Prediction of $2 \times 2$ tables of change from repeat cluster sampling of marginal counts

Steen Magnussen

Abstract: Repeat cluster sampling of a binary $(0,1)$ attribute at time 1 ($Y_1$) and time 2 ($Y_2$) in a finite population of discrete units is considered. All clusters contain $m$ units and a cluster provides the marginal count of ones and zeroes at the two time points only. From these counts, we seek to predict a $2 \times 2$ table of the rates of no change ($\pi_{01} = E[Y_1Y_2]$, $\pi_{10} = E[(1 - Y_1)(1 - Y_2)]$) and change ($\pi_{11} = E[Y_1(1 - Y_2)]$, $\pi_{00} = E[(1 - Y_1)(1 - Y_2)]$). Two predictors are proposed; one is derived from the temporal correlation of marginal counts and the second from the odds ratio of no change that maximizes a (pseudo-) likelihood of a non-central, hypergeometric distribution. The bias of the first is positive when there is a positive intracluster correlation of $Y_1$, $Y_2$, and $Y_1Y_2$, while the bias of the second is negative when the odds ratio of no change is >1. A proposed combined estimator worked well in three examples of change analysis with paired, classified Landsat images of forest cover type and cluster sampling with $3 \times 3$ arrays of 30 m × 30 m units (pixels). $2 \times 2$ tables obtained from marginal counts were superior, in terms of mean absolute error, to estimates based on a direct unit-by-unit count when the time 2 image had a root mean square registration error of 0.5 pixel relative to the time 1 image. The proposed method is intended for settings where a direct unit-by-unit estimation of the $2 \times 2$ table is either compromised or when data (by design) consist of marginal counts from a repeat cluster sampling.

Résumé : L’auteur examine l’échantillonnage en grappes d’un attribut binaire $(0,1)$ répété au temps 1 ($Y_1$) et au temps 2 ($Y_2$) dans une population finie d’unités discontinues. Toutes les grappes contiennent $m$ unités et un groupe fournit le dénombrement marginal de zéros et de uns seulement aux deux points dans le temps. À partir de ces dénombrements, nous voulons prédire une table $2 \times 2$ des taux d’absence de changement ($\pi_{01} = E[Y_1Y_2]$, $\pi_{10} = E[(1 - Y_1)(1 - Y_2)]$) et de changement ($\pi_{11} = E[Y_1(1 - Y_2)]$, $\pi_{00} = E[(1 - Y_1)(1 - Y_2)]$). Deux variables explicatives sont proposées : l’une découle de la corrélation temporelle des dénombrements marginaux et l’autre d’un rapport de cotes d’absence de changement qui maximise la (pseudo) vraisemblance de la distribution hypergéométrique non centrale. Le biais de la première variable explicative est positif quand il y a une corrélation intra-grappes entre $Y_1$, $Y_2$ et $Y_1Y_2$ tandis que celui de la seconde variable explicative est négatif quand le rapport de cotes d’absence de changement est supérieur à un. Un estimateur combiné, proposé par l’auteur, fonctionne bien pour trois exemples d’analyse de changement sur des paires d’images Landsat classifiées selon le type de couvert forestier et avec un échantillonnage en grappes avec des ensembles de $3 \times 3$ de 30 m × 30 m (pixels). Sur la base de l’erreur moyenne absolue, les tables $2 \times 2$ obtenues par dénombrement marginal étaient supérieures aux estimations basées sur un dénombrement direct unité par unité lorsque la racine carrée de l’erreur d’enregistrement de l’image au temps 2 était de 0,5 pixel par rapport à l’image au temps 1. La méthode proposée est applicable aux dispositifs pour lesquels l’estimation directe unité par unité de la table $2 \times 2$ est compromise ou à ceux dont les données représentent des dénombrements marginaux obtenus lors d’un échantillonnage en grappes répété.

Introduction

Monitoring changes in land cover, land use, and land status has gained global importance in the context of sustainable resource use, stewardship issues, and modeling (Brown 2002; Coomes et al. 2002; Anderson 2002; Corona et al. 2002; Parr et al. 2002). Estimates of rates of no change and change among land cover, land use, and land status classes provided by monitoring are pivotal statistics for natural resource managers.

Change data for a single class can be regarded as coming from a repeat measurement of a binary attribute (e.g., $Y = (1,0)$) in a finite population of discrete units. A repeat recording of the binary attribute at time 1 ($Y_1$) and time 2 ($Y_2$) generates four possible observations: two of no change ($Y_1 = Y_2 = (0,1)$) and two of change ($1 \rightarrow 0$ or $0 \rightarrow 1$). A $2 \times 2$ table with the observed count of each type of observation provides a succinct summary of the change process in the population; the counts are sufficient statistics (Fleiss 1981; Agresti 1992). Rates of change and various statistics of associations are derived from this table for further insight into the change process (Fleiss 1981; Agresti 1992).

An unbiased estimation of this $2 \times 2$ table requires, of course, that the same unit is measured on both occasions. A repeat observation of the exact same units may, however, pose a logistical or a technical challenge or simply be impossible if individual units are not identified in the first place. Monitoring of change by remote sensing, where the population unit is an image pixel, provides an example of less than perfect temporal integrity of the unit of observation. Periodic changes in data and image definitions, registration errors, point spread function of the sensor, and resampling...
of image pixels to achieve a common resolution (Pratt 1991; Lunetta and Elvidge 1999; Kennedy and Cohen 2003) can introduce a bias to estimates of change based on a presumed unit-by-unit count (Raffy 1994; Franklin 2001). A related challenge arises in the estimation of classification accuracy, when data are collected at two different scales and subject to location errors (Magnussen 1997; Foody 2000; Patil and Taillie 2003). In other settings, a $2 \times 2$ table of population averaged by estimates is desired but compromised by design because of the lack of a unit-by-unit observation of change. For example, a monitoring of forest health for the presence or absence of a stress indicator on a fixed number of trees at two points in time in a fixed set of sample locations but with no tree-by-tree observation of change (Parr et al. 2002; Mizoue and Dobbertin 2003; O’Laughlin and Cook 2003) would fall into this category.

When the temporal integrity of the observational unit is compromised, a direct estimate of the $2 \times 2$ table of change will be biased or infeasible. Under those circumstances, a method for obtaining a reliable, model-based prediction of the $2 \times 2$ table of change would be desirable. The success of a model will depend on its ability to mitigate the impact of both lack of temporal and spatial integrity of a unit of observation. Repeat sampling with spatially contiguous clusters with a fixed number of units (two or more) would be more immune against errors in unit locations, since the average location error of the cluster declines with the size of the cluster. The downside of counting by multi-unit clusters is clear: the change status of individual units is lost and we would need a model to predict the $2 \times 2$ table of change. This study proposes such a model for populations, where the odds ratio of no change is $\geq 1$. The estimation problem is one of finding a suitable bivariate distribution from two marginal distributions (of counts) (Long and Krzysztofowicz 1995; Cox 2003). Application of the method is demonstrated with three examples of predicting a $2 \times 2$ table of change in a land cover class from paired, classified Landsat images. When the location of a unit (pixels) at the second measurement is subject to a coregistration error, the predicted tables of change obtained via cluster sampling are shown to be less biased than those obtained by a presumed pixel-by-pixel estimation of change.

**Materials and methods**

**Population statistics and notation**

Estimates of the rates of change and no change in a binary attribute ($Y = (0,1)$) between two points in time ($t_1$ and $t_2$) in a population consisting of $N$ units are the statistics of primary interest in the analyses that follow. A population unit is a countable, distinct element about which information ($Y$) is wanted (Särndal et al. 1992). In this study, a population is a forested area within known boundaries. A population can be completely tessellated into $N$ population units of equal size and shape. Here, a unit is a square surface area with a side of 30 m. This unit is equivalent to a pixel in a Landsat image. $N$ is constant between $t_1$ and $t_2$. Let $Y_t$ denote the value of the binary attribute of a unit at $t_1$, and $Y_t$ denote the attribute value of the same unit at $t_2$. Define the four possible outcomes of observing $Y_1$ and $Y_2$ on a single unit as follows: $Z_{11} = Y_1 Y_2$ for the case of remaining 1; $Z_{10} = Y_1 (1 − Y_2)$ for changing from 1 to 0; $Z_{01} = Y_2 (1 − Y_1)$ for changing from 0 to 1; and $Z_{00} = (1 − Y_1)(1 − Y_2)$ for remaining 0. Repeat observations of a single unit is equivalent to observing $Z_{ij}$. Clearly, $Z_{11} + Z_{10} + Z_{01} + Z_{00} = 1$ for all units. We are interested in estimating the expectations of $Z_{ij}(j = (1,0))$. Denote these expectations as $\pi_{11}$, $\pi_{10}$, $\pi_{01}$, and $\pi_{00}$, respectively. A unit-by-unit census of $Y_1$ and $Y_2$ in a population with $N$ units produces the following $2 \times 2$ table of change:

<table>
<thead>
<tr>
<th></th>
<th>$Y_2 = 1$</th>
<th>$Y_2 = 0$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1 = 1$</td>
<td>$N_{11}$ ($\pi_{11}$)</td>
<td>$N_{10}$ ($\pi_{10}$)</td>
<td>$N_{1+}$ ($\pi_{1+}$)</td>
</tr>
<tr>
<td>$Y_1 = 0$</td>
<td>$N_{01}$ ($\pi_{01}$)</td>
<td>$N_{00}$ ($\pi_{00}$)</td>
<td>$N_{0+}$ ($\pi_{0+}$)</td>
</tr>
<tr>
<td>Total</td>
<td>$N_{++}$ ($\pi_{++}$)</td>
<td>$N_{++}$ ($\pi_{++}$)</td>
<td>$N_{++}$ ($\pi_{++}$)</td>
</tr>
</tbody>
</table>

A simple, random sample with observation of $Y_1$ and $Y_2$ on each of $n$ population units would produce a sample estimate of this table. Note that for a fixed set of margin totals, the entire table could be estimated from knowledge of any one of the change rates $\pi_{11}$, $\pi_{10}$, $\pi_{01}$, or $\pi_{00}$. The other three are uniquely determined by the constraints imposed by the fixed marginal totals.

In populations with an odds ratio of no change $\geq 1$, the probability $\pi_{11} \geq \pi_{1+} \times \pi_{+1}$ and the correlation between the random variables $Y_1$ and $Y_2$ is positive (Prentice 1986).

$$\rho(Y_1, Y_2) = \frac{\pi_{11} - \pi_{1+} \pi_{+1}}{\sqrt{\pi_{1+}(1 - \pi_{+1})\pi_{+1}(1 - \pi_{+1})}}$$

Hence, a model-based prediction of this correlation and sample-based estimates of $\pi_{1+}$ and $\pi_{+1}$ would allow a prediction of $\pi_{11}$, and hence, the desired $2 \times 2$ table of change.

The tendency of a unit to maintain its attribute value between $t_1$ and $t_2$ is commonly expressed by the odds ratio of no-change $\theta$ (Fleiss 1981, chap. 5.3).

$$\theta = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}}$$

For $\theta \geq 1$, the chance that a unit stays the same is greater than the chance that a unit will change between $t_1$ and $t_2$. Again, if one had a model prediction of $\theta$ and, as before, sample-based estimates of $\pi_{1+}$ and $\pi_{+1}$, one could predict $\pi_{11}$ via eq. 3, and hence, the complete $2 \times 2$ table of change.

---

© 2004 NRC Canada
where the correct solution is the one that satisfies $\pi_{i_1, i_2} \leq \min(\pi_{i_1}, \pi_{i_2})$. The odds ratio and the coefficient of correlation are exchangeable as predictors of $\pi_{i_1}$ when the population odds ratio of no change is $\geq 1$. In fact, each of the two measures of association can be expressed as a function of the other and the two marginal rates $\pi_{i_1}$ and $\pi_{i_2}$(not shown).

### Cluster sampling of population units

As argued in the Introduction, when a direct observation of $Z_{ij}$ is compromised (or infeasible), a suggested alternative is to tessellate the population into $M$ equal-sized clusters of $m$ units each ($N = M \times m$) and observe at $t_1$ the sum

$$n_{i_1} = \sum_{j=1}^{m} Y_{ij} = \sum_{j=1}^{m} Z_{ij}^{i_1} + Z_{ij}^{i_2}$$

in the $i$th cluster ($i = 1, \ldots, M$) for $t_2$ the sum

$$n_{i_2} = \sum_{j=1}^{m} Y_{ij} = \sum_{j=1}^{m} Z_{ij}^{i_1} + Z_{ij}^{i_2}$$

with the obvious extension of notation. The above expansions of the counts $n_{i_1}$ and $n_{i_2}$ merely serve to underline that they share the cluster-specific sum of the binary variable $Z_{i1}$.

The number of clusters to observe is either $M$ or a random sample of size $M^* \leq M$. Data from this repeat cluster sampling of marginal counts only allow design-unbiased estimation of the marginal rates $\pi_{i_1}$, $\pi_{i_2}$, $\pi_{i_1,i_2}$, and $\pi_{00}$. Estimates of $\pi_{i_1}$, $\pi_{i_0}$, $\pi_{01}$, and $\pi_{00}$ must be obtained from a model-based prediction of one of these four rates. The natural pivot is $\pi_{i_1}$.

From the cluster sampling (or census), estimates of $\pi_{i_1}$, $\pi_{i_2}$, $\pi_{i_1,i_2}$, and $\pi_{00}$ are obtained in the obvious way.

$$\hat{\pi}_{i_1} = \frac{1}{m \times M} \sum_{i=1}^{M} n_{i_1} \Rightarrow \hat{\pi}_{i_1} = \frac{1}{m \times M} \sum_{i=1}^{M} n_{i_1}$$

$$\hat{\pi}_{i_2} = \frac{1}{m \times M} \sum_{i=1}^{M} n_{i_2} \Rightarrow \hat{\pi}_{i_2} = \frac{1}{m \times M} \sum_{i=1}^{M} n_{i_2}$$

Equation 6 extends naturally to linear combinations and products of binary traits. The expected value of $\rho(n_{i_1,n_{i_2}})$ can be found analytically

$$E(\rho(n_{i_1,n_{i_2}})m) = \frac{\pi_{i_1}[1 + (m - 1)\phi_{i_1}] - \pi_{i_1}\pi_{i_2}}{\sqrt{\pi_{i_1}[1 - \pi_{i_1} + (m - 1)\phi_{i_1}]\pi_{i_2}[1 - \pi_{i_2} + (m - 1)\phi_{i_2}]}$$

To arrive at the expression in eq. 7, I made extensive use of the fact that $E(Z_{ij}^i Z_{ij}^j) = E(Z_{ij}^i Z_{ij}^j)$ and $i, j, r, s \in I_{i_1}$, where $I_{i_1}$ is the open set of positive integers. The expectation in eq. 7 showed that the correlation of marginal counts was an unbiased estimator of $\rho(Y_{i1}, Y_{i2})$ if and only if $\phi_{i_1} = \phi_{i_2} = \phi_1 = 0$. This was confirmed for clusters of sizes 9, 16, and 25 by assuming that the distribution of $n_{i_11}$, $n_{i_01}$, and $n_{00}$ was multinomial with parameters $m$ and $\pi = (\pi_{i_1}, \pi_{i_0}, \pi_{01}, \pi_{00})$. For example, the expected, maximum, absolute bias for a cluster of size 9 was 2.665 $\times 10^{-15}$ for $\pi = (0.2163, 0.1995, 0.2171, 0.3670)$. Nelder’s simplex method was used to solve this maximization problem (Nelder and Mead 1965). Note that imposing a proper, discrete, multivariate distribution on $n_{i_11}$, $n_{i_01}$, and $n_{00}$ shifted the solution away from the intuitive $\pi_{ij} = 0.25, i,j = (1,0)$. An extreme bias can occur when the number of changes from 1 to 0, and 0 to 1 is balanced within each cluster; that is $n_{i_1} = n_{i_2}$, which ensures $\rho(n_{i_1,n_{i_2}}) = 1$ and $0 \leq \rho(Y_{i1}, Y_{i2}) \leq 1$ when $n_{i_11}, n_{i_01}$, and $n_{00}$ are distributed as a censored, multinomial distribution (i.e., censored to satisfy the marginal constraint). A maximum bias $\geq 1$ occurs for $\pi = (0.1796, 0.2543, 0.2543, 0.3118)$. Cluster-specific random effect on the expectations of $Y_{i1}$, $Y_{i2}$, as captured by, for example, a beta-binomial distribution (Rosner and Milton 1988) complicates matters considerably, since the overdispersion is now the resultant of two distinct but inseparable effects (Ahn and Odom-Maryon 1995).

Examples of the impact of overdispersion on $\rho(n_{i_1,n_{i_2}})$ are illustrated in Fig. 1, where it is assumed that the expected unit attribute values of $Y_{i1}$, $Y_{i2}$, and $Y_{i1}Y_{i2}$ are fixed and...
that unit values in a cluster are positively correlated. Here, the correlation between two unit values is determined only by the distance between the units. A first-order, auto-regressive model with a correlation of $\rho_1$ between attribute values for units separated by a distance of 1 (city-block distance; Upton and Fingleton 1985) was assumed for illustrative purpose. The same autocorrelation function applies to $Y_1$, $Y_2$, and $Y_{11}Y_{22}$ (i.e., $\phi_1 = \phi_2 = \phi_{11}$). The graph for $m = 9$, $\pi_1 = 0.15$, $\pi_{11} = 0.18$, and four values of $\pi_{11}$, for which $\rho(Y_1,Y_2)$ is a constant 0.05, 0.30, 0.54, and 0.79, illustrates how $\rho(Y_{11},Y_{22})$ increases as $\rho_1 \to 1$. While the inflation is moderate (<10%) for $\rho(Y_1,Y_2) > 0.5$, it becomes quite serious (>30%) for low values of $\rho(Y_1,Y_2)$. A correction for this bias would be possible if the magnitude of $\phi_{11}$ could somehow be estimated (e.g., $\phi_{1+}$ and $\phi_{+1}$ can be estimated directly from the data via eq. 6). Quantification of the relationship among the three overdispersion parameters would be exceedingly complex and not feasible from just the marginal counts (Diggle 1988). The assumption that $\phi_{11}$ is, for example, the average or the minimum of $\phi_{1+}$ and $\phi_{+1}$ incurs the potential of additional bias in $\rho(Y_{11},Y_{22})$.

In consequence of the above, a prediction of $\pi_{11}$ via eq. 5 will, in most cases, have a positive bias. Since the bias is potentially serious, a prediction of $\pi_{11}$ via eq. 5 best serves as an estimate of the upper limit of $\pi_{11}$. An estimate of the variance of the prediction $\bar{\pi}_{11}$ is obtained from (Lloyd 1999, eq. 4.28)

\[ \bar{\pi}_{11} = \text{Var}(\bar{\pi}_{11}) = \text{Var}[E(\bar{\pi}_{11} | \pi_{11}, \pi_{11}, m)] \\
+ E[\text{Var}(\bar{\pi}_{11} | \pi_{11}, \pi_{11}, m)] \]

where $\bar{\pi}_{11}$ is the conditional prediction of $\pi_{11}$ for the $i$th cluster. Specifically

\[ \bar{\pi}_{11} = (\delta(n_{1+}, n_{+1}) \bar{n}_{1+}^0 (1-\pi_{11}) \bar{n}_{+1}^0 (1-\pi_{+1}) + \pi_{1+}^0 \pi_{+1}^0) \]

and (Fleiss 1981, eq. 5.21)

\[ \text{Var}(\bar{\pi}_{11} | \pi_{11}, \pi_{11}, m) \approx \left( \sum_{i=0}^{m} \sum_{s=0}^{m} (\bar{\pi}_{11}^i + 0.5)^{-1} \right)^2 \]

Variance estimators for $\bar{\pi}_{11}$, $\bar{\pi}_{11}$, and $\bar{\pi}_{11}$ are obtained in an analogous manner.

An odds ratio prediction of $\pi_{11}$

The basic problem in predicting a $2 \times 2$ table of change from marginal counts is, of course, that one or more tables may exist for a given set of margins. For example, if $m = 9$ and a marginal count of units for which $Y_1 = 1$ is $n_{1+} = 4$ and the number of units for which $Y_2 = 1$ is $n_{+1} = 6$, then the following four $2 \times 2$ tables of change counts would be consistent with the observed counts:

\[
\begin{pmatrix}
1 & 3 \\
5 & 0
\end{pmatrix}, \begin{pmatrix}
2 & 2 \\
4 & 1
\end{pmatrix}, \begin{pmatrix}
3 & 1 \\
3 & 2
\end{pmatrix}, \text{ and } \begin{pmatrix}
4 & 0 \\
2 & 3
\end{pmatrix}
\]

For any marginal count, the possible tables are found by setting $n_{+1}$ equal to the set of integers $u = (\max(0, n_{1+} - n_{1+}), \ldots, \min(n_{1+}, n_{+1}))$. Clearly, as $m$ increases, the cardinality of $u$ increases at an exponential rate. If there a limited population odds ratio of no change exists (0, see eq. 2), then these four tables would not have the same likelihood of ap-
pearing in the population. Let \( \mathcal{S}_i(m, n_{i+}, n_{i+}) \) denote the set of all possible \( u' \times 2 \times 2 \) tables of counts consistent with the observed marginal counts \( n_{i+} \) and \( n_{i+} \) in the \( i \)th sample cluster. Candidate values of \( n_{i+} \in \mathcal{S}_i \) are denoted by \( \tilde{n}_{i+} \). Conditional on \( m, n_{i+}, \) and \( n_{i+} \), the probability mass distribution of \( \tilde{n}_{i+} \) is non-central hypergeometric (Agresti 1992). The distribution is also known as the generalized or tilted, hypergeometric distribution (Lloyd 1999). Hence, we can write the likelihood of a table in \( \mathcal{S}_i \) as

\[
\Pr(\tilde{n}_{i+}) = \frac{\binom{m-n_{i+}}{n_{i+}-\tilde{n}_{i+}} \cdot \binom{n_{i+}}{\tilde{n}_{i+}} \cdot \binom{n_{i+}}{n_{i+}-\tilde{n}_{i+}}}{\sum_{u' \in \mathcal{S}_i} \binom{m-n_{i+}}{n_{i+}-u'} \cdot \binom{n_{i+}}{u'} \cdot \binom{n_{i+}}{n_{i+}-u'}}
\]

The probability statement in eq. 11 suggests an estimation of \( \theta \) by maximum likelihood methods. To this end, it was assumed that the tables in \( \mathcal{S}_i \) conditional on the margins are independent realizations of “pseudo” data. A pseudo-likelihood was computed for the observed counts on the basis of eq. 11 and maximized over \( \theta \) (Gauss-Newton method, Gallant 1987). Note that the longitudinal aspect of the data is not factored into the likelihood in eq. 11 (Irons et al. 2000) nor are random cluster effects in \( \theta \) recognized. Thus, a downward bias in \( \theta \) is expected (Cox 2003).

A pseudo maximum likelihood estimate of \( \theta(\hat{\theta}) \) serves to estimate \( \pi_{i1} \) via eq. 3. Call this estimate \( \hat{\pi}_{i1} \). The variance of this prediction (\( \text{Var}(\hat{\pi}_{i1}) \)) is estimated as outlined in eq. 8 for a correlation-based prediction. Cluster-specific predictions of \( \pi_{i1} \) conditional on the marginal counts in the \( i \)th cluster were obtained via eq. 3 after an obvious extension of notation. Denote a cluster specific prediction by \( \hat{\pi}_{i1} \). The first term of the variance estimator in eq. 8 is straightforward to compute, and Strawderman proposed an approximately ‘exact’ estimator of the second term, the variance of \( \hat{\pi}_{i1} \).

\[
\text{Var}(\hat{\pi}_{i1}|n_{i+},n_{i+},m) = \frac{1}{m^2} \sum_{(\lambda_n, \tilde{\lambda}_n) \in \Omega} \tilde{\lambda}_{n_{i+}} \hat{\theta}^{-1}(1 + \hat{\theta} \tilde{\lambda}_n)^{-2}
\]

where the set \( (\lambda_n, \tilde{\lambda}_n)_{n_{i+} \in \mathcal{S}_i} \) are the roots of the polynomial-generating function \( \psi(z) \) of the non-central, hypergeometric distribution.

\[
\psi(z) = \sum_{n_{i+} \in \mathcal{S}_i} \binom{n_{i+}}{\tilde{n}_{i+}} \binom{m-n_{i+}}{n_{i+}-\tilde{n}_{i+}} z^{n_{i+}}
\]

**Combining predictors of \( \pi_{i1} \)**

The expected opposite direction of bias in \( \hat{\pi}_{i1} \) and \( \hat{\pi}_{i1} \) suggests that a linear combination of the two may be less biased than either one. An extensive simulation study (not shown) with a beta-binomial distribution of \( Y_{11}, Z_{11} \), and \( Z_{01} \) (0.2 \( \leq \phi_{1+}, \phi_{11}, \phi_{10} < 0.6 \), respectively), covering a wide array of expected values, and \( m = 9, 16, \) and 25 suggested that a combined estimator with a relative weight of 2/3 given to the correlation-based prediction and a relative weight of 1/3 given to the prediction based on odds ratio would, on average, minimize the root mean square of prediction errors. All presented predictions represent this linear combination of the two predictors. Standard error of a combined prediction was obtained as the standard error of a linear combination of two perfectly correlated random variables (Snedecor and Cochran 1971).

**Examples**

Multitemporal, cover-type-classified Landsat images of three study sites are used to demonstrate application of the proposed combined method of predicting \( \pi_{i1} \) from repeat cluster sampling of marginal counts of pixels (units) classified to one of \( K \) cover types. Each cover type, in turn, is considered as a binary attribute (1 = presence, 0 = absence). Hence, \( K \) cover-type-specific predictions of \( \pi_{i1} \) are made on each site. A direct pixel-by-pixel count of the four possible change attributes \( Z_{ij} \) \((i,j=(1,0))\) over all \( N \) pixels in a site would generate \( K \) census 2 \times 2 tables of cover-type-specific change. Correlation-based and odds-ratio-based predictions were then generated following a tessellation of the study site into \( M \) equal-sized square clusters of size \( m = 9, 16, \) and 25 and obtaining the counts \( n_{i+} \) and \( n_{i+} \) for each of the \( M \) clusters.

The first data set represents a 11 664-ha forested area (129 600 pixels in a 360 \times 360 array) near Hinton, Alberta, Canada. An unsupervised, maximum likelihood classification of two coregistered Landsat images from 1985 to 1990 to 16 cover types generated the data set (Goodenough et al. 2000). Seven rare cover types (<0.2%) were present in one but not both images. No 2 \times 2 table of change is predicted for these types. The second data set represents a 1990 and 1999 Landsat image of a 10 900-ha forested area (121 104 pixels in a 348 \times 348 array) near Prince George, British Columbia, Canada. A repeat (independent), unsupervised classification (clustering) assigned each pixel in the two images to 1 of 15 cover types (Wulder et al. 2002). Again, five rare types were only present in one but not both images. The third data set represents a 1989 and 1999 classified Landsat image of a 14 982-ha rural area in the state of Selangor, Malaysia, south of Kuala Lumpur. A classification system of eight classes capturing information on both land use and characteristics of rubber plantations was used.

All classifications include errors. Cover-type-specific accuracies could be anywhere from 0.50 to 0.95 (Congalton 1991). Classification errors invariably bias estimates of change (Woodcock et al. 2001). The impact of classification errors on change estimates is beyond the scope of this study.

**Registration errors in population units**

A motivating rationale for the prediction of change from repeat cluster sampling of marginal counts was the expected increase in spatial and temporal fidelity of count data from clusters of units as opposed to a unit-by-unit count. To demonstrate the potential of a serious bias in change estimates derived from a unit-by-unit count when the location of some units at time \( t_1 \) is in error relative to their location at time \( t_2 \), a location error was introduced into the \( t_2 \) images. Specifically, one in four pixels in the \( t_2 \) image was shifted by one row or one column relative to their position at \( t_1 \). Clusters of \( m \) pixels were shifted at a time, since a registration error is generally constant within blocks of about 10 \times 10 pixels (Kennedy and Cohen 2003). The shifting of \( t_2 \) positions was done in a systematic manner. Every fourth cluster was shifted cyclically in one of the four cardinal directions.
Thus, the \( t_2 \) image has a registration root mean square error of 0.5 pixel, a registration error that is frequently encountered in rural and forested areas (Martin and Howarth 1989; Pratt 1991; Lunetta et al. 1991; Congalton 2001). The results of the impact of registration error have been averaged over the 16 possible shifts of the \( t_2 \) units.

**Results**

A cluster size of \( m = 9 \) produced the overall best results, as expected. Only results for \( m = 9 \) are presented here. The quality of \( \pi_{11} \) predictions from repeat cluster sampling of marginal counts depends critically on the quality of the predicted correlation between \( Y_1 \) and \( Y_2 \) and the predicted odds ratio of no change. Figure 2 shows, for each study site, the cover-type-specific binary correlation plotted against the combined prediction. In Hinton, the predicted and observed correlations were strongly correlated \( (R^2 = 0.94, \text{slope} = 1.00, \text{intercept} = -0.04 \ (P = 0.18)) \), and the linear least squares prediction error was 0.05. Registration errors in \( t_2 \) counts had little impact on this relationship as opposed to their impact on the binary correlation, which on average, dropped by 17%. Results from Prince George were more mixed despite a reasonably strong correlation \( (R^2 = 0.93) \); the standard error of prediction was 0.07 with a clear trend towards a positive bias for correlations between 0.1 and 0.3 and a negative bias for stronger correlations (>0.4). Again, errors in \( t_2 \) pixel positions had little impact on this relationship, whereas the binary correlation dropped by about 22% after introducing errors in the \( t_2 \) pixel locations. The strongest relationship between the observed and the predicted binary correlations was seen in the Selangor study, with an \( R^2 \) of 0.95 and a prediction error of 0.04. A location error in \( t_2 \) had only a minor effect on either relationship; for example, the binary correlation dropped by about 8%.

Predicting the population odds ratio of no change was, as expected (Ahn and Odom-Maryon 1995; Cox 2003), more difficult, at least in Hinton (Fig. 3), where the standard error of prediction was a considerable 4.2, despite a favourable, ordinary least squares trend line with a slope of 1.03 and a nonsignificant intercept of 0.06. Results from Prince George and Selangor were more encouraging. A strong linear relationship \( (R^2 = 0.92) \) and prediction errors of approximately 1.2 shows that reasonable predictions can be made. Positional errors in \( t_2 \) had only a slight (<10%) impact on either the observed or the predicted odds ratios.

Predicted 2 × 2 tables of change for four cover types on each site are presented in Tables 1, 2, and 3. The included cover types are random selections from cover types with \( \pi_{11} > 0.02 \). Lower valued predictions had large relative prediction error (25%–94%), which sharply curtailed their utility. Predicted values of \( \pi_{11} \) matched fairly closely the observed values. On average, the predictions were 0.001 too high, with a mean absolute prediction error of 0.004 (8%).

---

**Fig. 2.** Observed binary correlation \( \rho \) for three sites plotted against (i) combined predictions from repeat cluster sampling \( \tilde{\rho} \) (solid boxes), (ii) combined predictions with registration errors in \( t_2 \) data (open boxes), and (iii) direct count estimates with \( t_2 \) registration errors (shaded boxes). A dashed one-to-one line is given for reference.
The relationship between predicted and observed values across sites was linear, with $R^2 = 0.98$, a slope of 1.01 ($\pm 0.02$), and a nonsignificant intercept of $-0.003$ ($P = 0.17$). In comparison, the average of the estimated, relative prediction error of $\pi_{11}$ (last column in Tables 1, 2, and 3) was 9.5%, an indication that these model-based estimates of error were reasonable. A two-sided $t$ test of the paired differences between the predicted and the count-based estimates of $\pi_{11}$ yielded a $t$ value of 1.57 ($P = 0.06$); thus, there is only weak support for the null hypothesis of no difference. Note that all four entries in a predicted $2 \times 2$ table have the same standard error of prediction; relative errors, however, rise sharply as the entry value drops.

Introduction of a registration error in the $t_2$ data incurred a mean absolute relative bias of 11% in the pixel-by-pixel count estimate of $\pi_{11}$. The largest bias was seen in Hinton and the smallest in Selangor. The prediction approach to estimation of $\pi_{11}$ was, as indicated by a lower average absolute bias of 7%, more robust against registration errors than a unit-by-unit count. The absolute prediction bias was for most (five out of six) cover types lower than the bias incurred by a direct count and only about half as large in relative terms.

The impact of a registration error should be a function of the number of nonchanging pixels that at $t_1$ have one or more neighbours with a different attribute. The maximum miscount of change pixels ($1 \rightarrow 0$ and $0 \rightarrow 1$) happens when all these pixels have a registration error of 1 pixel. For Hinton, Prince George, and Selangor, this maximum bias (averaged over cover types) was 45%, 38%, and 18%, respectively. Change estimates for scattered, low frequency cover types and cover types that occur in narrow, string-like structures are most sensitive to registration errors. With only $1/4$ of the pixels in error, and assuming that miscounts are split evenly between the two types of change, a rough estimate of the expected bias would be to take $1/8$ of these maxima. Indeed, the two sets of estimates seem to corroborate.

**Discussion and conclusions**

A unit-by-unit repeat observation of a binary attribute is the only option that ensures a design-unbiased estimation of a $2 \times 2$ table of change. When this option is not feasible by design or the fidelity of a unit of observation is compromised, the indirect approach presented here with repeat cluster sampling of marginal counts followed by model-based predictions of the table of change may be a viable alternative. Estimation methods for data consisting of a mixture of direct and marginal counts have been presented elsewhere (Antelman 1972). Model-based predictions will minimally provide an upper and lower limit for $\pi_{11}$ and at best provide predictions with a moderate bias, a bias that could be less than the bias incurred by a direct count when unit fidelity is compromised. Relative prediction errors for rare categorical classes (<0.2%) will be high, thus limiting the utility of the proposed method to more frequent classes. The three examples with classified, remotely sensed data posed a tough challenge for the prediction method. The within-cluster correlations of $Y_1$, $Y_2$, and $Y_1Y_2$ was strong because of a prevalent strong spatial autocorrelation of forest cover types (Magnussen and de Bruin 2003), and the local variation in...
change rates (Nakajima et al. 1996) created strong over-dispersal effects. As the expected prediction bias increased with the size of the cluster, the cluster size had to be kept small. In this study, an increase in cluster size from 9 to 16 increased the bias by about 3% (not shown); a cluster of size 25 would increase the bias by about 10%, with no tangible benefits in terms of buffering against registration errors. Cluster size was held constant by design in this study. In practice, however, cluster size may vary. A stratified approach to prediction is suggested for this scenario. A prediction of \( \pi_{ij} \) is obtained for each actual cluster size, and the final prediction is obtained as a weighted average of size-specific predictions. Appropriate weights should reflect both the frequency of a cluster size and the negative impact of cluster size on the prediction error.

The problem of predicting a 2 × 2 table of change from marginal totals is difficult except for the rare case of no over-dispersal. The entries in the table of change can be considered as interrelated Bernoulli processes (Antelman 1972). Their joint distribution remains intractable (Irons et al. 2000), and random, between-cluster effects would confound the estimation problem further (Rosner and Milton 1988; Ahn and Odom-Maryon 1995; Wang and Louis 2003; Cox 2003). Predictions via the cumulative distribution functions

### Table 1. Observed and predicted 2 × 2 tables of change rates in Hinton, Alberta.

<table>
<thead>
<tr>
<th>Type</th>
<th>( \pi(n_{ij}) )</th>
<th>( \pi(n_{ij}, n_{ij}) )</th>
<th>( \pi(Y_{1}Y_{2}) )</th>
<th>( \pi(n_{ij}, n_{ij}) )</th>
<th>( se(\pi_{11}) )%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>293 81</td>
<td>295 80</td>
<td>268 106</td>
<td>290 85</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>164 461</td>
<td>163 463</td>
<td>190 436</td>
<td>168 457</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>2</td>
<td>25</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>109 165</td>
<td>100 173</td>
<td>93 181</td>
<td>98 176</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>102 624</td>
<td>111 615</td>
<td>119 607</td>
<td>114 612</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>9</td>
<td>16</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>29 116</td>
<td>34 111</td>
<td>24 121</td>
<td>25 119</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>31 824</td>
<td>26 829</td>
<td>37 818</td>
<td>24 821</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>32 58</td>
<td>36 54</td>
<td>26 64</td>
<td>27 63</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>51 860</td>
<td>47 864</td>
<td>57 854</td>
<td>56 855</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** MAD, the mean absolute error of the 2 × 2 table entries; \( se(\pi_{11}) \)% is an estimate of the relative prediction error of \( \pi_{11} \). All rates are multiplied by 1000 and rounded to nearest integer.

* Estimate from a direct unit-by-unit count of \( Y_1 \) and \( Y_2 \).
* Predicted from a cluster count of \( n_{ij} \) and \( n_{ij} \).
* Estimate from a direct unit-by-unit count subject to a coregistration error in \( Y_1 \).
* Predicted from a marginal count subject to a coregistration error in \( Y_2 \).

### Table 2. Observed and predicted 2 × 2 tables of change rates in Prince George, British Columbia.

<table>
<thead>
<tr>
<th>Type</th>
<th>( \pi(Y_{1}Y_{2}) )</th>
<th>( \pi(n_{ij}, n_{ij}) )</th>
<th>( \pi(Y_{1}Y_{2}) )</th>
<th>( \pi(n_{ij}, n_{ij}) )</th>
<th>( se(\pi_{11}) )%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>65 81</td>
<td>63 82</td>
<td>56 90</td>
<td>65 81</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>48 806</td>
<td>50 804</td>
<td>57 797</td>
<td>49 805</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>35 5</td>
<td>39 0</td>
<td>29 11</td>
<td>33 7</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>4 955</td>
<td>0 959</td>
<td>10 949</td>
<td>6 953</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>72 47</td>
<td>75 44</td>
<td>67 52</td>
<td>77 43</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>104 777</td>
<td>102 780</td>
<td>109 772</td>
<td>100 781</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>20 62</td>
<td>24 58</td>
<td>17 65</td>
<td>23 59</td>
<td>15.7</td>
</tr>
<tr>
<td></td>
<td>65 853</td>
<td>61 857</td>
<td>68 850</td>
<td>68 856</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>—</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** MAD, the mean absolute error of the 2 × 2 table entries; \( se(\pi_{11}) \)% is an estimate of the relative prediction error of \( \pi_{11} \). All rates are multiplied by 1000 and rounded to nearest integer.

* Estimate from a direct unit-by-unit count of \( Y_1 \) and \( Y_2 \).
* Predicted from a cluster count of \( n_{ij} \) and \( n_{ij} \).
* Estimate from a direct unit-by-unit count subject to a coregistration error in \( Y_1 \).
* Predicted from a marginal count subject to a coregistration error in \( Y_2 \).
of cluster counts at \( t_1 \) and \( t_2 \) via, for example, Clayton’s family of copulas (Genest and Boies 2003) showed some promise. However, determining a priori the error component (lack of fit) due to intracluster correlations proved difficult. Markov Chain Monte Carlo methods for estimating the joint distribution of the table entries also failed, as no complete set of proper conditional distributions could be defined (Kaiser and Cressie 2000). Improvements of the presented prediction method will most likely come from progress in predicting the conditional overdispersion parameter \( \phi_{11} \), since the bias in the correlation-based predictions would be sharply curtailed if good estimates of \( \phi_{11} \) were available. Establishing the relationship between overdispersion and the correlation of marginal counts was a necessary first step in that direction. The next will be to quantify the effect of among-cluster variation on these parameters. Improved guidelines for choosing the best weights for combining the correlation-based and the odds-ratio-based predictions are also needed. Generally, a large overdispersion in the marginal counts lowers the confidence in the marginal correlation. Posterior analysis showed that the suggested 2:1 weighting was not far from optimal (sensu minimum absolute bias in \( \pi_{11} \)) on any site.

The proposed prediction method also has a potential for application in the assessment of the classification accuracy of a classified, remotely sensed image when the scale and location of reference and image data effectively prevent a perfect unit-by-unit match of observations (Foody 2002; Steele et al. 2003; Cihlar et al. 2003; Duerr and Dietz 2000). Extensions to multiphase sampling (Patterson and Williams 2003; Cihlar et al. 2003; Duerr and Dietz 2000). Improvements of the presented prediction method will most likely come from progress in predicting the conditional overdispersion parameter \( \phi_{11} \), since the bias in the correlation-based predictions would be sharply curtailed if good estimates of \( \phi_{11} \) were available. Establishing the relationship between overdispersion and the correlation of marginal counts was a necessary first step in that direction. The next will be to quantify the effect of among-cluster variation on these parameters. Improved guidelines for choosing the best weights for combining the correlation-based and the odds-ratio-based predictions are also needed. Generally, a large overdispersion in the marginal counts lowers the confidence in the marginal correlation. Posterior analysis showed that the suggested 2:1 weighting was not far from optimal (sensu minimum absolute bias in \( \pi_{11} \)) on any site.

The proposed prediction method also has a potential for application in the assessment of the classification accuracy of a classified, remotely sensed image when the scale and location of reference and image data effectively prevent a perfect unit-by-unit match of observations (Foody 2002; Steele et al. 2003; Cihlar et al. 2003; Duerr and Dietz 2000). Extensions to multiphase sampling (Patterson and Williams 2003) with ground cluster-plots partially covering the area of nine Landsat image pixels also holds promise.

In a multiclass system with \( K > 2 \) classes, a prediction of \( K \times 2 \times 2 \) tables of change may be less important than the prediction of the complete \( K \times K \) table of change. The rate of change between two specific classes may have important ecological and (or) economic consequences. Extending the current methodology to the prediction of a \( K \times K \) table is feasible. The solution involves the prediction of a \( 3 \times 3 \) table of change for every \( K(K – 1)/2 \) distinct pairs of classes, with all classes not in the pair lumped into a single, complementary, third class. Each \( 3 \times 3 \) table of change is then predicted from the three distinct \( 2 \times 2 \) tables that can be formed from each \( 3 \times 3 \) table. Expectations of and constraints on each cell in a \( 3 \times 3 \) table can be written as an over-determined system of equations for which a least-squares solution exists. Research is currently underway to explore this extension.

### Acknowledgements

Data for the examples were kindly provided by Dr. A. Dyk, Dr. M.N. Suratman, and Dr. M.A. Wulder. Their helpful and prompt assistance is greatly appreciated. Three anonymous referees and the Associate Editor of the Journal are thanked for their constructive critique and numerous suggestions of improvements to an earlier version of this manuscript.

### References


---

### Table 3. Observed and predicted \( 2 \times 2 \) tables of change rates in Selangor, Malaysia.

<table>
<thead>
<tr>
<th>Type</th>
<th>( \pi(Y_1, Y_2) )</th>
<th>( \hat{\pi}(n_{11}, n_{12}) )</th>
<th>( \pi(Y_1, Y_2') )</th>
<th>( \hat{\pi}(n_{11}, n_{12}') )</th>
<th>( \text{se}(\hat{\pi}_{11}) ) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>31 113</td>
<td>35 109</td>
<td>30 114</td>
<td>33 111</td>
<td>9.0</td>
</tr>
<tr>
<td></td>
<td>[49 807]</td>
<td>[45 811]</td>
<td>[50 807]</td>
<td>[47 809]</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td></td>
<td>4</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>61 81</td>
<td>63 79</td>
<td>57 84</td>
<td>62 80</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>[147 712]</td>
<td>[145 714]</td>
<td>[150 708]</td>
<td>[146 713]</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td></td>
<td>2</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>95 180</td>
<td>97 178</td>
<td>91 183</td>
<td>97 178</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>[111 614]</td>
<td>[109 616]</td>
<td>[115 611]</td>
<td>[109 616]</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td></td>
<td>2</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>44 24</td>
<td>46 22</td>
<td>41 26</td>
<td>45 23</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>[174 758]</td>
<td>[172 760]</td>
<td>[177 755]</td>
<td>[173 759]</td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td></td>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** MAD, the mean absolute error of the \( 2 \times 2 \) table entries; \( \text{se}(\hat{\pi}_{11}) \)% is an estimate of the relative prediction error of \( \pi_{11} \). All rates are multiplied by 1000 and rounded to nearest integer.

- Estimate from a direct unit-by-unit count subject to a coregistration error in \( Y_1 \) and \( Y_2 \).
- Predicted from a cluster count of \( n_{11} \) and \( n_{12} \).
- Estimate from a direct unit-by-unit count subject to a coregistration error in \( Y_2 \).
- Predicted from a marginal count subject to a coregistration error in \( Y_2 \).


ginal shapes in binary random intercept models using a bridge
Woodcock, C.E., Macomber, S.A., Pax-Lenney, M., and Cohen,
W.B. 2001. Monitoring large areas for forest change using
Landsat: generalization across space, time and Landsat sensors.

Wulder, M., Cranny, M., and Dechka, J. 2002. An illustrated meth-
odology for land cover mapping of forests with Landsat-7
Canadian Forest Service, Internal Report, Victoria, BC.