MONITORING LANDSCAPE LEVEL PROCESSES USING REMOTE SENSING OF LARGE PLOTS

Raymond L. Czaplewski
Mathematical Statistician

USDA Forest Service Research
Rocky Mountain Forest and Range
Experiment Station
240 W. Prospect Street
Fort Collins, CO 80526 U.S.A.

ABSTRACT

Global and regional assessments require timely information on landscape level status (e.g., areal extent of different ecosystems) and processes (e.g., changes in land use and land cover). To measure and understand these processes at the regional level, and model their impacts, remote sensing is often necessary. However, processing massive volumes of remotely sensing data can be infeasible if high resolution data are required for very large regions. Remote sensing of sample plots, rather than a census of the entire area, can solve certain problems. Statistical aspects of remote sensing for large plots are described, concentrating on methods needed to produce sample estimates, combine time series of ancillary estimates from other sources, calibrate for misclassification bias, and combine remotely sensed data with model predictions. These methods might improve spatial and temporal accuracy, and test our understanding of processes that are captured in landscape level models.

KEY WORDS: Kalman filter, composite estimator, classification error, calibration, landscape models, landscape models, spatial autocorrelation, spatial heterogeneity.

1. INTRODUCTION

A simple example is used in the following discussion, where status is defined as the proportion forested area, and ground-based field work is used to determine if a point is truly forested. This hypothetical example uses the sampling frame and certain design features proposed by the U.S. Environmental Protection Agency as part of its Environmental Monitoring and Assessment Program (EMAP). This is a cooperative program among several agencies of the United States Government, including the USDA Forest Service. The EMAP sampling frame is composed of a triangular grid. Each 640 km^2 hexagon on this grid contains a 40 km^2 hexagon (i.e., a 1/16 sample by area) that is observed using Landsat data and high altitude aerial photography. However, the statistical models readily apply to sampling frames used in other programs, such as that proposed by the Food and Agricultural Organization of the United Nations for monitoring and assessment of the world's tropical forests, or that proposed by Czaplewski *et al.* (1987) for updating forest inventory estimates made by the USDA Forest Service.

Underlining statistical models for estimated status of the stratum and each sample unit are presented in Section 2. Section 3 gives a calibration model for measurement error that occurs when true status can not be perfectly classified with remotely sensing. Section 4 presents an illustration of the statistical composite estimator, which combines estimates from different sources. In Section 5, the composite estimator is used to combine calibrated, remotely sensed estimates from many sample units into an estimate of stratum status. The Kalman filter, which is a more general composite estimator, is introduced in Section 6. In Section 7, the Kalman filter is used to update estimates of status for each sample unit, and combine them into an estimate of status for the stratum. Section 8 uses the Kalman filter to combine ancillary estimates for aggregations of sample units. The cycle of landscape monitoring and process modeling is discussed in Section 9.

2. SAMPLE FRAME

Consider a sampling frame composed of a grid of large sample units that are well suited for monitoring using remote sensing. The frame is confined to a contiguous, homogeneous geographic area, i.e., a stratum. Estimates for status of multiple strata could be summed for regional or global assessments if definitions for true status are shared among strata.

2.1 Example sampling frame

Assume a large, homogeneous, geographically contiguous stratum is comprised of a known number of cells (n), e.g., 640 km^2 hexagons. A stratum might be the coastal plain in the southeastern United States, which is 320,000 km^2 in size; the number of cells n would be (320,000/640)=500. Each 640 km^2 cell is sampled with a single 40 km^2 sample unit, which is centered on the 640 km^2 cell. Each 40 km^2 sample unit is treated as a permanent plot, and each is periodically observed over time using remote sensing.

2.2 Statistical model for status

The status of the stratum (e.g., proportion forest) is denoted as unknown nonrandom variable X. Since the stratum is assumed spatially homogeneous, status of each sample unit in the stratum is assumed equal to the stratum status X. Deviation of the observed status of sample unit i from the stratum status is the unknown random variable W_i. The statistical model for estimated status of sample unit i is

$$\underline{X}_i = X + W_i, \text{ for } i = \{1, 2, \dots, n\} \qquad (1)$$

$\underline{X}_i$ is assumed an unbiased estimate of X, i.e., $E[W_i] = 0$. Variance of W_i, denoted var(W_i), is assumed heterogeneous among the n sample units in the stratum, i.e., var(W_i) does not necessarily equal the var(W_j) for i not equal to j. Deviations among sample units are not assumed independent, i.e., $E[W_i W_j]$ might be nonzero. No other distributional assumptions are made.

Invited paper presented at international symposium "Global and Environmental Monitoring: Techniques and Impacts", International Society for Photogrammetry and Remote Sensing, Victoria, British Columbia, Canada, September 17-21, 1990

3. MISCLASSIFICATION IN REMOTE SENSING

A portion of deviation W_i will be caused by errors in measuring the status (e.g., proportion forested) of sample unit i.

3.1 Misclassification error model

Let a measurement or calibration model for the unknown true status X_i of sample unit i be

$$X_i = H_f \; Y_i + H_o \; (1-Y_i), \tag{2}$$

where Y_i is the imperfect remotely sensed estimate of proportion forest in sample unit i, and H_f is the known conditional probability that a point is truly forest given that it is classified as forest by the remote sensing process. Similarly, $(1-Y_i)$ is the remotely sensed proportion measurement of other cover, and H_o is the conditional probability that a point is truly forest given it is classified as other cover by the remote sensing process. When classification accuracy of the remote sensing process is high, H_f will nearly equal 1, and H_o will nearly equal 0.

3.2 Misclassification bias

The remotely sensed estimate Y_i is a biased estimate of true status X_i of sample unit i if classification errors occur. Solving (2) for Y_i,

$$Y_i = (X_i - H_o) \; / \; (H_f - H_o). \tag{3}$$

The remotely sensed estimate Y_i in equation (3) will not equal the true status X_i unless H_f equals 1 and H_o equals 0, i.e., perfect classification accuracy.

3.3 Estimation of calibration model

In practice, the values of H_f and H_o are not perfectly known. Rather, H_f and H_o are assumed the same for all sample units in the stratum, and their values are estimated using a finite sample of reference points for which the remotely sensed and true status are known. For example, reference points might be available for M systematically located 0.4 ha forest inventory plots, which are measured in the field by USDA Forest Service crews, where the field classification is considered to be without error. Under certain conditions, the location of these field plots can be accurately registered to remotely sensed images, so that both remotely sensed and true classifications are available for a small sample of point plots. This would provide the necessary sample of reference points to make estimates ($\underline{H}_f$ and $\underline{H}_o$) of the true conditional probabilities (H_f and H_o).

Consider the statistical sampling model:

$$\underline{H}_f = H_f + J_f, \quad \underline{H}_o = H_o + J_o, \tag{4}$$

where J_f and J_o are random variables that equal the differences between the true and estimated conditional probabilities. $\underline{H}_f$ might be estimated from the M_f 0.4 ha Forest Service plots classified as forest using remote sensing equation (5):

$$\underline{H}_f = [(H_f)_1 + (H_f)_2 + \ldots + (H_f)_{M_f}]/M_f, \tag{5}$$

where $(H_f)_i = 1$ if 0.4 ha Forest Service plot i is truly forest given it is classified as forest using the remote sensing procedure, and $(H_f)_i = 0$ otherwise. Similarly, $\underline{H}_o$ might be estimated from the M_o 0.4 ha Forest Service plots that are classified as other cover using the remote sensing procedure,

$$\underline{H}_o = [(H_o)_1 + (H_o)_2 + \ldots + (H_o)_{M_o}]/M_o, \tag{6}$$

where $(H_o)_i = 1$ if 0.4 ha plot i is truly forest given it is classified as other cover using remote sensing, and $(H_o)_i = 0$ otherwise.

The following is an estimate ($\underline{X}_i$) of the status of sample unit i, using the estimated conditional probabilities (H_f and H_o) of correct and incorrect remotely sensed classifications from (5) and (6), and the known remotely sensed status Y_i (Tenenbein 1972):

$$\underline{X}_i = \underline{H}_f \; Y_i + \underline{H}_o \; (1-Y_i). \tag{7}$$

3.4 Variance of calibrated estimate

From equations (2), (4), and (7), the unbiased estimate $\underline{X}_i$ of the true status X_i of sample unit i, given the imperfect remotely sensed measurement Y_i of the same sample unit, is

$$\begin{aligned}
\underline{X}_i &= (H_f + J_f) \; Y_i + (H_o + J_o) \; (1-Y_i), \\
&= [H_f \; Y_i + H_o \; (1-Y_i)] + [J_f \; Y_i + J_o \; (1-Y_i)], \\
&= X_i + [J_f \; Y_i + J_o \; (1-Y_i)].
\end{aligned} \tag{8}$$

The estimate $\underline{X}_i$ of the status of sample unit i in (8) contains uncertainty propagated from the imperfect model for classification error. Since the estimated conditional probabilities (H_f, J_f) are assumed unbiased, $E[J_f] = E[J_o] = 0$, and measurement Y_i is a known nonrandom constant, then the variance of the estimate in (8) is

$$\begin{aligned}
\mathrm{var}(\underline{X}_i) &= [J_f \; Y_i + J_o \; (1-Y_i)]^2, \\
&= E[J_f^2] \; Y_i^2 + E[J_o^2] \; (1-Y_i)^2, \\
&= \mathrm{var}(\underline{H}_f) \; Y_i^2 + \mathrm{var}(\underline{H}_o) \; (1-Y_i)^2.
\end{aligned} \tag{9}$$

If it is assumed that there are no registration errors between field points and the remote sensing imagery, then the random errors J_f and J_o are caused solely by sampling error. The sampling variances $\mathrm{var}(\underline{H}_f)$ and $\mathrm{var}(\underline{H}_o)$ can be estimated from the simple randomized sample of $M=M_f+M_o$ plots using the binomial distribution:

$$\mathrm{var}(\underline{H}_f) = \underline{H}_f \; (1-\underline{H}_f) \; / \; M_f, \tag{10}$$

$$\mathrm{var}(\underline{H}_o) = \underline{H}_o \; (1-\underline{H}_o) \; / \; M_o. \tag{11}$$

From equations (9), (10), and (11), the variance of the estimated status of sample unit i is

$$\mathrm{var}(\underline{X}_i) = \frac{\underline{H}_r\ (1-\underline{H}_r)\ Y_i^2}{M_r} + \frac{\underline{H}_o\ (1-\underline{H}_o)\ (1-Y_i)^2}{M_o} \quad (12)$$

The estimated variance $\mathrm{var}(\underline{X}_i)$ in equation (12) is nearly zero when there is near perfect classification accuracy ($\underline{H}_r$ -> 1 and $\underline{H}_o$ ->0), or when there are a large number of plots (M_r and M_o) used to estimate the conditional misclassification probabilities $\underline{H}_r$ and $\underline{H}_o$.

3.5 Heterogeneity caused by calibration

The variance $\mathrm{var}(\underline{X}_i)$ of estimated status $\underline{X}_i$ for each sample unit i in (12) differs among sample units because the remotely sensed measurement Y_i will differ among sample units. Measurement error can produce heterogeneous variance among sample units. This is one motivation for composite estimation in Section 5.

3.6 Lack of independence caused by calibration

Propagated errors from the stratum level calibration model will cause a lack of independence among the estimated status of sample units within that stratum. From equation (8), the covariance between the unbiased estimates $\underline{X}_i$ and $\underline{X}_j$ of status of sample units i and j is

$$\mathrm{cov}(\underline{X}_i, \underline{X}_j) = \mathrm{E}[\{J_r Y_i + J_o(1-Y_i)\}\{J_r Y_j + J_o(1-Y_j)\}].$$

$\mathrm{E}[J_r^2] = \mathrm{var}(\underline{H}_r)$ in (10), and $\mathrm{E}[J_o^2] = \mathrm{var}(\underline{H}_o)$ in (11). Since J_r and J_o are independently estimated from different plots, $\mathrm{E}[J_r\ J_o] = 0$, and

$$\mathrm{cov}(\underline{X}_i, \underline{X}_j) = \mathrm{var}(\underline{H}_r) Y_i Y_j + \mathrm{var}(\underline{H}_o)(1-Y_i)(1-Y_j).$$

Errors propagated from the stratum level calibration model will cause dependence among estimated status of sample units within that stratum. This covariance will be positive, and will differ among sample units because the remotely sensed estimates Y_i and Y_j differ. This is one motivation for Section 5.4, which considers dependent sample units.

4. COMPOSITE ESTIMATORS

Composite estimators are often used in forestry (Gregoire and Walters 1988), including sampling with partial replacement. A composite estimator (Fig. 1) combines two estimates, each of which is weighted inversely proportional to its variance (or in the multivariate case, its covariance matrix). The weights can be derived using maximum likelihood, minimum variance, or Bayesian theory. If all assumptions are reasonable, error in a composite estimate is less than error in either prior estimate.

5. ESTIMATED STRATUM STATUS

If the variances $\mathrm{var}(W_i)$ for the calibrated estimates of sample unit status were homogeneous for the entire stratum, then the mean of the n sample units would be the minimum variance unbiased estimate of the stratum status. If $\mathrm{var}(W_i)$ is heterogeneous, the sample mean would

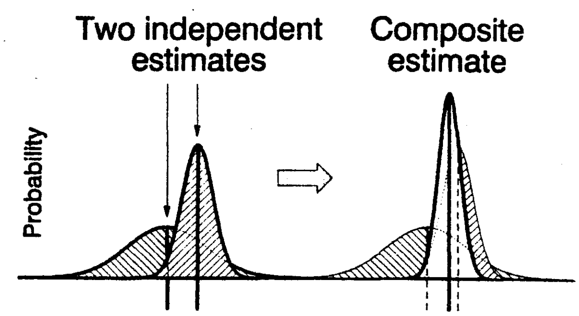

Fig. 1 Probability densities for two independent estimates. These are weighted inversely proportional to their variances, and combined into a single, more precise, composite estimate.

remain unbiased, but would not be the minimum variance estimate. As an alternative, the composite estimator (Maybeck 1979) could weight the estimated status $\underline{X}_i$ of each sample unit inversely proportional to its variance.

5.1 Estimated stratum status, two sample units

Estimates of stratum status from two sample units ($\underline{X}_1$, $\underline{X}_2$) can be combined using the composite estimator into an estimate $\underline{X}_{\#2}$ of stratum status X. The composite estimator uses weights, $(1-G_2)$ and G_2, that are inversely proportional to the variances of estimates $\underline{X}_1$ and $\underline{X}_2$:

$$\underline{X}_{\#2} = (1-G_2)\underline{X}_1 + G_2\underline{X}_2, \quad (13)$$

$$G_2 = \underline{\mathrm{var}}(\underline{X}_1) / [\underline{\mathrm{var}}(\underline{X}_1) + \underline{\mathrm{var}}(\underline{X}_2)]. \quad (14)$$

Estimated variances $\underline{\mathrm{var}}(\cdot)$ are used rather than their true, but unknown variances. In this case, estimate $\underline{X}_{\#2}$ in (13) would not be optimal with respect to minimum variance (Maybeck 1979). However, estimate $\underline{X}_{\#2}$ will be nearly optimal if $\mathrm{var}(\underline{X}_1)$ and $\mathrm{var}(\underline{X}_2)$ are accurate estimates, and assuming estimates $\underline{X}_1$ and $\underline{X}_2$ are independent. Even if estimate $\underline{X}_{\#2}$ is suboptimal, $\underline{X}_{\#2}$ is unbiased; from equations (1) and (13),

$$\underline{X}_{\#2} = (1-G_2)(X + W_1) + G_2(X + W_2),$$

$$= X + (1-G_2)W_1 + G_2 W_2. \quad (15)$$

Since random deviations W_1 and W_2 have expected values of zero in (1), and G_2 is a nonrandom constant, then $\mathrm{E}[\underline{X}_{\#2}] = X$ in (15), and estimate $\underline{X}_{\#2}$ is unbiased. Variance of the composite estimate using the first two sample units is

$$\mathrm{var}(\underline{X}_{\#2}) = (1-G_2)^2\ \mathrm{var}(\underline{X}_1) + (G_2)^2\ \mathrm{var}(\underline{X}_2). \quad (16)$$

5.2 Efficiency of the composite estimate

The statistical efficiency of the composite estimator is compared to the efficiency of the mean of estimates $\underline{X}_1$ and $\underline{X}_2$. Substituting G_2 from equation (14) into equation (16), variance of the composite estimate is given in (17)

$$\mathrm{var}(\underline{X}_{*2}) = \frac{\mathrm{var}(\underline{X}_2)^2\mathrm{var}(\underline{X}_1)+\mathrm{var}(\underline{X}_1)^2\mathrm{var}(\underline{X}_2)}{[\mathrm{var}(\underline{X}_1) + \mathrm{var}(\underline{X}_2)]^2},$$

$$= \frac{\mathrm{var}(\underline{X}_1)\ \mathrm{var}(\underline{X}_2)}{[\mathrm{var}(\underline{X}_1) + \mathrm{var}(\underline{X}_2)]}, \tag{17}$$

$$= G_2\ \mathrm{var}(\underline{X}_1). \tag{18}$$

The mean $(\underline{X}_1+\underline{X}_2)/2$, is equivalent to equation (13) with $G_2 = 1/2$. The mean is unbiased, as shown in (15). Variance of the mean is

$$(1/2)^2\mathrm{var}(\underline{X}_1) + (1/2)^2\mathrm{var}(\underline{X}_2) =$$

$$\frac{\mathrm{var}(\underline{X}_1)\ \mathrm{var}(\underline{X}_2)}{[\mathrm{var}(\underline{X}_1) + \mathrm{var}(\underline{X}_2)]} + \frac{[\mathrm{var}(\underline{X}_1) - \mathrm{var}(\underline{X}_2)]^2}{4\ [\mathrm{var}(\underline{X}_1) + \mathrm{var}(\underline{X}_2)]}. \tag{19}$$

Under heterogeneity, $\mathrm{var}(\underline{X}_1)$ does not equal $\mathrm{var}(\underline{X}_2)$, and the variance of the mean in (19) is larger than the variance of the composite estimate in (17). Therefore, the composite estimate is more efficient than the mean. Under homogeneity, $\mathrm{var}(\underline{X}_1)$ equals $\mathrm{var}(\underline{X}_2)$, and the variance of the mean in (19) is identical to the variance of the composite estimate in (17).

5.3 Estimated stratum status, all sample units

The composite estimate of stratum status in (13) uses only the sample units from two cells. However, there are n units sampled in the stratum, which can be incorporated by sequentially applying the composite estimator. The composite estimate from the first two cells $\underline{X}_{*2}$ in (13) is combined with the estimate from the third sample unit $\underline{X}_3$ using the variance $\mathrm{var}(\underline{X}_{*2})$ of the composite in equation (16):

$$\underline{X}_{*3} = (1-G_3)\underline{X}_{*2} + G_3\underline{X}_3, \tag{20}$$

$$G_3 = \mathrm{var}(\underline{X}_{*2})\ /\ [\mathrm{var}(\underline{X}_{*2}) + \mathrm{var}(\underline{X}_3)], \tag{21}$$

$$\mathrm{var}(\underline{X}_{*3}) = (1-G_3)^2\ \mathrm{var}(\underline{X}_{*2}) + (G_3)^2\ \mathrm{var}(\underline{X}_3). \tag{22}$$

Then, composite estimate $\underline{X}_{*3}$ is combined with the estimate from the fourth sample unit. This is repeated until estimates from all n cells are combined into a single estimate for status of the entire stratum. The sequence is inconsequential.

As in equation (15), the final composite estimate is unbiased. As in (17) and (19), the composite estimate using all n cells has smaller variance than the mean, and the composite estimate is identical to the mean under homogeneity.

5.4 Lack of independence among sample units

Sequential application of the composite estimator will produce a minimum variance estimate if all random deviations W_i are mutually independent. However, spatial autocorrelation and other factors (Section 3.6) can cause dependence among deviations. Consider two sample unit estimates $(\underline{X}_1,\underline{X}_2)$ that are combined to produce an estimate of stratum status, i.e., equation (13). If estimates $\underline{X}_1$ and $\underline{X}_2$ are dependent, variance of the combined estimate $\underline{X}_{*2}$ in equation (13) is

$$\mathrm{var}(\underline{X}_{*2}) = (1-G_2)^2\ \mathrm{var}(\underline{X}_1) + (G_2)^2\ \mathrm{var}(\underline{X}_2)$$
$$+\ 2\ (1-G_2)(G_2)\ \mathrm{cov}(\underline{X}_1,\underline{X}_2), \tag{23}$$

where $\mathrm{cov}(\underline{X}_1,\underline{X}_2)$ is the covariance between sample unit estimates $\underline{X}_1$ and $\underline{X}_2$. If these two estimates are combined using the composite estimator, as in equations (13), (14), and (16), then the composite estimate is unbiased, as shown in equation (15), even if the errors are correlated.

Variance of the mean will be larger than the variance of the composite estimate when estimates $\underline{X}_1$ and $\underline{X}_2$ are not independent (Section 5.5), unless there is a strong negative covariance between estimates $\underline{X}_1$ and $\underline{X}_2$ such that $\mathrm{var}(\underline{X}_1)+\mathrm{var}(\underline{X}_2)$ is less than $-2\mathrm{cov}(\underline{X}_1,\underline{X}_2)$. However, $\mathrm{cov}(\underline{X}_1,\underline{X}_2)$ is frequently positive. Spatial patterns in landscapes tend to produce positive covariances between proximate sample units, as will propagated errors for stratum level calibration models (Section 3.6) and prediction models (Section 7.2).

The composite estimator can be sequentially applied, as described in (20) through (22). However, the final composite estimate will not be a minimum variance estimate. A more general formulation of the composite estimator (Maybeck 1979) will have minimum variance with correlated errors using the following weight in (13):

$$G_2 = \frac{\mathrm{var}(\underline{X}_1) + \mathrm{cov}(\underline{X}_1,\underline{X}_2)}{\mathrm{var}(\underline{X}_1) + \mathrm{var}(\underline{X}_2) + 2\ \mathrm{cov}(\underline{X}_1,\underline{X}_2)}. \tag{24}$$

However, this is unsatisfactory for sequential composite estimation with all n estimates; the final estimate depends upon the sequence in which the n estimates are combined. Section 5.5 presents a possible solution, using a vector weight analogous to the scalar weight in (24).

5.5 Vector weighting in the composite estimator

Let $\underline{\mathbf{X}}=(\underline{X}_1|\underline{X}_2|\ldots|\underline{X}_n)'$ be the $(n \times 1)$ vector of estimates from n sample units; $\mathbf{W}=(W_1|W_2|\ldots|W_n)'$ be the $(n \times 1)$ vector of deviations of n sample units from the stratum status X; $\underline{\mathbf{C}}$ be the $(n \times n)$ estimated covariance matrix for sample unit deviations, $\underline{E}[\mathbf{W}\ \mathbf{W}'] = \underline{\mathbf{C}}$, where iith element of $\underline{\mathbf{C}}$ is $\underline{\mathrm{var}}(\underline{X}_i)$ and the ijth element is $\underline{\mathrm{cov}}(\underline{X}_i,\underline{X}_j)$; and $\mathbf{1}$ be a $(n \times 1)$ vector of ones. Vector representation of equation (1) is

$$\underline{\mathbf{X}} = \mathbf{1}\ X + \mathbf{W}. \tag{25}$$

For n=2, the unbiased composite estimator in (13) and (16) can be expressed in matrix from as

$$\underline{X}_{*2} = \mathbf{G}'\ \underline{\mathbf{X}}, \tag{26}$$

$$\mathrm{var}(\underline{X}_{*2}) = \mathbf{G}'\ \underline{\mathbf{C}}\ \mathbf{G}, \tag{27}$$

where the (2×1) vector $\mathbf{G} = [(1-G_2)|G_2]'$ contains the weight applied to each estimate $\underline{X}_i$

in $\underline{X}$. Weights in (24) produce minimum variance estimates given heterogeneous, correlated errors and $n = 2$, and can be expressed in matrix form as

$$\mathbf{G}_c = (\mathbf{1}\,\mathbf{1}' - \mathbf{I})\,\underline{\mathbf{C}}\,\mathbf{1} \,/\, [\mathbf{1}'\,(\mathbf{1}\,\mathbf{1}' - \mathbf{I})\,\underline{\mathbf{C}}\,\mathbf{1}\,],$$

$$= \frac{(\mathbf{1}\,\mathbf{1}' - \mathbf{I})\,\underline{\mathbf{C}}\,\mathbf{1}}{(n-1)\,\mathbf{1}'\,\underline{\mathbf{C}}\,\mathbf{1}}, \qquad (28)$$

where $\mathbf{I}$ is the $(n \times n)$ identity matrix. Based on *ad hoc* rationale for $n > 2$, the weights in equation (28) can be used in estimation equation (26), where $\mathbf{G} = \mathbf{G}_c$, for combining estimates from many sample units ($n > 2$) into an unbiased estimate of the stratum status. If the sample units are independent and have homogeneous variance, then s^2, $\underline{\mathbf{C}} = \mathbf{I}\, s^2$, and (28) becomes

$$\mathbf{G}_c = \frac{(\mathbf{1}\,\mathbf{1}' - \mathbf{I})\, s^2\, \mathbf{1}}{(n-1)\,\mathbf{1}'\, s^2\, \mathbf{1}} = \frac{(n-1)\,\mathbf{1}}{(n-1)\,n} = \frac{\mathbf{1}}{n}. \qquad (29)$$

The mean of the sample units can be expressed as matrix equation (26), where the $(n \times 1)$ weighting vector is

$$\mathbf{G}_m = \mathbf{1} \,/\, n. \qquad (30)$$

If the sample units are independent with homogeneous variance, then the composite estimate in (26) and (28) is identical to the mean because $\mathbf{G}_c$ from (29) is identical to $\mathbf{G}_m$ from (30).

5.6 Efficiency of vector composite estimator

When errors are heterogeneous and dependent, the composite estimator in (26) and (28) is expected to be more efficient than the mean of the sample units. From (27) and (28), variance $\mathrm{var}(\underline{X}_c)$ of the composite estimate $\underline{X}_c$ of stratum status X is

$$\mathrm{var}(\underline{X}_c) = \mathbf{G}_c'\,\mathbf{C}\,\mathbf{G}_c. \qquad (31)$$

The weighting vector $\mathbf{G}_m$ for the mean in equation (30) may be rewritten as

$$\mathbf{G}_m = \frac{\mathbf{1}\,\mathbf{1}'\,\underline{\mathbf{C}}\,\mathbf{1}}{n\,\mathbf{1}'\,\underline{\mathbf{C}}\,\mathbf{1}} = \frac{\mathbf{1}\,\mathbf{1}'\,\underline{\mathbf{C}}\,\mathbf{1} - \underline{\mathbf{C}}\,\mathbf{1} + \underline{\mathbf{C}}\,\mathbf{1}}{n\,\mathbf{1}'\,\underline{\mathbf{C}}\,\mathbf{1}},$$

$$= \frac{(\mathbf{1}\,\mathbf{1}' - \mathbf{I})\,\underline{\mathbf{C}}\,\mathbf{1}}{n\,\mathbf{1}'\,\underline{\mathbf{C}}\,\mathbf{1}} + \frac{\underline{\mathbf{C}}\,\mathbf{1}}{n\,\mathbf{1}'\,\underline{\mathbf{C}}\,\mathbf{1}}, \qquad (32)$$

From (27), (28), and (31), variance $\mathrm{var}(\underline{X}_m)$ of the mean estimate $\underline{X}_m$ of stratum status is

$$\mathrm{var}(\underline{X}_m) = \frac{\mathbf{G}_c'\,\mathbf{C}\,\mathbf{G}_c\,(n-1)^2}{n^2} + \frac{\mathbf{1}'\,\underline{\mathbf{C}}'\,\mathbf{C}\,\underline{\mathbf{C}}\,\mathbf{1}}{n^2\,(n-1)^2},$$

$$= \frac{\mathrm{var}(\underline{X}_c)\,(n-1)^2}{n^2} + \frac{\mathbf{1}'\,\underline{\mathbf{C}}'\,\mathbf{C}\,\underline{\mathbf{C}}\,\mathbf{1}}{n^2\,(n-1)^2}. \qquad (33)$$

Since $(n-1)^2/n^2$ in (33) will nearly equal to one for typical sample sizes n, variance $\mathrm{var}(\underline{X}_c)$ of the composite estimate in (31) will be smaller than variance $\mathrm{var}(\underline{X}_m)$ of the mean estimate in (33), unless there are large negative covariances among sample units in covariance matrix $\mathbf{C}$ such that the scalar $(\mathbf{1}'\underline{\mathbf{C}}'\mathbf{C}\underline{\mathbf{C}}\mathbf{1})$ in (33) is negative.

6. KALMAN FILTER

Misclassification in remote sensing will bias estimated status of individual sample units. A calibration model can correct for this bias, but the calibration model will introduce uncertainty into our remotely sensed estimate of the status of each sample unit (Section 3). Additional uncertainty is introduced by changes in land use, land management, and vegetation succession in each sample unit. If a deterministic model could predict these changes after the remotely sensed imagery is acquired, then this model can estimate status of each sample unit over time.

Predictions of the deterministic model can be incorporated into statistical estimates of stratum status using the Kalman filter (e.g., Gregoire and Walters 1988). Dixon and Howitt (1979) describe how the Kalman filter can be applied to sampling with partial replacement in continuous forest inventories. Czaplewski (1990) presents a simple tutorial example of estimating forest cover over time using the Kalman filter.

The Kalman filter is portrayed in Fig. 2. One unbiased estimate is made at time t (e.g., a calibrated remotely sensed measurement). The other unbiased estimate (e.g., a previous, calibrated remotely sensed measurement) is made at time t-1, but is updated for expected changes between times t and t-1 using the deterministic model. Variance for the updated estimate includes effects of errors in the previous estimate that are propagated over time, and prediction errors from the deterministic model.

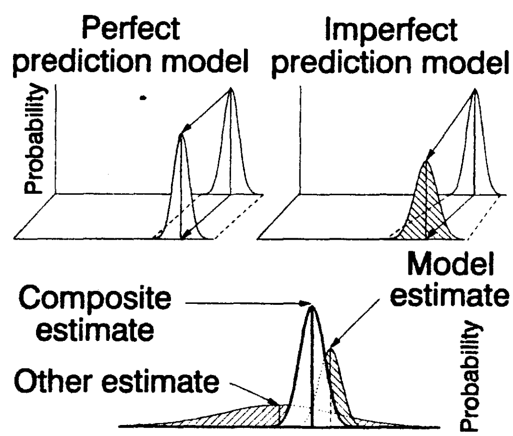

Fig. 2 Probability density for a Kalman estimate that is a composite of measurement data at time t and a prior estimate at time t-1, which is updated using a prediction model. Given a perfect prediction model, only estimation error at t-1 is propagated to time t. More realistically, the prediction model is imperfect, and a prediction error also occurs. The Kalman filter combines measurements and model predictions into a composite estimate, weighted inversely proportional to their variances.

The Kalman filter is usually applied to a time series of measurements (Fig. 3). With each new measurement, a composite estimate is made, which serves as new initial conditions for the next prediction from the deterministic model (e.g., Fig. 3, year 4).

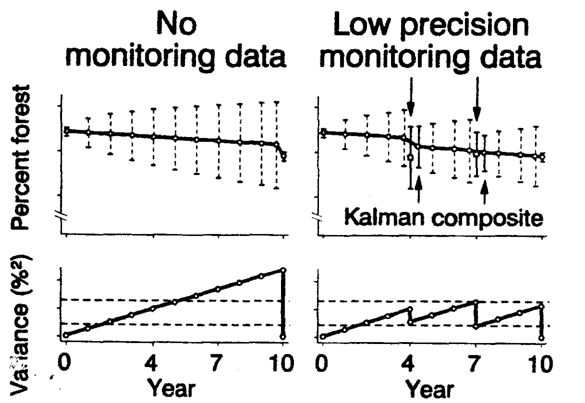

Fig. 3 Kalman estimates and confidence intervals for percent forest. In this example from Czaplewski, *et al.* (1988), intensive forest inventories were conducted in years 0 and 10; lower precision monitoring data were gathered in years 4 and 7.

The Kalman filter is a multivariate estimator (Maybeck 1979). It can simultaneously estimate multiple state variables, such as proportions of different vegetation cover types. Measured rates of change can be statistically combined with rates of change predicted from the deterministic model. The Kalman filter can model correlated errors among the state variables and rate coefficients, correlated prediction errors from the deterministic model, and random errors in measurement data.

6.1 Verification of the Kalman filter

Two independent estimates disagree or "diverge" in that neither estimate is likely given the other (Fig. 4). Contradictory estimates can be combined, but the resulting composite estimate can be biased. Discrepancies are can be caused by biased estimates of the error distribution (either location or spread) of the measurement at time t, or the estimate at time t-1 that is updated to time t using the deterministic prediction model.

It is possible that bias exists in the current measurement. For example, calibration equations are needed to correct for misclassification bias, as discussed in Section 3. Also, bias might exist in the estimated variance of errors from the prediction model; direct estimates of prediction variance require known differences between model predictions and the true status of the system. As an alternative, adaptive filters modify initial variance estimates until disagreements are within acceptable bounds (Fig. 4), often using a time series of residuals (Sorenson 1985). As accuracy of model predictions increases, the weight placed on model predictions will increase, and as will accuracy of the Kalman filter.

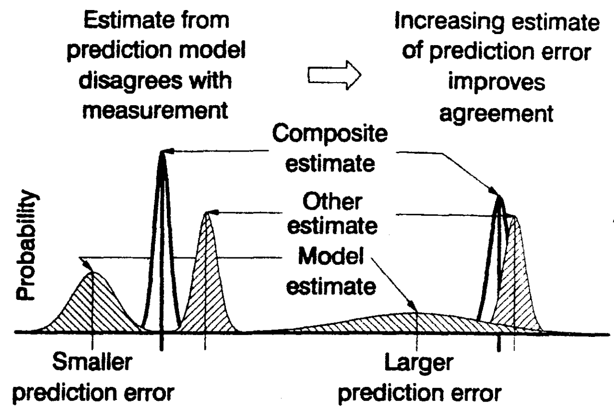

Fig. 4 Expected probability densities for two estimates that disagree. Adaptive filters assume the estimated variance of model prediction error is inaccurate, and change this estimate until the disagreement is within acceptable bounds.

7. KALMAN FILTER APPLIED TO SAMPLE UNITS

Consider the following hypothetical example, in which a 1/4 subsample of the sample units in the stratum are observed using remote sensing with imagery acquired at time t in an interpenetrating design. (Similar examples could be based on other intensities, such as 1/7, 1/9, 1/12, etc.) The biased estimate of the status of each sample unit in the 1/4 subsample is corrected using a stratum level calibration estimator, as in Section 3. An estimate of stratum status at time t is made with the 1/4 subsample using composite estimation (Section 5). A different 1/4 subsample of sample units is observed using remote sensing and imagery acquired at time t+1. An estimate of stratum status at time t+1 might be made using only this second subsample, as described in Sections 3 and 5.

The estimate for time t+1 might be improved using the sample units in the first 1/4 subsample, which were observed at time t. However, changes between times t+1 and t have probably occurred in the status of each sample unit in the first 1/4 subsample. If a model were available to predict these changes, then estimates from the 1/4 subsample observed at time t might be combined with the 1/4 observed at time t+1 into an estimate of the stratum status at time t+1, using the composite estimator presented in Section 5.

7.1 Updating estimates for one sample unit

Predicted true status $\underline{X}_{t+1}$ (e.g., proportion forest) of one sample unit at time t+1 is

$$\underline{X}_{t+1} = \underline{P}_f \underline{X}_t + \underline{P}_o (1-\underline{X}_t). \tag{34}$$

$\underline{X}_t$ is the estimated status of the sample unit at time t, $\underline{P}_f$ is the estimated conditional transition probability that a point is truly forest at time t+1, given it is was forest at time t, and $\underline{P}_o$ is the estimated conditional transition probability that a point at time t+1 is truly forest, given it was other cover at time t. Transition probabilities $\underline{P}_f$ and $\underline{P}_o$ are predicted from the deterministic model. The predicted status $\underline{X}_{t+1}$ of the one sample unit at time t+1 in equation (34), given its estimated status $\underline{X}_t$ at time t, is analogous to the calibrated estimate in equation (7), given an imperfect (biased) remotely sensed estimate.

If the deterministic model is perfect, variance var($\underline{X}_{t+1}$) of estimated status $\underline{X}_{t+1}$ of the one sample unit at time t+1 is

$$\mathrm{var}(\underline{X}_{t+1}) = \mathrm{var}(\underline{X}_t)\ P_r^2 + \mathrm{var}(\underline{X}_t)\ P_b^2, \tag{35}$$

as portrayed in Fig. 2. More realistically, the deterministic model is imperfect, and there is additional error (U_t) in predicting change between time t and t+1. Assuming additive, independent prediction errors, variance of the updated estimate for the one sample unit is

$$\mathrm{var}(\underline{X}_{t+1}) = \mathrm{var}(\underline{X}_t)P_r^2+\mathrm{var}(\underline{X}_t)P_b^2+\mathrm{var}(U_t), \tag{36}$$

as portrayed in Fig. 2. One fundamental problem will be estimating the variance of the prediction errors var(U_t) between times t+1 and t. This is discussed in Section 7.3.

7.2 Stratum estimates for each time period

The stratum level prediction model (i.e., transition probabilities $\underline{P}_r$ and $\underline{P}_b$) in equations (34) and (36) could update the estimated status of each sample unit in the 1/4 subsample observed at time t. These estimates might be directly combined with those from the other 1/4 subsample observed at time t+1, using the composite method presented in Section 5. The resulting stratum estimate at time t+1 would include measurements from 1/2 of the sample units.

At time t+2, the estimated status of each sample unit in the 1/4 subsample observed at time t+1, and the 1/4 subsample observed at time t and updated to time t+1 using equations (34) and (36), could be updated to time t+2 using

$$\underline{X}_{t+2} = \underline{P}_r\ \underline{X}_{t+1} + \underline{P}_b\ (1-\underline{X}_{t+1}), \tag{37}$$

$$\mathrm{var}(\underline{X}_{t+2})=P_r^2\mathrm{var}(\underline{X}_{t+1})+P_b^2\mathrm{var}(\underline{X}_{t+1})+\mathrm{var}(U_{t+1}). \tag{38}$$

These updated estimates from the 1/4 subsamples observed at times t+1 and t might be directly combined with those from the 1/4 subsample observed at time t+2, using the composite method in Section 5. The resulting stratum estimate at time t+2 would include measurements from 3/4 of the sample units.

The same method might be applied at time t+3 to estimate stratum status using all sample units. Most weight in the composite estimator would be placed on the 1/4 subsample observed at time t+3 because a prediction model is not needed to update estimated status of sample units within this subsample, and there would be no prediction errors; least weight would be placed on the subsample observed at time t because their status has not been directly observed for 4 time periods, and variance from prediction errors in updating estimated the status of the sample units would be greatest for this 1/4 subsample.

7.3 Variance of prediction errors

Variance of prediction errors from the deterministic model , i.e., var(U_t) = var(U_{t+1}) = var(U), are needed in (36) and (38) to update status estimates for sample units, which are combined into an estimate for the stratum. The variance of sampling errors from transition probabilities estimated using permanent ground plots (from other agencies or more detailed field sampling within the same monitoring system) might serve as initial estimates of prediction error variance. Initial estimates of prediction error variance for a process level landscape model might be made with data used to fit the model. These initial estimates are likely biased (i.e., too small) because the deterministic model is extrapolated over time or space. Stratum estimates from Section 7.2 can be compared to independent stratum estimates from other monitoring systems, and the adaptive methods discussed in Section 6.1 used to refine estimates of prediction error.

Direct estimates of prediction error variance from the deterministic model would be available through remote sensing of permanent sample units. For example, new imagery is acquired at time t+4 for the same 1/4 sample observed at time t. Misclassification bias in the estimated status of each sample unit at time t+4 is corrected using the calibration model in Section 3. A second estimate of the status of each sample unit in the 1/4 subsample at time t+4 is available from the deterministic prediction model, using the observed status at time t as initial conditions (Section 7.2). A sample estimate for variance of prediction errors between times t and t+4 can be made using the known differences between these two estimates at time t+4 for each sample unit. The remotely sensed estimate of these sample units at time t+4 would then be used as new initial conditions in the deterministic model to predict status at time t+5 and later.

This requires matrix representation of the statistical model, as in equation (25). The matrix solution for estimating var(U) would be complicated by covariances among prediction errors, use of the same calibration model at times t and t+4, or spatial autocorrelations. Approximations might be needed, but verification procedures introduced in Section 6.1 could protect against unreliable approximations.

8. KALMAN FILTER APPLIED TO CELLS

Each 40 km^2 sample unit may be considered a sample of the surrounding 640 km^2 cell, with a sample size of one. Estimates for aggregations of cells might utilize composite estimation (Section 5), treating the estimate for a 40 km^2 sample unit as an estimate of the entire 640 km^2 cell. This can reduce proliferation of stratification criteria from the calibration models and deterministic prediction models, and use of ancillary estimates from independent sources.

8.1 Combining independent ancillary estimates

Ancillary statistical estimates from independent sources can improve efficiency and temporal detail using composite estimation. For example, the USDA Forest Service and the USDA Soil Conservation Service both produce areal estimates of the extent of forestlands for geographic areas that might include one-hundred or more 640 km^2

cells. These independent estimates provide a source of useful data for landscape level monitoring. However, the estimates from the two systems are made with different definitions of forest, in different years, and can be contradictory. Differences in schedules among independent monitoring systems can be accommodated by annual estimates (Section 7); different classification systems can be accommodated by calibration.

Calibration for misclassification error in remote sensing requires plots for which reference and remotely sensed classifications are known (Section 3). Calibration for "misclassification" error caused by differences in classification systems requires plots which are independently classified by two independent monitoring systems. If sample units from other monitoring systems are accurately registered to the 40 km^2 sample units, then multivariate calibration models can estimate the quantitative statistical relationship between areal estimates from another agency, and areal estimates from the landscape monitoring system. These estimates that are "calibrated" for differences in definitions might be further calibrated using the calibration model for remote sensing errors.

This would allow several agencies to share areal estimates applicable to aggregations of sample units, while maintaining their own classification systems. These shared estimates might be made at the level of individual 640 km^2 cells using small area estimation, which takes the form of a composite estimator. Shared statistical estimates might improve the efficiency and compatibility of participating monitoring systems, without major disruptions to any one existing system. However, statistical calibration is not a panacea. Calibration will propagate statistical errors (Section 3.4), but these can be are minimized by making the independent classification systems as compatible as possible.

8.2 The cell as a stratum with sample size one

A stratum is a contiguous, homogeneous geographic area. However, calibration models superimpose additional stratification criteria, such as Landsat scene boundaries, or sets of sample units photointerpreted by one individual. These differences are needed to correct for different misclassification probabilities, and these criteria can change over time. When ancillary data from other monitoring systems are combined, a stratum is further subdivided by the geographic criteria used by each other system It is likely the number of strata will eventually approach the number of sample units.

The estimated status of a 640 km^2 cell might be considered a combination of the estimated status of the one 40 km^2 sample unit it that cell, and ancillary estimates from other agencies, which apply to aggregations of cells (Section 8.1). Estimation error associated with each 40 km^2 sample unit includes propagated and correlated errors from a regional calibration model (Section 5), propagated and correlated prediction errors from a regional deterministic prediction model (Section 7), and sampling error from use of one 40 km^2 sample unit in the cell. Sampling error might be estimated using aggregations of 40 km^2 plots and assuming independence and homogeneity, or geostatistical methods, such as Kriging and spatial correlograms.

9. LANDSCAPE DETECTION AND EVALUATION MONITORING

One objective might be monitoring "environmental health." "Detection" monitoring might use quantitative indicators of response and exposure to classify each 40 km^2 sample unit as "healthy" or "unhealthy". Unhealthy sample units could be further subclassified as to probable cause during "evaluation monitoring". Sample units classified based on their health can be used to make areal estimates of environmental health for regional assessments. Therefore, there is interest in individual sample units that might not be necessary if statistical estimates of regional status were the sole objective.

This is analogous to a psychologist's judgment (i.e., detection) whether a patient in a random sample (i.e., a sampling unit) is mentally ill (i.e., unhealthy) based on blood chemistry and psychological profile tests (i.e., response indicators), and history of chemical abuse or family mental health problems (i.e., exposure indicators); diagnosing probable cause(s) for the patient's condition (i.e., evaluation); and making an estimate of the suspected prevalence of various types of mental illnesses in the population (i.e., assessment) using a large sample of patients.

Quantitative indicators are needed to identify unhealthy sample units. Causal hypotheses might be suggested by exploratory statistical methods, such as scatter plots or principal components analyses, or geostatistical methods that might show similar spatial associations in unhealthy sites and indicator values. Hypotheses might be more difficult to formulate if landscape processes are nonlinear, with time lags and feedback mechanisms that obscure direct cause and effect relationships. Process oriented deterministic models contain a collection of individual hypotheses regarding landscape structure and function. If exposure indicators associated with individual sample units are included among driving variables for a landscape level model, and the model can predict response indicators that are measured on sample units, then aggregate hypotheses in the deterministic model can be scientifically tested.

The residual difference between model predictions and direct observations represents model prediction error, i.e., lack of agreement in predicting measurements of landscape structure and function. A model and direct measurements are imperfect caricatures of a system, and prediction errors are expected. However, residuals are expected to be random if the model and measurements are reliable. If spatial or temporal patterns exist in the residuals, then important processes are not included in the model, or there are unrecognized problems with the measurement process.

Such an unexpected situation should trigger a search for hypotheses that might explain the apparent nonrandom patterns. If the prediction model, rather than measurements, is judged to be the problem, alternative hypotheses might be incorporated in the prediction model, and tested with independent monitoring data. Therefore, analysis of data from a landscape monitoring system, and predictions from a landscape model, can be a crucial step in the cycle of hypotheses development, hypothesis testing, and hypotheses refinement to help understand the condition and functioning of landscapes.

10. DISCUSSION

The true status of spatially fixed sample units or cells are expected to have heterogeneous variance and lack independence, caused by landscape level processes such as regional land use practices, climatic patterns, and physiographic gradients. Therefore, no new complications are introduced by heterogeneous and dependent errors propagated from regional calibration and deterministic prediction models applied to 40 km^2 sample units, or small area estimation techniques for ancillary data applied to 640 km^2 cells.

It is frequently assumed that sampling errors associated with a systematic sample of plots in space are independent and identically distributed. These unrealistic assumptions will not bias estimates of stratum status, but there would be loss of efficiency, and bias in the estimated covariance matrix for stratum stratus estimates. Biased estimates of the covariance matrix might adversely affect important tests of hypothesis, and stepwise regression analytical models. Therefore, heterogeneity and lack of independence among should be expressed in the statistical models.

Additional statistical details need development before the hypothetical example in this paper could be implemented. This example is univariate, where status is defined as proportion of forest. More detailed categories would be required in a true landscape monitoring system, and the estimators in this paper would have to be developed for the multivariate case. Estimating model prediction error with remeasurements of permanent plots would require multivariate roots of polynomial matrix equations. Combining ancillary data from other monitoring sources would require multivariate, small area estimation techniques to estimate status of individual cells. It is assumed that the stratum is homogeneous, but multivariate spatial trends in status might be expected. Multivariate geostatistical methods used to estimate spatial trends and heterogeneous variance among sample units must deal with propagated heterogeneity and dependence from multivariate calibration and deterministic prediction models. Multivariate logit transformations, or the multivariate Dirichlet distribution might be needed to better deal with skewed error distributions for proportion estimates that approach zero.

The procedures outlined in this paper might have conceptual appeal to some, but they have never been put into operation within a broad scale, landscape level, environmental monitoring system. More work is needed to verify their applicability and feasibility. Alternatives, such as an interpenetrating design without the model based Kalman filter might be less risky, but could be less efficient, and might be incapable of testing deterministic models to improve understanding of system dynamics. Contingency plans should be made in case a design based or model based approach is found unacceptable.

REFERENCES

Czaplewski, R.L., 1990. Kalman filter to update forest cover estimates. In State of the Art Methodology of Forest Inventory, Syracuse, NY-USA, in press.

Czaplewski, R.L., Catts, G.P., and Snook, P.W., 1987. National land cover monitoring using large, permanent photoplots. In Land and Resource Evaluation for National Planning in the Tropics, Chetumal, Mexico, pp. 197-202.

Czaplewski, R.L., Alig, R.J., and Cost, N.D., 1988. Monitoring land/forest cover using the Kalman filter, a proposal. In Forest Growth Modelling and Prediction, Minneapolis, MN-USA, pp. 1089-1096.

Dixon, B.L., and Howitt, R.E., 1979. Continuous forest inventory using a linear filter. Forest Science, 25:675-689.

Gregoire, T.G. and Walters, D.K., 1988. Composite vector estimators derived by weighting inversely proportional to variance. Canadian Journal of Forest Research, 18:282-284.

Hay, A.M., 1988. The derivation of global estimates from a confusion matrix. International Journal of Remote Sensing, 9:1395-1398.

Maybeck, P.S., 1979. Stochastic Models, Estimation and Control, Vol. 2. Academic Press, New York. 423pp.

Sorenson, H.W., 1985. Kalman filtering: Theory and Applications. IEEE Press Inc., New York. 457pp.

Tenenbein, A., 1972. A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. Technometrics 14:187-202.