[9] J. R. Wang, E. T. Engman, J. C. Shiue, M. Rusek, and C. Steinmeier, "Microwave backscatter from agricultural fields observed by Shuttle Imaging Radar B," *IEEE Trans. Geosci. Remote Sensing,*, vol. GE-24, no. 4, pp. 510–516, 1986.
[10] M. C. Dobson and F. T. Ulaby, "Preliminary evaluation of the SIR-B response to soil moisture, surface roughness, and crop canopy cover," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-24, no. 4, pp. 517–527, 1986.
[11] E. T. Engman and J. R. Wang, "Evaluating roughness models of radar backscatter," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-25, no. 6, pp. 709–713, 1987.
[12] P. P. Batlivala and M. C. Dobson, "Soil moisture experiments (Kansas): Documentation of radar backscatter and ground truth data," RSL Tech. Rep. 264-7, Remote Sensing Lab., Univ. Kansas Center for Research, Inc., Mar. 1976.
[13] T. Mo, J. R. Wang, and T. J. Schmugge, "Estimation of surface roughness parameters from dual-frequency measurements of radar backscattering coefficients," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-26, no. 5, pp. 575–580, 1988.
[14] M. Autret, R. Bernad, and D. Vidal Madjar, "Theoretical study of the sensitivity of the microwave backscattering coefficient to the soil surface parameters," *Int. J. Remote Sensing,* vol. 10, no. 1, pp. 171–179, 1989.
[15] E. G. Njoku and J. A. Kong, "Theory for passive microwave remote sensing of near-surface soil moisture," *J. Geophys. Res.,* vol. 82, pp. 3108–3118, 1977.
[16] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes, "Microwave dielectric behavior of wet soil—Part II: Dielectric mixing models," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-23, pp. 35–46, 1985.
[17] F. T. Ulaby, P. P. Batlivala, and M. C. Dobson, "Microwave backscatter dependence on surface roughness, soil moisture and soil texture, Part I: bare soil," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-16, pp. 286–295, 1978.

Estimating the Area Fraction of Geophysical Fields from Measurements Along a Transect

Jeffrey R. Key

Abstract— **Methods are presented for estimating the fractional area coverage of geophysical phenomena from measurements taken along a transect. These methods are most useful for assessing potential errors in sampling strategies but may also be used for the analysis of data. The procedure provides a means to compute confidence interval estimates of the true area fraction when the autocovariance function for the geophysical field is known or assumed. Another approach that does not require *a priori* knowledge of the underlying autocovariance function is described for the special case of linear featurees modeled as a Poisson line process.**

I. Introduction

Valuable information about many geophysical phenomena is collected along lines by aircraft, ships, submarines, and ground personnel. For example, airborne LIDAR and radiometer data are used to infer cloud optical properties and surface characteristics; ice draft and width distributions of linear openings in sea ice ("leads") have been measured using upward-looking submarine sonar; field personnel report the fractional coverage of vegetation types along ground transects; and image data are often sampled along lines to reduce processing time. All these measurements are, however, one dimensional and provide no direct information about the area coverage of clouds, sea ice leads, or vegetation classes. In models that are two-dimensional in nature the spatial coverage of these parameters is of great importance.

The purpose of this communication is to describe a method of estimating the fractional area coverage of geophysical variables from measurements made along a line. The basic ideas are taken from theorems developed in the field of stochastic geometry but have not previously been presented as a general tool for remote sensing and geophysical studies. First the general lineal method is presented. With this method the underlying spatial structure of the geophysical field described by its autocovariance function is employed. Applications to cloud and sea ice lead patterns in simulated satellite imagery are then presented. In cases where the geophysical field can be adequately modeled by a known stochastic process, other methods of estimating the areal coverage may be available. One such example is given for leads modeled as a Poisson line process.

Manuscript received September 29, 1992; revised April 30, 1993. This work was supported by ONR Grant N00014-90-J-1840 and NASA Grant NAGW-2407.

The author is with the Cooperative Institute for Research in Environmental Sciences, Division of Cryospheric and Polar Processes, University of Colorado, Boulder, CO 80309-0449.

IEEE Log Number 9212139.

II. General Lineal Method

The general expression for the estimate of the fractional area coverage, p', of a geophysical parameter whose actual fractional coverage is p, regardless of the spatial structure of that parameter is

$$p' = \mu_U^{-1} \int_U I(\boldsymbol{x}) d\boldsymbol{x}$$

where $I(\boldsymbol{x})$ is the indicator function for the underlying function $q(\boldsymbol{x})$ at location $\boldsymbol{x}$ and the factor μ_U is needed for normalization and depends on the shape of the structuring element U (e.g., it may be the length of a line or the area of a square). The indicator function takes on a value of 1 if $q(\boldsymbol{x})$ satisfies some condition and 0 otherwise. For example, if a thresholding procedure is used to determine whether or not each pixel in an image or each data point in submarine sonar data represents some phenomenon, then $I(\boldsymbol{x}) = 1$ if the data value passes the threshold test and $I(\boldsymbol{x}) = 0$ otherwise. Again, this applies to any structuring element U.

Following Stoyan *et al.* [1] the expected value of p', $E(p')$, is p and its variance is

$$\mathrm{var}(p') = E(p' - p)^2 = \mu_U^{-2} \int_U \int_U k_I(||\boldsymbol{x} - \boldsymbol{y}||) d\boldsymbol{x} d\boldsymbol{y}$$

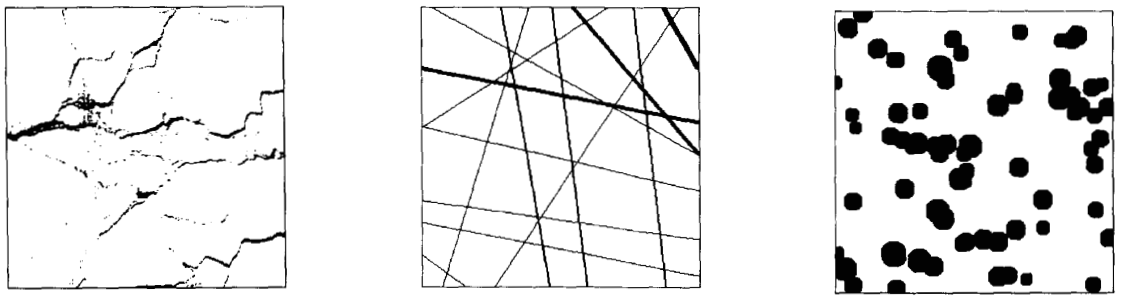
where k_I is the autocovariance function of the indicator function I:

$$k_I(r) = E[I(\boldsymbol{x}) I(\boldsymbol{x} + r)] - E[I(\boldsymbol{x})]^2.$$

The displacement or lag $r = ||\boldsymbol{x} - \boldsymbol{y}||$ is such that the covariance depends only on the distance between the two points and not on direction. The assumptions are that the geophysical field $q(\boldsymbol{x})$ is stationary and isotropic.

Now we consider the case where the structuring element is a line. For measurements along an array of N parallel lines, each of length L, the unbiased element of the fractional area coverage is

$$p' = \frac{l}{NL} \tag{1}$$



(a) (b) (c)

Fig. 1. (a) Binary image based on a Landsat MSS band 4 scene of the ice pack north of Alaska in March 1988. Field-of-view is 80 m; area covered is approximately (24 km)^2. (b) A simulated lead network modeled as a Poisson line process with thick lines. Field-of-view is 137.5 m; image size is approximately (42 km)^2. (c) A simulated cloud field based on a random disk model. Field-of-view is 137.5 m; image size is approximately (42 km)^2.

where l is the total length of NL where $I(\boldsymbol{x}) = 1$. Extending the work of Rothrock and Thorndike [2], the estimation variance is

$$\mathrm{var}(p') = N^{-1}L^{-2}\int_{\mathcal{L}}\int_{\mathcal{L}} k_I(|x-y|)dxdy \tag{2a}$$

$$= 2N^{-1}L^{-2}\int_{\mathcal{L}}(L-r)k_I(r)dr \tag{2b}$$

where $\mathcal{L}$ is the set containing the N lines and $r = |x-y|$, the distance between two locations x and y on the line.

Exponential covariance is a reasonable model for many geophysical parameters and is used here:

$$k_I(r) = p(1-p)e^{-\alpha r} \qquad (r, \alpha \geq 0) \tag{3}$$

where α describes the dependence of the covariance on the distance r. The parameter α can be determined from observed autocovariances by rewriting (3) in linear form as

$$\ln[k_I(r)] = \ln[p(1-p)] - \alpha r \tag{4}$$

and estimating α from the data through a least squares regression.

In the case of exponential covariance the estimation variance in (2) is

$$\mathrm{var}(p') = \frac{2p(1-p)\left(1 + \dfrac{e^{-\alpha L}-1}{\alpha L}\right)}{\alpha NL} \tag{5a}$$

$$\approx \frac{2p(1-p)\left(1 - \dfrac{1}{\alpha L}\right)}{\alpha NL} \tag{5b}$$

III. Application

One real and two simulated images are used in the application of the above methods. The real image is a Landsat Multispectral Scanner (MSS) band 4 (0.5–0.6 μm) scene of the Beaufort Sea, March 1988. The binary image produced by applying a threshold to the original grey-scale image is shown in Fig. 1(a). The pixel size is 80 m; image size is 24 × 24 km, a subset of a full Landsat scene. Next, a network of leads is simulated as a Poisson line process. The mean spacing between lines (leads) is 3000 m and their orientations are random. The lines are assigned thicknesses (widths) following the negative exponential density function:

$$f_W(w) = \frac{1}{\lambda}e^{-w/\lambda}$$

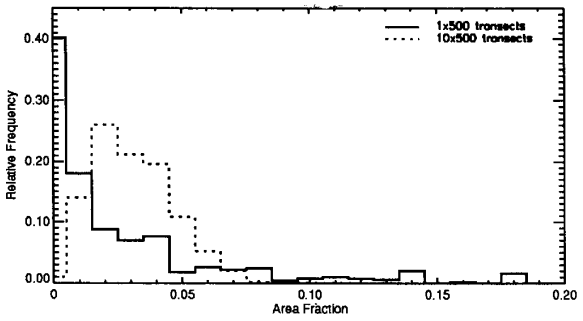
where w is lead width and λ is the mean width. For the simulation $\lambda = 200$ m. One realization of the Poisson line process is shown in Fig. 1(b), again as a binary image, where the pixel size is 137.5 m. Last, a cloud field is simulated as an ensemble of disks whose diameters are approximately normally distributed (in a true Gaussian distribution negative diameters would be possible) and whose center locations follow a binomial point process. This model is appropriate for cumuliform clouds but is obviously not applicable to stratiform cloud decks. One realization is shown in Fig. 1(c).

TABLE I
Regression-Estimated Parameters of the Autocovariance Function for the Images in Fig. 1

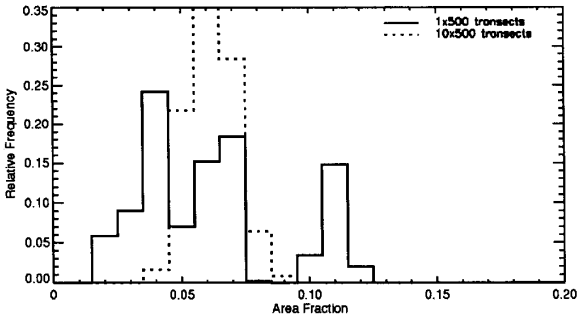
Figure	$p(1-p)$[1]	α	R
1a	0.029	0.319	-0.99
1b	0.078	0.554	-0.96
1c	0.218	0.121	-0.92

[1] Estimated with (4).

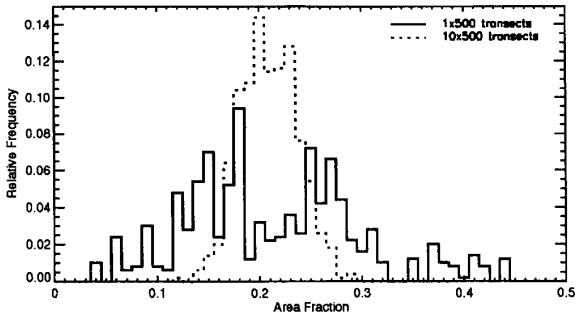
As stated earlier, the expected value of p' is p; i.e., the mean of the sampling distribution of sample proportions is the same as the population or true mean. The variance of the sample proportions is given by (2) in the general case and (5) for exponential covariance. How well does this theory compare to observations? As the first step, the "true" autocovariance functions were estimated from six random, horizontal transects through each image. The α coefficient in (4) was computed for each of the six transects and then averaged. With the simulated clouds and leads the horizontal transects should adequately represent the two-dimensional structure since the patterns are isotropic. This is less true, however, with the Landsat image where a preferred orientation is apparent. Table I gives the results of the least squares fit of the exponential autocovariance function to the observed autocovariances in the images. Listed are the regression-estimated variance (the antilog of the y-intercept in (4)), α, and the correlation coefficient. Next the distributions of sample fractional area estimates and their first two moments (mean and variance) were determined by computing p' with (1) for each of 500 single, random, horizontal transects through each image (i.e., the number of transects used to calculate p' in (1), N, is 1 for each of 500 samples). Distributions of p' were also computed for sets of ten such parallel transects giving



(a)



(b)



(c)

Fig. 2. Relative frequency histograms of the distribution of estimated area fraction for different total line lengths. Plots correspond to the images in Figs. (a)–(c). (a) Frequency of area fraction Landsat. (b) Frequency of area fraction simulated leads. (c) Frequency of area fraction simulated clouds.

5000 transects ($N = 10$ for each of 500 samples). Fig. 2 shows how the distribution of the estimates changes as a function of the number of transects or, effectively, total transect length. For single transects (solid lines) a broad range of p' are possible.

Table II gives the true area fraction p, the means and variances of the estimated area fraction p' from the observed distributions in Fig. 2, and the variance computed from (5b), where $L = 304$ pixels (same units as r in (4)). The true area fraction is the proportion of pixels in the binary image where $I(x) = 1$. It is apparent that the theoretical variance of the estimate given in (5) is generally applicable for the simulated leads and clouds in Figs. 1(b) and (c). For the Landsat data, however, the variance of the estimate for the 500-set simulation is three times as large as that computed with (5) or a factor of 1.7 for the standard deviation. This is due to the anisotropic nature of the lead network in the image and the large variability in the autocovariance function computed for individual transects. A

TABLE II
ACTUAL AND ESTIMATED FRACTIONAL AREA COVERAGES FOR FIGS. 1(a)–(c) USING ONE AND TEN TRANSECTS

		N = 1x500		N = 10x500	
Figure	p	E(p')	Var(p')	E(p')	Var(p')
1a	0.035	0.035	2.07e-3	0.036	2.35e-4
			(6.84e-4)[1]		(6.84e-5)
1b	0.067	0.067	7.56e-4	0.067	7.62e-5
			(7.38e-4)		(7.38e-5)
1c	0.215	0.218	7.28e-3	0.214	7.78e-4
			(8.93e-3)		(8.93e-4)

[1] Values in parentheses are computed with (5b).

two-dimensional autocovariance function and a modification of (5) may be needed in such cases.

The fact that the distributions of area fraction estimates tends toward Gaussian as N increases suggests a method for hypothesis testing and confidence interval estimates. If a normal distribution is assumed to apply, then the probability that a particular area fraction estimate comes from a population with area fraction p can be determined. This is perhaps most useful for confidence interval estimates of the true area fraction, defined as

$$p' - z(\beta/2)\mathrm{sd}(p') \quad \text{to} \quad p' + z(\beta/2)\mathrm{sd}(p')$$

where $\mathrm{sd}(p') = [\mathrm{var}(p')]^{1/2}$ is the standard deviation (sd) of p' and $1 - \beta$ is the level of confidence. Since the variance depends on an unknown p, p' is used as an estimate in (5). As an example, suppose that for Fig. 1(b) samples of one and ten transects are taken and for both $p' = 0.05$. Assuming the autocovariance structure given in Table I and computing the variance of the estimate with (5b) the confidence interval estimate of the true area fraction at the 90% confidence level is [0.011, 0.089] for a single transect and [0.038, 0.062] for a sample of ten transects. Neither of these intervals contains the true area fraction $p = 0.067$. Of course, the probability of obtaining such a p' from a population with a true fraction of 0.067 is very small, particularly for the set of ten transects (0.01 as opposed to 0.24 for a single transect), so that this example is improbable but useful for illustration. If, on the other hand, we obtain a sample p' of 0.07 then the confidence interval estimate of p is [0.024, 0.116] for a single sample transect and [0.056, 0.084] for a set of ten. Both contain the true area fraction but the larger sample size gives a much smaller range for the estimate.

The shortcoming of this approach is that the autocovariance function must be known. It may be possible to estimate it from the data itself if the sample size is large enough, although this is somewhat circular. In some cases it may be possible to infer a covariance structure by assuming a simple model of the geophysical variable, as done in [2] for sea ice floes. Even so, some knowledge of the field is needed; in their case the diameter of the floes. If, however, some basic autocovariance can be assumed for different cloud types, sea ice leads, etc., then the above procedure is certainly useful for planning sampling studies, and probably applicable to data analysis as well.

IV. Special Case: Poisson Procedures

For certain stochastic processes it is possible to determine the fractional area coverage from measurements along a line without any *a priori* knowledge of the process. Here such a possibility is given for a Poisson line process like the one used above as a model of leads.

For a Poisson process the area fraction is related to the intensity[1] τ of the process and the mean "area" of the objects ζ:

$$p' = \Pr[O \text{ is covered}] = 1 - e^{-\tau\zeta} \tag{6}$$

where O is an arbitrary origin. The area measure corresponds in units to the intensity measure; e.g., for leads the intensity is the number of points per unit distance and ζ is mean lead width.

The area fraction can be now estimated from lineal measurements through the use of the line (lead) thickness (width) distribution. The area term in (6) is the overall mean line thickness, W, defined as

$$W = \pi^{-1} \int_0^{\pi} w(\theta) d\theta$$

where $w(\theta)$ is the mean thickness of lines with orientation θ ($0 \leq \theta \leq \pi$) [4]. This applies to lines oriented isotropically; i.e., with a uniform distribution such that $f_\Theta(\theta) = \pi^{-1}$ where f_Θ is the probability density function for line (lead) orientations. For anisotropic thick lines then

$$W = \int_0^{\pi} w(\theta) dF_\Theta(\theta)$$

where $dF_\Theta(\theta) = f_\Theta(\theta) d\theta$, and F_Θ is the cumulative distribution function for orientations. A method for determining the actual lead width distribution, and hence W, from the width distribution measured along a transect has been presented in [3].

As an example of the use of (6), the lead network in Fig. 1(b) was generated with $\tau = 1/3$ (3-km mean spacing) and $W = 0.2$ km. This gives a p' estimate using (6) of 0.064 compared to the value of 0.067 reported in Table II. The discrepency is a function of the image creation and thresholding process, where all leads must fill an entire pixel.

In practice the intensity of the process is not known. For leads modeled as a Poisson line process an estimate of τ can be obtained from the transect data, where the points of intersection of the transect with leads constitute a Poisson process of intensity $2\tau/\pi$. The accuracy of this estimate depends on the size of the region over which the measurements are made. For Fig. 1(b) estimates of τ range from 0.19 to upwards of 0.45 which results in an estimate of p' in the range of 0.037 to 0.086. There is, of course, some variability in the estimate of W as well, which is discussed in [3].

V. Summary

A general method has been presented that allows for the assessment of potential errors in estimating the fractional coverage of geophysical variables from measurements along a line. Potential applications include the analysis of field data from aircraft, ships, and submarines as well as data collected on along ground transect by field personnel. For image processing the primary use of transect measurements is in the sampling of very large data sets. By application to fields of clouds and sea ice fractures it was shown how the variance of the estimate of area fraction depends on the spatial structure and the number and length of transects in the sample. With a single, short transect the estimated fractional coverage has a large variance. With large samples the sampling distribution of sample proportions tends towards normal with a mean equal to the population or true mean, so that confidence interval estimates and hypothesis tests are possible.

The shortcoming of the approach is that the autocovariance function must be known. If, however, it is assumed that some basic autocovariance structure exists for different cloud types, sea ice leads, etc., possibly as a function of the time of year and/or geographic location, the general lineal method is a useful tool. Even if such *a priori* knowledge of the geophysical field is not available, the method allows for the assessment of sampling errors and the design of sampling strategies in a general sense. In cases where the spatial structure of a geophysical variable can be described by a particular stochastic process such as a Poisson process, other methods of estimating the area coverage may be available.

[1] The *intensity* of a stochastic process is commonly called the *density* of the process. The former term is used here in order to avoid confusion with the concept of probability density.

Acknowledgment

The author thanks H. Maybee and S. Peckham, and D. Rothrock for useful discussions.

References

[1] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*. New York: Wiley, 1989.
[2] D. A. Rothrock and A. S. Thorndike, "Measuring the sea ice floe size distribution," *J. Geophys. Res.*, vol. 89, no. C4, pp. 6477–6486, 1984.
[3] J. Key and S. Peckham, "Probable errors in width distributions of sea ice leads measured along a transect," *J. Geophys. Res.*, vol. 96, no. C10, pp. 18417–18423, 1991.
[4] R. E. Miles, "Random polygons determined by random lines in a plane," *Proc. Nat. Acad. Sci.*, vol. 52, pp. 901–907, 1964.

LAI Inversion Using a Back-Propagation Neural Network Trained with a Multiple Scattering Model

James A. Smith

***Abstract*— Standard regression methods applied to canopies within a single homogeneous soil type yield good results for estimating leaf area index (LAI) but perform unacceptably when applied across soil boundaries. In contrast, the neural network reported here generally yielded absolute percentage errors of < 30%. The network was applied, without retraining, to a Landsat TM.**

I. Introduction

Current and projected satellite sensor systems, e.g., the Moderate Resolution Imaging Spectrometer (MODIS) [1], are able to obtain global and repetitive observations at high temporal sampling rates and many studies have demonstrated the utility of vegetation indices at continental scales for estimating photosynthetic processes and leaf area index (LAI) in plant communities [2]. Others, however, have

Manuscript received May 17, 1993.
The author is with the NASA Goddard Space Flight Center, Greenbelt, MD 20771.
IEEE Log Number 9212140.