Maximum Likelihood Forest Canopy Profile Estimation

Paul C. Van Deusen*

Abstract
A maximum likelihood estimator (mle) for the foliage density function (fdf) is developed that allows for a custom fdf to model the forest canopy profile. The estimator derived here depends on the user providing a functional form for the fdf. The unknown fdf parameters are estimated from a sample of heights-to-first-leaf taken at random locations from ground level under the canopy. The method presented here could be viewed as a discrete version of the well known MacArthur and Horn (1969) estimator (MHE). The mle has several advantages over the MHE. The mle allows the user to incorporate some prior knowledge of the canopy profile via the fdf and it provides a covariance matrix for the parameter estimates. The most important advantage is that sample data never result in division by zero for upper canopy heights as can happen with the MHE. Finally, a simulated comparison indicated that the mle was less variable than the MHE.

Key Words: Foliage sampling, foliage density function, MacArthur and Horn estimator

1. Introduction

Understanding the structure of the forest canopy is important due to its role in tree growth rates (Waring et al. , 1981), nutrient and carbon cycling (Radke and Bolstad , 2001; Baldocchi and Harley , 2006), and its contribution to biomass (Lefsky et al. , 2005; Seidel et al. , 2011).

Foliage density estimation methods are often derived from work in MacArthur and Horn (1969) and the following estimator of the integrated value of the fdf, $D(h)$, which is the total foliage between two heights, h_1 and h_2,

$$\int_{h_1}^{h_2} D(\mathrm{h})\mathrm{dh} \ = \ ln\{\phi(h_1)/\phi(h_2)\} \tag{1}$$

where $\phi(h)$ is the probability of no leaves over the first h meters. This method relies on sample frequencies to estimate $\phi(h)$ and therefore $D(h)$ is never specified. This is a non-parametric method.

An alternative maximum likelihood estimator is developed (mle) that requires the user to provide an assumed fdf, which gives the expected number of leaves per unit of height, M. For example, suppose $D(h) = d$, a constant, then there will be d leaves per M where M could be 1 meter or 1 foot (1 foot=0.3 meters), and would typically be in the same units as height. If d=0.1 leaf per M then we expect to have 1 leaf every $10M$. In any event, a specific functional form with unknown parameters, β, must be specified for this method. Call this function $D_\beta(h)$ to differentiate it from the non-parametric $D(h)$.

The problem should also be constrained by specifying the maximum canopy height, say H. Then the expected number of leaves, D, along a random vertical line from height 0 to H would be

$$D \ = \ \int_0^H D_\beta(\mathrm{h})\mathrm{dh} \tag{2}$$

It makes sense to choose a constrained function such that $0 \leq D_\beta(h) \leq 1$ regardless of the specific parameter values.

*National Council for Air and Stream Improvement, 60 East St, Mount Washington, 01258, USA. PVandeusen@ncasi.org

1.1 Materials and methods

For the purpose of this study, think of the forest canopy as consisting of a ceiling at height H with leaves suspended at random locations between heights 0 and H. A random vertical line can be sampled to obtain the heights where it passes through a leaf. Some of these lines will intersect no leaves and still be consistent with $D_\beta(h)$, which causes the canopy to be discontinuous.

A sample of n lines can be taken from under the forest canopy to provide observations that can be used to estimate the parameters of the fdf. The only reliable measurement available from the forest floor is often the height to the first leaf, so the minimum sample dataset consists of $f_1, ..., f_n$ from a sample of n vertical lines.

In order to develop a sample likelihood function, one needs a statistical mechanism for simulating/modeling leaves along a line. Leaves along a vertical string or line through the canopy are well represented by a discrete process, since there can only be a finite number of distinct leaves on a line. Therefore, each line is partitioned into bins of width δh such that $H/\delta h$ is an integer. For example, $H = 100$ and there can be 100 bins such that $\delta h = M$. There is no reason that δh must equal M, or that δh be constant along the line, but it simplifies the notation to assume constant width bins.

The probability of bin i having a leaf is

$$p_i = \frac{\delta h}{M} \int_{m_i - \delta h/2}^{m_i + \delta h/2} D_\beta(\mathrm{h})\mathrm{dh} \tag{3}$$

where m_i is the height to the midpoint of bin i. The $\delta h/M$ term adjusts the fdf in eq (3) for the situation where the bin width, δh is not equal to M. Adjustment is necessary because an fdf is specific to a unit of length.

Record height to first leaf, f_s, for each sample in the field and determine the bin, b_s, that this falls into back in the office. The likelihood for a sample-line is the probability of no leaf until b_s,

$$\tilde{p}_s = \prod_{j=1}^{j<b_s} (1 - p_j)\, p_{b_s} \tag{4}$$

The likelihood for a line with no leaves is $\prod_{j=1}^{J} (1 - p_j)$, where J is the total number of bins along the line.

The log-likelihood for the sample of size n is the log of the product of the individual sample-line probabilities shown in eq (4)

$$L(\beta) = \sum_{s=1}^{n} log\,(\tilde{p}_s) \tag{5}$$

where $L(\beta)$ is a function of the parameters to be estimated, i.e. the parameters in the assumed fdf.

1.2 Maximizing the likelihood

The maximum likelihood estimate (Lehmann and Casella , 1998) is the value, $L(\hat{\beta})$, where the gradient (first derivative) of the log likelihood equals 0. The gradient of (eq 5) is

$$\frac{\partial L}{\partial \beta} = \sum_{s=1}^{n} \frac{\tilde{p}'_s}{\tilde{p}_s} \tag{6}$$

where $\tilde{p}'_s$ denotes the derivative with respect to β.

The Hessian (second derivative) of the log likelihood with respect to β is useful for the maximization process and to provide an asymptotic variance estimate. The Hessian of eq (5) is

$$\frac{\partial^2 L}{\partial \beta} = \frac{\tilde{p}_s \tilde{p}_s'' - \tilde{p}_s'^2}{\tilde{p}_s^2} \tag{7}$$

Minus the expected value of the inverse of the Hessian is referred to as the information matrix, $I(\beta)$, which provides an asymptotic covariance matrix for the maximum likelihood parameter estimates, $\hat{\beta}$.

The first and second analytical derivatives are complicated enough to justify using a function maximization procedure based on numerical derivatives. The *nlm* procedure from R (R Development Core Team , 2010) is used for the example application. The nlm procedure minimizes a function and must be supplied with minus the log-likelihood to maximize the function. Therefore, the inverse of the Hessian matrix from the nlm procedure is an approximate covariance matrix for the parameter estimates.

1.3 Example applications

For the example applications a logistic function represents the fdf,

$$D_{a,b}(x) = \frac{1}{1 + exp(-(a + bx))} \tag{8}$$

where x=h/H. This results in a scale free fdf that predicts the instantaneous probability of a leaf at relative height, h/H, and is constrained to be between 0 and 1. The result is multiplied by H to scale it to the actual measurement units.

The probability computation for eq (3) involves the integral of eq (8),

$$\int D_{a,b}(x)\mathrm{d}x = F_{a,b}(x) = \frac{log\,(exp(a + bx) + 1)}{b} \tag{9}$$

Equation (3) is then evaluated as,

$$p_i = \frac{\delta h}{M}\{F_{a,b}([m_i + \delta h/2]/H) - F_{a,b}([m_i - \delta h/2]/H)\} * H \tag{10}$$

Vertical line samples from the canopy are simulated by setting $H = 100$ and setting the parameters of eq (9) to $a = -4.6,\ b = 2.1$. The parameter settings came from fitting eq (8) to the following data:
$x = \{0.1, 0.2, 0.3, .04, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and
$y = \{0.01, 0.01, 0.02, 0.02, 0.03, 0.04, 0.04, 0.05, 0.06\}$,
where x is relative height and y is the instantaneous probability of a leaf. This simulates a canopy where the probability of finding a leaf increases with height. The expected number of leaves in a line is, for this simulated canopy, $H * \int_0^H D_{a,b}(h)\mathrm{d}h = 3.28$.

Samples of size $n = 100$ are simulated by generating vertical lines of length $H = 100$ and breaking each line into 100 bins. A leaf was assigned to the bin if a uniform random number was less than the probability of a leaf being in that bin as computed with eq (10). The simulation was repeated 1000 times. The results (Table 1) indicate that the mle closely recovers the original $a = -4.6$ and $b = 2.1$ parameter values.

The simulation variance column (Table 1) is the variance computed from the 1000 simulated sample estimates. The simulation variances agree well with the mean of the inverse Hessian results computed for each simulated sample, which is what would be used for an actual application of this method.

Table 1: Simulation results for 1000 replications of a sample of size 100 of random vertical lines in a canopy where canopy height is 100. The objective of the simulation was to estimate the parameters in eq (8). The simulation mean is the mean of the replications. The simulation variance is computed directly from the 1000 replicated parameter estimates. The Hessian variance is the mean of the variance estimates provided by the inverse Hessian matrix for each replication. The min and max values from the 1000 replications are given.

	Simulation		Hessian	Simulation	
Parameter	Mean	Variance	Variance	Min	Max
a	-4.614	0.0415	0.0410	-5.332	-4.052
b	2.145	0.1722	0.1668	0.980	3.537

1.4 Comparison with MacArthur and Horn estimator

The (MacArthur and Horn , 1969) estimator for the number of leaves between 2 heights, h_1 and h_2, is $ln(\phi(h_1)/\phi(h_2))$, which is the log of the ratio of the probabilities of no leaves up to h_1 and h_2. The discrete version of $\phi(h)$ for vertical lines divided into bins is,

$$\tilde{p}_h = \prod_{j=1}^{j<b_h} (1 - p_j) \tag{11}$$

where b_h is the bin that contains the height of interest.

The required probabilities for the (MacArthur and Horn , 1969) estimator are simulated using eq (11) and logistic equation (8) with parameters $a = -4.6, b = 2.1$. The actual number of leaves between h_1 and h_2 is given by eq (10) as, $\{F_{a,b}(h_2/H) - F_{a,b}(h_1/H)\} * H$. The comparison for selected h_1, h_2 pairs is given in Table (2). The comparisons are made from the lower end of the first bin until the mid-point of the second bin in an effort to compensate for the effect of the so-called actual number being based on discretizing and then comparing to the MacArthur and Horn (1969) continuous calculation. The results (Table 2) indicate that the MH method performs well when it is supplied with exact probabilities.

Table 2: Comparing the MacArthur and Horn (1969) estimate of number of leaves between heights h_1 and h_2 with the actual value as computed with equation (8) with parameters $a = -4.6, b = 2.1$.

h_1, h_2	Actual	MH	MH/Actual
0,24.5	0.318	0.312	0.981
25,49.5	0.532	0.524	0.985
50,74.5	0.886	0.880	0.993
75,99.5	1.461	1.468	1.005
0,100	3.280	3.354	1.022

1.4.1 Simulated comparison of MacArthur-Horn with Maximum Likelihood

Considered here is a simulated application of the MHE versus the mle with the same simulation method that was used above to demonstrate the properties of the mle. The example application from MacArthur and Horn (1969) is followed and estimates are obtained of the expected number of leaves between several height pairs (h_1, h_2) using the MHE which can be compared to the mle using sample sizes $n = 16$ and $n = 100$.

The bias is small for both methods (Table (3), but the ratio of mean squared error (mse) indicates that the mse is 2-3 times larger for the MHE method. Another disadvantage of the MHE approach is that the estimator can result in division by 0 for some height pairs. This occurs when all samples had their first leaf before h_2 causing the sample based estimate of $\phi(h_2) = 0$. This was avoided by not going above $h_2 = 50$ in the simulation. Therefore, the simulation actually understates the mse of the MacArthur and Horn (1969) method.

Table 3: Comparing the MacArthur and Horn (1969) and maximum likelihood estimators for number of leaves between heights h_1 and h_2 with sample sizes $n = 16$ and $n = 100$. A simulation was performed with 1000 replications to compute the mean, bias and mse for each method. The samples were simulated with equation (8) with parameters $a = -4.6, b = 2.1$.

		—n=16—			—n=100—		
h_1, h_2		10,25	25,50	10,50	10,25	25,50	10,50
	mean	0.220	0.584	0.804	0.216	0.550	0.766
mle	bias	0.003	0.032	0.035	-0.001	-0.003	-0.003
	mse	0.007	0.034	0.061	0.001	0.004	0.007
	mean	0.227	0.595	0.822	0.218	0.560	0.778
MHE	bias	0.010	0.042	0.052	0.001	0.007	0.008
	mse	0.020	0.085	0.107	0.003	0.011	0.015
MHE/mle	mse	3.07	2.52	1.75	2.85	3.27	1.96

2. Discussion

A maximum likelihood method was developed for estimating the parameters of a foliage density function (fdf), which gives a profile of the leaf density in a forest canopy. The first example application showed that the method can recover the unknown parameters in a user supplied fdf. It was then demonstrated that the (MacArthur and Horn , 1969) estimator (MHE) for the number of leaves between two heights works well when the exact probabilities that it requires are supplied. However, the MHE probabilities must come from sample data, which can produce highly variable estimates as pointed out by MacArthur and Horn (1969). The mle method presented here uses the sample data to estimate a small number of fdf parameters, and then any aspect of the canopy profile can be estimated from the fdf.

The second simulated application demonstrated that the mse of the MHE is 2-3 times greater than the mle for the lower half of the canopy. Our simulation understated the variance of the MHE, because it avoided problematic upper canopy heights where the MHE can fail due to division by 0. The simulated comparison was based on estimating the expected number of leaves between two heights. The expected number of leaves above a point in the forest could be used for the same purpose as leaf area index (Radke and Bolstad , 2001). For example, it can serve as a relative index of photosynthetic capacity. Therefore, the most valuable estimate from the fdf is likely to be the number of leaves from ground level up to total canopy height, H. In order to get this estimate with the MHE, one would have to ensure that the sample has at least one observation with no leaves, as pointed out by MacArthur and Horn (1969). The mle does not impose this requirement.

A downside of the mle method, as presented here, is that it reduces the continuous height measurement to the first leaf into discrete bins. Therefore, any height measurement that falls within a bin is treated the same. In practice, this may not be important given the fact that leaves are not actually at fixed locations, since branches bend and leaves flutter in the wind. A small bin width of 1 foot or 100 cm should not downgrade the precision of the final estimates by a practically meaningful amount.

In fact, the discrete version of $\phi(h)$ in eq (11) will converge to the continuous result in MacArthur and Horn (1969) as bin width is reduced. The MHE is derived from the following recursive equation for the probability of no leaves up to $h + dh$,

$$\phi(h + \mathrm{d}h) \;=\; \phi(h)(1 \;-\; D(h)\mathrm{d}h) \tag{12}$$

where the "bin width" would be $\mathrm{d}h$ and the probability of a leaf being in the bin is $D(h)\mathrm{d}h$. It is clear that eq (11) could also be written in recursive form, and the bin width could be decreased until eq (11) converges to eq (12).

Methods for estimating an fdf could be relevant for LiDAR (Popescu and Wynne , 2004; Naesset et al. , 2013) applications that seek to profile forest canopies. The fdf methods profile from below, whereas LiDAR profiles from above. As such, the methods developed here and by MacArthur and Horn (1969) could have wider application after some additional development.

3. Conclusions

The (MacArthur and Horn , 1969) method for estimating the number of leaves between two heights has the advantage of being non-parametric. It is not necessary to specify the fdf, $D(h)$. However, the MHE requires sample estimates of the probability of no leaves to heights, h_1 and h_2 for at least one height pair. The results depend on the accuracy of these estimates and some samples lead to division by zero. Another issue is that it is not clear how to place a confidence interval on the MHE.

An alternative maximum likelihood estimator was presented that requires the user to specify a functional form for the fdf, $D_\beta(h)$. The sample data are then used to estimate the unknown parameters, β. An asymptotic covariance matrix is available for the parameter estimates using standard maximum likelihood theory. The, fdf can then be used to estimate various aspects of the canopy profile, including the number of leaves between any pair of heights, h_1 and h_2. A simulation indicated that the mle is considerably less variable than the MHE.

References

Baldocchi, D.D. and Harley, P.C. 2006. Scaling carbon dioxide and water vapour exchange from leaf to canopy in a deciduous forest. II. Model testing and application. Plant, Cell & Environment 18(10): 1157-1173.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C, Espirito-Santo, F.D.B., and Hunter, M.O. 2005. Estimates of forest canopy height and aboveground biomass using ICESat. Geophysical Research Letters 32, L22S02, doi:10.1029/2005GL02397.

Lehmann, E.L.; Casella, G. (1998). Theory of Point Estimation, 2nd ed. Springer. ISBN 0-387-98502-6.

MacArthur, R.H., and Horn, H.S. 1969. Foliage profile by vertical measurements. Ecology, 50: 802-804.

Naesset, E., Gobakkena, T., Bollandsas, O.M., Gregoire, T.G., Nelson, R., and Goran Stahl, G. 2013. Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. Remote Sensing of Environment 130: 108-120.

Popescu, S.C and Wynne, R.H. 2004. Seeing the Trees in the Forest: Using Lidar and Multispectral Data Fusion with Local Filtering and Variable Window Size for Estimating Tree Height. Photogrammetric Engineering & Remote Sensing 70(5):589-604.

R Development Core Team. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org

Radtke, P.J. and Bolstad, P.V. 2001. Laser point-quadrat sampling for estimating foliage-height profiles in broad-leaved forests. Canadian Journal of Forest Research 31:410-418.

Seidel, D., Fleck, S., Leuschner, C., and Hammett, T. 2011. Review of ground-based methods to measure the distribution of biomass in forest canopies. Annals of Forest Science 68(2):225-244.

Waring, R.H., Newman, K., and Bell, J. 1981. Efficiency of tree crowns and stemwood production at different canopy leaf densities. Forestry 54(2):129-137.