Dynamic population segregation by genetics and morphometrics in Antarctic minke whales

Toshihide Kitakado[1], Tore Schweder[2], Naohisa Kanda[3], Luis A. Pastene[3] and Lars Walløe[2]

[1]Tokyo University of Marine Science and Technology, 4-5-7 Konan, Minato-ku, Tokyo 108-0075, Japan
[2] University of Oslo, P.O. Box 1103, 0317 Blinder, Oslo, Norway
[3] Institute of Cetacean Research, 4-5 Toyomi-cho, Chuo-ku, Tokyo, 104-0055, Japan
Contact e-mail: kitakado@kaiyodai.ac.jp

ABSTRACT

An integrated approach for estimating longitudinal segregation of two populations using different sources of information, genetic and morphometric data is introduced. A soft boundary is allowed to vary by year and sex although baseline populations are assumed. A joint conditional likelihood function derived from the two sources is defined for the estimation of mixing proportions. The method is applied to the extensive data for Antarctic minke whales taken by the JARPA and JARPA II surveys during the austral summers from 1989/90 to 2010/2011 in Antarctic areas III-E, IV, V and VI-W. The mixing proportion is modeled by a linear logistic model with population-specific parameters for the two sets of data. The genetic and morphometric data in the western areas (Areas III-E and IV-W) and the eastern areas (Areas V-E and VI-W) are used for estimating parameters for the two putative populations, I-stock (western population) and P-stock (eastern population), respectively. It was observed that the morphometric data shown dominated information compared to the genetic data and it helped convergence in the optimization. However, the inclusion of the morphometric data altered the estimation results and tended to give softer boundaries. On the whole, the result indicates that the spatial distribution of the two populations has a soft boundary in Area IV-E and V-W, which depends on the year. It also suggested possible sex differences along the boundary.

INTRODUCTION

Minke whales are feeding in large numbers in the circumpolar Southern sea during the austral summer. Recently, the abundance levels of minke whales in the Antarctic Ocean were estimated based on the data derived from the IDCR-SOWER program (Okamura and Kitakado 2012, Bravington and Hedley 2012, and IWC 2013). In addition, the estimates were made for areas III-E, IV, V and VI-W from line transect surveys carried out in the JARPA program (Hakamada and Matsuoka, 2014). These surveys show larger variation in abundance estimates than what is consistent with the nominal standard deviations representing sampling variability. This might partially be due to shifting oceanographic and/or feeding conditions in the areas from year to year. This might also have an impact on the yearly changes in the underlying stock structure.

The population structure for Antarctic minke whales has been investigated by using genetic and morphometric data obtained by the JARPA sampling surveys from 1987/88 to 2004/2005. Pastene et al. (2006) found genetic separation by longitude although the genetic divergence between populations is small. The paper also showed that Area V-W is a mixing area of two stocks but the transition area may change by year. In addition, a cluster analysis on a subset of the body measurements also indicates that there are two populations that feed in this region of the Antarctic Ocean (Hakamada 2006). These two papers contributed to a concluding paper, Pastene (2006).

Schweder et al. (2011) and Kitakado et al. (2012) also investigated the population structure through an integrated method by simultaneously using the two sources of information on genetic and morphometric data for the JARPA data and these papers indicated that there is a soft boundary in Areas IV-E and V-W between the two different populations and the boundary significantly depends on the year. The result also suggested that the boundary might be sex-specific although it showed greater variations from year to year.

Now additional extensive data on genetic and body measurements have also been collected from harvested Antarctic minke whales from the JARPA II sampling surveys. These data would not only expand the range of years to monitor the yearly changes but also contribute to obtaining information on baseline populations for the mixture analyses.

In this paper, we extend our previous analyses by updating the data sets from the JARPA and JARPA II. Genetic information was separately used, but the results with morphometric data were also provided. We assume two breeding populations softly separated by longitude and other covariates according to a linear logistic mixing model with year-specific intercept.

MATERIALS

Microsatellite data were obtained from analyzing up to twelve sets of primers EV1, EV104, GT023, GT211, GT195, DlrFCB14, AC045, AC082, AC087, AC137, CA234 and GT129. More detailed information on the laboratory procedures are described in Kanda et al. (2014). Also up to 10 measurements of body length and of other lengths between points on the body of the minke whale were available. Each sampled individual is recorded by date and location of sampling as well as biological information on sex and maturity. Figure 1 shows sampling locations of minke whales in 1989/90 - 2010/2011 seasons. Table 1 summarizes the sample size by year. The samples obtained in 1987/88 and 88/89 were not used because of lack of microsatellites.

Here, we denote genetic markers on up to 12 microsatellite loci and the morphometric measurements denote as $\{(a_{2l-1}, a_{2l}); l = 1, 2, ..., 12\}$ and $v = (v_1, \cdots, v_{10})$. Measurements of body length and of other lengths between points on the body of the minke whale, as shown on Figure 2, were transformed into the 9 allometric measures against the body length as $m_j = \log(v_{j+1} / v_1)$ $(j = 1, 2, ..., 9)$. The plots of the morphometric data against the longitude for mature individuals shown in Figure 3 suggest that Eastern individuals tend to have higher values in all but a couple of the measurements. To take into account the allometric change with growth, observed maturity status for each individual is used.

STATISTICAL MODELLING

Population mixture models

Consider the case of at most two different populations of Antarctic minke whales. Let y represent year and x longitude of a sampled individual. Longitude is here measured in degrees with origo at 180 degrees (E or W), and with negative numbers to the west and positive to the east. With $x < 0$ the whale was taken at $180 + x$ degrees east, and for a positive x it was taken at $180 - x$ degrees west. For example, x= -20 and 20 mean 160E and 160W, respectively.

The probability that an individual belongs to the Eastern population (P-stock in Pastene 2006) is assumed to follow the linear logistic form,

$$P(I = E) = p = \frac{e^{\alpha_y + \beta x}}{1 + e^{\alpha_y + \beta x}}. \qquad (1)$$

Here $\beta > 0$, and $M50_y = -\alpha_y / \beta$ is defined as the longitudinal point of 50% mixing by year. Of course, the

probability that an individual belongs to the Western population (I-stock) is $P(I=W)=1-P(I=E)$. The model with different mixing proportions by year and sex is straightforward. Table 2 summarizes a list of representative models employed in this study.

To see the effectiveness of such longitudinal segregation, we also estimate area-wise mixing proportions by year without assuming the longitudinal covariate.

Likelihood components: overview

Each individual contributes likelihoods from its microsatellite and its body measurements, when available. The parameters in the likelihood components depend on population indexed by *I=E,W.*

The microsatellite alleles at locus l, G_{2l-1}, G_{2l}, are assumed independent and identically multinomially distributed with probabilities $P(G=a)=\gamma_{la}^{I}$ for the set of alleles *a* observed at the locus, and independent across loci. The 9 morphometric measurements *M* are assumed multivariate normally distributed with mean vector μ^{I} depending on both the population and sex and covariance matrix Σ_s depending only on sex. The two different observations, *G* and *M*, are further assumed independent within an individual, and aslso a set of observations, *G* and *M*, are assumed to be independent between individuals. Note that the morphometric data from immature animals are not used for the analysis, and therefore the estimated parameters for the immature animals based on both data sets are usually the same as those from the genetic data unless parameters common to the groups (mature males, mature females, and immature animals) are incorporated.

Likelihood contribution from microsatellites

Disregarding genotyping errors, and dropping the index for population, a homozygote (a,a) in a locus l has probability $g_{l,aa}=\gamma_{l,a}^{2}$, while a heterozygote (a,b) has probability $g_{l,ab}=2\gamma_{l,a}\gamma_{l,b}$. The likelihood of the observed microsatellite alleles of an individual is then

$$L_{micro}=(1-p)\prod_{l=1}^{12} g^{W}_{l,a_{2l-1}a_{2l}}+p\prod_{l=1}^{12} g^{E}_{l,a_{2l-1}a_{2l}}\,. \qquad (2)$$

Likelihood contribution from Morphometrics

Within population I and sex *s*, the morphometric measurements are assumed multi-normally distributed with mean depending on population and sex and with covariance matrix only depending on sex or possibly also on population. With the covariance matrix not depending on population, the likelihood of the vector m of the 9 observed measures of an individual whale of sex s is

$$L_{morph}=(2\pi)^{-9/2}\begin{bmatrix}(1-p)\,|\Sigma_s|^{-1/2}\exp\left(-\frac{1}{2}(m-\mu_s^{W})'\Sigma_s^{-1}(m-\mu_s^{W})\right)\\ +p\,|\Sigma_s|^{-1/2}\exp\left(-\frac{1}{2}(m-\mu_s^{E})'\Sigma_s^{-1}(m-\mu_s^{E})\right)\end{bmatrix}. \qquad (3)$$

Joint likelihood

The likelihood of an observation for each individual, conditional on population, is by independence

$$L^{I}=L^{I}_{micro}(\gamma^{I})L^{I}_{morph}(\mu^{I}_{Male},\mu^{I}_{Female},\Sigma_s),\quad I=E,W \qquad (4)$$

where

$$L^I_{micro}(\gamma^I) = \prod_{l=1}^{12} g^I_{l,a_{2l-1}a_{2l}} \qquad (g_{l,ab} = 2\gamma_{l,a}\gamma_{l,b}),$$

$$L^I_{morph}(\mu^I_{Male}, \mu^I_{Female}, \Sigma_s) = (2\pi)^{-9/2} |\Sigma_s|^{-1/2} \exp\left(-\frac{1}{2}\left(m-\mu^W_s\right)'\Sigma_s^{-1}\left(m-\mu^W_s\right)\right)$$

for each of the two populations, $I = E, W$.

The mixing over populations is done at the individual level. Conditioning on "year", "longitude" and "group" (mature male, mature female and immature animals), and other covariates the likelihood for the observed individual is

$$L = p(\alpha_y, \beta)L^E + \left\{1 - p(\alpha_y, \beta)\right\} L^W . \qquad (5)$$

Here, we assume that western areas (Areas III-E and IV-W) and eastern areas (Areas V-E and VI-W) are respectively occupied by baseline populations (I- and P-stocks). Therefore, the genetic and morphometric data in these two areas are used only for estimating parameters for the two putative populations, which are used as known fixed parameters in the joint likelihood function for mixture parameters. A reason of employing this baseline assumption is to provide stabilized baseline populations themselves to reduce the possibility of occasional disturbance by third or even fourth populations visiting our study areas. A sensitivity test was conducted to examine the influence of the baseline assumption by assuming narrower baseline areas (Area III-E and VI-W for the two populations, respectively).

The method is then to maximize the product of all the individual likelihood components. This scheme is carried out by using the software ADMB for various models for mixing.

RESULTS

Table 2 summarizes the information on model specifications and model selection. The number of parameters means the effective mixing parameters estimable by the data. The result indicates that the inclusion of longitudinal covariate tends to improve the model fitting.

In the estimation based only on the genetic data, the best model is Model (M50$1 + Slope$1), which does not possess any "year" covariate. Neither the "year" nor "group" effect was statistically significant. In fact, Figure 6a showed the estimated longitudinal mixing under Model (M50$Year + Slope$1), which indicated less variable mixing patterns across years. However, this might be attributed to a non-significant masked effect by "group". Figure 6b showed "group"-specific yearly variation, which indicated a sharper mixing slope than the combined one and potentially different mixing patterns among groups.

The "year" effect became clearer when the morphometric data were employed (see Table 2). The result indicated that the morphometric data showed dominated information compared to the genetic data and it helped convergence in the optimization. Meanwhile, as shown in Figures 7a and 7b, the inclusion of the morphometric data more or less altered the estimation results and tended to give softer boundaries.

Figures 8 and 9 showed the pattern of yearly changes in the longitude at 50% mixture (M50). The parameter M50 ranged over Areas IV-E to V-W centering around x = 130E to 140E (-50 to -40 in our definition).

Model selection by AIC does not clearly say, but on the whole, the result indicates that the spatial distribution of the two populations has a soft boundary in Area IV-E and V-W, which depends on the year. It also suggested possible sex differences along the boundary.

DISCUSSION

We have introduced our integrated analyses for estimating longitudinal segregation of two populations using genetic and morphometric data. It was demonstrated that the spatial distribution of the two populations has a soft boundary between E and W populations (so-called I- and P-stocks) and it may depend on the year. The results also suggested that the boundary may be sex-specific. Some further covariates including environmental indicators might improve the fitness to the data.

The reason that we integrated these sorts of information for inferring population structure is to use all the available data relevant to the structure. More correctly, we originally thought the morphometric data would work by supplementing the genetic data. However, as shown in our previous paper, for the case of Antarctic minke whales, the likelihood contribution from the morphometric data is much greater than that from the genetic data. This was the case in this updated analysis. This is attributed to the populations being weakly differentiated. The situation of weakly differentiation is very common for fishery populations, and therefore, we believe our methods are of course promising to investigate cetacean population structures but also useful for other species.

As shown in Figures 6 and 7, the estimated smooth mixture functions may not be compatible with our baseline assumption (Areas III-E and IV-W for I-stock and Areas V-E and VI-W for P-stock). To see the impact of the baseline assumption on the estimation result, a sensitivity test with a different baseline assumption (Areas III-E and VI-W are baseline line areas for I- and P-stock, respectively) was conducted with both the genetic and morphometric data (see Figure 10). Some difference in the estimation results was observed, which means that a slight sensitivity to the baseline assumption may occur. It is hard to confirm that the assumed baseline is the true baseline or not, but more reliable analysis with some more valid baseline assumptions may become possible if the samples outside the JARPA and JARPAII areas are available. This warrants further investigation with a broader longitudinal range of data.

For handling morphometric data of immature animals, a model expressing allometric change along with growth can be developed (see Figure 11). This model has three parameters; a mean of allometric quantity in the mature state, a logarithm of length at which the shift of phases of allometic relationship between immature and mature occurs, and a slope parameter expressing allometric changes in the immature state. We noted that the second parameter relevant to the shift in phase is shared with all the (nine) measurements, which means the parameter is assumed to be common to the measurements:

$$E[m_j] = \begin{cases} a_j + b_j(\log v_1 - c) & \text{for} \quad \log v_1 < c \\ a_j & \text{for} \quad \log v_1 \geq c \end{cases} \tag{6}$$

The model considers a change in allometric relation between immature and mature stages. The allometric model is originally derived by the famous formulas;

$$\begin{aligned} v_j &= \alpha_{\mathrm{Imm}} v_1^{\beta_{\mathrm{Imm}}} \qquad \text{for} \quad \text{immature animal} \\ v_j &= \alpha_{\mathrm{Mat}} v_1^{\beta_{\mathrm{Mat}}} \qquad \text{for} \quad \text{mature animal} \end{aligned} \tag{7}$$

with the assumption of $\beta_{\mathrm{Mat}} = 1$, for which no allometric change is supposed to happen in the mature state, and the connectivity at the change point. The parameter β_{Mat} might not necessarily be 1, so we will also be able investigate the following general case:

$$E[m_j] = \begin{cases} a_j + b_{\mathrm{Imm},j}(\log v_1 - c) & \text{for} \quad \text{immature animal} \\ a_j + b_{\mathrm{Mat},j}(\log v_1 - c) & \text{for} \quad \text{mature animal} \end{cases} \tag{8}$$

Furthermore, the change point is not necessarily common to the measurements but may differ between spine and cranial bones, and hence this is also to be tested in the model. Unfortunately, this investigation has not been completed yet because of missed communication in the data set between the data owner and analysts, but this

modelling is highly worth conducting.

As mentioned earlier, microsatellites could not be read for some individuals at some loci. We investigated several models for such drop out. The question of whether this is caused by allelic drop out is of specific interested. The presence of allelic drop out would weaken the information content in the recorded microsatellite data, and thus affect its likelihood function. If also the probability of allelic drop out depends on population, the estimated probabilities of population assignment would be biased. The size of such biases could potentially be investigated by simulation, something which is not done here. Under individual and locus drop out, i.e. all loci dropping out, Hardy-Weinberg equilibrium can be tested by contrasting the number of homozygotes to its expected value under the null hypothesis of equilibrium. Consider one locus l. The number of homozygotes recorded for this locus is T_l. The estimated probability for a given individual to be homozygous is the mixture of the estimated population specific probabilities of homozygotisity for an individual i, $q_i = \sum(\hat{p}_i \hat{g}_{aa}^E + (1-\hat{p}_i)\hat{g}_{aa}^W)$. Individuals are independent with respect to homozygosity, and $E(T_l) = \sum q_i$ and $\mathrm{var}(T_l) = \sum q_i(1-q_i)$. There is also independence between loci. The mean and variance under the null hypothesis of the test statistic $T = \sum T_l$ are thus obtained by summing the locus specific quantities. The normal approximation should be excellent in our large data set, and Hardy-Weinberg equilibrium is tested by homozygote excess in the usual way. This test might be regarded as a conditional test since the data for estimating the null distribution are ancillary to T. This kind of exercise is also under way.

Response to recommendations

The Review Panel suggested some short-term recommendation for improving this paper:

(1) *notation in the paper should be improved to avoid using the same index for individual and stock;*

[We slightly changed our notation so that the readers are not confused.]

(2) *the estimates of the parameters and their standard errors should be reported – this information could be used to identify which of the morphometric measurements is most informative regarding stock mixing;*

[In our estimation, a conditional likelihood approach given the parameters of the baseline populations was used, and therefore the standard errors of morphometric measurements are not useful to identify which of the morphometric measurements is most informative regarding stock mixing. Instead of reporting the standard errors, we showed boxplots of the data to compare the difference in the morphometric data between baseline populations (Fig A1). We also exhibited some estimated parameter values with their SEs (Table A1). Unfortunately, the sex-specific mixture patters were less precisely estimated at this stage, but it could be improved by incorporating the random effects across years. See also Fig A2.]

(3) *diagnostic statistics for the fits to the morphometric data should be presented;*

[This was not conducted in the current model, but diagnostic plots in previous analyses were reasonably well (Kitakado et al. 2012; SC/64/IA4).]

(4) *the meaning of the statement: 'The variables G and M are further assumed independent within an individual, and also between individuals' should be clarified – it is not clear whether this means that matrix G is assumed independent of matrix M, or that the different elements of M are assumed to be independent (i.e. is the Σ diagonal?);*

[We changed the sentence as “The two different observations, *G* and *M*, are further assumed independent within an individual, and also a set of observations, *G* and *M*, are assumed to be independent between individuals”. Note that the covariance matrix Σ is not necessarily diagonal.]

(5) the parameters for each model should be documented to clarify how many of the parameters of the mean function and variance-covariance matrix (morphometric data) as well as the baseline allele frequencies (genetic data) are estimated parameters; and

[We added Table A2 for showing the number of parameters of the baseline allele frequencies and those of the mean and variance-covariance matrix for the morphometric measurements. The number of parameters in the mixture was originally shown in Table 2.]

(6) the discussion should explore the fact that one of the main results of this study—that morphometric data had a stronger influence on results than the genetic data—differs from that found in many other such studies, and the authors should suggest a possible biological explanation for this result.

[Our initial consideration was that the Antarctic minke whales are highly migrated species and so weakly differentiated in terms of genetics, which might be a reason that the genetic data are relatively less informative compared to the morphometric data. However, we agreed that further convincing explanation must be required in addition to the statistical works.]

The Review Panel also kindly suggested the following longer-term recommendation especially for improving the estimation precision for mixture parameters.

(1) the model be formulated as a random effects model (Bayesian or maximum likelihood). This may eliminate some of the problems associated with lack of convergence for some of the more complicated models – this might also reduce some of the large inter-annual fluctuations in mixing proportions, which a plot produced during the Workshop (Fig. 3) shows are generally very imprecise;
(2) more flexible functions for the relationship between longitude and proportion should be considered; and

(3) the benefits of applying the integrated model to all data (i.e., data aggregated by sex and maturity state) should be re-evaluated since, for example, there are clear between-sex differences in some of the morphometric measurements such as V7 (fig. 3 of SC/F14/J29) - a model with sex-specific values for the mean and variance functions for the morphometric data, but sex-independent distribution proportions should be explored’.

[We have estimated sex-specific parameters for genetic allele frequencies and morphometric mean and covariance matrix for each baseline population.]

The authors have also identified usefulness of incorporation of random effects models to expect a shrinkage effect on the magnitude of variability in the yearly changes of mixing pattern, which could also contribute to improve the precision of estimation. The authors therefore consider implementing the recommendation above with high priority.

REFERENCES

Bravington, M.V. and Hedley, S.L. (2012) Abundance estimates of Antarctic minke whales from the IWC IDCR/SOWER surveys, 1986-2002. Paper SC/64/IA13

Hakamada, T. (2006) Morphometric analysis on stock structure in the Antarctic minke whale based on JARPA samples. SC/D06/J10

Hakamada, T. and Matsuoka, K. (2014) Estimates of abundance and abundance trend of the Antarctic minke whale in Areas IIIE-VIW, south of 60S, based on JARPA and JARPAII sighting data (1989/90-2008/09). SC/F14/J3

International Whaling Commission. (2013) Report of the Scientific Committee. Annex G. Report of the Sub-Committee on In-Depth Assessments. J. Cetacean Res. Manage. (Suppl.) 14:195-213.

Kanda, N, Goto, M, Oikawa, H. and Pastene, L.A. (2014) A note on sampling and laboratory procedure protocols of the genetic work at the Institute of Cetacean Research. SC/F14/J27

Kitakado, T., Schweder, T., Kanda, N., Pastene, L.A. and Walløe, L. (2012) Progress report on the estimation of longitudinal mixing proportions for the Antarctic minke whales using genetic and morphometric measurements. SC/64/IA4

Okamura, H. and Kitakado, T. (2012) Abundance estimates of Antarctic minke whales using the OK method. Paper SC/64/IA2

Pastene. LA. (2006) What do we know about the stock structure of the Antarctic minke whale? A summary of studies and hypotheses. SC/D06/J12.

Pastene, LA., Goto, M. and Kanda, N. (2006) Genetic analysis on stock structure in the Antarctic minke whales from the JARPA research area based on mitochondrial DNA and microsatellites, SC/D06/J9.

Schweder, T., Kitakado, T., Kanda, N., Pastene, L.A. and Walløe, L. (2011) Dynamic population segregation by genetics and morphometrics in Antarctic minke whales. SC/63/IA7.

Table 1. Sample size by year and area.

Year	Mature male						Mature female						Immature animals					
	III-E	IV-W	IV-E	V-W	V-E	VI-W	III-E	IV-W	IV-E	V-W	V-E	VI-W	III-E	IV-W	IV-E	V-W	V-E	VI-W
1989/90		114	42					45	40					63	22			
1990/91				99	48					57	74					31	14	
1991/92		96	32					61	23					53	23			
1992/93				105	45					66	57					34	20	
1993/94		72	77					38	31					47	65			
1994/95				97	60					14	73					29	57	
1995/96	47	130	30				29	65	11				33	62	33			
1996/97				38	70	52				58	113	22				31	34	22
1997/98	60	88	52				19	23	5				32	85	74			
1998/99				83	92	26				62	18					38	35	35
1999/00	39	47	81				11	28	73				59	25	76			
2000/01				71	71	49				26	85	17				43	34	44
2001/02	43	53	50				31	43	76				36	58	50			
2002/03				48	83	45				36	84	11				16	64	53
2003/04	53	59	33				27	106	36				30	45	51			
2004/05				33	58	52				43	143	18				6	47	40
2005/06	48	216	27	73			37	153	13	38	1		45	130	34	37	1	
2006/07					74	51					272	16					58	34
2007/08	107	54	6	28			32	95	3	52			90	73	4	7		
2008/09				55	81	157				31	130	39				48	39	99
2009/10	81	33		76	2		113	49	0	37			51	21		35	8	
2010/11					16	25					10	84					20	15
Total	478	962	430	806	700	457	299	706	311	520	1060	207	376	662	432	355	431	342

Table 2. Description and comparison of models employed in this study. The highlighted models are not the best model in terms of AIC, but those are used for the graphical presentation of results in Figures 6 to 9. The number of parameters is that only for estimable parameters in the mixing proportions. Convergence was not reached in some cases due to an inadequate gradient and/or degenerated hessian matrix.

Data			Convergence	negloglike	Delta-negloglike	#para	AIC	Delta-AIC
Genetic	Area							
	1		OK	294734	40	2	589472	6
	Group		OK	294733	39	6	589478	12
	Year		OK	294724	30	21	589490	24
	Year*Group		Grad OK. Hess No	294711	17	66	589554	88
Genetic	M50	Slope						
	1	1	OK	294731	37	2	589466	0
	Group	1	OK	294731	37	4	589470	4
	Group	Group	OK	294730	36	6	589472	6
	Year	1	OK	294720	26	21	589482	16
	Year	Group	OK	294715	21	25	589480	14
	Year	Year	Hess OK, Grad No	294713	19	44	589514	48
	Year*Group	1	NO	294694	0	67	589522	56
	Year*Group	SM	NO	294715	21	69	589568	102
	Year*Group	Year*Group	NO	294732	38	132	589728	262
Both	Area							
	1		OK	98465.8	55.7	2	115.4	26
	Group		OK	98463	52.9	6	117.8	28.4
	Year		OK	98439.8	29.7	21	101.4	12
	Year*Group		OK	98416.3	6.2	66	144.4	55
Both	M50	Slope						
	1	1	OK	98455.3	45.2	2	94.4	5
	Group	1	OK	98453.9	43.8	4	95.6	6.2
	Group	Group	OK	98452.8	42.7	6	97.4	8
	Year	1	OK	98432.1	22	23	90	0.6
	Year	Group	OK	98429.8	19.7	25	89.4	0
	Year	Year	Hess OK, Grad No	98428.7	18.6	44	125.2	35.8
	Year*Group	1	OK	98410.1	0	67	134	44.6
	Year*Group	SM	No	98430.2	20.1	69	178.2	88.8
	Year*Group	Year*Group	No	98449.8	39.7	132	343.4	254

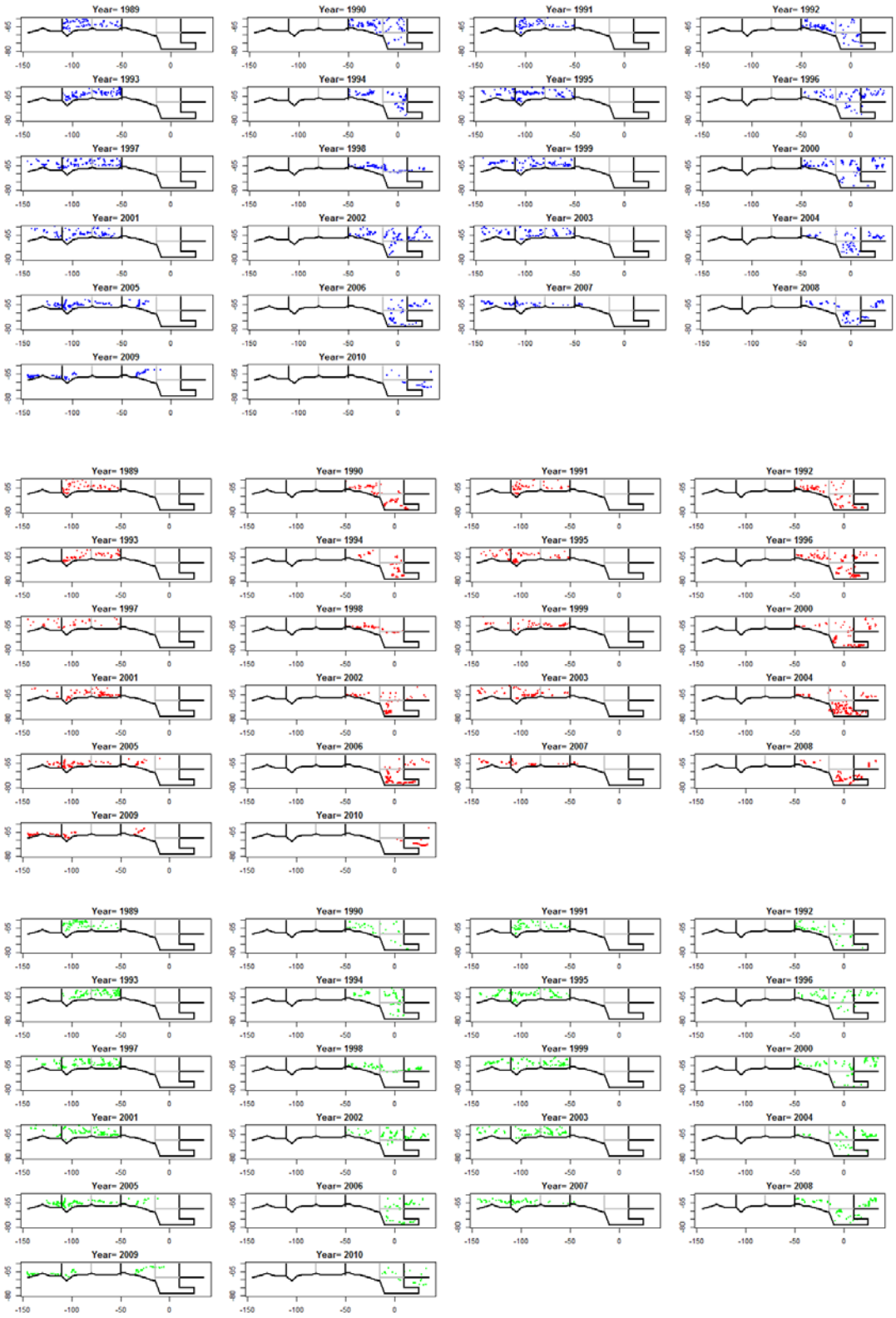
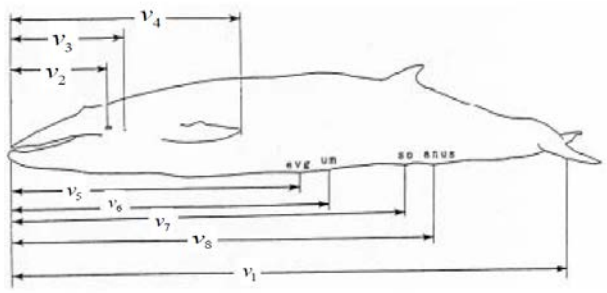


Figure 1. Locations of sampling by year. Blue, red and green dots show mature males, mature females and immature animals. The values on the horizontal axis showed the longitude (e.g. -100, -50 and 0 respectively mean 80E, 130E and 180E).



v1: Body length
v2: from the tip of snout to center of eye
v3: from the tip of snout to ear
v4: from the tip of snout to tip of flipper
v5: from the tip of snout to end of ventral gloves
v6: from the tip of snout to center of umbilicus
v7: from the tip of snout to sexual apparatus
v8: from the tip of snout to anus
v9: length of skull
v10: width of skull

Figure 2. Morphometric measurements for Antarctic minke whales used in this study. These measurements other than v_1 are transformed to the logarithms of allometric measures as $m_i = \log(v_{i+1} / v_1)$.

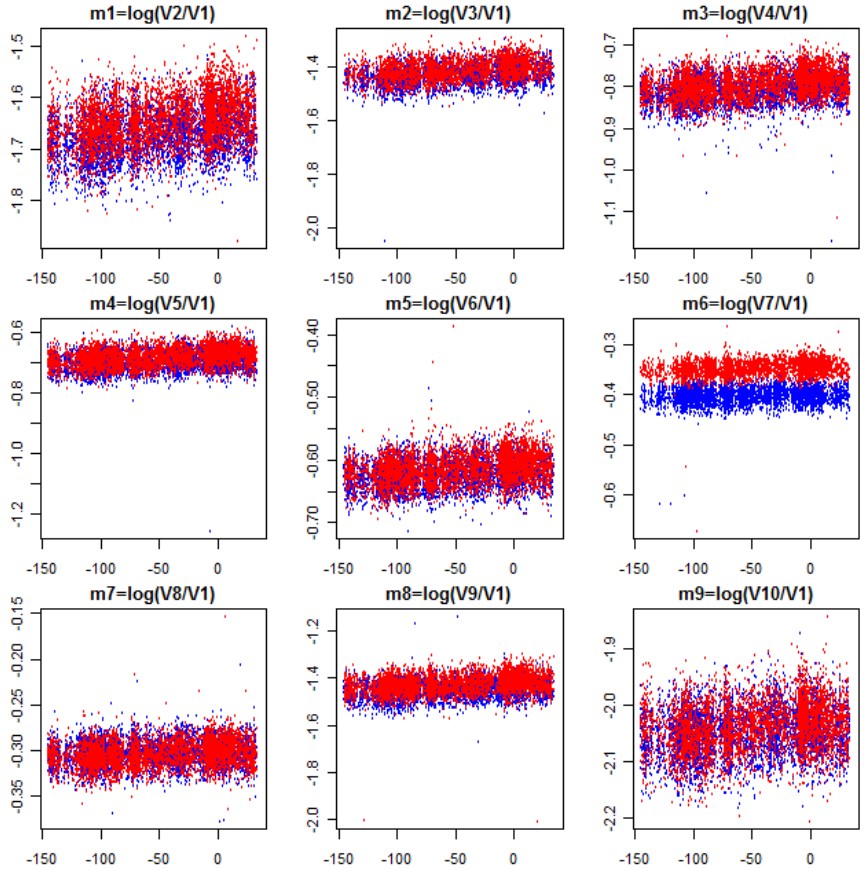
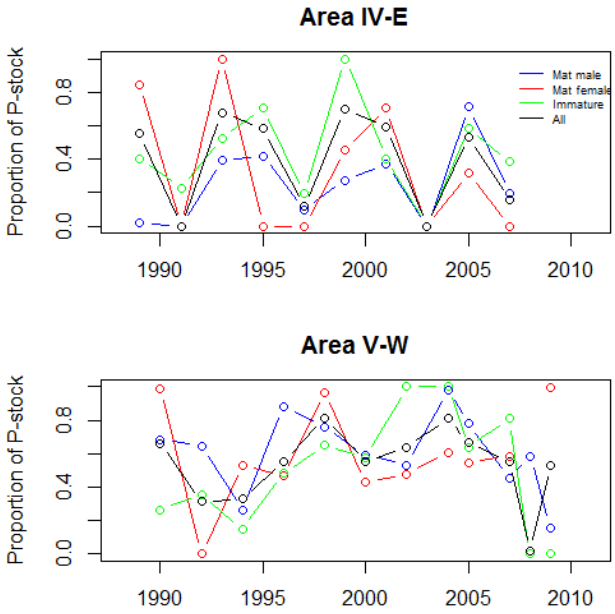


Figure 3. Morphometric data against the longitude where samples were taken (The blue and red dots show mature male and mature female individulas, respectively).

(a) Based only on the genetic data



(b) Based on the genetic and morphometric data

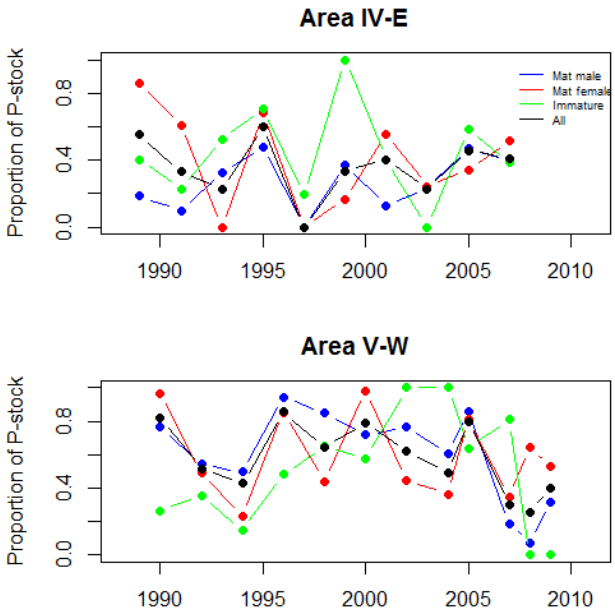


Figure 4. Yearly variation of mixing proportion by Area and Group (blue=mature males, red=mature females, green=immature animals, black=all animals).

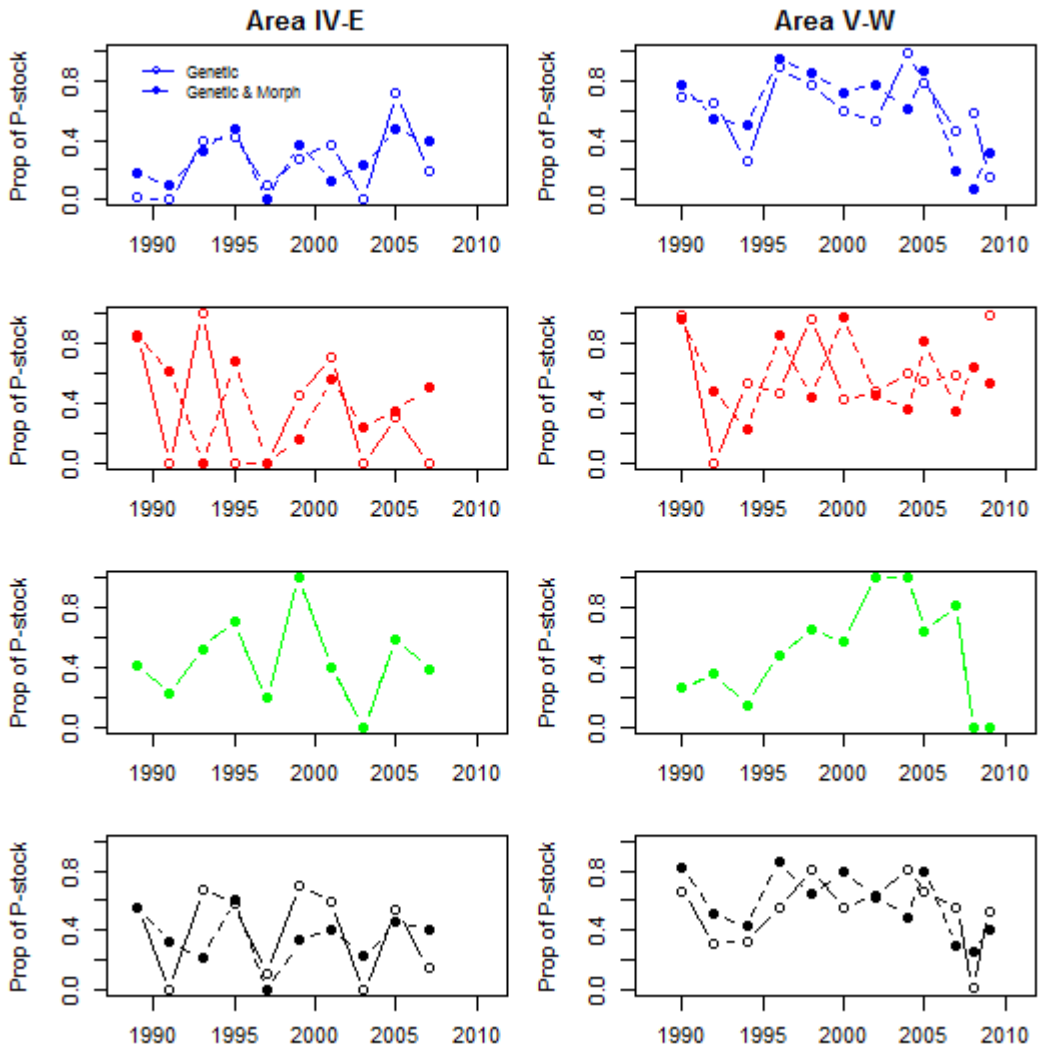


Figure 5. Comparison of area-wise mixing proportions estimated by the genetic data and by the genetic and morphometric data (genetic data only=solid line with open circles; genetic and morphometric data=shaded line with filled circles) for mature males and females as well as immature animals and all individuals ((blue=mature males, red=mature females, green=immature animals, black=all animals).

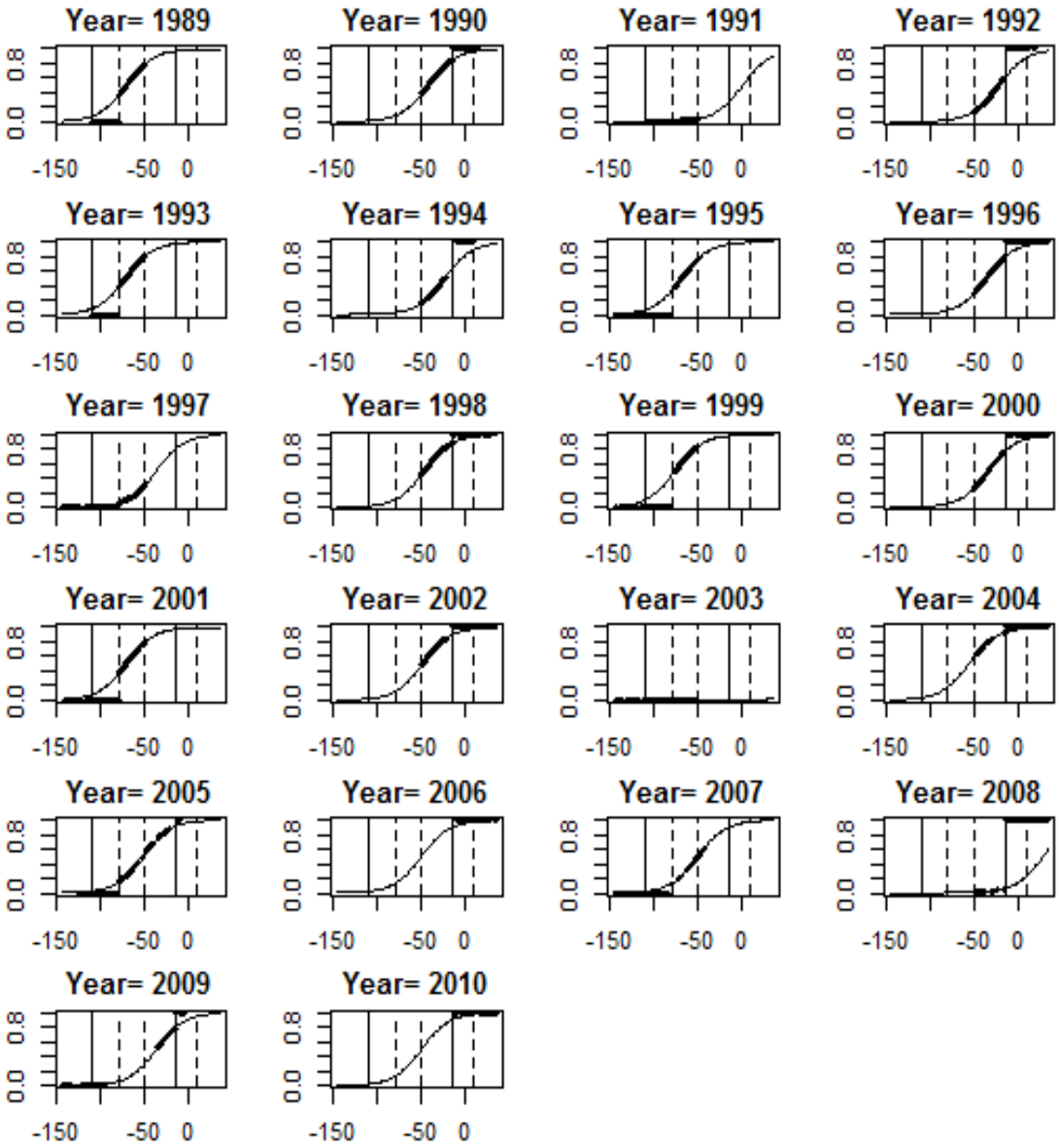


Figure 6a. Estimated mixing proportions against longitude for the Eastern population (P-stock) based only on the genetic data under Model (M50$Year + Slope$1) that the intercepts in the mixing proportions differ across years while the slopes are common to years and groups. Circles are estimated mixing proportions where samples were taken.

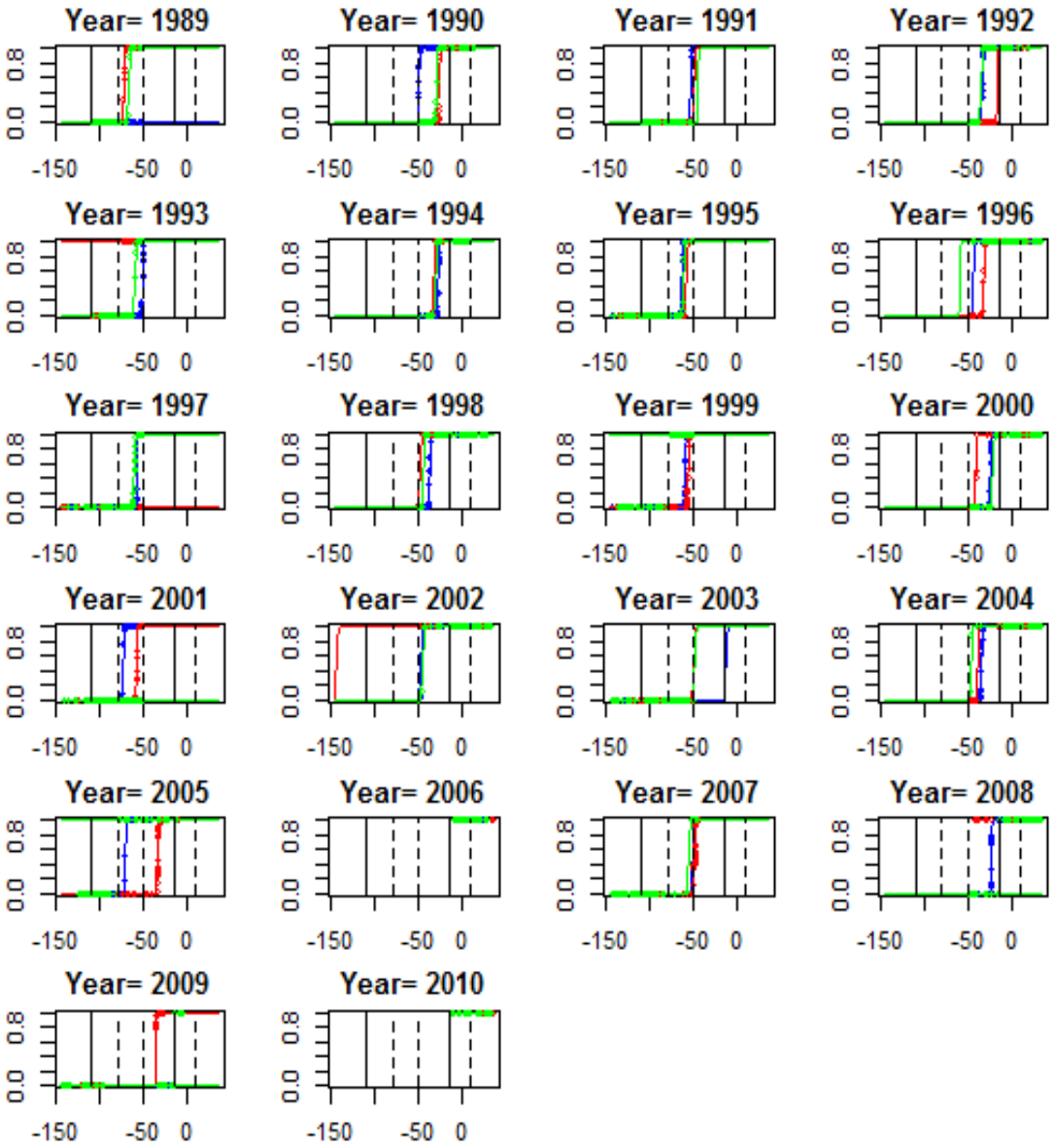


Figure 6b. Estimated mixing proportions against longitude for the Eastern population (P-stock) based only on the genetic data under Model (M50$Year*Group + Slope$1) that the intercepts in the mixing proportions differ across years and groups while the slopes are common to years and groups. The blue, red, and green lines are for mature male, mature female and immature animals, respectively. Circles are estimated mixing proportions where samples were taken.

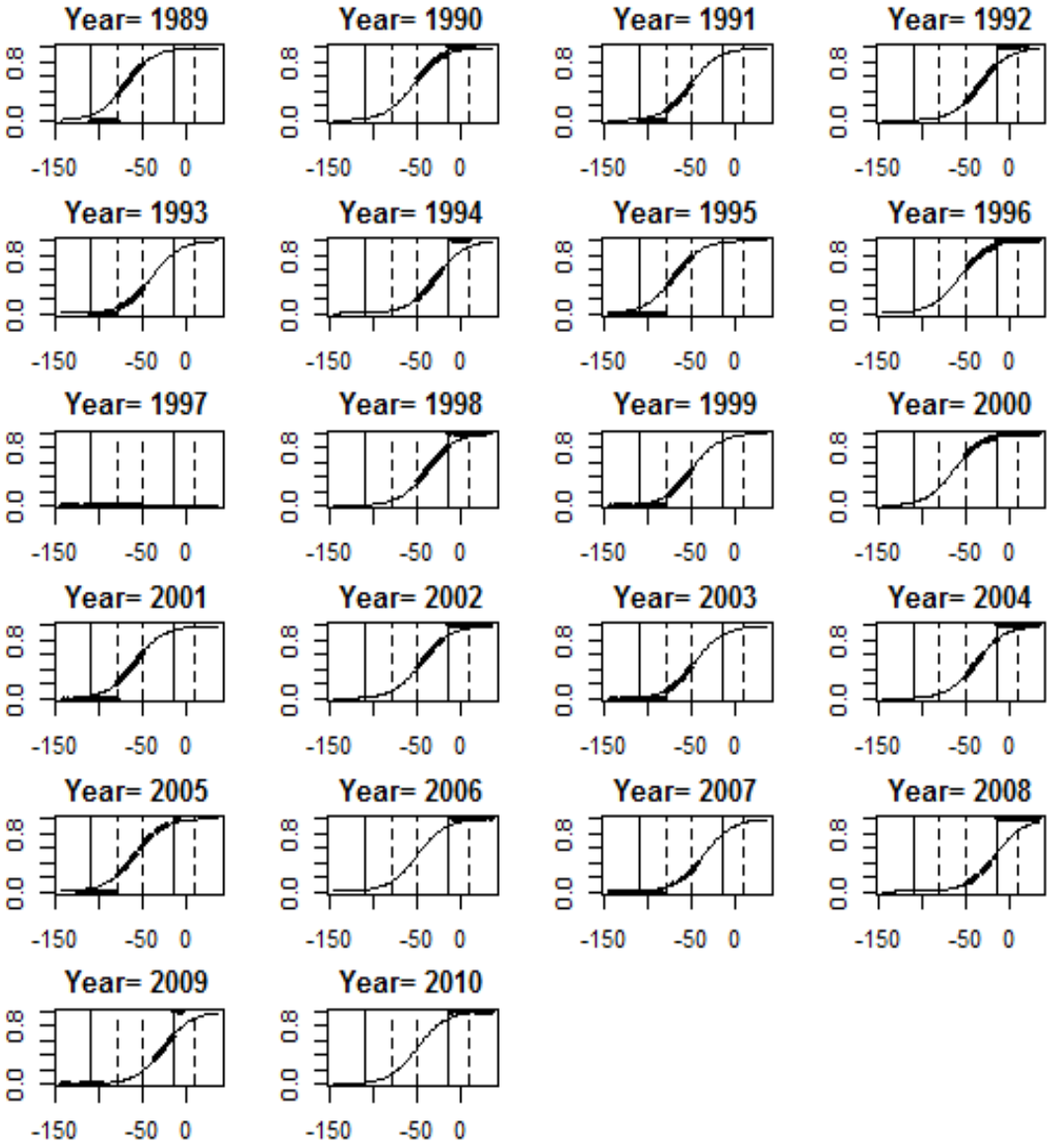


Figure 7a. Estimated mixing proportions against longitude for the Eastern population (P-stock) based on the genetic and morphometric data under Model (M50$Year + Slope$1) that the intercepts in the mixing proportions differ across years while the slopes are common to years and groups. Circles are estimated mixing proportions where samples were taken.

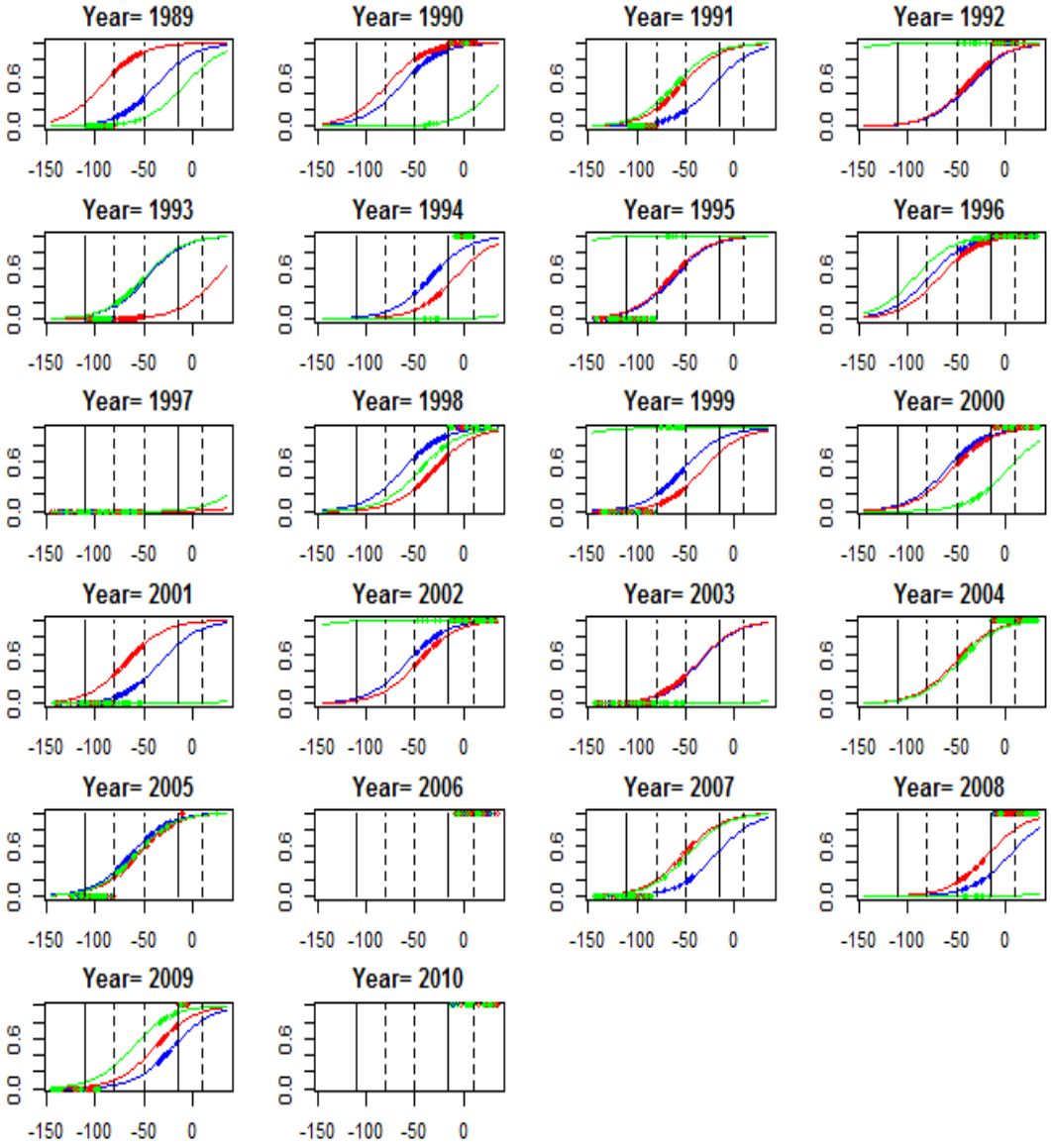


Figure 7b. Estimated mixing proportions against longitude for the Eastern population (P-stock) based on both the genetic and morphometric data under Model (M50$Year*Group + Slope$1) that the intercepts in the mixing proportions differ across years and groups while the slopes are common to years and groups. The blue, red, and green lines are for mature male, mature female and immature animals, respectively. Circles are estimated mixing proportions where samples were taken.

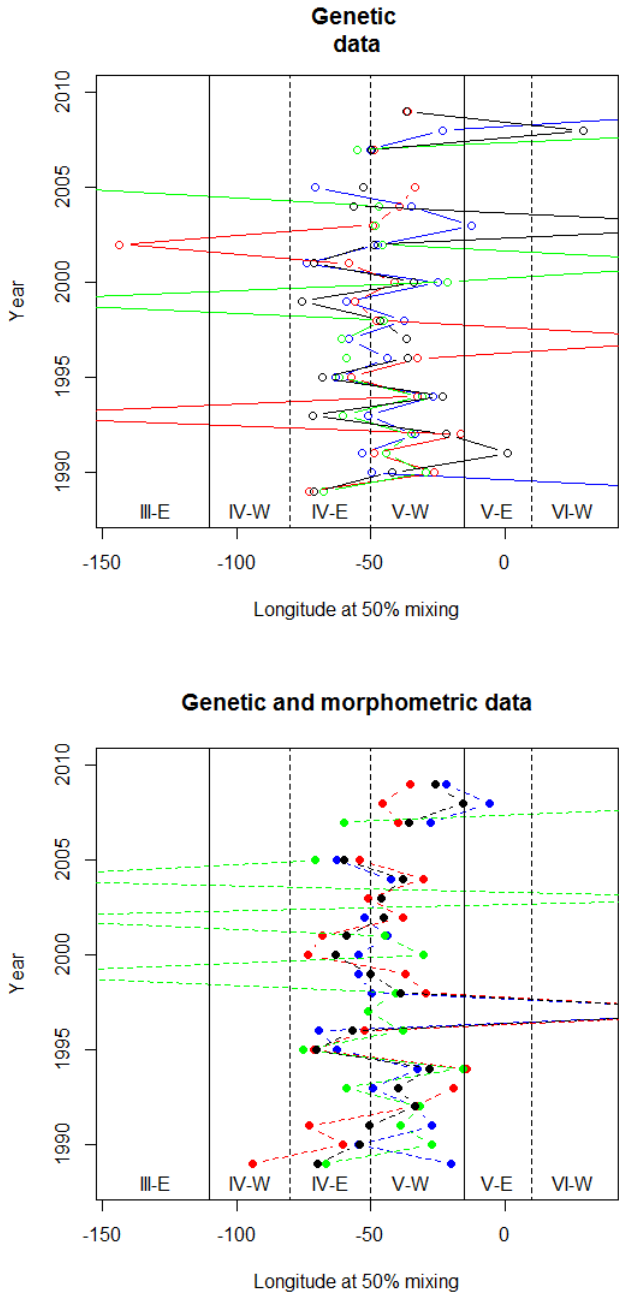


Figure 8. Yearly variation of longitudes at which 50% mixtures occur. The assumed model is Model (M50$Year*Group + Slope$1). The blue, red, and green dots and lines are for mature male, mature female and immature animals, respectively. The black ones are the estimates from Model (M50$Year + Slope$1).

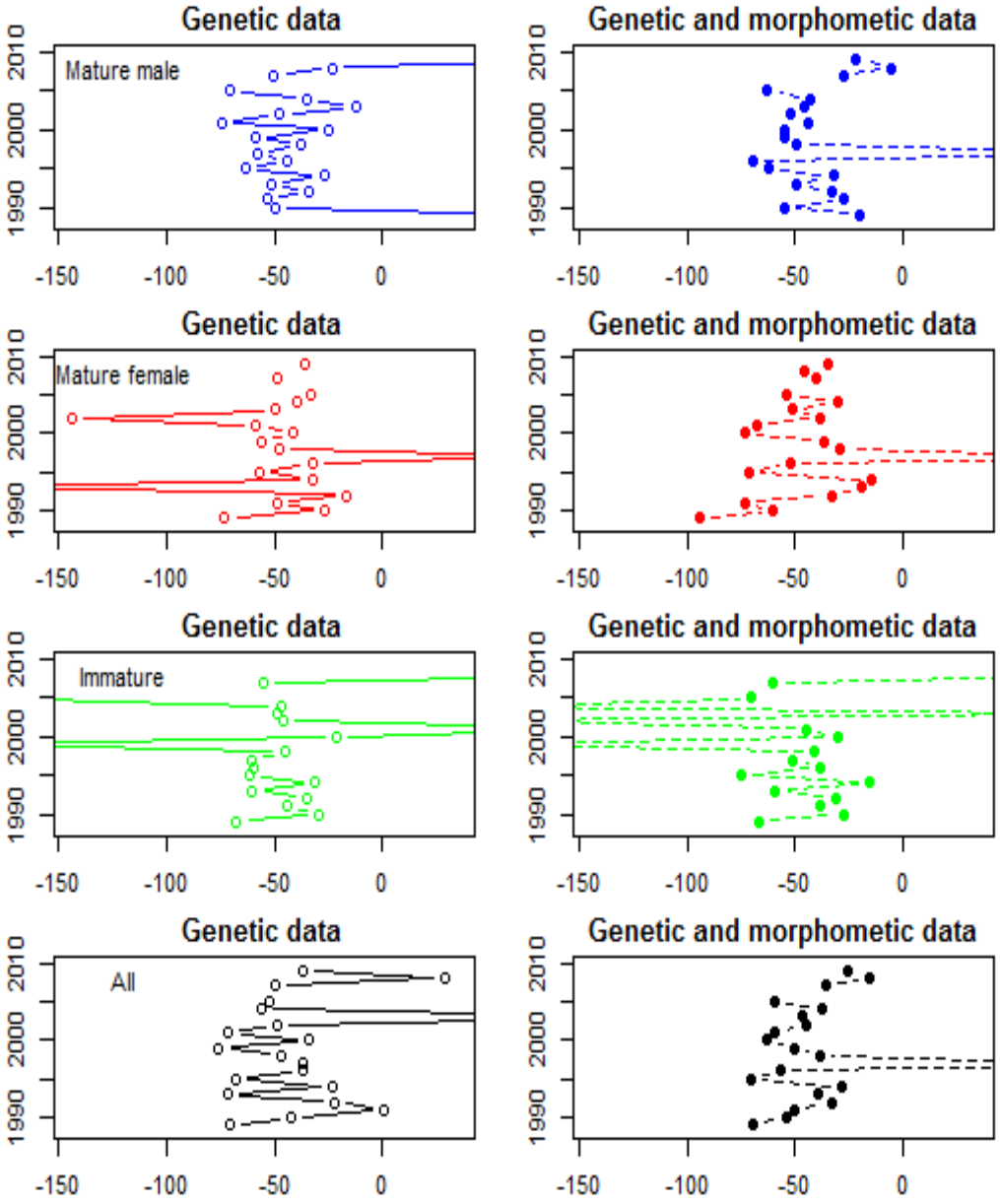


Figure 9. Yearly variation of longitudes at which 50% mixtures occur by group. The assumed models are same as in the Figure 8.

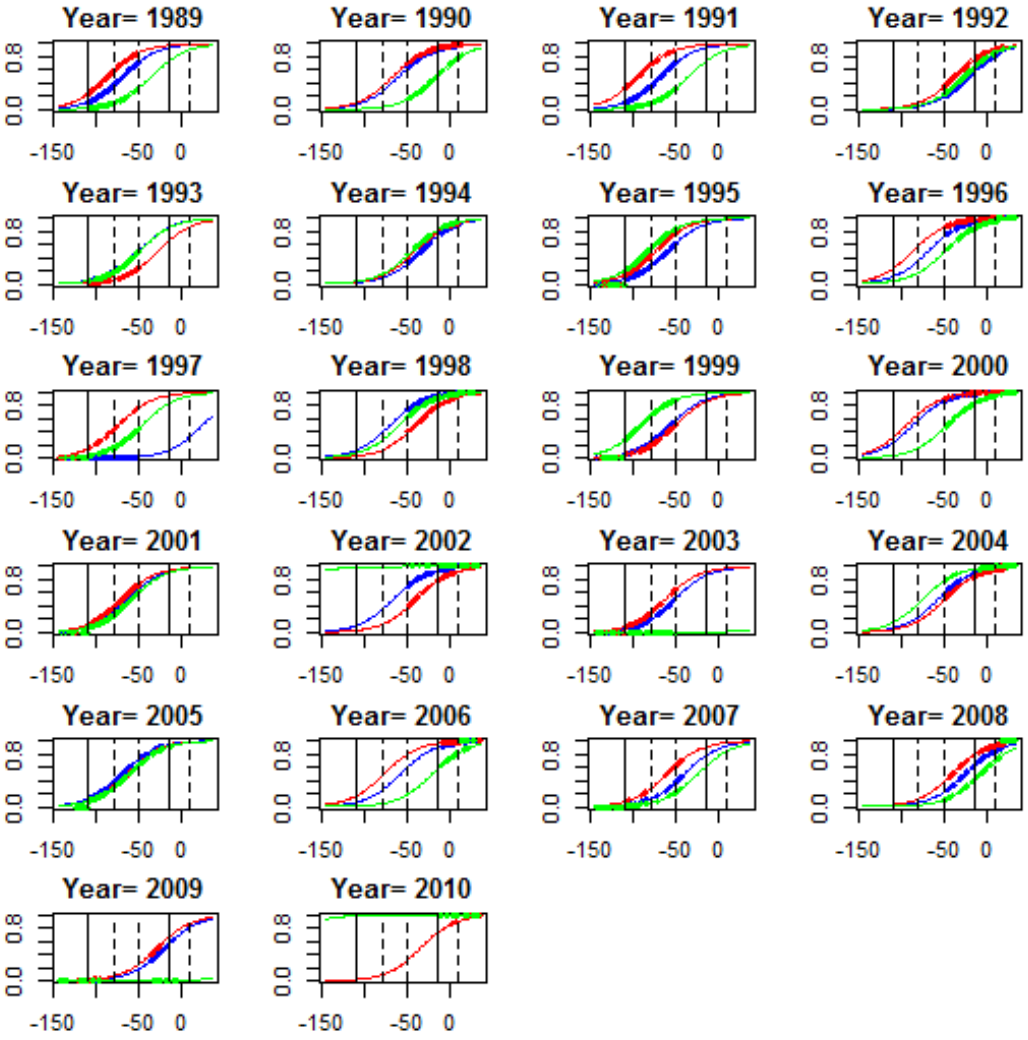


Figure 10. Estimated mixing proportions against longitude for the Eastern population (P-stock) based on both the genetic and morphometric data under Model (M50$Year*Group + Slope$1) with a different baseline assumption (III-E and VI-W are respective baseline populations).

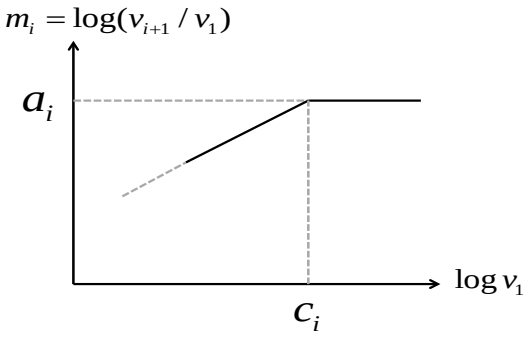


Figure11. An assumed allometric change in the immature period. (*c*'s are common to measurements)

Table A1. Parameter estimates and their standard errors for the parameter M50, longitude at which 50% mixtures occurs. See also Fig A2.

(a) Case with year-specific M50.

	Genetic data only		Genetic and morphometric data	
Year	Estimate	SE	Estimate	SE
1989	-71.03	18.92	-69.94	10.18
1990	-42.18	14.01	-54.41	10.46
1991	0.92	180.54	-50.62	12.88
1992	-22.10	16.41	-33.45	6.26
1993	-71.76	16.87	-39.86	11.46
1994	-23.19	16.84	-28.17	7.39
1995	-68.09	17.78	-70.38	13.47
1996	-36.39	14.10	-57.02	14.68
1997	-36.87	25.74	100.00	7.52
1998	-46.57	17.95	-38.76	7.55
1999	-75.70	17.57	-50.21	7.71
2000	-33.82	14.57	-63.29	17.55
2001	-71.32	15.43	-59.08	7.25
2002	-48.77	25.10	-45.28	10.14
2003	100.00	5.67	-46.30	11.42
2004	-56.34	23.06	-38.05	9.64
2005	-52.77	12.03	-59.84	8.50
2006	-50.00	#	-50.00	#
2007	-49.75	9.79	-35.61	8.56
2008	28.91	184.57	-15.79	13.35
2009	-36.79	13.14	-26.18	7.77
2010	-50.00	#	-50.00	#

(b) Case with year- and sex- specific M50 (convergence was not achieved based only on genetic data).

Mature male		Mature female		Immature animals	
Estimate	SE	Estimate	SE	Estimate	SE
-20.12	75.21	-94.23	21.53	-66.78	29.43
-54.81	13.09	-60.34	21.31	-27.30	19.04
-27.19	45.83	-73.19	15.76	-38.95	33.16
-33.46	7.42	-33.35	8.96	-31.58	27.09
-49.25	13.09	-19.12	37.27	-59.19	19.05
-32.47	7.17	-14.35	17.32	-15.49	43.50
-62.54	18.55	-71.43	25.68	-75.26	18.91
-69.45	58.13	-52.34	15.21	-38.27	29.78
100.00	16.38	100.00	30.49	-50.91	19.66
-49.64	15.56	-29.64	8.69	-40.81	28.04
-54.72	9.25	-36.96	15.20	-200.00	6.77
-54.76	20.56	-73.53	32.00	-30.61	28.38
-44.08	17.55	-67.92	8.00	-44.98	51.24
-52.47	19.20	-38.12	11.34	-199.99	130.85
-46.19	21.03	-51.25	11.40	99.98	311.51
-42.76	13.81	-30.53	12.89	-200.00	97.35
-62.77	12.65	-54.41	10.91	-70.55	26.66
-50.00	#	-50.00	#	-50.00	#
-27.69	15.07	-40.05	8.84	-60.18	29.22
-5.72	20.41	-45.50	14.60	100.00	3.50
-21.84	9.84	-35.30	10.53	99.99	155.45
-50.00	#	-50.00	#	-50.00	#

Table A2. The number of parameters of the baseline allele frequencies and those of the mean and variance-covariance matrix for the morphometric measurements.

Genetic parameters	
Locus	#alleles in each population
EV1	23
EV104	17
GT023	24
GT211	11
GT195	25
DlrFCB14	23
AC045	24
AC082	22
AC087	18
AC137	32
CA234	21
GT219	22

Morphometic parameters	
Parameter	Dimension of parameters
Mean vector (Male, Indian population)	9
Mean vector (Female, Indian population)	9
Mean vector (Male, Pacific population)	9
Mean vector (Female, Pacific population)	9
Variance-covariance matrix (Male)	45
Variance-covariance matrix (Female)	45

Note: We assumed that the two population (I and P) shaere the same sex-specific variance-covariance matrix.

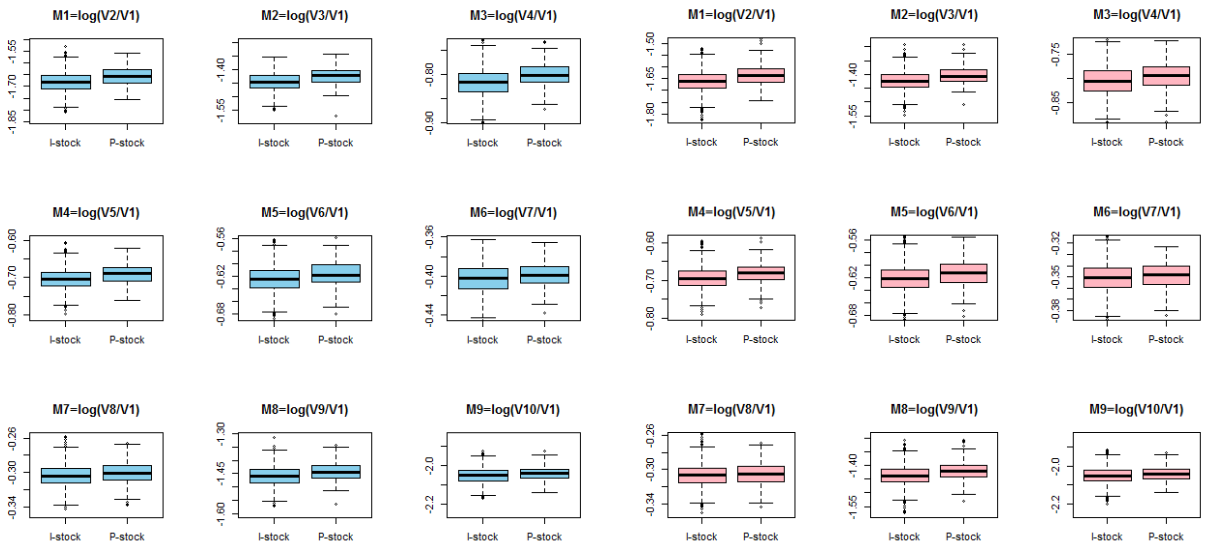


Fig A1. Boxplots of the data to compare the difference in the morphometric data between baseline populations (left = mature male; right = mature female).

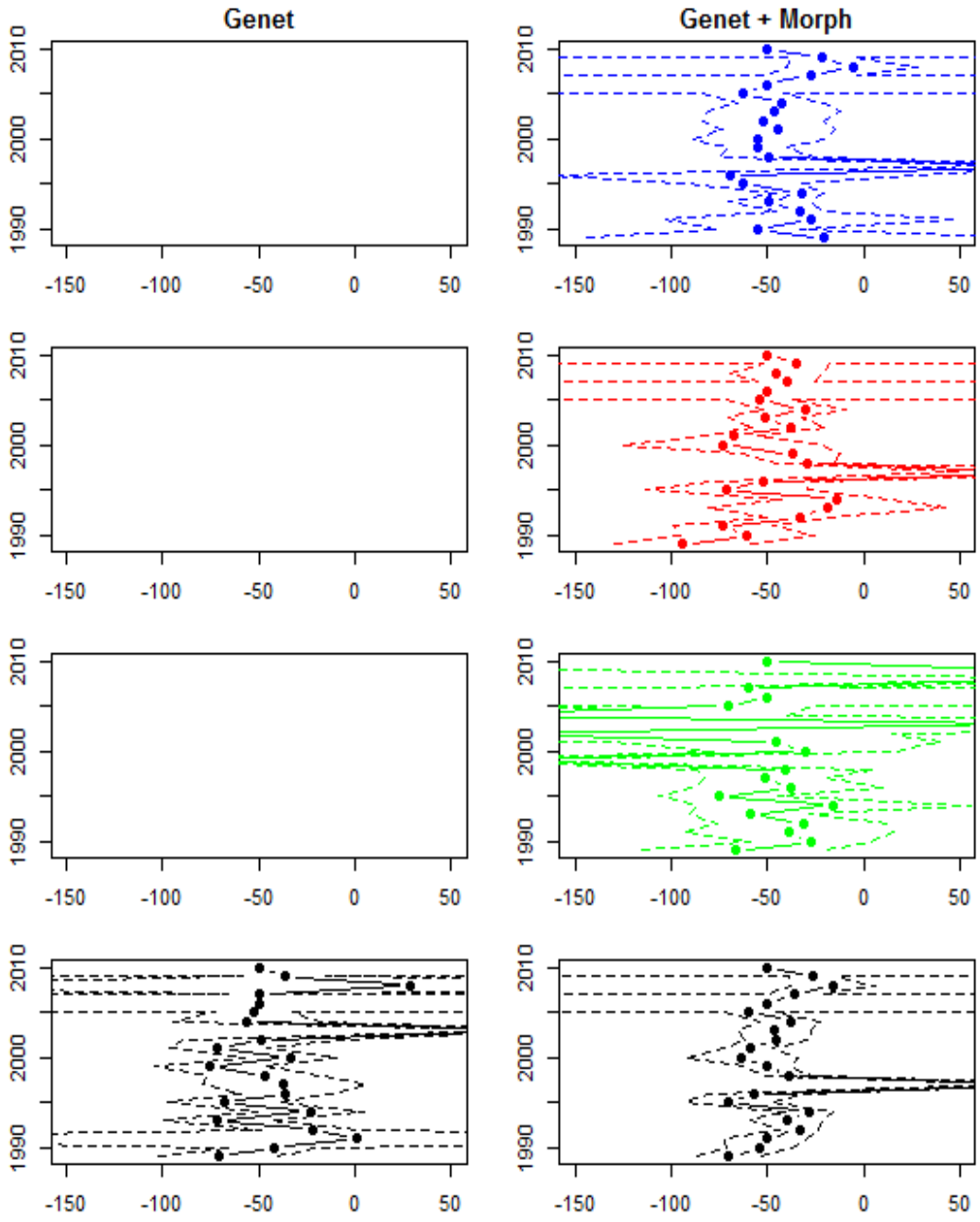


Figure A2 Point estimate and 90% confidence interval for the estimation of M50, **"logistic mixing proportion with longitudinal covariate",** based on **"only genetic data"** and "**genetic and morphometric data"**.