Journal of Environmental Management (2001) **61**, 381–385
doi:10.1006/jema.2000.0412, available online at http://www.idealibrary.com on IDEAL®

LETTER TO THE EDITOR

Estimating Markov transitions

Dear Sir,

The purpose of this letter is to comment on methods described by Brown *et al.* (2000) ('the paper'). Specifically it addresses: (1) issues in estimating a Markov transition matrix from a pair of images, (2) how to improve the paper's algorithm for 'standardizing' a transition matrix, and (3) how to generalize the estimation and standardization procedures from two-state transitions to multi-state transitions.

The paper describes what appears to be a novel approach to estimating transition probabilities between two images. I am concerned this approach may have been miscommunicated, because the procedure outlined in the paper estimates the correct probabilities only in unusual circumstances.

The setting is a pair of binary images. To be general, assume the image values are either 0 or 1. In the paper, for example, 0 would represent non-forested cover and 1 would represent forested cover. The second image, J, is assumed to result from a homogeneous Markov transition from the first image, I. Let the probability of a transition from 0 to 1 be p (originally, p_{nff}) and the probability of a transition from 1 to 0 be q (originally, p_{fnf}).

The procedure, described in the text surrounding the paper's Equations (4) and (5), uses a 5×5 square moving window to compute the proportions of forested cover (ones) within the images. Let these proportions, which vary from point to point, be f (for image I) and g (for image J). Under the hypotheses, f is given and g is the outcome of a random variable because it is determined by a Markov transition from I. The paper appears to estimate the Markov transition probabilities as:

$$\hat{p}=\frac{1}{N}\sum_{\alpha}(1-f)g \qquad (1)$$

and

$$\hat{q}=\frac{1}{N}\sum_{\alpha}f(1-g) \qquad (2)$$

In these sums, α ranges over all pixels common to the two images and N is the number of such locations. In the authors' words: '. . .we calculated the average value of $p_{\mathrm{nff}}[(1-f)g]$ and $p_{\mathrm{fnf}}[f(1-g)]$ across all the pixels in each sample site [image domain]'.

However, these expressions are poor estimators of p and q. Their expected values (conditional on image I) are readily computed from the expected values of g at each cell of the image. Since g is the proportion of ones at a pixel in J, its expected value is equal to (the proportion of ones in the I-window) * (expected number of ones that remain ones) + (the proportion of zeros in the I-window) * (expected number of zeros becoming ones) $=f^{*}(1-q)+(1-f)^{*}p$. Taking expectations in Equations (1) and (2) then gives:

$$Exp(\hat{p})=p+(\nu^2-2F)p+(F-\nu^2)(1-q) \qquad (3)$$

and

$$Exp(\hat{q})=q+(F-\nu^2)(1-p)+(\nu^2-1)q \qquad (4)$$

where

$$F\equiv\frac{1}{N}\sum_{\alpha}f$$

and

$$\nu^2\equiv\frac{1}{N}\sum_{\alpha}f^2$$

The bias in these estimators consists of everything after the initial terms on the right hand sides in Equations (3) and (4) above. The bias depends on the initial proportion of ones (F) in I and on their second moment (ν^2) as well as on the true values of p and q. It is large, sometimes larger even than p or q. Finally, the bias will, *a fortiori*, depend

0301–4797/01/040381+05 $35.00/0

on other land-cover variables such as proportions of developed (DEV1I) or agricultural (AGI) land in the initial image. These variables (DEV1I and AGI) are the only ones the paper found to be significant at an *overall* (not testwise) level of 5% or less. Therefore, it is possible that: (1) the authors' 'significant' results arise solely from bias in the transition probability estimator, or—with an exactly opposite effect—(2) the bias may have masked or weakened other important relationships among transition probabilities and land use variables. Therefore it is important to use higher-quality estimators of the transition probabilities.

It is tempting to correct the bias by solving Equation (2) for p and q in terms of the expected values and then replacing these expected values by the estimates (1). This would accommodate the original intuition expressed in the paper that using windowed averages might create error-resistant estimates while eliminating the bias in their estimators. I have obtained simulation results and preliminary theoretical results indicating there is merit in this approach. In particular, when the images exhibit spatial correlation, the bias-corrected estimators are more resistant to image mis-registration error than are the usual counting estimators of p and q (computed by evaluating the observed transitions pixel by pixel between I and J).

This casts significant doubt on the paper's discussion and conclusions regarding land-use and land-cover change in the Upper Midwest of the USA, because its results depend fundamentally on the estimated transition probabilities. For this reason, I hope that Brown *et al.* (2000) actually used a different estimation procedure than the one they described. They need to comment on this.

The paper concludes by suggesting, 'more than two class transitions could be modeled to allow for the inclusion of other land-cover categories'. The purpose of the remarks that now follow is to report on practical techniques for standardizing arbitrary-rank Markov transition matrices and to alert the reader to potential complications, even in the simplest case of transitions between binary images.

'Standardization' supposes that the Markov process, originally conceived of as occurring at discrete time intervals, actually is continuous in time, so that there is a matrix $P(t, dt)$ giving transition probabilities from the image at time t to the image at time $t+dt$. The assumption of homogeneity in time is that P does not actually depend on t, so we can write $P=P(dt)$.

For any two positive time intervals ds and dt, the Markov property implies that $P(dt)P(ds)=P(dt+ds)=P(ds+dt)=P(ds)P(dt)$. It is evident that $P(0)=1$ (the identity matrix). To an arbitrarily good approximation any two time intervals ds and dt are integral multiples of a very small interval de: $ds=(ds/de)de=k\,de$, say, and $dt=(dt/de)de=n\,de$, say. Then, assuming the function $dt \rightarrow P(dt)$ is continuous,

$P(dt)=P(n\,de)=P(de)^{n}$, which can be written equivalently as:

$P(de)=P(dt)^{1/n}$, implying

$P(ds) = P(k\,de) = P(de)^{k} = (P(dt)^{1/n})^{k} = P(dt)^{k/n} = P(dt)^{ds/dt}$.

Therefore the conversion from one time interval dt to another interval ds involves being able to obtain any desired power $s=ds/dt$ of the matrix $P(dt)$. (In the paper, $ds=10$ and $dt=x$ years, so $s=ds/dt=10/x$.) From now on I will drop the explicit dependence of P on dt and simply write $P=P(dt)$.

Several different approaches to standardization have been published in the context of land-cover change analyses. Some would avoid the problem by collecting all images at equal time steps or even by requiring the collection of additional images at intermediate time steps when the steps vary (Wood *et al.*, 1997). Among those that directly address the problem, one would multiply all off-diagonal terms of the matrix P by s and then re-normalize the rows to sum to unity. Another approach takes the complement of the sth power of the complement of each off-diagonal entry and adjusts the diagonal entry to enforce the sum-to-unity requirement for all rows (National Research Council, 2000). These are both approximations whose accuracy depends on all off-diagonal transitions being close to zero and s being close to 1. Neither approach agrees with the correct one of taking the sth power of P. Therefore I will describe the procedure in some detail, with remarks on the effective computation of P^{s}.

It is instructive to cast the results of the binary case into a form that more readily generalizes to higher-rank transition matrices. I begin with Equation (6) of the paper, which is:

$$P=\begin{pmatrix} 1-p_{\text{fnf}} & p_{\text{fnf}} \\ p_{\text{nff}} & 1-p_{\text{nff}} \end{pmatrix}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

(Although it makes no difference in the analysis that follows, note that this transition matrix is applied by multiplying it on the *right* of row vectors, despite the notation used in Equations (1) and (2) in the paper).

For simplicity let $p=p_{\text{nff}}$ and $q=p_{\text{fnf}}$, so that $a=1-q$, $b=q$, $c=p$, and $d=1-p$. By algebra it follows from the authors' Equation (9) that:

$$R^2=a^2+4bc-2ad+d^2=(p+q)^2$$
$$a+d-R=2(1-(p+q))$$

and

$$a+d+R=2$$

Therefore, the complicated Equations (7) and (8) in the paper simplify to:

$$P^s=\begin{pmatrix} 1-q^* & q^* \\ p^* & 1-p^* \end{pmatrix}$$

where,

$$p^*=p(1-\Delta^s)/(p+q)$$
$$q^*=q(1-\Delta^s)/(p+q)=(q/p)p^*$$
$$\Delta=(1-p-q)$$

and

$$s=10/x$$

Using these expressions will avoid potential numerical errors resulting from the excessive manipulation required by the authors' Equations (7) and (8).

Because in general $\Delta^s=(1-p-q)^s=\exp(s\ln(1-p-q))$, computation of arbitrary powers s requires $1-p-q>0$; that is:

$$p+q<1 \qquad (5)$$

is a necessary condition on the transition probabilities for standardization to work.

The remaining question is how to generalize decade standardization (or standardization to any time period) to transitions among more than two states. Specifically, let P be an irreducible Markov transition matrix. Write P in the form $P=1+Q$, (1 is the identity matrix having 1s on the diagonal and zeros elsewhere), so that:

$$Q=P-1$$

Again letting $s=10/x$ we wish to compute P^s. To do so, apply Newton's theorem on the binomial expansion (1676; first published in an appendix to his *Optics*, 1704):

$$P^s=(1+Q)^s=\sum_{k=0}^{\infty}\binom{s}{k}Q^k \qquad (6)$$

Here, $\binom{s}{k}$ is a generalized binomial coefficient: $\binom{s}{0}=1$ for any s (fractional or not) and $\binom{s}{k+1}=\binom{s}{k}*\frac{s-k}{k+1}$ for $k>0$

(This theorem, although originally stated for real numbers, is true in any real algebra where the infinite sum on the right hand side can be defined and converges. Therefore it is true for the algebra of square matrices with real coefficients using any matrix norm.)

The sum converges (coefficient by coefficient) provided P has no negative eigenvalues. Indeed, if P has a negative eigenvalue, say $\lambda<0$, then by definition there exists some real linear combination of states x for which $xP=\lambda x$. Taking the sth power, we would need $xP^s=\lambda^s x$, an impossibility because λ^s will generally be a complex number but not real.

This provides an immediate and practical test for the existence of arbitrary powers of a Markov transition matrix. The roots of the matrix's characteristic polynomial $p_P(\lambda)=\det(P-\lambda 1)$ are the eigenvalues, so they must all be non-negative. This occurs exactly when the coefficients of $p_P(\lambda)$ alternate in sign. Press *et al.* (1986) suggest that computing all eigenvalues directly is a better method than examining the coefficients of the characteristic polynomial. See Godement (1968) for rigorous background on eigenvalues, the characteristic polynomial, and reduction of matrices to canonical form.

Equation (9) provides a fast and practical method for computing arbitrary matrix powers. If s is a positive integer, the sum terminates after $s+1$ terms. Otherwise it is an infinite sum, but after s terms it begins alternating in sign. (s is the largest integer less than or equal to s.) Thereafter, the error will be no larger (in any matrix norm) than the norm of the last term computed. Thus, computation may cease as soon as every coefficient of Q^k is smaller than a predetermined accuracy threshold. In practice it may be wise to compute P^s as $P^k * P^{(s-k)}$, where $k=s$, using the binomial expansion for the second term only.

Finally, it is not hard to check that the binomial formula reduces to Equations (7) and (8) of the paper; the key is to note that when $P=\begin{pmatrix} 1-q & q \\ p & 1-p \end{pmatrix}$.

We immediately have $Q=\begin{pmatrix} -q & q \\ p & -p \end{pmatrix}$ and (for $k>0$) $Q^k=(-1)^k(p+q)^{k-1}\begin{pmatrix} -q & q \\ p & -p \end{pmatrix}$ follows by induction on k.

Thus, Q is a constant factor in the binomial expansion and ordinary algebra gives:

$$P^s=1+Q\left(\frac{1-(1-p-q)^s}{p+q}\right) \tag{7}$$

This agrees with the authors' Equations (7) and (8) as simplified above, demonstrating that the binomial expansion generalizes the paper's approach. It is unfortunate that I can discover no such simplification like Equation (10) above for general transitions among three or more states.

The concerns about negative eigenvalues and non-existence of standardized transition matrices are not just theoretical. Even in the simplest case of two states described in the paper, standardization of the transition matrix to fractional time periods can be impossible for several reasons: (1) Suppose the two states are 'wetlands' (or some other rare, vanishing land-cover type) and 'non-wetlands'. Early in the study period (the time interval comprising the two images), encroaching development may have eliminated all original wetlands; later in the study period growing environmental awareness may have caused new wetlands to be formed. The transition frequencies from wetlands to non-wetlands therefore will be close to unity ($q=1$) and from 'non-wetlands' to 'wetlands' will be some small positive value (p). It is evident that it is possible that $p+q>1$, violating condition (5).

(2) Suppose again that the states are wetlands, which originally occur with frequency F, and non-wetlands. Further suppose that the majority of wetlands pixels are isolated. Imagine that image registration is accurate to 1/2 the width of a single pixel—this is acceptable accuracy. Suppose, however, a systematic bias of 1/2 pixel down occurs in the first image and 1/2 pixel up occurs in the second, comparison image. This causes virtually all original wetlands pixels to appear to change to non-wetlands and for proportion F of the non-wetlands pixels to appear to change to wetlands. The estimated transition probabilities are therefore approximately $q=1$ and $p=F$, and again $p+q>1$.

(3) The value $p+q$ may exceed unity due to errors of classification and registration in either or both images, especially when one class is scarce.

The second and third examples highlight the interest and importance of propagating uncertainty in the estimated transition probabilities. Separating the uncertainty components in the Markov transition model might thereby enable one to standardize the matrix to arbitrary time intervals.

There is another problem that actually occurs in practice. It can be proven that the rows of P^s sum to unity for any power s where Equation (6) converges, which is one defining property of a Markov transition matrix, but when more than two states occur in the images there is no guarantee all coefficients will be positive, which is the other defining property. Indeed, the 7×7 transition matrix derived from Table 5 of Aspinall and Pearson (2000), another paper that appears in the same issue of the *Journal of Environmental Management*, exhibits several small negative coefficients when raised to any power between 0 and 1. These coefficients never get smaller than $-0{\cdot}0024$, but that is sufficiently negative to demonstrate the point.

(Aspinall and Pearson (2000) do not need to standardize their matrix, so this observation has no implications for their paper. I used their matrix only to demonstrate that the problem of negative coefficients is not a contrived one; it can arise with 'real-world' matrices.)

In some cases, slightly negative coefficients can simply be set to zero. A comparison of the 1/s power of the resulting matrix to the original matrix will help establish whether this procedure is acceptable: the two matrices should be identical to within the degree of error inherent in the coefficients. Again, quantitative assessment of uncertainty is needed to effect such a procedure. The Aspinall and Pearson table does not contain sufficient information to carry this out, although discrepancies between the row totals and the totals of the tabulated values indicate the table values include some error.

Very truly yours,

William A. Huber
Quantitative Decisions
Merion Station, PA 19066
Email:whuber@quantdec.com

References

Aspinall, R. and Pearson, D. (2000). Integrated geographical assessment of environmental condition in water catchments: linking landscape ecology, environmental modeling and GIS. *Journal of Environmental Management* **59**, 299–319, doi:10.1006/jema.2000.0372.

Brown, D. G., Pijanowski, B. C. and Duh, J. D. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management* **59**, 247–263, doi:10.1006/jema.2000.0369.

Godement, R. (1968). *Algebra, part VII*. Paris: Hermann.

National Research Council (US) Committee to Evaluate Indicators for Monitoring Aquatic and Terrestrial Environments (2000). *Ecological Indicators for the Nation, Appendix B*. Washington, DC: National Academy Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986). Eigensystems. In *Numerical Recipes*. Cambridge: Cambridge University Press.

Wood, E. C., Lewis, J. E., Gray Tappan, G. and Lietzow, R. W. (1997). *The Development of a Land Cover Change Model for Southern Senegal*. Presented at the Land Use Modeling Workshop, June 1997, sponsored by the USGS and NCGIA. http://ncgia.ncgia.ucsb.edu/conf/landuse97/papers/wood_eric/pecdoc.html (accessed 23 October 2000).