

Quantifying rates of inter-area movement using capture-recapture results: a comparison of methods

P. T. Stevick

College of the Atlantic, 105 Eden Street, Bar Harbor, ME 04856 USA

and

Aquatic Systems Group, University of Southern Maine, 350 Commercial Street,
Portland ME 04101 USA

peterstevick@aol.com

Abstract

The extent to which individual animals move between habitat units determines population spatial structuring, influences social and reproductive systems and has important implications for management. However, the extent of inter-area movement may be difficult to infer from available data. In order to compare the behavior of five indices which may be used to quantify rates of inter-area movement using capture-recapture data, these were applied in simulations to data generated from a simple model. The results suggest that some of the methods that have been presented for estimating movement of individual marine mammals between habitats are poor predictors of the underlying biological processes and that the Interchange Index in particular should not be used. The simulations provide considerable support for use of Hilborn's estimator of transition probability and indicate that it is the most powerful and useful technique over a broad range of circumstances. The Movement Index is more limited in predictive ability, but may be of use in those situations in which comparison of observation with a specific expectation is of interest.

INTRODUCTION

Marine mammals typically have a patchy distribution, often demonstrate strong fidelity to seasonal high-use habitat, yet are highly mobile (Stevick et al. 2002). The extent to which individual animals move between habitat units determines population spatial structuring (Stevick et al. 2006), influences social and reproductive systems (Whitehead et al. 1998) and has important implications for management

(Clapham et al. 2008). The extent of such inter-area movement may be difficult to infer from available data, however (Nichols & Kaiser 1999).

Capture-recapture studies based upon natural markings are commonplace in investigations of marine mammal populations. These provide an opportunity to estimate between-area movement rates based on re-sightings of individuals in multiple areas. Several methods have been used in cetacean studies to attempt to quantify movement rates, but most of these have not been tested. I present a comparison of five methods using a simple simulation model to test their relative performance.

METHODS

Indices

In this paper I compare the properties and behavior of five indices which may be of potential use for quantifying rates of inter-area movement using capture-recapture data. All of these indices can be calculated simply with only two samples.

The following terminology is used to designate samples:

a = number of individuals identified in area A

b = number of individuals identified in area B

c = number of individuals identified in both areas.

Note that sample sizes are not designated as n_i , nor the number of recaptures m_{ij} as this terminology may easily be confused with that commonly used in abundance estimation. The two should not be treated as equivalent. In abundance estimation it is assumed that all animals in the population are equally likely to be captured in the second sample whether they were included in the first sample or not (Pollock et al. 1990). In those cases for which inter-area exchange is of interest this is clearly not the case. Where site fidelity is high, for example, the capture probability of marked animals in the second sample is artificially low between distant areas, while the reverse will be true between adjacent areas.

The underlying process that all of the estimators are designed to represent is the probability that an animal that is present in one area when it is sampled is also present in a second area when it is sampled, or alternatively, the proportion of the animals present in one area when it is sampled that are also present during a second

when sampling is conducted there. The differing sampling intensities in the two areas, and thus differences in individual capture probability in these areas, along with differing numbers of individuals using each area, influence the ability of the different approaches to model this process.

Simple Proportion – The extent of movement may be presented as the proportion (P) of animals from the smaller of the two samples that is also identified in the other sample, $c/\min(a,b)$. This provides the simplest possible method of estimating the overall proportion of animals moving between areas, but is most unlikely to give a robust estimation of inter-area transition. I am not familiar with any instances where this technique has been used as an indicator of animal movement. It is included here simply as a baseline for comparison with the other methods.

Sub-set Index - Katona and Beard (1990) presented what they referred to as a Sub-set Index (S) for comparing movements of individual animals between areas. This is the sum of the proportion of recaptured animals in each of the two samples:

$$S = \frac{c}{a} + \frac{c}{b}.$$

Or it could be also be calculated as the mean of those proportions.

Being based upon the proportions of re-sightings in each sample, the theoretical basis of this method is clear. If presented as the mean, it takes the form of a proportion with values of the index bounded by 0 and 1, though the value is heavily dominated by the proportion of the smaller sample (P), and in practice it rarely exceeds half of the theoretical maximum value.

Movement Index - Another way to quantify movement between areas is to compare the observed number of animals identified in both areas to the expected number given the sample sizes if movement were random throughout the population (Baker et al. 1985, Whitehead & Glass 1985, Baker et al. 1986). Using this approach, Stevick et al. (2003) developed a Movement Index (M) based on the standardized deviate from the expected number of recaptures.

The probability of identifying an individual in a (random) sample is the reciprocal of the number of individuals in the population from which the sample is drawn, so:

$$p = \frac{1}{N}.$$

Where:

p = probability of identification

N = abundance

If animals mix randomly within the overall area in which the N animals occur, then the product of any two sample sizes and the individual probability of capture yields the expected number of re-sightings between those two samples:

$$ab\hat{p} = \frac{ab}{\hat{N}}.$$

Therefore, deviations from this expected value indicate deviation from this behavior pattern, *i.e.*, positive deviates indicate higher than expected exchange between samples. The expected value may be used to compare observed and expected values statistically (for example by χ^2) or as an index of movement. Because absolute deviates are influenced by the magnitude of the expected value, standardized deviates may be calculated as:

$$M = (\text{observed-expected})/\text{expected}^{0.5}.$$

Rates of movement that approximate random mixing will have a value of M near 0. This can be seen by noting that if mixing were random then these samples would then be appropriate for use in a Petersen estimator and the capture probability would be c/ab . Thus the expected number of re-sightings is:

$$ab\left(\frac{c}{ab}\right) = c.$$

So the observed and expected number of recaptures is the same leading to a deviate of 0. There are no theoretical limits on the range of this index; it can take large values both positive and negative.

Interchange Index – Several studies (Calambokidis et al. 1997, Urbán R et al. 2000, Calambokidis et al. 2001, Acevedo et al. 2007) have used an Index of Interchange (I) (also referred to as a Match Index (Calambokidis et al. 2001), though that term is not used here to avoid confusion with the Movement Index). This is presented as:

$$I = \frac{c}{ab}.$$

All values are positive, and are generally extremely small. In simulations (see below), the value rarely exceeded 0.01. Thus multiplying the index x1000 or x100 depending on observed rates clarifies presentation.

The origin and theoretical basis of this index is unclear. Calambokidis et al. (1997) and Urbán R et al. (2000) cite the index as being developed and used by Baker et al. (1985, 1986) The formulation is used in those papers for calculating expected values for chi-square analysis (similar to the approach used in calculating M), but it does not appear to be used in either as an index of movement. In cases where the samples involved meet the assumptions of capture-recapture abundance estimation it comprises the inverse of the Petersen estimator, and therefore the inverse of estimated abundance. This, in turn represents an estimate of the capture probability of an individual (see above). In comparing movement between two areas with limited exchange, however, the assumptions of capture-recapture theory are clearly not met, and the index simply represents the number of re-sightings per comparison of identified individuals. This rate may be influenced by factors other than movement; a high rate of successful comparison may indicate either a small population size or a high rate of movement (Calambokidis et al. 2001).

Hilborn's method – Hilborn (1990) developed a powerful and flexible technique for estimating population and movement parameters from mark-recapture data using maximum likelihood methods. This method has been further developed by Whitehead (2001). This approach can be used to directly estimate the transition probability between areas (t), and has been used to estimate the extent of movement between areas in cetacean studies (Whitehead 2001, Stevick et al. 2006). The method does not provide a closed-form estimator for t . However, t for movement between two areas can easily be estimated easily using any spreadsheet software.

The log-likelihood that c of the a animals sampled in area A were also identified from the b animals sampled in area B is given as the Poisson approximation (Whitehead 2001):

$$L = c \log(\hat{c}) - \hat{c} + \text{constant}$$

where the expected number of re-sightings given a total of n individuals present in areas A and B is:

$$\hat{c} = tab/n.$$

These equations provide a formula for calculating the likelihood given the observed values of a , b and c and any pair of values for t and n . All that is needed to estimate t for any set of observed parameters, therefore, is to calculate a large number of likelihood values, L , using randomly selected probability values for t , and randomly

selected estimates for the animals using the sampled areas for n . The result with the largest value for L provides the best parameter estimate for t (and n).

The method can be expanded to use observed sightings histories to estimate a wide range of parameters for geographically structured populations, including estimation of transition probability from multiple sites and multiple sampling occasions (Hilborn 1990, Whitehead 2001). It can even be used for analysis of movement over continuous space (Whitehead 2001). However, only the simplest case with two samples is considered here.

Simulations

In order to compare the behavior of these indices, they were applied in simulations to data generated from a simple model constructed in two stages using R software (R Core Development Team 2005). Initially the numbers of individuals using each of two areas, designated A and B, were selected randomly from a uniform distribution, and ranged from 30 to 500. The number of individuals occurring in both areas was calculated as the product of the mean sample size and a randomly selected transition probability, also selected from a uniform distribution. Instances where the number in common to the two areas exceeded the smaller regional abundance, the number in common to the two samples was set equivalent to the smaller sample. From each of these simulated areas, a sample of between 20 animals and the total abundance of the area was selected randomly (designated a and b respectively), and the recaptures between the two samples were identified (designated c).

The five indices were calculated for each of 5,000 replicates of the simulation. S was calculated as the mean, I was calculated $\times 100$ and M assuming an overall abundance of 1,000. Hilborn's estimator for t was determined based on selection of 5,000 random parameter values. Values for t were selected from a uniform distribution (0-1), while n was selected from a normal distribution with an unbiased estimate of the combined number of individuals in areas A and B for the mean and a CV of 0.2. These indices were then compared with the transition probability from the simulation using simple linear regression to see how well they represented the underlying movement process.

RESULTS & DISCUSSION

All five models show a strong positive relationship between the index value and the true transition probability. As expected with the very large sample size, all relationships were very highly significant. However, the indices by no means performed equally well (Table 1). Hilborn's method provided the most reliable indicator of the underlying movement process. As anticipated, the simple proportion provided a poor estimate of movement. However, the Interchange Index performed even less well.

There was a remarkably weak relationship between the transition probability from the simulation model and the Interchange Index, with by far the lowest R^2 and F values of any index considered; substantially lower even than a simple proportion. The Interchange Index produced many values that were well above the overall trend, leading to some large positive residuals (Figure 1). This method is influenced by regional abundance as well as movement. Indeed, given complete sampling and the same proportional rate of movement, an order of magnitude increase in the regional abundances results in a corresponding order of magnitude decrease in the index. This is the most probable cause of the poor predictive ability of this index.

Due to the exceedingly poor performance of the Interchange Index in predicting movement of animals compared with the other indices, the influence of abundance as well as movement on the index and questions concerning its mathematical and theoretical basis, the use of this index is clearly not advisable. Given the relatively widespread application of this method this is of some concern. Re-analysis of results that have been calculated using it may be warranted.

The Sub-set Index was a better predictor of animal movement than the simple proportion, and a far better estimator than the Interchange Index. However, it was substantially out-performed by the Movement Index and Hilborn's method, and there would not appear to be any particular advantage to its use.

The Movement Index performed well over a range of sample sizes and sampling intensities. Unlike the previous indices it compares observed movement against expectation based upon assumptions about abundance and individual behavior. This comparison of observation and expectation lends itself to certain types of statistical comparison in a manner which the other approaches do not. Additionally, when there are a small numbers of recaptures all other indices are constrained to 0,

while M provides information on how this level of movement compares with expectation given population and sample sizes (Figure 1).

Hilborn's estimator substantially out-performed the other methods. It is the only method that provides a direct estimate of t rather than of proportional movement, and provided a good approximation of the simulation value for t across a wide range of animal abundances and sampling intensities. It can also be adapted to work with situations of increasing spatial and temporal complexity (Whitehead 2001).

Calculation of M is made using an estimate of N , or the population size against which deviation from random movement is to be tested, local, regional or ocean-wide depending on the question being addressed, and t requires input values for n , the number of individuals using the two areas sampled. Thus these indices may be sensitive to the input values provided. Estimates of these values may be unavailable, and if available may be highly imprecise and are likely to also be biased.

Hilborn's method estimates n in addition to t , so with a large enough range of input values and a large enough number of replicates, the technique should not be influenced by the choice of input values. However, if the range of random values provided for n is severely biased or inadequately precise, the values for t and n that provide the maximum overall likelihood may not be selected together even over several thousand replicates. Setting the mean for n to twice and three times the actual value, while keeping the cv at 0.2 reduced the r^2 from 0.654 to 0.556 and 0.424 respectively. In simulations where a much lower precision (sd=0.5) was used for selecting n , the estimator also performed substantially less well ($r^2 = 0.471$ for an unbiased mean and $r^2 = 0.290$ for biasx2). A potential solution to the problem of uncertainty regarding appropriate input values for n is to estimate t iteratively, conducting an initial trial using a broad range of parameter values, and using these results to help select a more appropriate range of values for n (and also possibly t) to be used in the final estimation of t .

The value of M explicitly compares the degree of movement observed to that expected given random movement in a population of a given size (N), and is therefore contingent of the choice of input value for N . Fortunately, while variations in the abundance estimates used will influence the absolute value of the M , relative values are not particularly sensitive to the value of N selected, allowing comparison of M between areas to be relatively unaffected. Calculating M for N s of 8000, 1200 and 800 gave r^2 values of 0.490, 0.555 and 0.541 respectively. The circumstances in which this

index would appear to have advantages in comparison with Hilborn's method are primarily those in which the utility of conducting a comparison of observation and expectation would infer that N was known with some degree of precision.

The simulations conducted here suggest that some of the methods that have been presented for estimating movement of individual marine mammals between habitats are poor predictors of the underlying biological processes and that the Interchange Index in particular should not be used. The simulations provide considerable support for use of Hilborn's estimator of transition probability and indicate that it is the most powerful and useful technique over a broad range of circumstances. The Movement Index is more limited in predictive ability, but may be of use in those situations in which comparison of observation with a specific expectation is of interest.

ACKNOWLEDGEMENTS

S. Buckland, P. Hammond and H. Whitehead provided valuable comments on earlier versions of this manuscript. H. Whitehead's emphasis on the importance of Hilborn's method encouraged a more thorough investigation of that work. The continued interest of J. Calambokidis and R. Baird in these findings was instrumental in keeping the project active. I wish to thank N Stevick for her forbearance when I insist on committing time to these unfunded projects.

LITERATURE CITED

- Acevedo J, Rasmussen K, Félix F, Castro C, Llano M, Secchi E, Saborío MT, Aguayo-Lobo A, Haase B, Scheidat M, Dalla-Rosa L, Olavarría C, Forestell P, Acuña P, Kaufman G, Pastene LA (2007) Migratory destinations of humpback whales from the Magellan Strait feeding ground, southeast Pacific. *Mar Mamm Sci* 23:453-463
- Baker CS, Herman LM, Perry A, Lawton WS, Straley JM, Straley JH (1985) Population characteristics and migration of summer and late-season humpback whales (*Megaptera novaeangliae*) in southeastern Alaska. *Mar Mamm Sci* 1:304-323
- Baker CS, Herman LM, Perry A, Lawton WS, Straley JM, Wolman AA, Kaufman GD, Winn HE, Hall JD, Reinke JM, Östman J (1986) Migratory movement and population structure of humpback whales (*Megaptera novaeangliae*) in the central and eastern North Pacific. *Mar Ecol Prog Ser* 31:105-119
- Calambokidis J, Steiger GH, Straley JM, Herman LM, Cerchio S, Salden DR, Urbán R. J, Jacobsen J, vonZiegesar O, Balcomb KC, Gabriele CM, Dalheim ME, Uchida S, Ellis G, Miyamura Y, Ladrón de Guevara P, Yamaguchi M, Sato F,

- Mizroch SA, Schendler L, Rasmussen K, Barlow J, Quinn TJI (2001) Movements and population structure of humpback whales in the North Pacific. *Mar Mamm Sci* 17:769-794
- Calambokidis J, Steiger GH, Straley JM, Quinn TJI, Herman LM, Cerchio S, Salden DR, Yamaguchi M, Sato F, Urbán R. J, Jacobsen J, vonZiegesar O, Balcomb KC, Gabriele CM, Dalheim ME, Higashi N, Uchida S, Ford JKB, Miyamura Y, Ladrón de Guevara P, Mizroch SA, Schendler L, Rasmussen K (1997) Abundance and population structure of humpback whales in the North Pacific basin. Report No. Contract #50ABNF500113, SWFSC, NOAA, NMFS, P.O. Box 271, La Jolla, CA
- Clapham PJ, Aguilar A, Hatch LT (2008) Determining spatial and temporal scales for management: lessons from whaling. *Mar Mamm Sci* 24:183-201
- Hilborn R (1990) Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. *Can J Fish Aquat Sci* 47:635-643
- Katona SK, Beard JA (1990) Population size, migrations and feeding aggregations of the humpback whale (*Megaptera novaeangliae*) in the western North Atlantic Ocean. *Rep int Whal Commn, Special Issue* 12:295-305
- Nichols JD, Kaiser A (1999) Quantitative studies of bird movement: a methodological review. *Bird Study* 46:S289-298
- Pollock KH, Nichols JD, Brownie C, Hines JE (1990) Statistical inference for capture-recapture experiments. *Wildl Monog* 107:1-97
- R Core Development Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Stevick PT, Allen J, Bérubé M, Clapham PJ, Katona SK, Larsen F, Lien J, Mattila DK, Palsbøll PJ, Robbins J, Sigurjónsson J, Smith TD, Øien N, Hammond PS (2003) Segregation of migration by feeding ground origin in North Atlantic humpback whales (*Megaptera novaeangliae*). *J Zool, Lond* 259:231-237
- Stevick PT, Allen J, Clapham PJ, Hammond PS, Katona SK, Larsen F, Lien J, Mattila DK, Palsbøll PJ, Sears R, Sigurjónsson J, Smith TD, Vikingsson GA, Øien N (2006) Population spatial structuring on the feeding grounds in North Atlantic humpback whales. *J Zool, Lond* 270:244-255
- Stevick PT, McConnell BJ, Hammond PS (2002) Patterns of movement. In: Hoelzel AR (ed) *Marine mammal biology: an evolutionary approach*. Blackwell Science, Oxford, p 185-216
- Urbán R J, Jaramillo L A, Aguayo L A, Ladrón de Guevara P P, Salinas Z M, Alvarez F C, Medrano G L, Jacobsen JK, Balcomb KC, Claridge DE, Calambokidis J, Steiger GH, Straley JM, vonZiegesar O, Waite JM, Mizroch SA, Dalheim ME, Darling JD, Baker CS (2000) Migratory destinations of humpback whales wintering in the Mexican Pacific. *J Cetacean Res Manage* 2:101-110
- Whitehead H (2001) Analysis of animal movement using opportunistic individual identifications: application to sperm whales. *Ecology* 85:1417-1432
- Whitehead H, Dillon M, Dufault S, Weilgart L, Wright J (1998) Non-geographically based population structure of south Pacific sperm whales: dialects, fluke-markings and genetics. *J Anim Ecol* 67:253-262
- Whitehead H, Glass C (1985) The significance of the Southeast Shoal of the Grand Bank to humpback whales and other cetacean species. *Can J Zool* 63:2617-2625

Method	R-square	F	P	Outliers
Hilborn's <i>t</i>	0.6535	9428	***	6
Movement Index	0.5544	6219	***	5
Sub-set Index	0.4825	4661	***	0
Proportion	0.4122	3505	***	0
Interchange Index	0.2934	2075	***	44

Table 1. Regression results indicating the fit of different indices to the probabilities assigned in the simulation model. Outliers are defined as records having studentised residuals with an absolute value greater than four.

Figure 1. The relationship between four two-sample movement indices and the transition probability between areas derived from 5,000 replicates of a simulation model described in the text.

