_j} \quad (3)$$

where

R'_{ij} = the standardized-residualized indicator for tree i within species j ,

R_{ij} = the residualized indicator for tree i within species j ,

\bar{R}_j = the average for the residualized indicator for species j , and

s_j = the standard deviation for the residualized indicator for species j .

Stand-level CCV Indicators

Standardization of indicators makes it feasible to produce plot-level indicators (by averaging tree-level values across all trees on each plot). Plot-level averages (across all species) for 250 plots were calculated from the raw and standardized-residualized tree-level CCV values. Distributions of these stand-level means are shown in figures 2a and 2b. Note that the standardized residuals do not have a mean of 0 and standard deviation of 1 (as seen with tree level-indicators). This is because standardization was initially performed at the individual tree and species level, and species distributions differ across the plots. As with the tree-level indicators, residualization reduces the skewness coefficients of the stand-level indicator.

Further comparison of average raw stand-level values and their standardized-residualized counterparts was performed by classifying 250 plots into good and poor categories. For demonstration, a threshold for the poor class was set at the low or 5-percentile of the statistical distribution. Each pair of raw and standardized-residual values was then plotted on a scatter diagram, with the threshold value indicated by reference lines (figure 3). Agreement between raw values and their standardized residuals is attained for all plots located in the upper right and lower left quadrants. The other two quadrants represent opposite classifications by the indicators. The raw indicator and its standardized-residual counterpart classify the same plots as poor only twice, confirming that the raw and adjusted values are measuring different aspects of crown condition.

The reason for differences in classification is based on the adjustment potential of the regression models and its effect on the creation of the standardized-residual indicators. When using raw CCV values, stands with high percentages of trees with small crowns are classified as poor, even if crowns are normally small for those particular species in those types of stands. However, when using the adjusted values, stands are classified as poor only if they have high percentages of trees at the low end of their respective species statistical distributions after adjusting for tree and stand conditions.

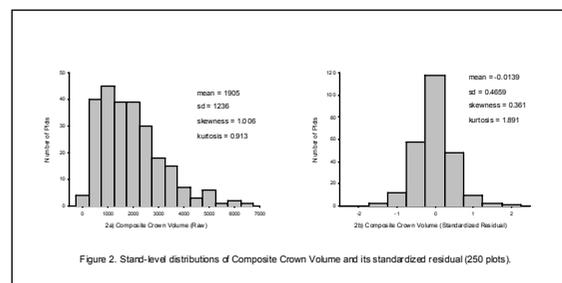


Figure 2. Stand-level distributions of Composite Crown Volume and its standardized residual (250 plots).

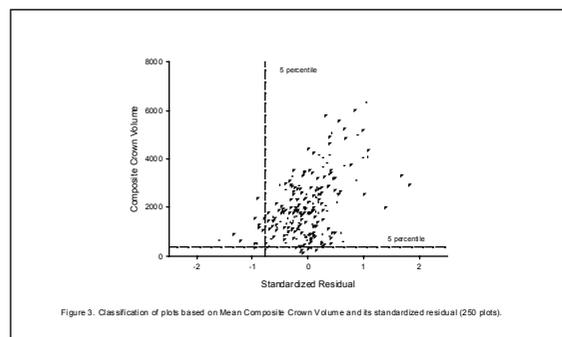


Figure 3. Classification of plots based on Mean Composite Crown Volume and its standardized residual (250 plots).

Conclusions

The utility of raw indicators can be extended by standardization across species. Analyses can be further enhanced with model residuals that adjust for the effect of tree size, stand density, and other parameters. Model residuals can also be enhanced by standardization. Combinations of these techniques result in a variety of analytical tools that can be tailored to address issues concerning forest health. The appropriateness of standardized, residualized, or raw values depends on the circumstances. Standardized values are useful when comparing or summing across individuals with different statistical distributions for the indicator of interest. Residualized indicators are useful for detecting deviations from expected values. Raw values are useful for detecting unadjusted net change.

Like raw indicator values, adjusted values can be correlated with other indicators (e.g., lichen diversity, or soil erosion) to establish biological thresholds. They can be correlated with other plot-based data or spatial overlays such as elevation, forest type, or physiographic class to determine if there are statistically significant differences between categories. They can be used for spatial analyses designed to detect clusters or gradients of unusually good or poor indicator values.