Article

Quantifying Variances of Line-Intercept-Sampling Estimators of Percentage Cover

Kaiguang Zhao^{*1}, Sorin Popescu^{*1} and Ross Nelson^{*2}

ABSTRACT

The line-intercept sampling (LIS) method has found important applications in such areas as forest and wildlife, ecological and biological sciences, and crop and agriculture fields. LIS is a sampling technique to make observations along line transects in order to make inferences of area properties. The placement of transects can be chosen in many different manners, i.e., randomly or systematically. The motivation of this study is to use LIS to infer regional information of forestry biophysical parameters based on the linear transects measurements of a profiling LiDAR system. However, there is no optimum method to properly derive a reasonable measure to the uncertainty of LIS estimates. As such, the study first developed a theoretical framework to describe the LIS estimation in two settings, one with fixed landscape configuration, and another with random configuration. The subsequent simulation of transect observation is realized for two categorical maps: the artificial one simulated by SIMMAP, and the real one classified from Landsat ETM+ multispectral imagery. The simulated samples were used to test four estimators. The methodology employed in this study provides a good starting point for practically implementing the quantification of variance estimates with LIS.

Keywords: line-intercept sampling, line-intersect sampling, profiling LiDAR, estimator, Monte Carlo

INTRODUCTION

The line-intercept method is a quite effective sampling technique, and it has found extensive applications in natural resource disciplines, such as the characterization of landscape pattern and the inventory of forestry tracts. Instead of intensive sampling over the whole study area, Line-Intercept Sampling (LIS) only measure samples along linear transects; the placement of transects over study sites can be done either randomly or systematically, with fixed or varying length for each transect, or sometimes in a prescribed configuration. Basically, LIS is used to observe a region consisting of randomly distributed particles or items by only sampling the particles intercepted by transects. According to the settings of

applications, the particles may represent different entities, i.e., shrubs, tree crowns, dens, lakes, plants, roads and signs, or patches in a landscape. The particles can be geometrically arbitrary-shaped. Many estimators have been developed so far to extend the line transect observation into the inference of area properties of particles, e.g., estimating the percentage cover of certain patch type based on the length proportion of transect intercepted with the given type of patch (KAISER, 1983; BUTLER and McDONALD, 1983).

LIS technique has a long history and it can be roughly dated back as early as 1868 when CROFTON (1868) discussed the local probability of straight lines drawn at random in a plane. CANFIELD (1941) first successfully introduced this technique to sample range vegetation and employed the random placement of transects and used the proportion of the sampled transect intercepted by the vegetation as an estimate of the ratio of the vegetation area to the whole study area. BAUER (1943) then compared the relative efficiency of the transect and quadrature methods of sampling vegetation for statistical analysis based on a series of simulated plant communities of known composition, and he showed that the transect sampling gave more accurate results than quadrature methods in the cases he investigated. In 1953, MCINTYRE further developed the LIS technique by using line transects to

Corresponding author: Kaiguang Zhao

^{*1} Spatial Sciences Lab., Dept. of Forest Sciences,
Texas A&M University, College Station, TX
77801, USA

^{*2} Biospheric Sciences Branch Code 614.4
NASA's Goddard Space Flight Center Greenbelt,
Maryland USA 20771

estimate plant density, say, the number of plants (particles) per unit area. Later, WARREN and OLSEN (1964) proposed a LIS-based estimator for the volume of logging waste and documented the first use of LIS in forestry applications. Instead of naming it line-intercepting sampling, the authors coined it as line-intersect sampling. The unbiasedness of their estimator was proved to be true under certain assumptions later by VAN WAGNER (1968). The theoretical work by KAISER (1983) presented two generic estimators for LIS and his theory tended to unify all the LIS results up to that time. He used geometrical probabilities to prove the unbiasedness of the proposed estimators, and provided examples in which his estimators degenerated to the commonly known estimators. KAISER's paper (1983) proved to be one of the most appreciated works in the literature of LIS. Recently, extensive applications of LIS have been seen in such areas as forest and wildlife, ecological and biological sciences, and crop and agriculture fields (e.g., KELLER *et al.*, 2004; KEANE *et al.*, 2005; NELSON *et al.*, 2005). Recent progress of LIS sampling theory can also be witnessed in the work of AFFLECK *et al.* (2005), among others.

With this study, the authors are interested in developing an on-the-fly portable airborne laser profiling system for forest inventory, based on an existing system called Portable Airborne Laser System (PALS) developed by NELSON *et al.* (2003). The proposed system is a profiling laser altimeter or LiDAR (Light Detection And ranging). A laser profiler usually emits laser pulses at near-nadir direction (without scanning)

and collects the returned pulses. The round-trip travel time of the laser pulse can be converted to ranging information based on the speed of light. As shown in Fig. 1, the profiling LiDAR can measure the forest canopy height profiles along a transect (NELSON, 1984). The schematic in Fig. 1.a suggests that two profiles are involved to characterize the stand structure: the top one depicting the profile of top canopy surface, and the bottom one describing the terrain topography. Studies found that the forest stand characteristics and biophysical parameters can be derived from this profiling measurement (MACLEAN and KRABILL, 1986). Transects flown over a study site in Fig. 2 illustrate the basic idea on how a profiler collects samples and what the placement of transects looks like. Unlike most current commercially available scanning LiDAR systems capable of continuously sampling the study area (LEFSKY *et al.*, 2002; POPESCU *et al.*, 2002; POPESCU *et al.*, 2003), the proposed profiler can produce only data along transects which are typically almost parallel and 2-4km apart.

In order to infer information on the area-based variables, LIS needs to analyze the profiling transect data. However, before applying LIS, relevant information should be able to be extracted from profiler altimeter data along transects, preferably by some automatic data processing algorithms. These may include but not be limited to such algorithms as those for reconstructing ground topography, deriving canopy height profiles, and stratifying the canopy height profiles into segments with respect to the types of land cover. Central to

Fig. 1 An example of a forest canopy profile which might be obtained via a laser profiling instrument (see also Nelson *et al.*, 1984)

Fig. 2 Parallel transects flown by a profiling LiDAR (profiler) over a study site: the dark lines are the flight lines of the laser profiler, the background is a false-color QuickBird image.

this study, the stratification of profiling transects is essentially important for the purpose of estimating percentage cover of each land cover. To achieve this, advanced signal processing techniques such as wavelet analysis is effective in designing automatic algorithms. No elaboration on these aspects is made in this paper, and details on the implementation of relevant algorithms is planned to appear in a future publication.

Airborne profiling measurements have been used to inventory forestry resources with LIS sampling techniques, particularly over large areas. NELSON *et al.* (2005) exploited the possibility of using over 1,300km of systematic profiling transect measurements acquired over the State of Delaware, USA, to do state-level multiple resources inventory by the LIS method, and their estimates of merchantable volume fell into the 21% of US forest service estimates at the county level, and 1% statewide. Total above-ground dry biomass estimates were within 22% of USFS estimates at the county level and with 16% statewide. They also used profiling measurements to estimate the area percentage of impervious and open water area for three counties of Delaware, and found that results were comparable to estimates from other sources. By using the same dataset, NELSON *et al.* (2005) again utilized the LIS method to help estimate the extent of Delmarva Squirrel Fox habitat. Their results have demonstrated the feasibility of profiling laser measurements for regional resources inventory. However, although their estimations are accurate enough to be comparable to the known estimates by other means, their report on the variance of the LIS estimates only provides an approximation to the true variance; also, they claimed that such estimated variance gives an upward conservative estimation of the true value (NELSON *et al.*, 2005).

Because the variance or standard error of an estimator is an indicator of the accuracy or, on the other hand, the uncertainty of the estimator; without good knowledge of the estimator variance, estimates will be less valuable. KAISER (1983) gave for his estimators a definition formula of the variance which includes two components, namely the variance term and the covariance term. He mentioned that even in simple cases the formula is mathematically intractable mostly due to the unknown form of the covariance component. He finally suggested obtaining a number (m) of independently and identically distributed estimates based on m independently chosen transects of certain length L , then pooling these m estimates by arithmetic averaging as a final estimate and taking the sample variance as the estimated variance. In terms of the intractability of the formula for covariance, DEVRIES (1986) had the same argument, and he offered a similar formula to evaluate the variance in the cases of varying transect length, where the length of transect is used as weighting in the calculation. NELSON *et al.* (2005) restated a successive difference estimator based on the study of LINDBERG (1926); and he also envisioned relying on a large Monte Carlo experiment to obtain a valid estimate of the estimator's variance for the systematic airborne LIDAR

samplings. As such, this paper attempts to investigate the uncertainty of LIS estimates by quantifying its variance, which will serve as a crucial step in appropriately analyzing and interpreting profiling LiDAR measurements.

Specifically, the paper will use both the computer-simulated landscape and the real land cover map classified from a Landsat ETM+ image to simulate transect observations in a systematic manner. From a series of simulated data, four estimators for cover percentage of certain patch type are tested and their variances are also constructed. The paper simplifies the demonstration by only considering a binary landscape, i.e., background and patches of interest for the computer-generated landscape, and forestry/non-forestry for the land cover map. Furthermore, for practical conveniences, the following discussion will term pixels of value 1 as forestry and 0 as non-forestry in both the artificial and real landscape maps. The results can be easily generalized to the multiple-type cases. Also, this paper formulates a theoretical framework to evaluate estimator variances under certain assumptions.

METHODS

Computer-simulated Landscape

A landscape simulation package SIMMAP, which implemented a modified random simulation method (SAURA and MARTÍNEZ-MILLÁN, 2000), was utilized to generate a categorical landscape map with 2 patch types and an extent of 900×900 pixels (Fig. 3). The default values for initial probability and neighborhood criteria were used, and the minimum patch size was set to 90 pixels. The forest cover percentage is the main metric of inference interest.

Fig. 3 The computer-generated categorical map and a mask of the boundary of study area, and the resulting landscape after applying the mask to the original square region.

Usually, the shape of the study region, i.e., a county or state, may be arbitrary. Hence, in order to mimic reality, a mask was arbitrarily delineated on-screen in ENVI (ITT Visual Information Solutions, Inc.) and used to clip out the region of interest, as shown in Fig. 3.

Classified Map From Remotely Sensed Data

A Landsat ETM+ multispectral image over Huntsville, East Texas, USA, was used for this study. The unsupervised ISO-DATA algorithm was used to classify the image into 7 classes, which were then grouped into two categories: forest and non-forest. Because the inference of interest does not depend on the exact type of classification and the absolute true value of relative to the real landscape is of no interest in this study, the step of accuracy assessment was ignored. Likewise, a mask was applied to discard the pixels out of a delineated boundary in order to mimic a natural shape of study area. The resulting map is depicted in Fig. 4.

Calculating the Total Area of a Given Class

To simplify the illustration, again it is assumed that the landscape contains only two categorical classes, i.e., non-forest and forest. The interest is in estimating the total forested area. In practice, when a profiler system is flown, the flight direction is typically set and fixed; the sampling transects are parallel and spaced almost equally. A coordinate system was chose such that the x axis is perpendicular to the direction of sampling transects and the y axis is selected as the tangent to the left-most boundary (Fig. 5). For a given transect at x , its

Fig. 4 The classified Landsat ETM+ image with two classes (green for forest and black for non-forest), and a mask of the study area boundary, and the resulting landscape after applying the mask to the original square region.

length $l(x)$ is equal to the distance between the upper and lower boundary points connecting the transect, excluding the portion that may lie outside the study area, if any (i.e., the second transect from the left in Fig. 5); and the proportion of it intercepted by forest patches is denoted as $r(x)$. Note that a transect may intercept many patches. Then, the total area of all forest patches can be evaluated by:

$$a = \int_0^s r(x) l(x) dx \quad (1)$$

where s represents the span of study area along x axis.

To clarify notations, lowercase letters will denote variables or functions that are fixed or prescribed, while the uppercase is used to denote a random variable or function.

Estimators when the Configuration of Landscape is Fixed yet Unknown

Before investigating a study site, its extent is usually known by referring to archived data source, most likely a GIS database. In addition, suppose that the orientation of all sampling transects are the same and already determined. Typically, this is the case for the profiling LiDAR survey. Hence, $l(x)$ is known and can be calculated. However, the $r(x)$, which is the proportion of transect intercepted by the patches of a certain type, e.g., forest, is fixed for a given study site, but remains unknown to investigators. If the placement of transects is uniformly random over the x -direction span $[0, s]$ and the property of $r(x)$ is unknown, the following estimator $\hat{a}_1$ can be effective:

Fig. 5 A hypothetical region of study where the black solid lines are transects for LIS and will be used to infer information on the colored patches, i.e., the cover percentage. The transect at x touches the upper and lower boundary points of the region, and has a length of $l(x)$; the transect intercepts with patches, and denote the proportion intercepted to its total length is $r(x)$.

$$a = s \int_0^s r(x) l(x) \frac{1}{s} dx \Rightarrow \hat{a}_1 = s \frac{\sum_{i=1}^n r(x_i) l(x_i)}{n} \quad (2)$$

where n is the number of transects observed, and the variance is given as,

$$\begin{aligned} \text{Var}[\hat{a}_1] &= \frac{s^2}{n} \text{Var}_{l(X)}[r(X) l(X)] \\ &= \frac{1}{n} \left[s \int r^2(x) l^2(x) dx - \left(\int r(x) l(x) dx \right)^2 \right] \end{aligned} \quad (3)$$

where $\text{Var}_{l(X)}[r(X)l(X)]$ means the variance of $r(X)l(X)$ when assuming X is uniformly distributed, i.e., $X \sim 1/s$. Since the above variance formula requires evaluating $r(x)$ over $[0, s]$, it cannot be used directly. Alternatively, the sample values of $r(x)$ at the observed transect should be substituted to calculate the sample variance of $r(x)l(x)$ as the estimated variance, that is,

$$\hat{\text{Var}}_{nm}[\hat{a}_1] \approx \frac{s^2 \sum_{i=1}^n [r(x_i) l(x_i) - \overline{r(x) l(x)}]^2}{n(n-1)} \quad (4)$$

where $\overline{r(x) l(x)} = \sum [r(x_i) l(x_i)] / n$ is simply the sample mean.

Intuitively, the longer $l(x)$, the more contribution to the total area it is supposed to bring. Rather than being randomly placed, transects can be located by the probability of $l(x)/m$, where m is the total area of the whole study region. This means that there are more transects placed over x 's with larger $l(x)$. According to such a design, a natural estimator should be,

$$a = m \int_0^s r(x) \frac{l(x)}{m} dx \Rightarrow \hat{a}_2 = m \frac{\sum_{i=1}^n r(x_i)}{n} \quad (5)$$

where m , again, is the total area of the whole landscape such that $m = \int_0^s l(x) dx$ and the corresponding variance will be,

$$\begin{aligned} \text{Var}[\hat{a}_2] &= \frac{m^2}{n} \text{Var}_{l(x)/m}[r(X)] \\ &= \frac{1}{n} \left[m \int r^2(x) l(x) dx - \left(\int r(x) l(x) dx \right)^2 \right] \end{aligned} \quad (6)$$

where $\text{Var}_{l(x)/m}[r(X)]$ is the variance of $r(X)$ by assuming $X \sim l(x)/m$. Similarly, the observed samples of $r(x)$ can be used to obtain an approximation to the true variance, and it follows that,

$$\hat{\text{Var}}[\hat{a}_2] \approx \frac{m^2 \sum_{i=1}^n [r(x_i) - \overline{r(x)}]^2}{n(n-1)} \quad (7)$$

Estimators when the Configuration of Landscape is Supposed to be Random

In contrast to the aforementioned setting, here the configuration, i.e., $r(x)$, is supposed to be an unknown yet random function of x and was denoted as $R(x)$. Consequently, the total area of forest cover (patch type of interest) is a random variable instead of an unknown constant. It can be computed by the counterpart to Eq. 1,

$$A = \int_0^s R(x) l(x) dx \Rightarrow \mu_A = E[A] = \int_0^s \mu_R(x) l(x) dx \quad (8)$$

where $\mu_R(x)$ is the mean function of $R(x)$, and in most cases may be not a constant function. Of particular note is that mathematically $R(x)$ cannot be a Gaussian random function because for any x , $R(x)$ is bounded over $[0, 1]$; however, for practical convenience, $R(x)$ may be approximately gaussian. In this random setting of landscape configuration, the corresponding parameter of interest is $E[A]$, i.e., the mean of A . To simplify, assuming that $R(x)$ has a mean μ_R independent of x , the $E[A]$ and $\text{Var}[A]$ then can be obtained respectively by:

$$\mu_A = \mu_R \int_0^s l(x) dx = m \mu_R \quad (9)$$

$$\text{Var}(\mu_A) = \int_0^s \int_{-x_1}^{s-x_1} C_R(x) l(x+x_1) l(x_1) dx dx_1 \quad (10)$$

where $C_R(x)$ is the covariance function of $R(x)$, and Eq. 10 holds if and only if $R(x)$ is widely-sense stationary, which, in most cases, should not be too rigid a requirement. By further assuming the mean-ergodicity of $R(x)$, a reasonable estimator of μ_A can be expressed as,

$$\hat{\mu}_{A,1} = m \hat{\mu}_R = \frac{m \sum_{i=1}^n R(x_i)}{n} \quad (11)$$

and its variance is,

$$\text{Var}[\hat{\mu}_{A,1}] = \frac{m^2}{n^2} 1' \cdot \Sigma_R \cdot 1 \quad (12)$$

where again, m is the area of whole landscape, $R(x_i)$'s are the sampled value at n transects that can be either systematically or randomly placed; and Σ_R is the variance-covariance matrix of $R(x_i)$, and can be constructed from $C_R(x)$ which, for the (i,j) entry of Σ_R , gives $C_R(|x_i - x_j|)$. Unfortunately, the structure of $C_R(x)$ is usually unknown; therefore, in practice prior knowledge about $C_R(x)$ may be entailed.

Another estimator of μ_A is given in Eq. 11, and Eq. 12 is for the associated variance.

$$\hat{\mu}_{A,2} = s \frac{\sum_{i=1}^n R(x_i) l(x_i)}{n} \quad (13)$$

$$\begin{aligned} \text{Var}[\hat{\mu}_{A,2}] &= \frac{s^2}{n^2} [l(x_1), l(x_2), \dots, l(x_n)] \Sigma_R [l(x_1), l(x_2), \dots, l(x_n)]' \\ &\quad (14) \end{aligned}$$

The unbiasedness of $\hat{\mu}_{A,2}$ depends on both the properties of $R(x)$ and the sampling schemes of x_i . For instance, if $R(x)$ assumes a constant mean, $\hat{\mu}_{A,2}$ is unbiased when x_i 's are randomly sampled; conversely, suppose that $\mu_R(x)$ is dependent on x , $\hat{\mu}_{A,2}$ is biased when transects are observed at fixed locations x_i 's.

Simulation Schemes

First, systematic samples of transects, which are observed in a manner similar to the profiling LiDAR, are drawn randomly. The spacing between adjacent transects are purposely designed to be randomly distributed around the average spacing since the flight lines of profiling LiDAR are not exactly equally spaced. However, randomly sampled transects are employed to obtain more appropriate Monte Carlo estimates of uncertainty; specifically, for each of the computer-generated landscape and classified categorical map, a range of transect numbers are used; and at a given number of transects, 30 realizations are simulated. The sample variance of each 30 realizations is calculated to be compared with those obtained by the proposed variance formula.

RESULTS

The proportion of transects intercepted by forest patches, $r(x)$, is plotted against x in Fig. 6a and 6b respectively for the artificial landscape generated by computer and the real landscape created from ETM+ imagery. The curves suggested that in both cases, the $r(x)$ appears to be randomly fluctuating and there exists some spatial correlation to a certain degree as also shown in Fig. 7a and 7b which show the sample covariance function calculated from the corresponding $r(x)$'s. It is clearly seen that the correlation monotonically decreases

as the lags increase. These two sample covariance functions were plugged as $C_R(x)$ into the formula in Eqs. 10, 12, and 14 developed above, to compute the sample variance of the estimators.

Fig. 8 shows the forest cover percentage estimated by four different estimators, at different choices of transect number. The forest cover percentage is simply the ratio of estimated forest area to the total area. The four estimators are respectively the one used by NELSON *et al.* (2005), and three of the four discussed above except the second one. They are respectively denoted by "Nelson" or "N", "a1", "A1" and "A2" as labeled in the legends of Fig. 8-11. The plotted values of estimates in Fig. 8 and 10 only represent one randomly selected realization out of 30. The true forest percentages are 0.3684 and 0.2281 for the simulated artificial and classified real landscape maps, respectively. In both cases, the estimated values tend to come closer to the true value as more transects are observed. Moreover, an overall trend of overestimation is observed for all the four estimators (Fig. 8a) in the case of simulated landscape, although occasionally the estimates are lower than the true value at some transect numbers, and the estimator of "N" and "A1" seem to consistently give the estimation closest to the true value. On the other hand, an opposite trend was noticed for the classified ETM+ image where overall, all four estimators were apt to underestimate the percentage cover.

A way of calculating estimator variance is to take the

(a) $R(x)$ for the computer-generated landscape

(b) $R(x)$ for the classified Landsat ETM+ Image

Fig. 6 The proportion of transects intercepted by forest patches, $r(x)$, is plotted against x (pixels) in Fig. 5a and 5b respectively for the artificial landscape generated by computer and the real landscape created from ETM+ imagery.

(a) Sample covariance function for the computer-generated landscape

(b) Sample covariance function for the classified ETM+ image

Fig. 7 The sample covariance function calculated from the $r(x)$'s of Fig. 5, respectively for the artificial landscape (a), and the classified land cover map (b).

sample variance of 30 realizations. This is quite often employed in Monte Carlo simulation and supposed to offer a quite reasonable approximation to the true variance inherent in the sampling design. Fig. 9 suggests that for both landscapes the variances of four estimators almost coincide; but, to be more precise, for the simulated landscape, the estimated standard errors (SE) from the simulation with “Nelson” estimator is the lowest with “a1” and “A2” providing the highest estimates, while for the classified image, the estimated SE of “Nelson” and those of “a1” and “A2” are very close especially at a large number of transects. In addition, another way to calculate the estimator variance is to directly use the aforementioned formula which is derived under certain assumptions, with results given in Fig. 10. As with the

Monte Carlo simulated estimates, the SEs calculated with variance formula appear to follow a similar trend except that “a1” and “A2” have different estimates due to the difference in assumptions. Moreover, in contrast to the decreasing variance with more transects for Nelson’s estimator and a1, A1 and A2 exhibit an almost invariant pattern.

To further compare the two ways of computing variance, the ratio of the formula-derived SE to the estimates by Monte Carlo simulation is presented in Fig. 11. Assuming that the simulation-derived sample variance is a more reasonable estimate, the closer to 1 the ratio approaches, more valid the formula for variances is, or a more reasonable assumption is made as to $r(x)$. As shown in Fig. 11, the resulting plots reveal that in the case of simulated landscape, the ratios fluctuated

(a) Estimated percentage cover for simulated landscape

(a) Estimated standard error by Monte Carlo simulation for simulated landscape

(b) Estimated percentage cover for ETM+ image

(b) Estimated standard error by Monte Carlo simulation for ETM+ image

Fig. 8 The forest cover percentage estimated by four different estimators as a function of the number of transects used. The four estimators are the ones used by NELSON *et al.* (2005), the first estimator a1 discussed in “Estimators when the configuration of landscape is fixed yet unknown”, and the two estimators A1 and A2 discussed in “Estimators when the configuration of landscape is supposed to be random”. (a) is the case for artificial landscape, and (b) for the classified remote sensing imagery.

Fig. 9 The estimated variance by sample variance of the simulation: (a) is the case for artificial landscape, and (b) for the classified remote sensing imagery.

(a) Estimated standard error by Variance Formula for simulated landscape

(b) Estimated standard error by Variance Formula for ETM+ image

Fig. 10 The estimated variance calculated by the variance formula: (a) is the case for artificial landscape, and (b) for the classified remote sensing imagery.

around 1.0 for all four variance estimators, e.g., with a range from 0.72 to 1.41 for the Nelson's estimator; However, for the real landscape of classified image, "A1" and "A2" provided inflated estimates relative to reference values of Monte carlo simulation, with "A2" producing the most upward estimates and "A1" following next. In contrast, Nelson's estimator and "a1" have fluctuating ratios more closer to 1, which suggests that in this case these two estimators offer more realistic estimates of variance.

DISCUSSIONS AND CONCLUSION

By mimicking the way a profiling LiDAR collects linear transect observations, the simulation of random realizations supposes to offer the possibility of examining the variability of simulated samples as well as quantifying the variance of a given estimator for a certain quantity. In this study, estimators for the percentage cover of given classes is tested, more importantly, the variance of these estimates are obtained by

(a) Ratios of SE between formula and Monte carlo for simulated landscape

(b) Ratios of SE between formula and Monte carlo for ETM+ image

Fig. 11 The ratio of the formula-derived variance (shown in Fig. 9) to the sample variance of simulation (shown in Fig. 8) as a function of number of transects used. The closer to 1 the ratio is, the more valid the variance formula is. (a) is the case for artificial landscape, and (b) for the classified remote sensing imagery.

either taking sampling variance of the simulation or using the proposed formula as expressed in Eqs. 4, 7, 12 and 14.

Two settings for deriving the estimator and variance formula are assumed, when the landscape configuration is fixed and unknown, or random. Instead of using geometrical probability as in KAISER (1983), in this study, Monte carlo integration approach is referred to for developing estimators of percentage cover; and the basic ideas are illustrated by Eqs. 1, 2, 5 and 8. The estimator ($\hat{a}_1$ or a_1) of the first setting appear to be quite similar to the one used by NELSON *et al.* (2005), but difference exists that $\hat{a}_1$ is an estimator for area of interested patches which must be divided by the already known value of total area to obtain percentage cover, while the estimator in NELSON *et al.* (2005) directly provides estimates of percentage cover. The estimators for the random setting, however, tend to have a much more inflated variance that's much larger than the reference value of sample variance obtained by simulation; and this may either be caused by treating the systematical

sampling as random when evaluating the variance, or be an artifact of introducing randomness to depict the fixed but unknown landscape, i.e., $r(x)$. Furthermore, in practice the inference of the second setting can be only made for the mean since $R(x)$ in itself is assumed to be random. As such, the simulation of drawing transects from a fixed landscape is not too methodologically rigorous; but this appeared not to be a problem as demonstrated by the results for the simulated landscape.

In fact, the apparent disparity between the formula-based and simulation-based variance estimates for “A1” and “A2”, as indicated in Fig. 11. b for the classified image, can be most likely attributed to the departure of $R(x)$ from the underlying assumptions made for deriving the variance estimators. It becomes clear that a drift is present in $r(x)$, as depicted in Fig. 6. b, which shows a high tail at large x ; as a result, the incorporation of this overall trend into the calculation may mislead the estimation of covariance function as in Fig. 7.b since covariance function is supposed to capture only the random components of underlying stochastic process. Consequently, the inappropriate knowledge about covariance function will result in unreasonable variance estimates, as also noted for the variance estimator in KAISER (1983). In this study, estimates of covariance function were computed from $r(x)$. However, in practice, $r(x)$ remains unknown and is of inference interest based on the available samples observed at transects. Because of the limited number of transects, it is less possible to do estimation on covariance function using only these few samples. This prevents the effective use of variance estimators corresponding to the random setting. A possible remedy is to fit a parametric covariance model of known form to the available observation; specifically, the parameter relevant to the magnitude is simply estimated by the sample variance of $r(x)$'s, and parameters relevant to effective correlation distance may be inferred by analyzing the observations along the transect direction (i.e., the y axis in Fig. 5), due to continuous measurements available in this direction. However, the development of an efficient and reliable fitting procedure needs further investigation in future studies.

In our derivation, it is assumed that the geometry of study areas is known, which means that given a direction, $l(x)$ can be calculated. This is typically the case as GIS layers about study areas are often readily available prior to the flight mission. By changing the observation direction, not only $l(x)$ but also $r(x)$ changes; therefore, the orientation of transects has great influence on the properties of relevant estimation; and also there may exist a preferable direction for high-precision estimates. For example, an extreme scenario is hypothesized in Fig. 12 where the dotted half of the square region represents forested area. If a single vertical transect is randomly observed (the left of Fig. 12), the sampled percentage can take either 1 or 0 with equal probabilities; however, when observed horizontally, the true value 1/2 is always observed. To this end, prior knowledge, if available,

can be used to guide the selection of appropriate direction; otherwise, the direction that could minimize the variability of $l(x)$ may be an option. In fact, when $l(x)$ is constant, all the four estimators as well as the one used in Nelson et al. (2005) essentially produce the same estimates, but they do have varying variance estimates due to the different assumptions. In deriving the LIS estimate or the corresponding variance estimator such as $\hat{d}_1$ and $\hat{\mu}_2$, only the transect length at observed locations is used. To compensate this, the primary author suspects that incorporating all information on $l(x)$ should increase the precision of estimates.

As a final note, although the aforementioned derivation is based on binary categorical landscapes, the proposed estimators can be equally applied to estimate percentage cover for landscapes with multiple categories without any modification. Moreover, they can also be extended for estimating the mean value of variable that varies continuously across the study area, e.g., regional biomass. To illustrate this point, suppose that the variable is a spatial-explicit function denoted by $f(x, y)$; then, by assuming the same geometry as in Fig. 1, the mean value of the variable averaged over the study area is simply obtained by an integration as follows,

$$\bar{f} = \frac{1}{m} \int_0^s \int_{y_l(x)}^{y_u(x)} f(x, y) dy dx = \frac{1}{m} \int_0^s f_y(x) l(x) dx \quad (15)$$

where $f_y(x) = \int_{y_l(x)}^{y_u(x)} f(x, y) dy$. $y_l(x)$ and $y_u(x)$ are the y coordinates respectively for the lower and upper boundary points at x , and m is the size of study area. In eq. 15, $f_y(x)$ plays the same role as the $r(x)$ in eq. 1. Hence, after the samples of $f_y(x)$ are observed at a number of transects that are parallel to y , $\bar{f}$ can be estimated by using the same estimators as previously discussed.

The logic to formulate the estimators for LIS is a little different from the mostly used geometrical probability approach by KAISER (1983). Hence, it still remains undetermined whether a logically sound reasoning line is followed to deal with the LIS problems in this study, or not. Nevertheless, the attempt to rely on Monte carlo integration random simulation for evaluating LIS estimators points a bright direction towards the ease of practical implementation of LIS schemes other than the mathematical intractability.

Fig. 12 An extreme hypothetical landscape of square shape where the dotted half represents forested area, for the purpose of demonstrating the effects of transect orientation : vertically-placed transects (the left) and horizontally-placed transects (the right).

LITERATURE CITED

- AFFLECK, D. L. R., GREGOIRE, T. G. and VALENTINE, H. T. (2005): Design unbiased estimation in line intersect sampling using segmented transects. *Environmental and Ecological Statistics* **12**(2): 139-154
- BAUER, H. L., (1943): The statistical analysis of chaparral and other plant communities by means of transect samples. *Ecology* **24**: 45-60
- BUTLER, S. and McDONALD, L. L., (1983): Unbiased systematic sampling plans for the line intercept method. *Journal of Range Management* **36**: 463-8
- CANFIELD, R. H., (1941): Application of the line interception method in sampling range vegetation. *Journal of Forestry* **39**: 388-394
- CROFTON, M. W., (1868): On the theory of local probability, applied to straight lines drawn at random in a plane. *Philosophical Transactions of the Royal Society of London* **158**: 181-199
- KAISER, L., (1983): Unbiased estimation in line-intercept sampling. *Biometrics* **39**: 965-976
- KEANE, R. E., REINHARDT, E. D., SCOTT, J., GRAY, K. and REARDON, J., (2005): Estimating forest canopy bulk density using six indirect methods. *Can. J. For. Res.* **35**: 724-739
- KELLER, M., PALACE, M., ASNER, G. P., PEREIRA, R. and SILVA, J. N. M., (2004): Coarse woody debris in undisturbed and logged forests in the eastern Brazilian Amazon. *Global Change Biol.* **10**: 784-795
- LEFSKY, M. A., COHEN, W. B., PARKER, G. G. and HARDING, D. J., (2002): Lidar remote sensing for ecosystem studies. *Bioscience* **52**: 19-30
- LINDBERG, J. W., (1926): Zur Theorie Derr Linientaxierung. *Acta Forestalia Fennica* **31**(6): 3-9
- MACLEAN, G. A. and W. B. KRABILL, (1986): Gross-merchantable timber volume estimation using an airborne LIDAR system. *Canadian Journal of Remote Sensing* **12**(1): 7-18
- MCINTYRE, G. A., (1953): Estimation of Plant Density Using Line Transects. *Journal of Ecology* **41**: 319-330
- NELSON, R., KELLER, C. and RATNASWAMY, M., (2005): Locating and estimating the extent of Delmarva fox squirrel habitat using an airborne LiDAR profiler. *Remote sensing of environment* **96** (3-4): 292-301
- NELSON, R. F., KRABILL, W. B. and MACLEAN, G. A., (1984): Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment* **15**: 201-212
- NELSON, R. F., KRABILL, W. B. and TONELLI, J., (1988): Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment* **24**: 247-267
- NELSON, R., PARKER, G. and HOM, M., (2003): A portable airborne laser system for forest inventory. *Photogrammetric Engineering & Remote Sensing* **69** : 267
- NELSON, R., SHORT, A. and VALENTI, M., (2004): Measuring biomass and carbon in Delaware using airborne profiling LiDAR. *Scand. J. For. Res.* **19**: 500-511
- POPESCU, S. C., WYNNE, R. H. and NELSON, R. F., (2002): Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Com Elect Agr* **37**: 71-95
- POPESCU, S. C., WYNNE, R. H. and NELSON, R. E., (2003): Measuring individual tree crown diameter with LiDAR and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*.
- SAURA and MARTINEZ-MILLÁN, (2000): Landscape patterns simulation with a modified random clusters method. *Landscape Ecology* **15**(7): 661-678
- VAN WAGNER, C. E., (1961): The line intersect method in forest fuel sampling. *Forest Science* **14**(1): 20-26
- WARREN, W. G. and P. F. OLSEN., (1964): A line intersect technique for assessing logging waste. *Forest Science* **10**: 267-276

(Received 13 January 2007)

(Accepted 6 May 2007)