Comparative performance testing of spatially explicit genetic analytical methods

KAREN K. MARTIEN AND DAVID P. GREGOVICH
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, CA 92307, USA
Contact e-mail: Karen.Martien@noaa.gov

ABSTRACT

An accurate understanding of population structure is critical to the success of any management effort. Although there have been a number of genetic methods developed to detect population structure, these methods have not been systematically tested to determine which are most effective in a management setting. In this study we conducted comparative performance tests of three genetic analytical methods – Wombling, the Monmonier algorithm, and the cluster algorithm of Waples and Gaggiotti. Performance of these methods was evaluated with respect to how well they detect population genetic structure and use this information to construct appropriate management units for use in the International Whaling Commission's Revised Management Procedure (RMP). Trials of each genetic method were performed in a simulated management setting across a range of population structure scenarios. The scenarios varied with respect to the number and sizes of populations and the annual rate of dispersal among them. The methods generally detected populations and managed them appropriately when the annual dipsersal rate was low ($5X10^{-6}$/year). At intermediate dispersal rates ($5x10^{-5}$ and $5x10^{-4}$), there was a large difference in the performance of the methods, with the Monmonier algorithm and the Waples and Gaggiotti clustering method performing very well and Wombling performing poorly. None of the methods was able to detect population structure when the annual dispersal rate was $5X10^{-3}$. Consequently, populations were frequency over-harvested in these trials. Nonetheless, our results indicate that the Monmonier algorithm and the Waples and Gaggiotti clustering algorithm may prove to be useful tools for defining management units for use with the RMP. Further testing to fully characterize the performance of these methods is required before final conclusions can be drawn.

INTRODUCTION

The proper management of species subject to human-caused mortality requires an understanding of population structure. The most commonly used tool in evaluating population structure is Genetic data (Palsbøll *et al.*, 2007; Taylor *et al.*, submitted). New analytical methods for inferring population structure from genetic data are published routinely. Most methods were created to address evolutionary questions and are not necessarily well suited applied studies aimed at defining management units (MUs). Though comparative performance testing of analytical methods has become increasingly common (Abdo *et al.*, 2004; Chen *et al.*, 2007; Latch and Rhodes, 2006; Waples and Gaggiotti, 2006), few methods have been tested across the range of dispersal rates relevant to managers. The Testing of Spatial Structure Methods (TOSSM) project was designed to fill this gap (IWC, 2004). The purpose of the TOSSM project is to evaluate the performance of a range of methods across a range of different types of population structure and levels of connectivity relevant to management of wild populations.

The Revised Management Procedure (RMP) is the management scheme that the International Whaling Commission (IWC) adopted for managing commercial harvest of large whales. The RMP relies on a Catch-Limit Algorithm (CLA) for calculating catch limits (IWC, 1994). Martien *et al*. (2008b) showed that in order for the CLA to prevent over-harvest of

undetected populations, it is necessary to separately manage populations exchanging dispersers at annual rates as high as $5x10^{-3}$. Detecting this level of population structure is likely to be very challenging for most analytical methods (Morin *et al.*, In press; Palsbøll *et al.*, 2007; Taylor *et al.*, submitted; Waples and Gaggiotti, 2006). One of the goals of the TOSSM project is to evaluate analytical methods to determine their utility in defining MUs for management under the RMP.

In this study, we examine the performance of three analytical methods for identifying population structure – 'Wombling' (Crida and Manel, 2007; Womble, 1951), the Monmonier algorithm (Monmonier, 1973), and the clustering method described by Waples and Gaggiotti (2006). Wombling uses the multi-locus genotypes and spatial locations of individual samples to identify populations. It determines both the number of populations present in the study area and the locations of the boundaries between them. The clustering method of Waples and Gaggiotti uses allele frequency data from pre-defined sampling sites to determine the number of independent gene pools in the study area. It does not use spatial information, and therefore has the potential to define geographically discontiguous populations. The Monmonier algorithm also uses allele frequency data from pre-defined sampling sites to divide the study area into groups. However, like Wombling, it utilizes spatial data. The Monmonier algorithm does not include a mechanism for determining how many groups should be defined. Rather, the number of groups is an input to the algorithm, which then simply determines the boundary location(s).

Though both the Waples and Gaggiotti algorithm and Monmonier algorithm have been subjected to comparative performance testing (Dupanloup *et al.*, 2002; Waples and Gaggiotti, 2006), none of these methods have been tested at the relatively high dispersal rates relevant to management under the CLA, and none have been tested in a management context. In this paper, we present preliminary results of performance tests of these three methods. Performance tests were performed using the TOSSM package (Martien *et al.*, 2008a), which is an R package developed as part of the TOSSM project. The TOSSM package can be used to simulate management of populations under the RMP. Within the package, management units (MUs) are defined by a user-defined Boundary Setting Algorithm, or BSA. We developed BSAs based on the three methods examined in this paper, and evaluated their ability to define MUs that protected populations from over-harvest. Though preliminary, our results shed light on the utility of these three methods as management tools.

METHODS

We used the TOSSM package to examine the performance of each of these three BSAs. Trials were performed across a range of population structure scenarios (Table 1, Figure 1). These scenarios involved one, two, or three adjacent populations (Figure 1A-D), with annual dispersal rates among populations varying from $5x10^{-6}$ to $5x10^{-3}$. The carrying capacity (K) of the entire study area (summed across populations) was 7500 and the maximum sustainable yield rate ($MSYR_{1+}$) was 4% for all simulations. Either six or twelve contiguous sampling polygons, from which genetic samples were collected, covered the entire study area. Twelve sampling polygons were used in the case of the Monmonier-BSA (Figure 1F), which will not accept as input sampling sites situated along a line. Trials involving the other two BSAs used six linearly arranged sampling polygons (Figure 1E). In all other respects— spatial situation of populations, density and distribution of animals, and number of samples collected—the simulations were identical for all three BSAs.

Table 1. Population structure scenarios used in trial simulations.

Scenario	*#populations*	*Population K*	*Dispersal Rates*
Panmixia	1	7500	NA
'2-even'	2	3750/3750	$5x10^{-6}$, $5x10^{-5}$, $5x10^{-4}$, $5x10^{-3}$
'2-uneven'	2	750/6750	$5x10^{-6}$, $5x10^{-5}$, $5x10^{-4}$, $5x10^{-3}$
'3-even'	3	2500/2500/2500	$5x10^{-6}$, $5x10^{-5}$, $5x10^{-4}$, $5x10^{-3}$

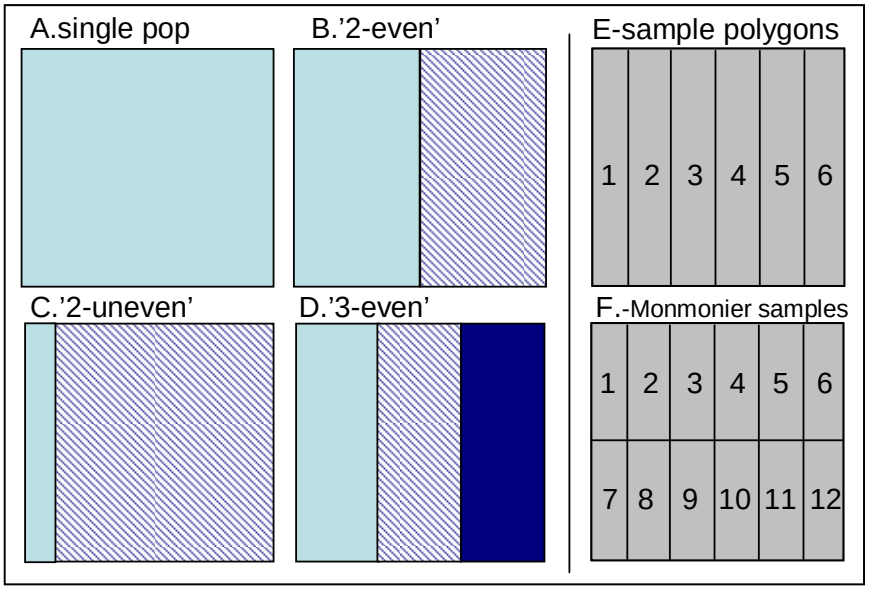

Figure 1. Depiction of populations (A-D) and sampling polygons (E-F) used in the simulations.

Each simulation consisted of a single year in which historic harvest took place, followed by 100 modern management years. In the historic harvest year, the initial abundance of the left-most population (always referred to as population 1) was reduced to 0.3 of its carrying capacity. In trials involving a single population, the entire population was reduced to 0.3*K*. This initial depletion mimics the historic depletion that has taken place in many cetacean populations worldwide. Also in the first simulation year, a total of 600 genetic samples were collected. These samples were evenly distributed across the sampling polygons and formed the basis for genetic analysis and MU definition by the BSAs.

In the 100-yr modern management period, populations were harvested in each MU defined by the BSA according to the catch limits calculated by the IWC's Catch-Limit Algorithm (CLA; IWC, 1994). Abundance was estimated every fifth year of the simulation, and the catch limit re-calculated accordingly. The parameter controlling the relationship between the abundance estimates and their coefficient of variation (CV) was set such that when all populations were at carrying capacity, the CV equaled 0.037. This CV level is unrealistically low (Taylor *et al.*, 2007). It was chosen because Martien et al. (2008b) showed that for nearly all population scenarios we examined, this CV level results in failure to adequately conserve populations if population structure goes undetected. Thus, this parameter setting ensures the performance of BSAs as reported in our results will depend on their ability to accurately define MUs.

The conservation performance of the BSAs was assessed relative to one of the performance metrics used when developing the CLA, namely, the probability that a population initially depleted to 0.3*K* recovers to greater than 0.54*K* after 100 years of managed harvest. Recovery was evaluated only for population 1, which was the most vulnerable population in our simulations. Population 1—

henceforth referred to as *P1*—was the population subjected to historic harvest in simulations. *P1* was pressured further by a spatial bias in harvest. In all simulations, the study area was divided into ten equally-sized harvest intervals. Harvest was begun in the left-most harvest interval, and then proceeded to the right as each interval was extirpated. In addition to the spatial bias in both historic and managed harvest, the carrying capacity, *K*, of P1 also represented only 1/10 of the total *K* in the '2-uneven' trials (Table 1). The combination of these factors leads to a high vulnerability of *P1* to overharvest, an important situation to simulate from a conservation management perspective.

Hyptest-BSA

The hyptest-BSA is based upon a method originally proposed by Waples and Gaggiotti (2006) as a simple way to test for gene pool independence. It treats sampling sites as putative independent populations, performing pairwise tests for genetic differentiation between the each of the sites. Sampling sites that are not significantly differentiated are combined into the same 'network'. Two sampling site can share a network vicariously through intermediate sampling sites (ie. if A is connected to B, and B is connected to C, then A, B, and C are in the same network). A particular sampling site forms its own network if all pairwise G-tests in which the site is included are significant.

Hyptest-BSA defines each network as a separate MU. Within the BSA, genetic differentiation between sampling sites was assessed using multi-locus G-tests conducted using the R package 'hierfstat' (Goudet *et al.*, 1996). The threshold at which tests are considered statistically significant can be passed as an argument to the BSA. For all simulations reported here, we used the default value of `alpha` = 0.05.

Wombsoft-BSA

Wombsoft-BSA is based on the work of Womble (1951), who proposed a method for identifying spatial regions of maximum change amongst multiple organismal traits, be they morphometric or genetic. The 'wombsoft' package (Crida and Manel 2007) uses this 'wombling' method with further refinements (Fan and Gijbels, 1996). The wombsoft package takes as input genetic data (either haploid or diploid, with the number of loci chosen by the user), and computes a systemic function from the degree of allele frequency change across all loci at each grid cell on a map. A binomial test then assesses the significance of potential genetic boundaries.

In testing the wombsoft package, multiple iterations were initially run with simulated genetic data to tune the input parameters used. Arguments for the various wombsoft functions were set at the values that optimized performance. The bandwidth *H*— the relative size of the entire study area used to define the systemic function at each grid cell—was set at 10. A percentile threshold, p_B, above which the systemic function at a grid cell will be considered as a candidate boundary, was set (at the wombsoft authors' suggestion) to 0.3. The significance level for binomial tests was set at 0.05. An attempt to reduce border effects from any potential lack of data at the study area periphery was made by using wombsoft's `DataMirror` function, which duplicates multilocus datapoints in an area around the border defined by radius *m* (set to 10). Though the current version of Wombsoft-BSA always uses these parameter values, in future versions of the BSA these parameters will be user-specifiable arguments.

The information supplied by the `BinomialTestCodominant` function of wombsoft includes a matrix of binomial test results by grid cell across the entire study area. In this matrix, patches of the study area isolated by boundaries significant at the 0.05 level were identified, and each unique patch was defined as a separate MU. The resulting wombsoft borders are not perfectly

linear but polygonal and vary in thickness. Wombsoft-BSA splits these border areas to create MUs that cover the study area without overlapping.

Monmonier-BSA

The Monmonier Algorithm (Dupanloup *et al.,* 2002; Monmonier, 1973) is a method of drawing boundaries through geographic space so as to maximize some measure of difference across the boundaries. In genetic applications, the goal is to maximize genetic differentiation across boundaries. The method begins by drawing a Voronoi diagram (Voronoi, 1908), which divides the entire study area into multi-sided Voronoi tiles. Each tile is centered on a sampling site and delineates the 'neighborhood' of that sampling site, such that every point within the tile is closer to that sampling site than it is to any other sampling site. The edges of the Voronoi tiles separate adjacent sampling sites.

Once the Voronoi diagram has been drawn, the genetic differentiation between all adjacent sampling sites (i.e., those separated by an edge) is calculated. All edges that terminate on the boundary of the study area are identified, and the one across which genetic differentiation is highest is chosen as the first boundary segment. The edges that abut the first boundary segment are examined, and the one across which differentiation is highest is chosen as the next boundary segment. The boundary propagates across the study area in this way until it meets another boundary segment or the study area boundary, at which point the boundary is complete. Once an edge has been chosen to as a boundary segment, the choice is never reconsidered. Thus, the final boundary defined by the algorithm may not maximize genetic differentiation between the groups it defines.

Monmonier BSA uses Wright's F_{ST} as the measure of genetic differentiation between sampling sites. F_{ST} is calculated with the R package 'eiaGenetics'[1]. The Voronoi diagram is calculated with the R package 'deldir' (Turner, 2008), using the centroid of each sampling polygon as the point location of the sampling site. The BSA does not include any means for determining how many boundaries should be drawn. The current version of the BSA generates a single boundary, which defines two MUs. Consequently, Monmonier BSA was only applied to the 2-population scenarios listed in Table 1. Future versions will allow the user to specify the number of MUs to be defined.

RESULTS

The three BSAs we examined varied substantially in their ability to define MUs that adequately conserved populations. Hyptest-BSA performed well at accurately defining the correct number of MUs in all trials except for those with the highest annual dispersal rate (Table 2). In trials with annual dispersal rates of $5x10^{-4}$ or lower, hyptest-BSA never defined too few MUs, though there was a slightly tendency for it to define too many MUS. This tendency was most pronounced in the '3-even' trials, for which hyptest-BSA defined too many MUs 15-20% of the time, sometimes defining as many as 5 MUs (Table 2). There was a dramatic dropoff in the performance of the hyptest-network BSA when the annual dispersal rate reached $5X10^{-3}$. In these trials, hyptest-BSA nearly always defined a single MU. The ability of hyptest-BSA to adequately protect *P1* (the population most vulnerable to overharvest) largely mirrored its ability to accurately determine the number of MUs that should be defined (Table 3). This reflects the fact that when hyptest-BSA defined the correct number of MUs, the MU boundaries nearly always corresponded perfectly to actual population boundaries.

[1] Written by and available upon request from Eric Archer (Eric.Archer@noaa.gov)

Table 2. Performance of hyptest- and wombsoft-BSAs in accurately determining the number of management units based on genetic data over 100 trials: (-) — number of MUs was underdefined, (acc) — number of MUs was accurately defined, (+) — number of MUs was overdefined.

		hyptest			***wombsoft***		
Scenario	*Dispersal rate*	-	*acc*	+	-	*acc*	+
Panmixia	NA	0	100	0	0	93	7
2-even	$5X10^{-6}$	0	95	5	0	99	1
	$5X10^{-5}$	0	99	1	4	95	1
	$5X10^{-4}$	0	98	2	92	7	1
	$5X10^{-3}$	98	2	0	91	8	1
2-uneven	$5X10^{-6}$	0	100	0	12	44	44
	$5X10^{-5}$	0	99	1	85	11	4
	$5X10^{-4}$	0	100	0	93	6	1
	$5X10^{-3}$	100	0	0	92	6	2
3-even	$5X10^{-6}$	0	81	19	0	99	1
	$5X10^{-5}$	0	86	14	51	48	1
	$5X10^{-4}$	0	87	13	98	2	0
	$5X10^{-3}$	88	11	1	100	0	0

Table 3. Percentage of trials in which *P1* went extinct or was depleted to less than 0.54K.

		hyptest		***wombsoft***		***Monmonier***	
		extinct	<0.54K	extinct	<0.54K	extinct	<0.54K
Panmixia	NA	0	0	0	0	--	--
2-even	5X10-6	0	0	0	0	0	0
	5X10-5	0	0	0	2	0	0
	5X10-4	0	0	0	39	0	0
	5X10-3	0	21	0	21	0	24
2-uneven	5X10-6	2	9	80	96	0	0
	5X10-5	3	9	100	100	0	0
	5X10-4	0	2	75	100	10	10
	5X10-3	0	100	0	100	0	78
3-even	5X10-6	0	0	0	0	--	--
	5X10-5	0	0	2	3	--	--
	5X10-4	0	0	67	92	--	--
	5X10-3	0	48	0	80	--	--

Wombsoft-BSA did not conserve *P1* as well as hyptest-BSAs. While wombsoft-BSA detected populations accurately in the '2-even' and '3-even' trials with annual dispersal rates of $5X10^{-6}$, performance was poor in most other trials, with wombsoft-BSA typically defining only a single MU regardless of the number of populations actually present (Table 2). When wombsoft-BSA did specify the correct number of MUs, inspection of the MU polygons created showed that the polygon boundaries often did not coincide with population boundaries. Instead, MUs were often created in areas where no genetic gradient would be expected. These MUs were oftentimes small and near the border of the study area. This poor correspondence between population boundaries and MU boundaries resulted in wombsoft-BSA failing to protect *P1* in nearly all of the '2-uneven' simulations (Table 3), even in instances where it defined the correct number of MUs.

The sensitivity of the wombsoft-BSA results to errors in boundary placement is illustrated in Figure 3, which shows the MUs defined in two replicates of the '2-uneven' trials for which annual dispersal rate was $5x10^{-6}$. In the first replicate (Fig. 3A), an MU was placed directly on top of and encompassing *P1*. In the second replicate (Fig. 3B), an MU was placed in roughly the correct position, but covering less than half of *P1*. A second MU covered the remainder of *P1* and the rest

of the study area. The final abundance of *P1* in the first replicate was 633 (*0.85*K). In the second simulation, *P1* was extirpated. In the second replicate, the protection *P1* gained by having a small MU largely contained within it was not enough to offset the fact that a large portion of *P1* was managed in an MU along with the much larger second population.

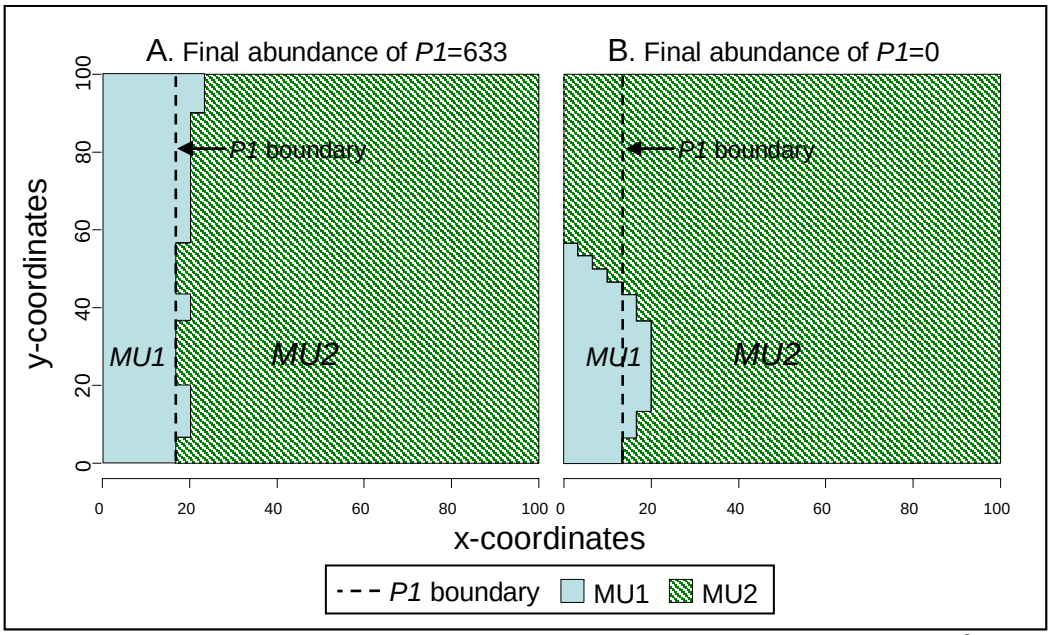

Figure 3. Examples of wombsoft-BSA MU placment relative to *P1* in '2-uneven'—$5X10^{-6}$ trials.

The Monmonier-BSA performed well at all but the highest dispersal rate (Table 3). Because Monmonier-BSA always defines 2 MUs, its performance was evaluated only with respect to the accuracy of boundary placement in the '2-even' and '2-uneven' trials. *P1* was not depleted to <0.54*K* in any of these trials at the two lowest dispersal rates. When the annual dispersal was $5x10^{-4}$, *P1* again recover to greater than 0.54*K* in all of the '2-even' trials and 90% of the '2-uneven' trials. Like the other two BSAs, monmonier-BSA generally failed to accurately define MUs at the highest dispersal rate ($5x10^{-3}$) (Table 3).

DISCUSSION

Both hyptest-BSA and monmonier-BSA performed well at defining MUs that prevented the over-harvest of *P1*. Hyptest-BSA met the performance objective of allowing *P1* to recover to greater than 0.54*K* in more than 90% of replicates, except when the annual dispersal rate was $5x10^{-3}$ (Table 3). In the trials consisting of a single population, hyptest-BSA accurately defined a single MU 100% of the time (Table 2). In the two-population trials ('2-even' and '2-uneven'), it nearly always defined the correct number of MUs when the annual dispersal rate was $5x10^{-4}$ or lower. In three-population trials ('3-even'), it showed a tendency to define more than three MUs. Though such an 'over-protection' error does not harm the conservation performance of the BSA, it can place an undue burden on whalers by reducing catch limits or increasing the effort required in order to fill their quotas. Though we did not collect catch and effort data in our simulations, these data are generated by the TOSSM package and need to be considered in future evaluations of all BSAs.

The performance of monmonier-BSA was comparable to that of hyptest-BSA in the '2-even' trials. However, it performed slightly better in the '2-uneven' trials. The version of

monmonier-BSA used in this study always defines two MUs. However, since hyptest-BSA defined two or more MUs in all of the '2-uneven' trials, the superior performance of monmonier-BSA in these trials reflects a lower rate of boundary placement errors rather than an advantage in determining the number of MUs. The lower rate of boundary placement errors is particularly noteworthy due to the fact that the monmonier-BSA trials involved twice as many sampling sites as the hyptest-BSA trials. The greater number of sampling sites means that monmonier-BSA was less constrained and therefore had more opportunities to make mistakes than hyptest-BSA. Additional trials of hyptest-BSA utilizing twelve sampling sites will be necessary in order to directly compare its performance to that of monmonier-BSA.

Though monmonier-BSA exhibited the best performance of the three BSAs we tested, it cannot be considered a complete method of defining MUs. Monmonier-BSA does not include any mechanism for determining the number of MUs that should be defined. Its performance in these trials is therefore predicated on the assumption that the correct number of MUs is known *a priori*. Further development of this BSA should therefore include testing its performance when combined with other methods intended for evaluating the demographic independence (or lack thereof) of groups defined by the Monmonier algorithm. For instance, the Monmonier algorithm could be combined with methods designed to estimate dispersal rates between putative populations, such as Migrate (Beerli and Felsenstein, 1999; 2001), BayesAss (Fisher *et al.*, 2002), or LAMARC (Kuhner, 2006).

None of the BSAs was able to accurately identify population structure when the annual dispersal rate was $5x10^{-3}$. A recent study examining the performance of the CLA when population structure was not correctly identified showed that detecting such a dispersal rate is necessary in order to avoid over-harvest of undetected populations (Martien *et al.*, 2008b). This conclusion is supported by the results of this study, which show that *P1* failed to recover to greater than 0.54*K* in most of the trials utilizing this dispersal rate. Though low enough to render the populations demographically independent (Palsbøll *et al.*, 2007; Taylor, 1997; 2005; Waples and Gaggiotti, 2006), an annual dispersal rate of $5x10^{-3}$ is high enough to largely counteract the effects of genetic drift. Consequently, populations exchanging dispersers at this rate will exhibit very low levels of genetic differentiation, and therefore be very difficult to detect using genetic methods (Morin *et al.*, In press; Palsbøll *et al.*, 2007; Taylor *et al.*, submitted). Though *P1* failed to recover to greater than 0.54*K* in the majority of trials with this dispersal rate, none of these trials resulted in the complete extirpation of *P1*. Rather, *P1* was 'rescued' by recruitment via dispersal from its neighboring population.

All three BSAs had lower performance in the '2-uneven' trials than they did in the trials where the population polygons were of equal size. Wombsoft-BSA performed particularly poorly in these trials. Unlike hyptest-BSA and monmonier-BSA, both of which work by maximizing differentiation between sampling sites, wombsoft-BSA searches for gradients in allele frequencies across the entire study area. These gradients may be harder to detect when they occur close to the edge of the study area, as was the case for the '2-uneven' trials (Fig. 1). It is possible that more thorough testing of alternative parameters to use in the WombSoft package would yield more accurate boundary placement in these trials.

Wombsoft-BSA is also disadvantaged relative to the other BSAs because it works at the level of the individual and disregards sampling sites. Monmonier-BSA and hyptest-BSA both assume that all animals within a sampling site belong to the same MU, and therefore only have the much easier task of deciding which MU each sampling site belongs to. Because there are only a limited number of ways that the sampling sites can be combined into MUs, the chance of errors is

greatly reduced. The advantages of working at the sampling site level rather than the individual level were magnified in our trials due to the fact that our sampling polygons covered the entire study area. Thus, hyptest-BSA and monmonier-BSA had only to assign sampling sites to MUs and did not have to divide up any area between sampling sites.

Because hyptest-BSA and monmonier-BSA require that samples be grouped *a priori* into sampling sites, they are vulnerable to errors in sampling site definition. For instance, if a sampling site included samples from two different populations, both hyptest-BSA and monmonier-BSA would be precluded from the outset from placing an MU boundary in exactly the correct location. Wombsoft-BSA does not suffer from this shortcoming.

Though none of the BSAs was able to detect population structure across the full range of dispersal rates required, our results suggest that hyptest-BSA and monmonier-BSA show promise as methods for defining MUs in species managed under the IWC's CLA. However, further testing of all of these BSAs is necessary before firm conclusions can be drawn. All of our trials were conducted using a relatively large total sample size of 600, which is larger than the sample sizes available for many species of large whales. Reduced sample size will reduce the methods' ability to accurately detect population structure. Conversely, with larger samples, some of the BSAs may be able to detect population structure at annual dispersal rates of $5x10^{-3}$ or higher, something they failed to do in our simulations. The performance of hyptest-BSA and monmonier-BSA is also likely to be influenced by the number and spatial configuration of sampling sites. In real studies, both the number and spatial complexity of sampling sites are likely to be higher than they were in our simulations, increasing the complexity of the problem the BSAs face and likely reducing their performance. Thus, further testing with more realistic sampling polygons is warranted.

ACKNOWLEDGMENTS

We would like to thank Stephanie Manel for help in implementing the WombSoft package, and Eric Archer for providing us with his eiaGenetics package. Eric Archer also provided helpful comments on this manuscript.

REFERENCES

Abdo, Z., Crandall, K.A. and Joyce, P. 2004. Evaluating the performance of likelihood methods for detecting population structure and migration. *Molecular Ecology* 13:837-51.

Beerli, P. and Felsenstein, J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763-73.

Beerli, P. and Felsenstein, J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* 98:4563-68.

Chen, C., Durand, E., Forbes, F. and Francois, O. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* 7:747-56.

Crida, A. and Manel, S. 2007. WOMBSOFT: an R package that implements the Wombling method to identify genetic boundary. *Molecular Ecology Notes* 7:588-91.

Dupanloup, I., Schneider, S. and Excoffier, L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11:2571-81.

Fan, J. and Gijbels, I. 1996. Local Polynomial Modelling and its Applications. London: Chapman and Hall.

Fisher, M.C., Rannala, B., Chaturvedi, V. and Taylor, J.W. 2002. Disease surveillance in recombining pathogens: Multilocus genotypes identify sources of human *Coccidioides* infections. *Proceedings of the National Academy of Sciences of the United States of America* 99:9067-71.

Goudet, J., Raymond, M., De Meeus, T. and Rousset, F. 1996. Testing differentiation in diploid populations. *Genetics* 144:1933-40.

International Whaling Commission. 1994. The Revised Management Procedure (RMP) for baleen whales. *Rep. Int. Whal. Commn.* 44:145-67.

International Whaling Commission. 2004. Report of the Workshop to design simulation-based performance tests for evaluating methods used to infer population structure from genetic data. *J. Cetacean Res. Manage.* 6:469-85.

Kuhner, M.K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768-70.

Latch, E.K. and Rhodes, E. 2006. Evidence for bias in estimates of local genetic structure due to sampling scheme. *Animal Conservation* 9:308-15.

Martien, K.K., Gregovich, D., Bravington, M.V., Punt, A.E., Strand, A.E., Tallmon, D.A. and Taylor, B.L. 2008a. TOSSM: an R package for assessing performance of genetic analytical methods in a management context. Paper SC/60/SD2 presented to the IWC Scientific Committee, June, 2008 (unpublished).

Martien, K.K., Gregovich, D.P. and Punt, A.E. 2008b. Evaluating the performance of the CLA when population structure is not correctly identified. Paper SC/60/SD3 submitted to the IWC Scientific Committee, June, 2008 (unpublished).

Monmonier, M.S. 1973. Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis* 3:245-61.

Morin, P.A., Martien, K.K. and Taylor, B.L. In press. Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*.

Palsbøll, P.J., Bérubé, M. and Allendorf, F.W. 2007. Identification of management units using population genetic data. *Trends in Ecology & Evolution* 22:11-16.

Taylor, B.L. 1997. Defining "population" to meet management objectives for marine mammals. pp. 49-65. *In*: A.E. Dizon, S.J. Chivers and W.F. Perrin, (eds.) *Molecular genetics of marine mammals*. Society for Marine Mammalogy.

Taylor, B.L. 2005. Identifying units to conserve. pp. 149-64. *In*: J.E. Reynolds III, W.F. Perrin, R.R. Reeves, S. Montgomery and T.J. Regan, (eds.) *Marine mammal research: conservation beyond crisis.* The Johns Hopkins University Press, Baltimore, Maryland.

Taylor, B.L., Martien, K.K. and Morin, P.A. submitted. Identifying units to conserve using genetic data. pp. *In*: I. Boyd, D. Bowen and S. Iverson, (eds.) *Marine mammal ecology and conservation: a handbook of techniques*. Oxford University Press, Oxford, UK.

Taylor, B.L., Martinez, M., Gerrodette, T., Barlow, J. and Hrovat, Y.N. 2007. Lessons from monitoring trends in abundance of marine mammals. *Marine Mammal Science* 23:157-75.

Turner, R. 2008. The deldir package.

Voronoi, M.G. 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques, deuxième mémoire, recherche sur les parallélloedres primitifs. *Journal of Reine Angewandte Mathematik* 134:198-287.

Waples, R.S. and Gaggiotti, O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15:1419-39.

Womble, W.H. 1951. Differential Systematics. *Science* 28:315-22.