Are reported trends in Antarctic minke whale body condition reliable?

WILLIAM K. de la MARE
Australian Antarctic Division, Channel Highway, Kingston, Tasmania. Australia, 7050.
Contact email: bill.delamare@aad.gov.au

ABSTRACT

Information on body condition of Antarctic minke whales collected under Japan's Special Permit Whaling Program (known as JARPA) were analysed using a particular form of multiple regression model. Using this model Konishi *et al.* (2008) report statistically significant declines in blubber thickness and other indicators of body condition over the 18 years of JARPA. However, these trends rely on the assumption that the particular multiple regression model used was appropriate. Several features of the regression model are biologically unlikely. This paper investigates whether the reported trends in body condition could have arisen from an inappropriate choice of statistical model. Biologically plausible models were used to generate simulated data consistent with the reported properties from JARPA, but which have no time trend in blubber thickness. Applying the multiple regression model used in the JARPA analyses to simulated data produced spurious and apparently statistically significant trends in blubber thickness. Further analyses show that features of the realised JARPA sampling design may preclude reliable inferences about trends in the body condition of Antarctic minke whales even if more biologically plausible models were to be fitted to the data.

KEYWORDS: ANTARCTICA, MINKE WHALES, BLUBBER, LONG-TERM CHANGE, ECOLOGICAL MONITORING, WHALING, SIMULATION

INTRODUCTION

For 18 years, starting in 1987, Japan conducted the Japanese Whale Research Program under Special Permit in the Antarctic (JARPA). One of the objectives of JARPA was "Elucidation of the effect of environmental changes on cetaceans" (GOJ, 1995). In 2008 Konishi *et al.* published an analysis of trends in body condition of Antarctic minke whales from specimens taken over the 18 years of the program. They concluded that "regression analyses clearly showed that blubber thickness, girth and fat weight have been decreasing for nearly two decades". However, that conclusion rests on the assumption that the regression model fitted to the data was appropriate.

A regression model is needed for analysing the JARPA data not only for statistical inference, but to correct for the heterogeneous manner in which the data were collected. Because the whales are taken by a single operation, they are taken on different dates at different places throughout the whaling season. The effect of date is particularly important because body condition improves throughout the whaling season because the whales are taken on their feeding grounds. Thus a statistical model is needed to account for the effects of date and location when making inferences about possible year trends.

I examine whether the choice of the particular multiple linear regression model used by Konishi *et al.* is appropriate considering the basic biology of body condition, and whether the model could lead to biased or spurious estimates of time trends in body condition through a failure to properly correct for the effects of sampling heterogeneity. Konishi *et al.* used blubber thickness, half girth and fat weight as body condition indicators. The analyses here will consider only blubber thickness from mature males. The results from Konishi *et al.* for males are similar to those for pregnant females and both sexes combined. This one case is representative of the potential problems of all the Konishi *et al.* analyses because they all use the same basic form of linear model to correct for heterogeneity and there is only one realised "sampling design" for all the cases.

The model presented by Konishi *et* al. (hereinafter referred to as the K model) is a multiple linear regression with Blubber Thickness (model variable names are capitalised) as the dependent variable as a linear function of Date, Diatom Coverage, Longitude, Year, Latitude and Body Length:

$$b = 0.01822 + 0.0157d + 0.2053f + 0.0029\phi - 0.0180y + 0.0189\theta + 0.0859l$$

where:

d = Date in terms of the number of days after November 30,
f = Diatom Coverage score (an integer [0..4]),
ϕ = Longitude in degrees east of the prime meridian,

y = Year of the program (1 corresponds to the 1987/88 whaling season),
θ = Latitude in degrees south of the equator and
l = Body Length in metres.

There are a number of features of the realised sampling design that are potentially important in terms of fitting a model to the body condition data. The realised sampling design in the JARPA studies does not consist of simple random sampling of whales in the study region in terms of times or locations. Year and Longitude are correlated because of the program operates in the western half of the JARPA region in odd years and the eastern half in the even years. There are in effect two segregated sampling programs one to the West of 130° E (IWC Management Areas IV) and one to the East (IWC Management Areas V), each being sampled in alternate years. From 1995/96 the eastern and western ends of the region were extended by 30° of longitude, thus extending into IWC Management Areas III and VI. The data are assigned to strata based on longitude and latitude, with an additional stratum for Prydz Bay. The longitudinal strata divisions divide IWC Management Areas in half (half-Areas). There was no spatial overlap in the two halves in same whaling season and so IWC Management Area and Year are confounded. Latitude and Longitude will also be correlated because of the much greater range of latitudes available in the Ross Sea. Collinearity in explanatory variables makes inferences from linear models more sensitive to underlying model structure and assumptions (Burnham and Anderson, 1998). The daily rate of blubber thickness growth is the most significant coefficient, but the time spans in which samples are collected are often short in many of the strata. This will lead to highly variable slope estimates in those strata if they were to be estimated separately.

Results are presented in Konishi *et al.* in a way that ignores that all half-Areas are not sampled in all years. Their plots of predicted blubber thickness calculated for the program's overall mean longitude means that virtually all of the data used to predict each year's blubber thickness were either west or east of the mean longitude, but never centred on it. It would be more justifiable to plot the predictions for the western half at the western mean longitude and likewise for the eastern half, and with the other explanatory variables at their western or eastern mean values as well.

The analysis of fat weight in Konishi *et al.* produces approximately the same year trend as for blubber thickness in terms of percentage reduction. However, to a first approximation, we would expect that the change fat weight would be proportional to the square of the blubber thickness. Thus if blubber thickness was reduced by 9% then fat weight would be expected to decline by around 17%.

There are other statistical issues in the Konishi *et al.* analyses that are not explored in detail here. One of these is the failure to account for the effect on the statistical significance and confidence intervals of using stepwise regression procedures. Statistical hypothesis testing (in the K model based on nominal statistical significance using F statistics) is a poor choice for model selection (Miller, 2002; Burnham and Anderson, 1998). Stepwise procedures generally result in confidence intervals that are too narrow because they neglect the probability associated with whether a given variable is selected through the stepwise procedure. Stepwise procedures are sequential hypothesis tests; the criteria for statistical significance should be adjusted to keep the significance for the series of tests at the nominal significance level (Burnham and Anderson, 1998, Miller, 2002).

Only one class of linear model is reported in Konishi *et al.*, however, even within the family of multiple linear regression models other possibilities exist, some of which are biologically more reasonable than the K model. Choosing models to fit to data should take into account likely biological processes (Hilborn and Mangel, 1997). Using only the K model is not consistent with such an approach.

Konishi *et al.* rightly recognise that correcting for the effects of the heterogeneous sampling pattern is essential. The critical question is whether the K model is robustly sufficient for that correction.

Biological implications of the K model

The K model has some properties that are not biologically plausible. The most important independent variable in the K model is the date of capture, whose linear coefficient (slope) represents an estimate of the daily growth in blubber thickness. The K model estimates this as a single constant coefficient applicable for all months, latitudes, longitudes and years. The other terms of the K model adjust the intercept of the growth line, not its slope. For example, the model predicts that there is a difference of 5.4 mm in blubber thickness on the date of arrival between the westernmost longitude and the easternmost longitude.

For blubber thickness, the daily growth rate would depend on feeding conditions. In a region with spatial variation in krill abundance both the slope and the intercept of the relationship between blubber thickness and date would vary over the region. Feeding conditions are unlikely to increase smoothly and monotonically with latitude and longitude, but are more likely to reflect locales and years of high and low productivity. Consequently, the K model assumption that animals grow in blubber thickness at the same rate everywhere but with different starting points depending on location is unlikely. Animals arriving in the Antarctic after migration would likely have reasonably similar blubber thickness early in the season, yet the K model predicts that the blubber thickness of new arrivals increases from west to east.

The K model assumes that the same growth coefficient applies to each month. Three cruise reports (Nishiwaki *et al.* 1994, 1995 and 1996) give plots of the moving average of blubber thickness by date. These plots suggest that the growth rate in blubber thickness is low early in the summer and accelerates towards the end of the feeding season. It would be reasonable to include in the analysis an interaction term that allows growth rate to depend on month. This issue is not addressed in the analyses that follow in this paper.

Because animals move systematically between strata during their annual migration, the JARPA strata are not strata in the usual statistical sense that a stratum is a sampling unit that has no members in common with other strata. The change in blubber thickness for a given period in a stratum is the effect of growth in blubber thickness of those animals continually present in the stratum combined with the effects of animals with different blubber thicknesses entering or leaving the stratum.

Fig.1 gives a schematic representation of how blubber thickness might be expected to change with date in the northern and southern strata during the cycle of migration and feeding. Period 1 applies to the main southwards migration. In the north stratum the apparent growth in blubber thickness is low because of the continual influx of new arrivals with low blubber thickness. In the south stratum the apparent intercept ($B_{0,S}$) is higher than that of the north stratum ($B_{0,N}$) because animals found in the southern strata could have been feeding on the northern strata during migration. In period 3, late in the season, in the north stratum the apparent growth in blubber thickness is high because of the influx of animals on their return migration after feeding in further south. In the south stratum the apparent growth rate slows if the fattest animals are the first to leave. The estimates of slope will likely depend on the time at which samples are collected in the different strata. The average slope over the whole season in the north strata could be more or less than that in the south stratum depending on both the their relative productivity and the migratory behaviour of the animals.

Another difficulty with the K model is the assumption that the effect of latitude on blubber growth is linear with the same slope everywhere given that the latitudinal range is very much greater in the Eastern Ross Sea (Area VE) than in the other half Areas. It would be reasonable, at least, to include an interaction term in the analysis that allows the Latitude coefficients to be different for each half Area. In any case, it is the proximity to the ice-edge that is likely to be important, but this is not used directly as an independent variable (although ice proximity is captured to some extent by the stratification into north and south, but perhaps not well in the Ross Sea with its more complex ice conditions)

It would also be reasonable allow for year-to-year variability in feeding conditions, which cannot be captured in a single linear year trend coefficient. Only two partial and uncorrected regression analyses of blubber thickness were reported in the cruise reports (Kato *et al.* 1989; Nishiwaki *et al.*, 1994) both in Area IV. These have slopes of 0.23 mm/day and 0.1 mm/day respectively. Ichii *et al.* (2006) report inter-annual variability in body fat condition (based on girth measurements) using JARPA data from IWC Management Area V.

The effect of diatom adhesion as a proxy for time spent on the feeding grounds is also fitted as a spatially and temporally invariant coefficient that adjusts the intercept of the blubber thickness versus date relationship. Although this may capture some of the effects of the acceleration of blubber thickness with month, it may not be reasonable to suppose that the same diatom coefficient applies in all strata and years. The K model treats the integer coding values (0 – 4) for the diatom adhesion as having an interval scale; the scaling may only be ordinal, and hence the analysis could also consider treating diatom film as a categorical variable.

The least plausible feature of the K model is the estimation of a single constant slope for the growth in blubber thickness with date when there are both biological grounds and some evidence for the growth rate to vary regionally and from year to year. The critical question is whether the actual spatial and temporal distribution of catches could lead to a bias in estimating year trends when fitting the K model when blubber thickness growth rates and intercepts vary over the region and over time.

DATA GENERATING MODELS

Since I do not have the raw data, I will use simulation studies to examine the combined properties of the JARPA program data collection and the K model for correcting for heterogeneity in the data collection. All calculations, simulations and model fitting are carried out using R (R Development Core Team, 2010). The R scripts and data tabulations used in these studies are available from the author.

Reproducing the sampling design

The analysis by simulation has to reproduce a reasonable approximation of the locations and times where the whales taken. I determined from the 18 JARPA cruise reports the locations of catches taken in each year and bounded these by simple boxes specified by latitude and longitude. These boxes were based on the strata identified in each year's cruise report. The first two years of the JARPA program were termed a feasibility study and the strata were defined in an ad-hoc manner and the areas sampled were small. After these first two years, the strata definitions became standardised as a division of the IWC management areas into four boxes defined by the combinations of east – west and north – south. The east-west divisions were defined by the mid-longitudes of the IWC management areas. The north-south divisions were different in each longitudinal sector, and not always at consistent latitudes from year to year because of inter-annual variability in sea ice. The north south division stratifies roughly according to proximity of the ice-edge (although less clearly in the Ross Sea [Area VE]). In Area IV, Prydz Bay was designated a separate stratum. Some catches in the Ross Sea were actually in IWC Area VI, south of the sea-ice barrier often found in the eastern Ross Sea. These samples were assigned to Area V. The range of dates on which whales were caught was also recorded. In the analyses that follow, the ad-hoc strata from the first two years of JARPA were reassigned to the corresponding stratum from the main years of the program. There are 11 unique Area-Stratum combinations.

Simulated catch locations are generated by randomly sampling latitudes and longitudes for each season from uniform distributions for each box, at times also sampled randomly from a uniform distribution within the range of dates reported for each box. The simulated catches are equal in number to the reported catches. Fig. 2 shows the catch locations from Konishi *et al.* However their figure does not show the catches taken north of 60°S. Fig. 3 shows one realisation of the simulated catch locations. The simulated sampling reproduces the main features of the JARPA catch locations.

The JARPA cruise reports tabulate the numbers of animals taken by sex, maturity and pregnancy. They also report the means and standard deviations of lengths of animals taken. In the early years (prior to 1993/94) the length information was combined for mature and immature animals. In the seasons 1993/4 to 1995/6 and in 1998/9 the length information was given for matures only. In all other years length information was given separately for matures and immatures. The missing length information for some years was replaced by the mean values from the closest equivalent times and strata. Each simulated catch is assigned a random length drawn from a normal distribution with the mean length and standard deviation tabulated for that date interval and stratum.

Generating values of Diatom adhesion

The information available to derive simulated observations for diatom adhesion consist of the statistical summaries given in Konishi *et al.* and the observation that whales have to be on the feeding ground for about one month before the first stage of diatom adhesion is apparent (Hart, 1935). Modelling the diatom adhesion requires a simulated value for the difference between the date of capture and the date of arrival on the feeding ground. The date of arrival is assumed to be normally distributed with a mean arrival date for males of December 1 and a standard deviation of ten days. The date of arrival is conditional on the date of capture, that is, a whale cannot be taken on the feeding grounds before its date of arrival. Thus, the date of arrival is randomly sampled from the specified normal distribution of arrivals, but truncated on the right at the date of capture. The formation of diatom score is derived from the following logistic function:

$$d = \frac{4}{1 + \exp\left(-k\left(t - t_{50}\right)\right)}$$

The values of d are rounded to the nearest integer. Consequently the value of d = 0.5 when t = 30, so that the rounded value will be equal to one on day 30. Subject to this condition, the value of t50 of 64.5 days was selected so that the generated values of d have a similar mean (simulated 1.69, reported 1.70) and standard deviation (simulated 1.28, reported 1.24) as given in the statistical summary of Konishi *et al.* The value of k is given by:

$$k = \frac{ln(7)}{t_{50} - 30}$$

The generated values of diatom adhesion are consistent with the observed values, although the simulated standard deviations are about 5% too large.

Model 1 for generating the Blubber thickness

The first model used for generating simulated data is the K model itself. Konishi *et al.* do not report the estimates for the standard deviation of the residuals. The total sum of squares (SST) for K model regression used here is calculated from the standard deviation reported for the blubber thickness (0.922 cm). This gives an SST of 2450 and using the reported R2 values gives a sum of squares due to error (SSE) for the published model of 1414. The standard deviation of the residuals derived from the SSE value is 0.701 cm.

Assuming the K model is correct, I generated simulated datasets to check that the data generating model has similar statistical properties to the published model. The results of 1000 simulations are summarised in Table 1. The data simulation reproduces the estimates of the coefficients of the original K model quite accurately.

Model 2 for generating the blubber thickness – separate slopes and intercepts model for blubber growth

As discussed above, a reasonable model would allow the growth rate of blubber to depend on feeding conditions that are unlikely to be simply proportional to latitude and longitude. Each stratum also may need a separate intercept because, for example, animals arriving in the more southerly strata will have been feeding before reaching them (see Fig. 1.). To examine these possibilities Model 1 as the data-generating simulator is replaced with the following model that allows for growth in blubber thickness to vary by Area-Stratum.

$$B_j(d,f,l) = c + a_j s_j + b_j s_j d + g_j s_j f + hl + \epsilon; \qquad j \in \{1 \cdots n\}$$

where:

$B_j(d,f,l)$ = blubber thickness (cm) in Area-Stratum j as a function of date, diatom film and length
c = constant (cm)
n = the number of strata
a_j = blubber thickness at day zero for a given stratum (cm)
s_j = an indicator variable (0,1) indicating the stratum in which the whale was taken
b_j = daily growth in blubber thickness for a given stratum (cm/day)
d = date of capture
f = index of diatom adhesion
g_j = the incremental correction to blubber thickness associated with the extent of diatom adhesion for a given stratum (cm)
h = coefficient relating blubber thickness to length (cm/m)
l = animal length
ϵ = an error term from a normal distribution (cm).

c is set to keep the mean blubber thickness close to that of Model 1, that is:

$$c = 0.1822 + 0.0029 \times 130 - 0.0180 \times 9 + 0.0189 \times 65$$

where the coefficients are from the fitted K model. The numbers 130, 9 and 65 above are the approximate mean longitude, year number and latitude respectively. The values of g_j are set at 13 times the growth coefficient for the corresponding stratum. This is consistent with the average ratio of daily growth coefficient to the diatom coefficient estimated from the K model.

Data generated using this model have no year trend in blubber thickness. Finding a significant year trend from the fitting the K model to data from this data-generating model would indicate that the K model may not be reliable for correcting for the effects of sampling heterogeneity present in the JARPA program.

Adding a variable year effect in blubber growth – Model 3

Model 2 assumes that the feeding conditions, and hence blubber thickness have no time trend. However, feeding conditions will also likely change with time due to various random or cyclic causes. These could include the effects of *el Niño* and the Indian Ocean Dipole (Saji *et al.* 1999) for example. Model 3 is designed to examine the effects that non-linear time effects might have on the properties of the estimates. Model 3 has variable feeding conditions that randomly affect all areas equally and simultaneously. The model is given by:

$$B_{j,y}\left(d,f,l\right)=c+\left(a_j+\alpha_y\right)s_j+\left(b_j+\beta_y\right)s_jd+g_js_jf+hl+\epsilon;\qquad j\in\{1\cdots n\}$$

where:

$B_{j,y}(d,f,l)$ = is the blubber thickness in Area-Stratum j in year y as a function of date, diatom film and length

β_y is a random variable with a normal distribution with mean = 0 and standard deviation = 0.0025. For simplicity, α_y is a random variable completely correlated with β_y, such that:

$$\alpha_y=\left(b_j+\beta_y\right)\times 80$$

Model 3 can also have the expected values of slopes and intercepts set according to either model 1 or model 2. The random yearly variations in slopes and intercepts are added to the slopes and intercepts in each stratum. In such cases there can be spatial differences in the expected slope and intercepts for growth that is then subject to year-to-year variation around the mean values.

Model 4 – blubber growth slopes and intercepts vary over both space and time.

Model 3 has each Area-Stratum or region being equally affected by a spatially homogeneous inter-annual variation in feeding conditions and hence blubber thickness. However, it is also likely that trends in feeding condition will vary in both space and differently from year to year. Such an effect might arise for example from the Antarctic polar wave (White and Peterson, 1996). To simulate this class of process, random variation in feeding conditions is allowed for through the following model:

$$B_{j,y}(d,f,l)=c+\left(a_j+\alpha_{j,y}\right)s_j+\left(b_j+\beta_{j,y}\right)s_jd+g_js_jf+hl+\epsilon;\qquad j\in\{1\cdots n\}$$

The values of $\beta_{j,y}$ that affect each half area and year are drawn from a normal distribution with mean = 0 and standard deviation = 0.0025

For simplicity:

$$\alpha_{j,y}=\left(b_j+\beta_{j,y}\right)\times 10$$

SPURIOUS YEAR TRENDS IN BLUBBER THICKNESS

The first question is whether there are plausible stratum-specific blubber thickness growth coefficients that would produce erroneous and apparently statistically significant trends in blubber thickness with year. To determine this, 1000 replicate analyses of fitting the K model to simulated data from Model 2 were carried out with the growth coefficients in each trial for each stratum drawn independently from a uniform distribution with a range 0.0057 to 0.0257 cm/day (the K model Date coefficient ± 0.01). The intercepts are additively adjusted with numbers drawn at random from a uniform distribution with a range ± 0.5. The fitted model does not undergo any stepwise regression selection, the full set of K model independent variables is used in each trial.

The standard deviation of the data-generating model's error distribution was adjusted until the average of the residual standard deviations estimated using the K model from the 1000 trials had the mean value 0.701, as estimated in the original K model fit. The required value of the generating error standard deviation was found to be 0.4847. This adjustment is required because lack of fit due to the variation of blubber coefficients by stratum contributes to the residual standard deviation estimated using the K model.

Fig. 4 shows the distributions of estimates of the year coefficient for 1000 stochastic trials. The lower pane shows the distribution of year coefficients that were reported as statistically significantly different from zero at the P=0.05 level. There were 653 apparently significant Year coefficients. Of these, 335 were negative, giving a mean blubber coefficient of -0.0129 cm/year. The K model estimate (-0.0180 cm/year) lies within the distribution of statistically significant negative coefficients. Table 2 shows the average values of the intercepts and slopes that produce spurious positive or negative Year coefficients. These values are modest departures from the K model coefficients.

Arrangements of slopes and intercepts that produce erroneous statistically significant year coefficients are quite common when the K model is fitted to data collected with the realised spatial and temporal pattern of catching of the JARPA program.

However, the K model also has statistically significant latitude and longitude coefficients. Recalling that the data-generating model has latitude and longitude coefficients equal to zero, the combined slopes and intercepts model would be contradicted if there were no arrangements of the coefficients that simultaneously produce statistically significant estimates of the three coefficients with similar sign and magnitude as the K model estimates.

There were 62 instances where all three coefficients were statistically significant with the required sign, with the distributions shown in Fig. 5. The distributions of the coefficients include the values estimated from the K model. The mean of the corresponding Year coefficients is -0.0128 cm/year (K model estimate = -0.180), Longitude mean is 0.0029/degree (K model = 0.0029) and Latitude = 0.0390 cm/degree (K model = 0.0189).

A monitoring program and method of analysis that could plausibly produce high numbers of false significant hypothesis tests of a linear year trend cannot be considered reliable.

MODEL 3 – RANDOM VARIATIONS IN YEARLY FEEDING CONDITIONS

Random variation in blubber growth coefficients (Model 3) should have little or no net trend in year effect, with the years of improved feeding conditions being balanced on average by the years with worse conditions. The first test is to examine the effects of year to year variation in growth characteristics in the absence of any spatial effects. That is, the mean values of the slopes and intercepts of the date coefficients are the same in every stratum and equal to their K model estimates. To maintain the mean residual from the K model fits at its observed value required a residual standard deviation of 0.586.

1000 trials of fitting the K model to Model 3 data gives the distributions of estimates given in Figs 6 and 7. The distribution of the estimates of Year coefficients (Fig. 6) includes the K model estimate. Statistically significant Year, Latitude and Longitude coefficients occur in 777, 339 and 720 instances respectively. Although the distributions of Latitude and Longitude coefficients include the K model estimates, these are towards the tails of the distributions.

The results show that there is a high propensity to give spurious statistically significant trends. This is due to a form of “pseudo-replication” (Millar and Anderson, 2004), i.e. not all of the observations are independent, but rather the residuals in each year are correlated. This means that the numbers of degrees of freedom associated with the tests for statistical significance are overstated.

Mixed effects model

One possible remedy for this form of pseudo-replication is to fit a mixed effects model (Millar and Anderson, 2004), that in this case would estimate at least the slopes and intercepts of the blubber growth as random effects, while estimating the remaining terms as fixed effects. Results from fitting this model to the same data as in the K model trial above leads to the distributions of the estimates of year coefficients are shown in Fig. 8. The statistical significance of a year trend was determined using an asymptotic likelihood ratio test based on the change in log-likelihood from fitting a model without a year trend with that from the nested model with the year trend included. The number of significant year coefficients has been reduced from 777 to 107. Clearly, the mixed effects model has substantially reduced the number of spurious significant estimates of year trends, although they are more frequent than would be expected given the nominal significance level of 5%.

The number of statistically significant estimates for the latitude and longitude were 55 and 98 respectively. These significance tests may be less reliable because they are based on the reported t statistics, assuming that number of degrees of freedom is high for the fixed effects estimates (rather than the likelihood ratios used for the year trend). Interestingly the distributions of these estimates, shown in Fig. 9, exhibit less spread compared with those from fitting the K model (Fig. 7). Random year effects on their own do not lead to high numbers of spurious significant estimates of longitude and latitude coefficients.

Combined spatial variation and region wide year effects

The narrow distributions of latitude and longitude effects in the preceding trials suggest that both spatial and temporal variability in blubber growth coefficients may be important. The next set of trials uses data generated from combining models 2 and 3, that is, the slopes and intercepts for the blubber growth coefficient are random variables with different expected values in each Area-Stratum. In each year a common random year effect is added to the blubber growth coefficient in each Area-Stratum. This means that the random effects are completely correlated across the region. As before the K model mean of the estimated residual standard deviation is maintained at 0.701. In this case this requires a pure error standard deviation = 0.2955.

Fitting the K model leads to 821 spurious significant year trends, with 856 significant latitude and 888 significant longitude coefficients. The distribution of estimated year coefficients in Fig. 10 shows that the K model estimate lies well within the distribution of spurious significant outcomes. The combined spatial and random year effects data model produces distributions of estimates of the latitude and longitude coefficients as shown in Fig. 11. Unlike the distributions from random year effects data alone, the K model estimates now lie well inside these distributions.

Fitting mixed effects model to the combined spatial and year effects.

The next set of trials examines the results from fitting the mixed effects model to the same data from the combined spatial variation and random year effects data generator. Fig. 12 shows the distribution of year trend estimates. Fig. 13 shows the estimates of longitude and latitude coefficients. There were now 79 statistically significant year trends, along with 887 and 858 significant latitude and longitude coefficients respectively. The latter are similar to the results from fitting the K model and reflect that in this model the spatial variation in blubber growth coefficient will often lead to a significant linear coefficient.

MODEL 4 - SPATIALLY UNCORRELATED YEAR EFFECTS

The first test with Model 4 is fitting the K model to data generated with pure random year effects combined with underlying spatial patterns in the expected value of the blubber growth coefficients. Unlike Model 3, the year effects are uncorrelated between half areas. Maintaining the K model residual estimates at a mean value of 0.701 requires a true residual error = 0.482.

Fitting the K model leads to the distributions of year effect estimates shown in Fig. 14. There were 696 spurious statistically significant year effects. 853 and 812 latitude and longitude coefficients shown in Fig. 15 were statistically significant even though neither effect is present as a pure linear trend.

Mixed effects fit with year effect as a uniform regional effect

The next test fits the mixed effects model with region wide year effects to the data generated with Model 4. The fitted mixed effects model assumes that the random year effects are completely spatially correlated, and so it does not correspond with the data generating model.

There were 55 significant year trends (Fig. 16), along with 844 and 787 significant latitude and longitude coefficients (distributions shown in Fig. 17). The mixed effects model that allows for random variation in blubber coefficient in each year across all Area/Strata has reduced the number of spurious significant year effects to approximately the nominal 0.05 probability level.

Fitting mixed effects model with Date as a random effect with a Year:Area-Stratum interaction.

The next mixed effects model allows for the blubber growth coefficients in each Area-Stratum - Year to be independent random variables. This fitted model assumes that the random year effects are spatially uncorrelated, and so it is consistent with the Model 4 data generating model.

Fitting this model to the same Model 4 data gives estimates of year trends as shown in Fig. 18. There were 173 year trends that are spuriously statistically significant, however only 8 of these were negative. This model does not fit a linear term for latitude or longitude since these effects are accounted for in the Year:Area-Stratum interaction.

In principle this fitted model is appropriate for data generated using model 4. However, the results show that the probability of a type 1 error is substantially above the nominal 5% level. The following section suggests reasons why this model may not work well with the realised sampling pattern of JARPA.

Fitting a Year/Stratum effects model to data from Model 4

An appropriate sampling design would allow for fitting models that estimates of the growth of blubber to be different in space and time, i.e. the sampling design should allow for an analysis using a "full model" with fixed effects for all the interaction terms. A full fixed effects linear model to use for the analysis of the blubber thickness is one which allows each stratum year combination to have a unique intercept and slope for the Date coefficient. This class of model allows for the estimation of patterns of year effects other than those of a simple linear trend. Such a Year/Stratum effects (YSE) model has a three-way interaction between Date, Area-Stratum and Year. This has the effect that each Area-Stratum can have a separate slope and intercept for the effect of date on blubber thickness in each year. Such a model is then consistent with data generated using model 4.

Fig. 19 shows the results of fitting the YSE model to Model 4 data using the realised sampling design of JARPA. The data are generated so that there is only a single realisation of the random year effects used in all 1000 replicates. The expected value for the blubber growth coefficient from the K model applies in all Area/Strata, that is, there are no spatial effects generated in blubber growth rates.

Clearly, as shown in Fig. 19, very unreliable corrected estimates (predictions) of blubber thickness could be produced. "Corrected" blubber thicknesses range from roughly -600 mm to +600 mm, which is more than an order of magnitude greater than the mean (36 mm). This indicates that some of the estimates of slopes and intercepts are highly variable. This variability swamps the prospect of obtaining reasonable estimates of the pattern of year effects in each stratum with the full interaction model. These results indicate that there are features of the realised sampling pattern of JARPA whaling that could undermine the ability to estimate and correct for sampling heterogeneity when fitting the YSE model.

Correcting for the effects of sampling heterogeneity requires reasonably precise estimates of the blubber growth coefficients. Fig. 20 shows the relationship (estimated by simulation) between the Coefficient of Variation (CV) for slope estimates from linear regressions for different sample sizes and ranges for the dates between which samples are measured, using a random normal error distribution in blubber thickness with a standard deviation of 0.701 cm. Obtaining slope estimates with CVs below 1.0 generally requires sampling over date ranges of more than 50 days and sample sizes of 30 or more.

Fig. 21 shows the sample sizes and date ranges for each Area-Stratum – Year combination in the JARPA samples of mature males. The majority of Area-Stratum-Year combinations have sampled date ranges substantially below 50, and many have sample sizes less than 30.

The standard error of slope estimates in a regression of Blubber thickness on the independent variable Date depends on the variance (contrast) in Date. The greater the variance in the Date, the lower will be the standard error of the slope. The greatest variance in Date is obtained by concentrating all the observations equally at both ends of the whaling season; so-called "dumbbell" designs. If we take the nominal length of the whaling season to be 115 days (Date range 0 to 114), the maximum variance in Date is 3249. However, dumbbell designs preclude the possibility of estimating departures from linearity of the relationship between the dependent and independent variable and so therefore it is reasonable to spread the sampling throughout the whaling season. The variance in Date for a uniform sampling scheme over the whaling season would be 1083.

In JARPA the intervals during which whales are taken are different in each year and stratum. Some years and strata have more than one interval in which whales are taken. In order to calculate the variance in Date I assume for a given year and stratum that the dates are sampled from a mixture of uniform distributions.

The standard error of prediction increases with larger distances between the mean Date of the sample and the Date at which Blubber thickness is predicted. The upper plot in Fig. 22 shows that the mean sampling date is often concentrated before or after the mean date of the season, and is systematically related to the Area-Stratum. In most cases the prediction date for the standardisation (mid-season) falls outside the range of dates sampled.

The lower plot in Fig. 22 shows that the variance in the distributions of dates sampled is usually very substantially below the "ideal" variance of 1083. Consequently, the use of regression predictions to standardise the blubber thickness to a given date will lead to highly variable results as shown in Fig. 19.

These features of the sampling design indicate why the YSE model cannot be fitted reliably to data with the JARPA realised sampling design. These features of the sampling are probably also implicated in the potential bias in estimates of the linear trend in blubber thickness using the K model.

From these analyses we can conclude that the sampling design realised in JARPA is very heterogeneous, such that the correction for sampling heterogeneity may be unreliable. The ability to use the most general of the plausible linear models to explore the possible changes in blubber thickness with time and space is substantially compromised. This is further exacerbated by splitting of the sampling in half in alternate years.

Balanced sampling

The final question is to corroborate whether the poor properties of the corrections (predictions) from the YSE model arise from the poor realised sampling pattern of JARPA. This is examined by generating data using the same version of model 4 as the preceding test, but this time with annual sample sizes in each half-Area of 52-54 mature males balanced among the strata. The sampling dates are drawn from a uniform distribution across the whole whaling season. In this test the whaling still occurs in each half of the JARPA region in alternate years. This would give a total sample size over the 18 years of JARPA of 2826 mature males, which is about the same as the realised sample size.

Fig. 23 shows the year effects predicted across years for the Area-Strata. The effects of variation are captured without the very large range in predictions shown in Fig. 19. A balanced sampling design would allow for the disentangling of the Area-Stratum effects from the year effects.

However, balancing the dates on which whales were taken over the whaling season and whale spatial distribution would be formidably difficult using only a single whaling fleet. Consequently achieving the results that are shown here to be theoretically attainable are unlikely to be feasible in practice.

CONCLUSION

The analyses of minke whale body condition reported in Konishi *et al.* (2008) cannot be considered reliable. This is because the K model is biologically implausible. Statistical significance will be overstated because of “pseudo-replication” if there is inter-annual variability in blubber growth rates. Applying the K model to data generated using biologically plausible models can result in statistically significant year trends in blubber thickness even though such trends are not present in the data. This is because the realised sampling design of the JARPA program is highly heterogeneous over space and time, and the K model is incapable of correcting for some of the important potential sources of heterogeneity.

Appropriate analyses would compare variants within the family of linear models that allow for interactions so that blubber growth can be different among years and locations. However, the realised sampling design is poor in terms of estimating the slopes of the blubber growth coefficient when it is not estimated as a single constant applicable to all years and strata. This is because the time spans during which whales are taken are insufficiently spread throughout the whaling season in many of the strata. This is compounded by the mean sampling date often being concentrated towards the beginning or end of the whaling season. Another complication is that the JARPA program was virtually two separate programs running in alternate years in the two halves of the JARPA region.

The JARPA objective of monitoring the Antarctic ecosystem by means of the analysis of the body condition of minke whales on their feeding grounds is demonstrated here as likely to be unreliable with the spatial and temporal distribution of sampling effort realised with JARPA.

REFERENCES

Burnham K. P. and Anderson D. R. 1998. *Model selection and inference: a practical information-theoretic approach*. Springer Verlag, New York. 353pp

GOJ. 1995. The 1995/96 Research Plan for the Japanese Whale Research Program under the Special Permit in the Antarctic. Paper SC/47/SH3 presented to the IWC Scientific Committee, May 1995 (unpublished). Government of Japan. 9pp

Hart T. J. 1935. On the diatoms of the skin film of whales, and their possible bearing on problems of whale movements. *Discovery Reports* 10:249-82

Hilborn R. and Mangel, M. 1997. *The ecological detective: confronting models with data.* Monographs in Population Biology 28:1-315. Princeton University Press, Princeton. 315pp

Ichii T, Shinohara N, Fujise Y, Nishiwaki S, Matsuoka K. 1998. Interannual changes in body fat condition index of minke whales in the Antarctic. *Mar. Ecol. Prog. Series* 175:1-12

Kato H, Hiroyama H, Fujise Y, Ono K. 1989. Preliminary report of the 1987/88 Japanese feasibility study of the special permit proposal for Southern Hemisphere minke whales. *Rep. int. Whal. Commn* 39:235-48

Konishi K, Tamura T, Zenitani R, Bando T, Kato H, Walloe L. 2008. Decline in energy storage in the Antarctic minke whale (*Balaenoptera bonaerensis*) in the Southern Ocean. *Polar Biol*. 31(12):1509-20

Millar, R. B. and Anderson, M. J. 2004. Remedies for pseudoreplication. *Fisheries Research* 70:397-407.

Miller A. J. 2002. *Subset selection in regression* (2nd edition). Monographs in Statistics and Applied Probability 95:1:237. Chapman and Hall/CRC, Boca Raton, Florida

Nishiwaki S, Ishikawa H, Itoh S, Matsuoka K, Yuzu S, Nagatome I, Yamagiwa D, Murase H, Tanifuji S, Miyakoshi H, Ono K. 1994. Report of the 1993/94 cruise of the Japanese whale research programme under special permit in the Antarctic Area IV. Paper SC/46/SH15 presented to the IWC Scientific Committee, May 1994 (unpublished). 42pp

Nishiwaki S, Ishikawa H, Itoh S, Shimamoto K, Mogoe T, Kawazu H, Machida S, Yamane T, Ono K, Ohkoshi C. 1995. Report of the 1994/95 cruise of the Japanese whale research programme under special permit (JARPA) in the Antarctic Area V. Paper SC/47/SH5 presented to the IWC Scientific Committee, May 1995 (unpublished). 38pp.

Nishiwaki S, Ishikawa H, Tohyama D, Kawasaki M, Shimamoto K, Yuzu S, Tamura T, Mogoe T, Hishii T, Yoshida T, Hidaka H, Nibe H, Yamashiro K, Ono K, Taguchi F. 1996. Report of the 1995/96 Japanese Whale Research Programme Under Special Permit in the Antarctic (JARPA) in Area IV and eastern part of Area III. Paper SC/48/SH12 presented to IWC Scientific Committee, June 1996 (unpublished). 34pp

R Development Core Team (2010) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Saji N. H., Goswami B. N., Vinayachandran P, Yamagata N. T. 1999. A dipole mode in the tropical Indian Ocean. *Nature* 401: 360–363

White W. B. and Peterson R. G. 1996. An Antarctic circumpolar wave in surface pressure, wind, temperature and sea-ice extent. *Nature* 380:699-702.

Table 1.

Comparison of simulation results with the K model coefficients.

Variable	**Coefficient**		**Standard error**	
	Reported	**Mean of simulations**	**Reported**	**Mean of simulations**
Constant	0.1822	0.1665	0.4461	0.4382
Date (Dec. 1 = 1)	0.0157	0.0157	0.0005	0.0012
Diatom (integers 0 – 4)	0.2053	0.2062	0.0107	0.0265
Longitude (E)	0.0029	0.0029	0.0003	0.0003
Year (87/88 = 1)	-0.0180	-0.0179	0.0028	0.0028
Latitude (S)	0.0189	0.0188	0.0054	0.0048
Body length (m)	0.0859	0.0880	0.0326	0.0365
Residual std deviation	0.701	0.701	-	-

Table 2.

Mean coefficients for Date and intercept in each Area-Stratum that produce apparently statistically significant positive or negative Year coefficients from fitting the K model to data from the separate slopes and intercepts simulator (Model 2). The intercept adjustment is added to the K model intercept when generating the data. The estimated single coefficient from the K model is 0.0157 cm/day, with intercept 0.1822 cm.

Area-Stratum	**Positive date coefficient (cm/day)**	**Positive date intercept adjustment (cm)**	**Negative date coefficient (cm/day)**	**Negative date intercept adjustment (cm)**
III East	0.0175	0.0981	0.0137	-0.1158
IV East-North	0.0154	-0.0107	0.0162	-0.0061
IV East-South	0.0118	-0.0672	0.0196	0.0927
IV Prydz Bay	0.0161	-0.0206	0.0151	-0.0149
IV West-North	0.0150	-0.0380	0.0163	-0.0102
IV West South	0.0165	-0.0281	0.0145	0.0585
V East-North	0.0144	-0.0188	0.0166	0.0340
V East-South	0.0144	-0.0184	0.0170	-0.0143
V West-North	0.0151	-0.0294	0.0160	0.0214
V West-South	0.0185	-0.0083	0.0135	-0.0035
VI West	0.0162	0.0773	0.0142	-0.0574

List of Figures

Fig. 15. Distributions of the estimates of Longitude and Latitude coefficients from the K model fit to data from Model 4 with combined spatial variation and spatially uncorrelated random year effects. The thick vertical lines are the corresponding K model estimates.

Fig. 16. Distributions of the estimates of year coefficients from the mixed effects model with simple random year effects fit to data from Model 4 with combined spatial variation and spatially uncorrelated random year effects. The thick vertical lines are the corresponding K model estimates.

Fig. 17. Distributions of the estimates of Longitude and Latitude coefficients from the mixed effects model with simple random year effects fit to data from Model 4 with combined spatial variation and spatially uncorrelated random year effects in blubber growth coefficients. The thick vertical lines are the corresponding K model estimates.

Fig. 18. Distributions of the estimates of year coefficients from the mixed effects model with Area-Stratum random year effects fit to data from Model 4 with combined spatial variation and spatially uncorrelated random year effects. The thick vertical lines are the corresponding K model estimates.

Fig. 19. Predictions from 1000 trials of fitting the YSE model to data generated from Model 4 without spatial variation in blubber growth and a single realisation of uncorrelated Area-Stratum random year effects. The sampling pattern is the realised pattern of JARPA whaling.

Fig. 20. The relationship between the Coefficient of Variation (CV) for slope estimates from linear regressions for different sample sizes and ranges for the dates (d) between which samples are measured, estimated by simulation with a random normal error distribution in blubber thickness with a standard deviation at the estimated value from the K model of 0.701 cm.

Fig. 21. Sample sizes and date ranges for each Area-Stratum – Year combination in the JARPA samples of mature males.

Fig. 22. The distributions of the means and variances of the sampling dates across years for each stratum combination, based on mixtures of uniform distributions. The thick horizontal line is the mean date of the season (57) or the variance for uniform sampling across the whole season (1083). In some years Area IIIE is to some extent a ‘dumbbell’ sampling pattern because the stratum is sampled early and late in the whaling season.

Fig. 23. Predictions from 1000 trials of fitting the YSE model to data generated from Model 4 without spatial variation in blubber growth and a single realisation of uncorrelated Area-Stratum random year effects using a balanced sampling pattern.

Figures

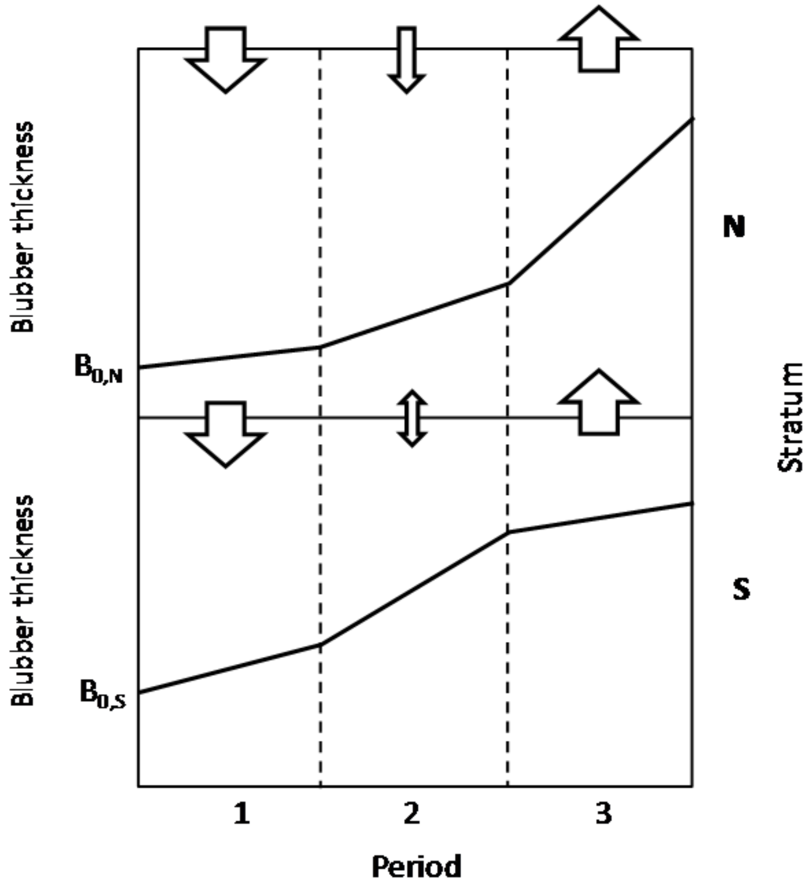


Fig. 1.

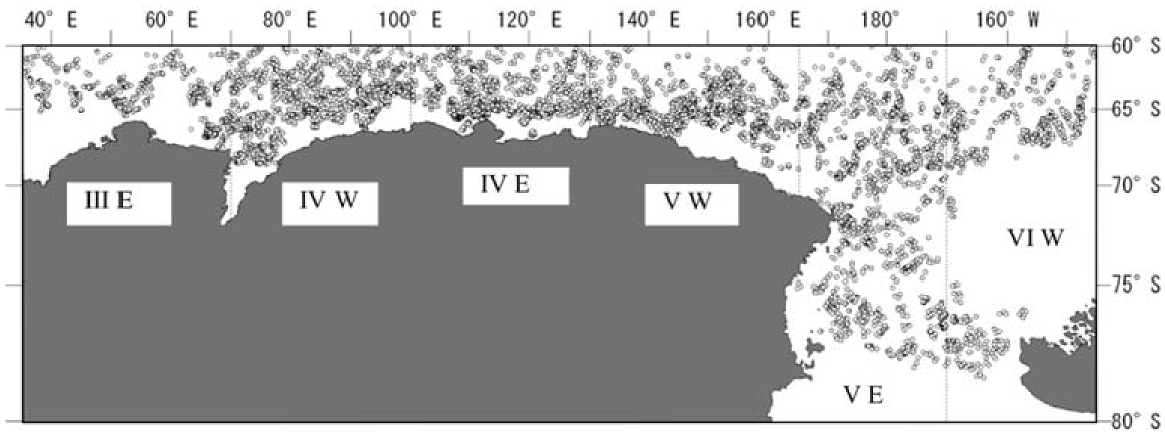


Fig. 2.

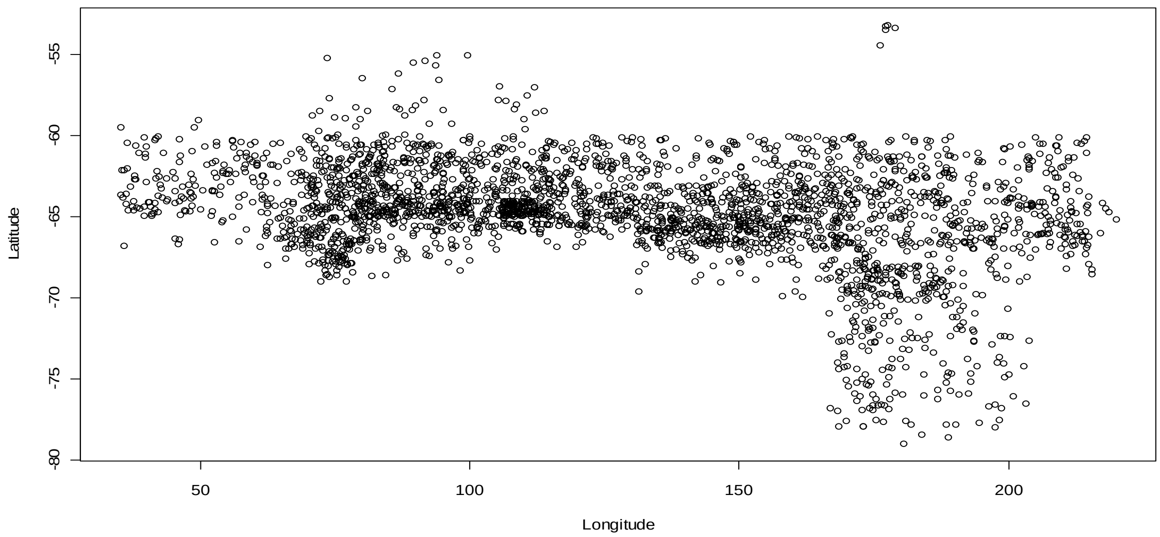

-55
-60
-65
-70
-75
-80
Latitude
50
100
150
200
Longitude

Fig. 3.

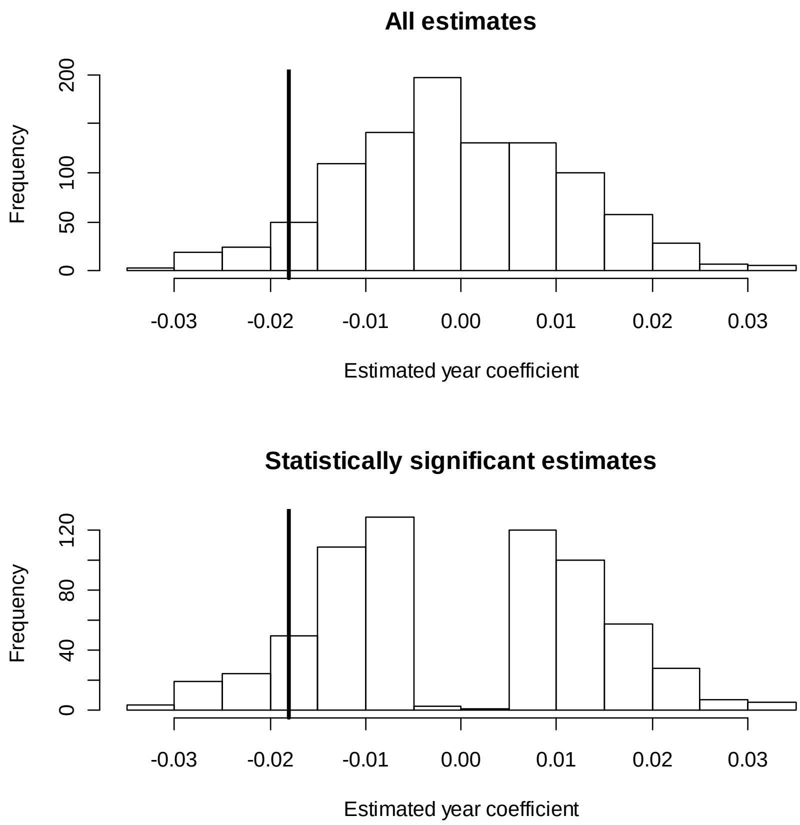

All estimates
Frequency
0
50
100
200
-0.03
-0.02
-0.01
0.00
0.01
0.02
0.03
Estimated year coefficient
Statistically significant estimates
Frequency
0
40
80
120
-0.03
-0.02
-0.01
0.00
0.01
0.02
0.03
Estimated year coefficient

Fig. 4.

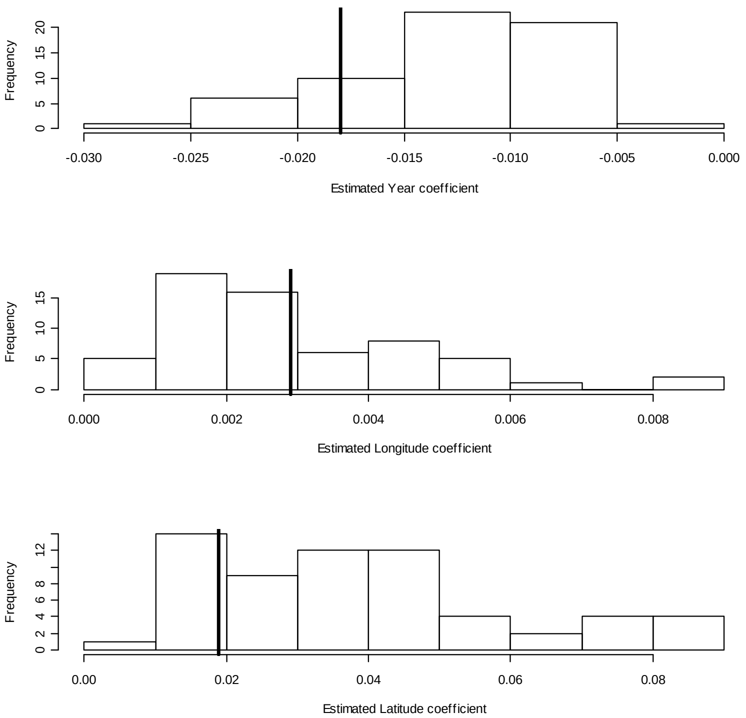

Frequency
0
5
10
15
20
-0.030
-0.025
-0.020
-0.015
-0.010
-0.005
0.000
Estimated Year coefficient
Frequency
0
5
10
15
0.000
0.002
0.004
0.006
0.008
Estimated Longitude coefficient
Frequency
0
2
4
6
8
12
0.00
0.02
0.04
0.06
0.08
Estimated Latitude coefficient

Fig. 5.

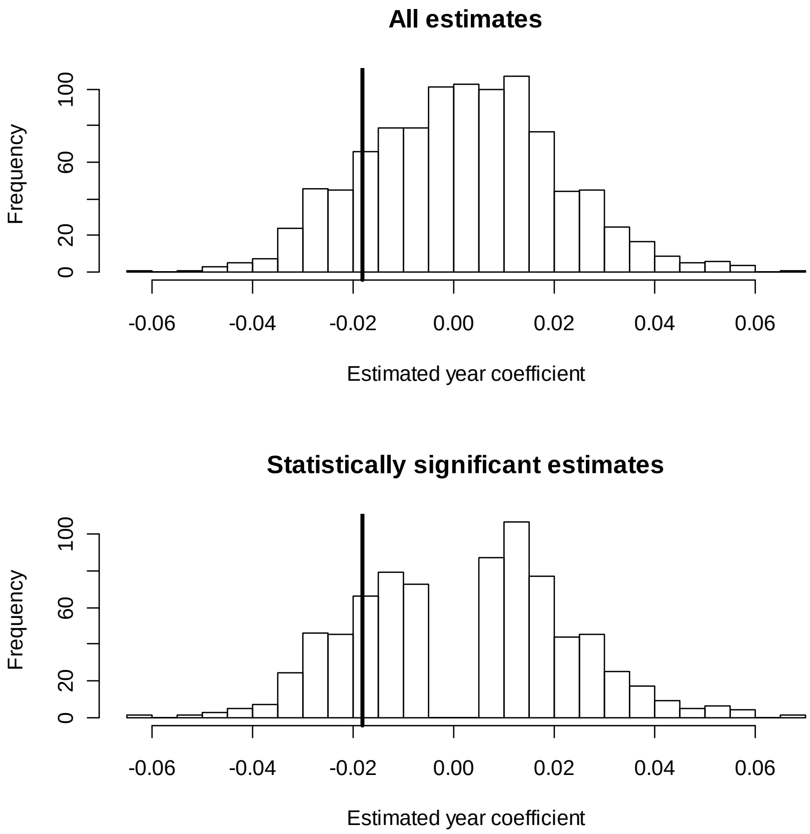
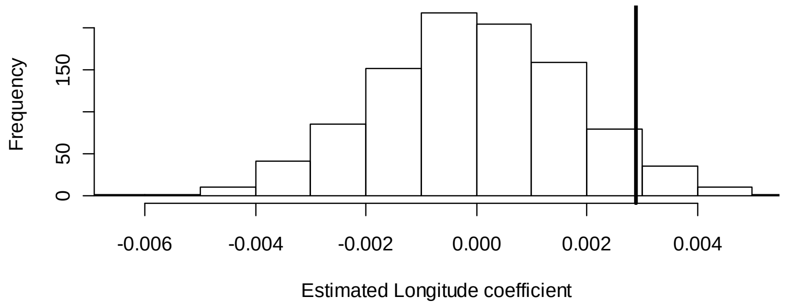
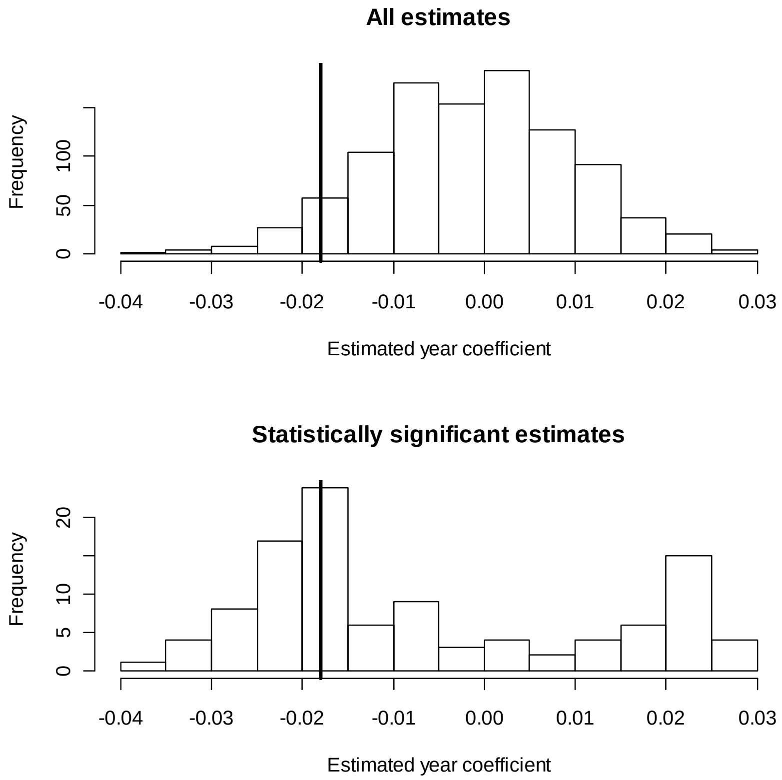

All estimates
Frequency
0
20
60
100
-0.06
-0.04
-0.02
0.00
0.02
0.04
0.06
Estimated year coefficient
Statistically significant estimates
Frequency
0
20
60
100
-0.06
-0.04
-0.02
0.00
0.02
0.04
0.06
Estimated year coefficient

Fig 6.



Frequency
0
50
150
-0.006
-0.004
-0.002
0.000
0.002
0.004
Estimated Longitude coefficient
Frequency
0
50
150
-0.04
-0.02
0.00
0.02
0.04
Estimated Latitude coefficient

Fig. 7.



All estimates
Frequency
0
50
100
-0.04
-0.03
-0.02
-0.01
0.00
0.01
0.02
0.03
Estimated year coefficient
Statistically significant estimates
Frequency
0
5
10
20
-0.04
-0.03
-0.02
-0.01
0.00
0.01
0.02
0.03
Estimated year coefficient

Fig. 8.

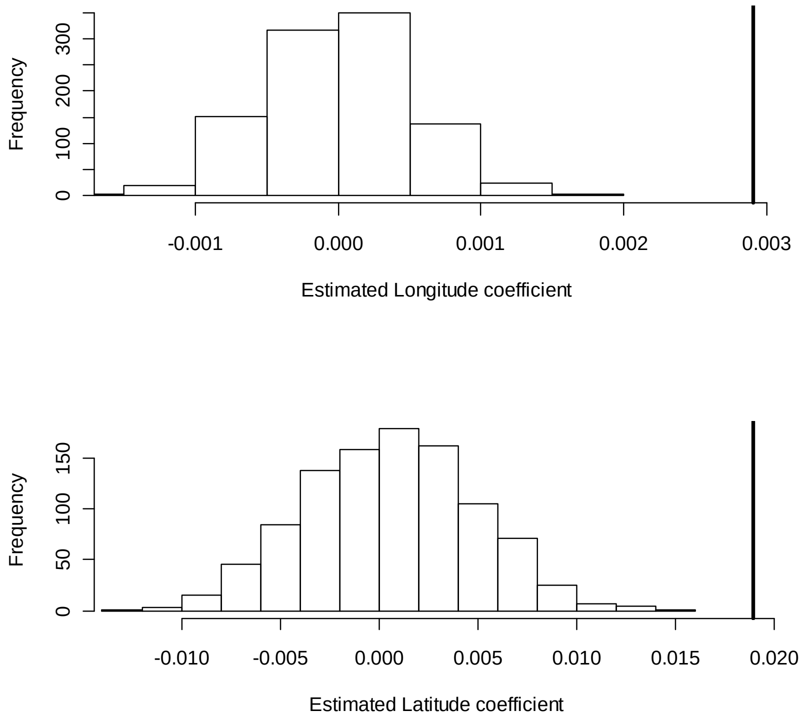

Frequency
0
100
200
300
-0.001
0.000
0.001
0.002
0.003
Estimated Longitude coefficient
Frequency
0
50
100
150
-0.010
-0.005
0.000
0.005
0.010
0.015
0.020
Estimated Latitude coefficient

Fig. 9.

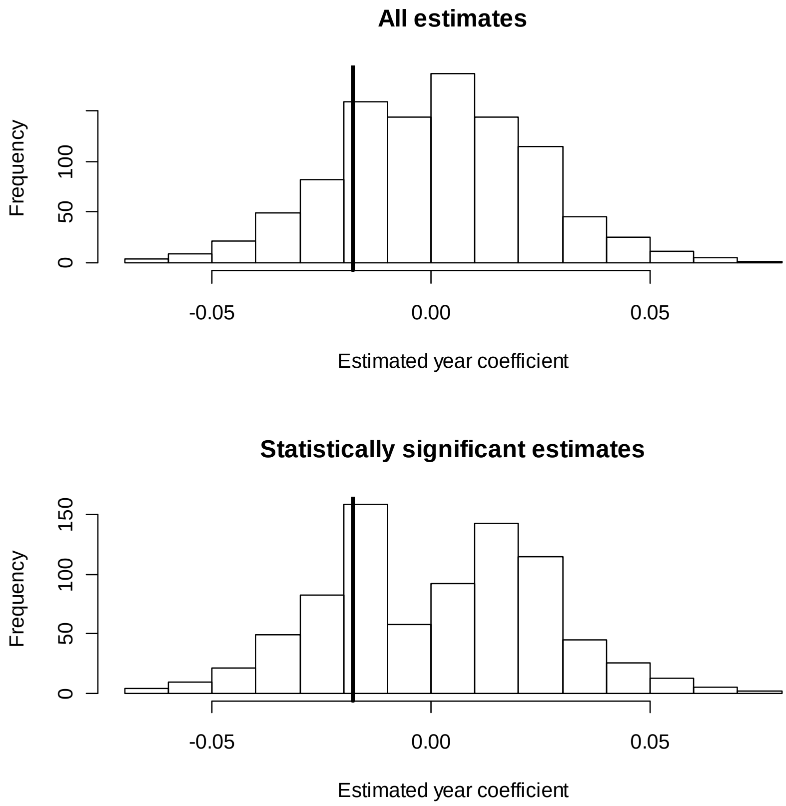

All estimates
Frequency
0
50
100
-0.05
0.00
0.05
Estimated year coefficient
Statistically significant estimates
Frequency
0
50
100
150
-0.05
0.00
0.05
Estimated year coefficient

Fig. 10

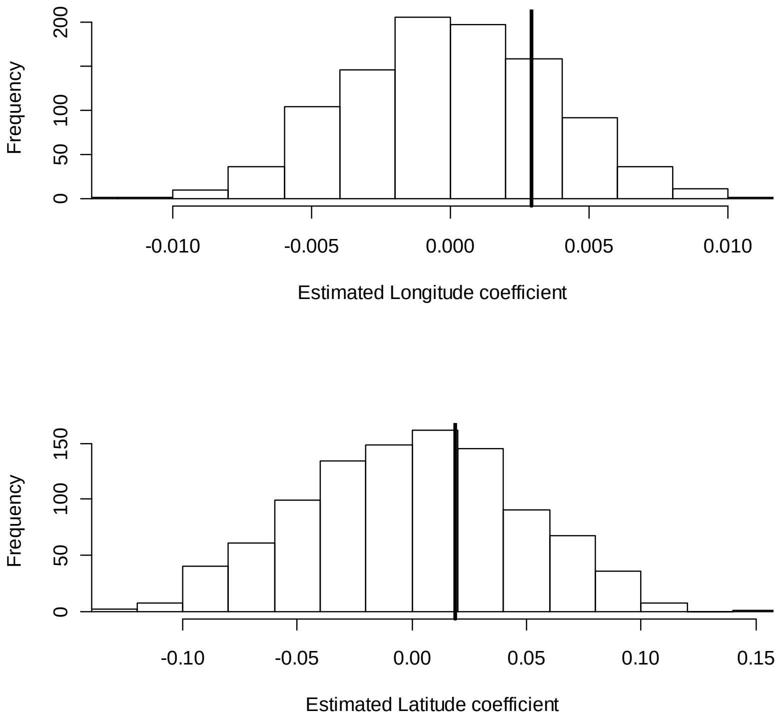

Frequency
0
50
100
200
-0.010
-0.005
0.000
0.005
0.010
Estimated Longitude coefficient
Frequency
0
50
100
150
-0.10
-0.05
0.00
0.05
0.10
0.15
Estimated Latitude coefficient

Fig. 11.

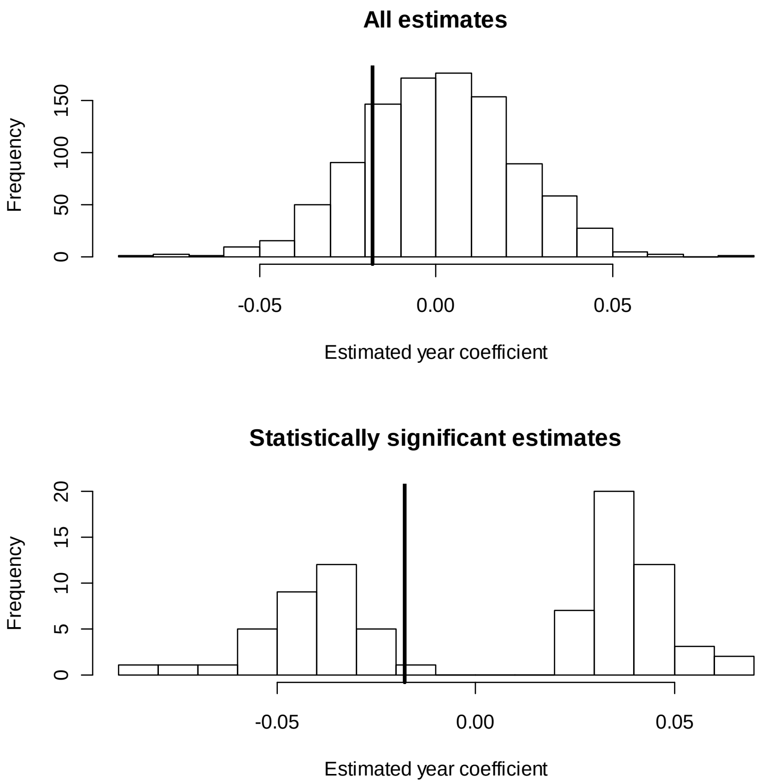

All estimates
Frequency
0
50
100
150
-0.05
0.00
0.05
Estimated year coefficient
Statistically significant estimates
Frequency
0
5
10
15
20
-0.05
0.00
0.05
Estimated year coefficient

Fig. 12.

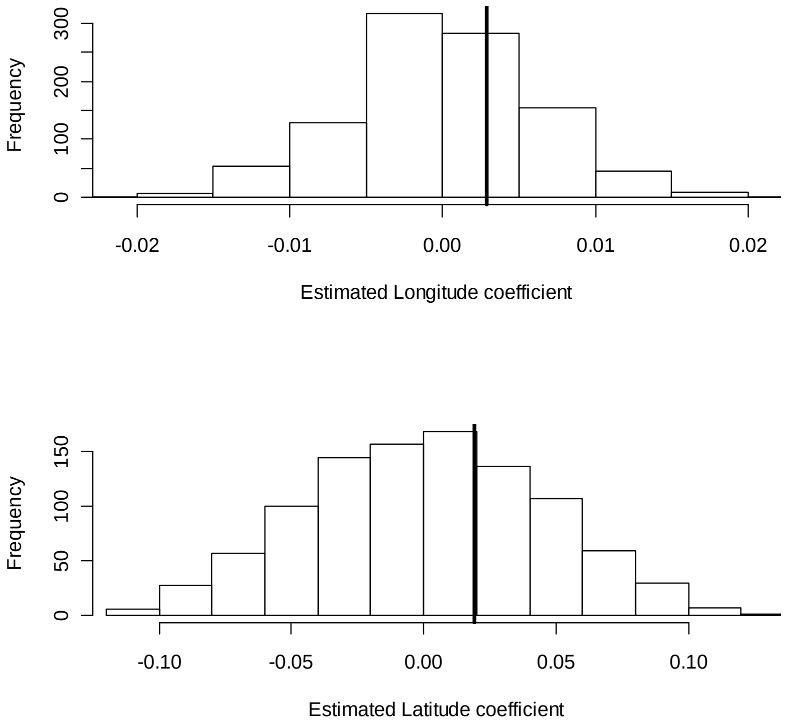

Frequency
0
100
200
300
-0.02
-0.01
0.00
0.01
0.02
Estimated Longitude coefficient
Frequency
0
50
100
150
-0.10
-0.05
0.00
0.05
0.10
Estimated Latitude coefficient

Fig. 13.

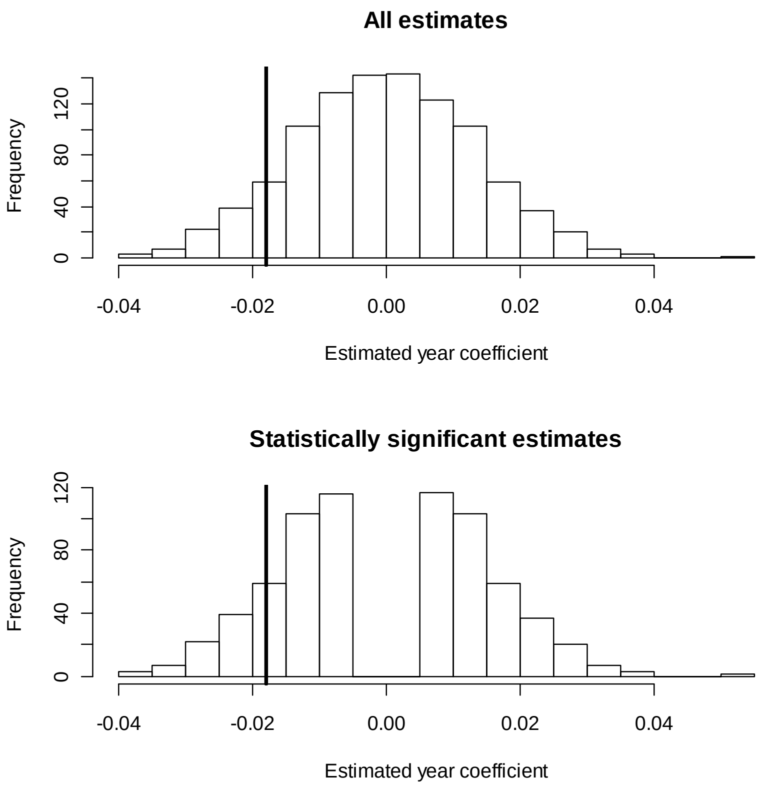

All estimates
Frequency
0
40
80
120
-0.04
-0.02
0.00
0.02
0.04
Estimated year coefficient
Statistically significant estimates
Frequency
0
40
80
120
-0.04
-0.02
0.00
0.02
0.04
Estimated year coefficient

Fig. 14

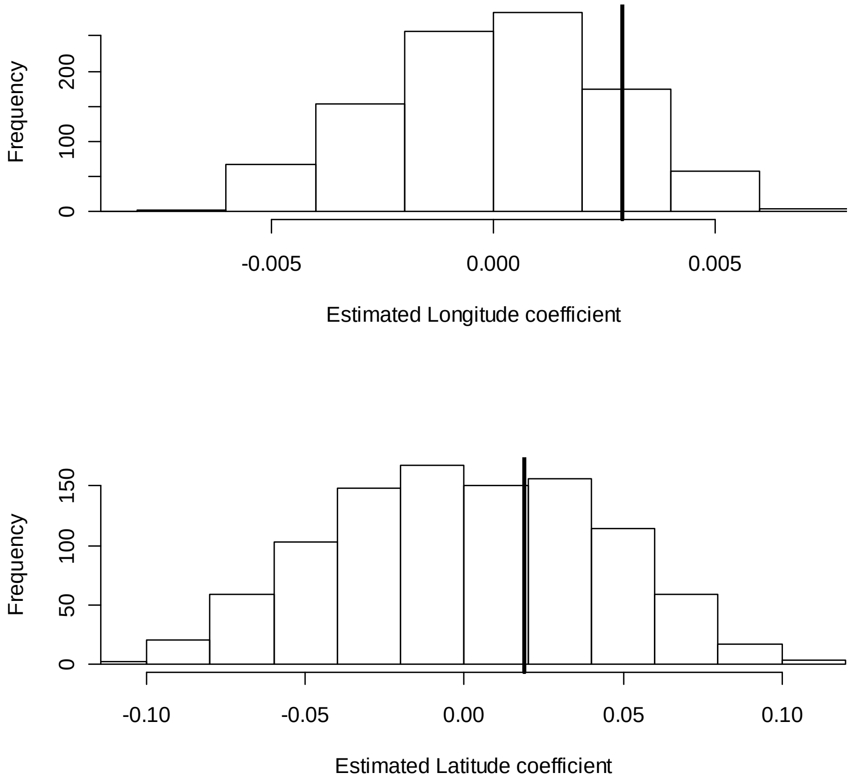

Frequency
0
100
200
-0.005
0.000
0.005
Estimated Longitude coefficient
Frequency
0
50
100
150
-0.10
-0.05
0.00
0.05
0.10
Estimated Latitude coefficient

Fig. 15

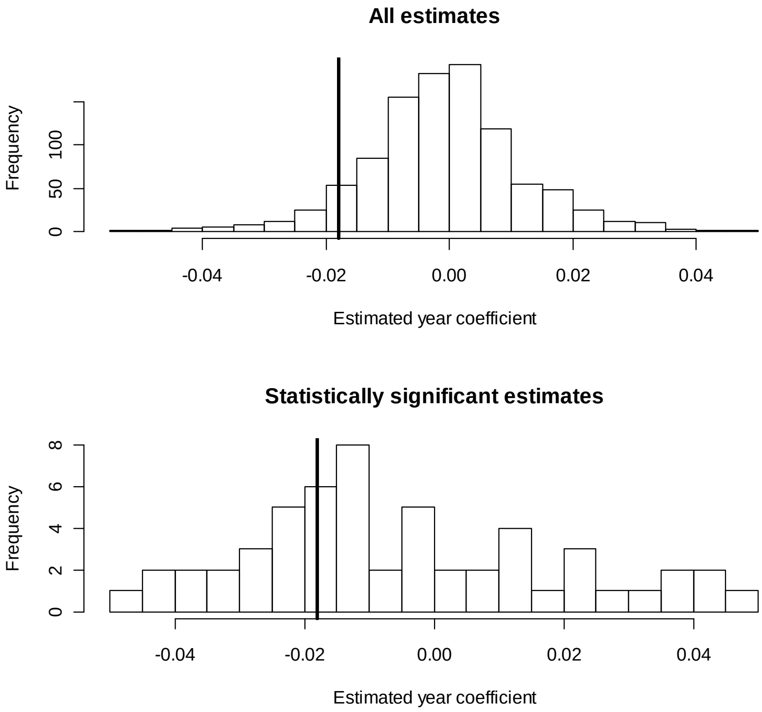

All estimates
Frequency
0
50
100
-0.04
-0.02
0.00
0.02
0.04
Estimated year coefficient
Statistically significant estimates
Frequency
0
2
4
6
8
-0.04
-0.02
0.00
0.02
0.04
Estimated year coefficient

Fig. 16

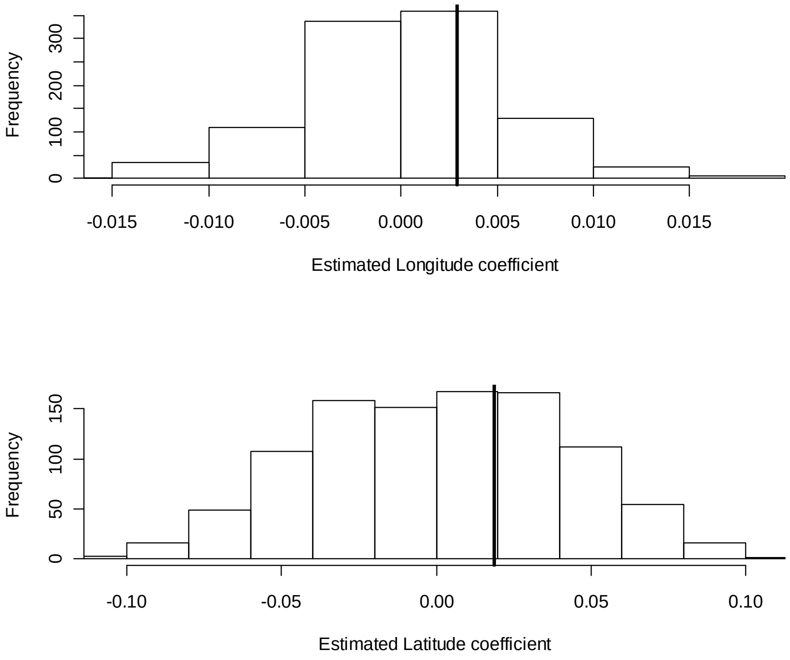

Frequency
0
100
200
300
-0.015
-0.010
-0.005
0.000
0.005
0.010
0.015
Estimated Longitude coefficient
Frequency
0
50
100
150
-0.10
-0.05
0.00
0.05
0.10
Estimated Latitude coefficient

Fig. 17

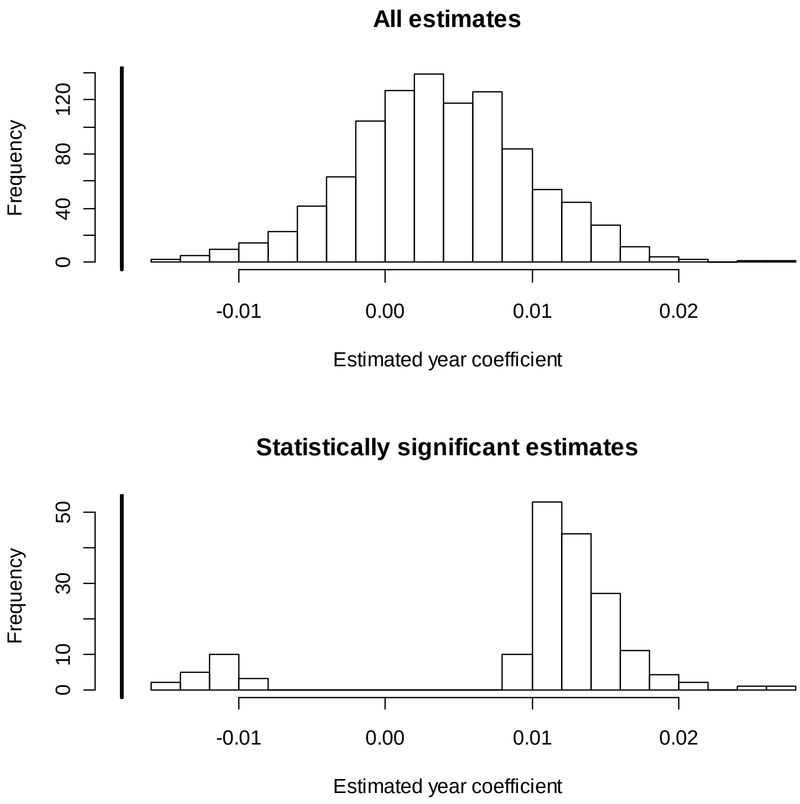

All estimates
Frequency
0
40
80
120
-0.01
0.00
0.01
0.02
Estimated year coefficient
Statistically significant estimates
Frequency
0
10
30
50
-0.01
0.00
0.01
0.02
Estimated year coefficient

Fig. 18

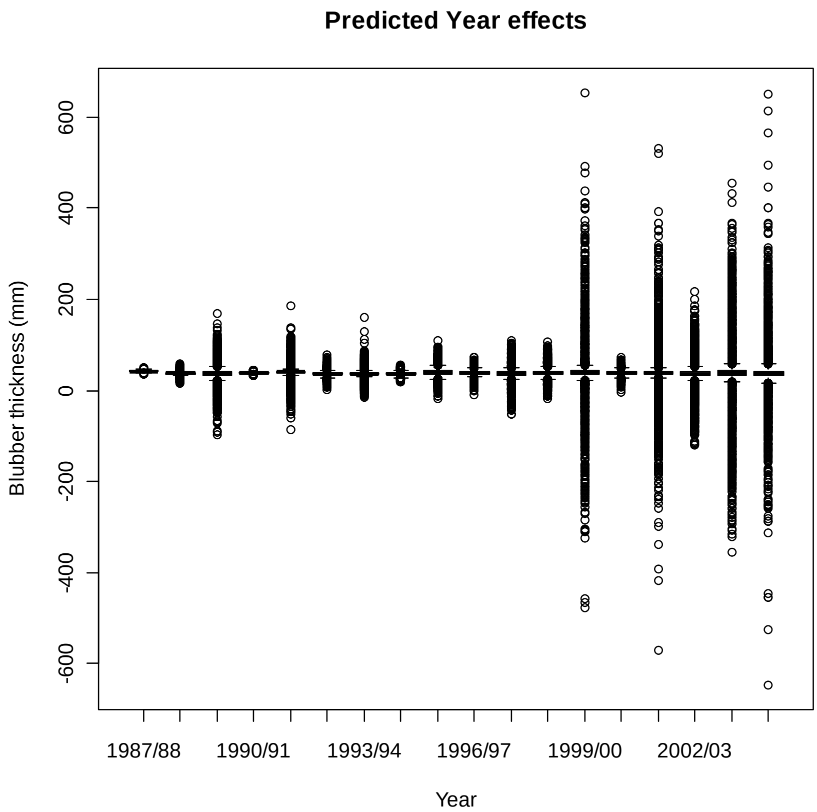

Predicted Year effects
Blubber thickness (mm)
600
400
200
0
-200
-400
-600
1987/88
1990/91
1993/94
1996/97
1999/00
2002/03
Year

Fig. 19.

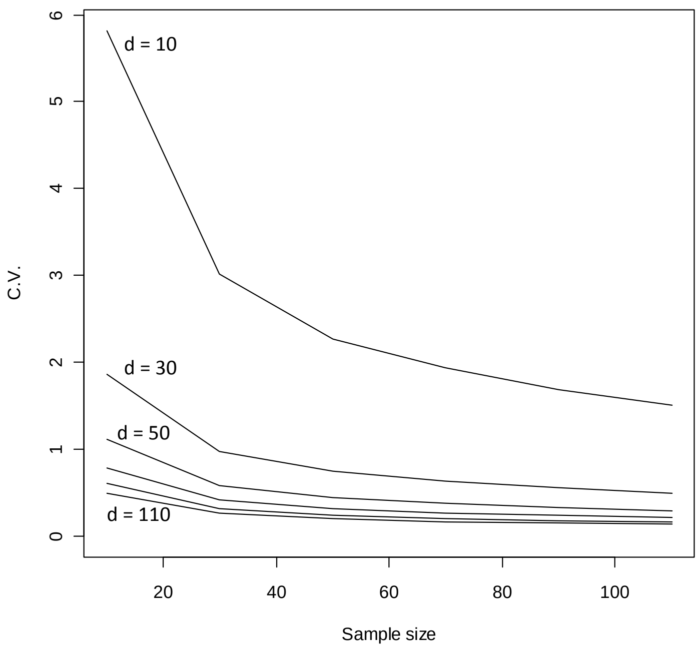

C.V.
6
5
4
3
2
1
0
d = 10
d = 30
d = 50
d = 110
20
40
60
80
100
Sample size

Fig. 20.

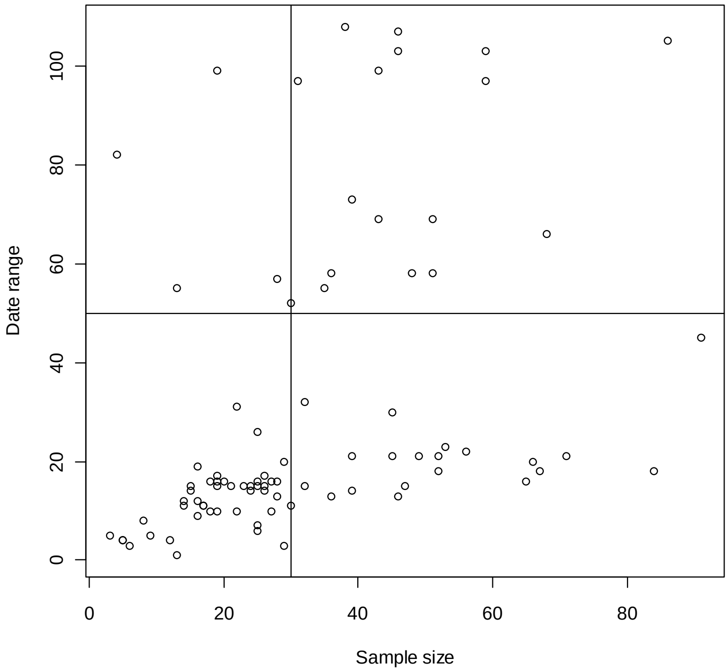

Date range
Sample size
0
20
40
60
80
100

Fig. 21.

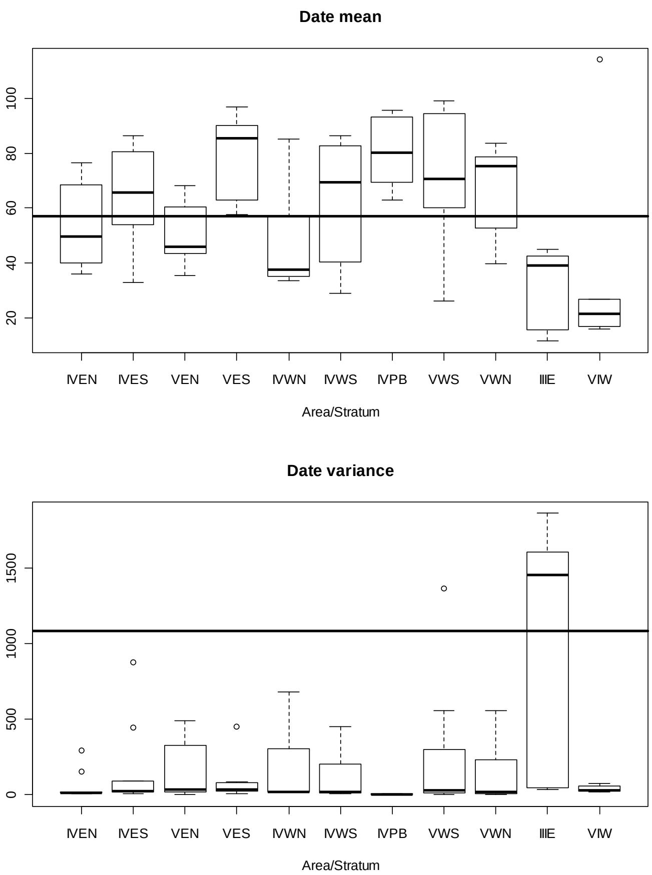
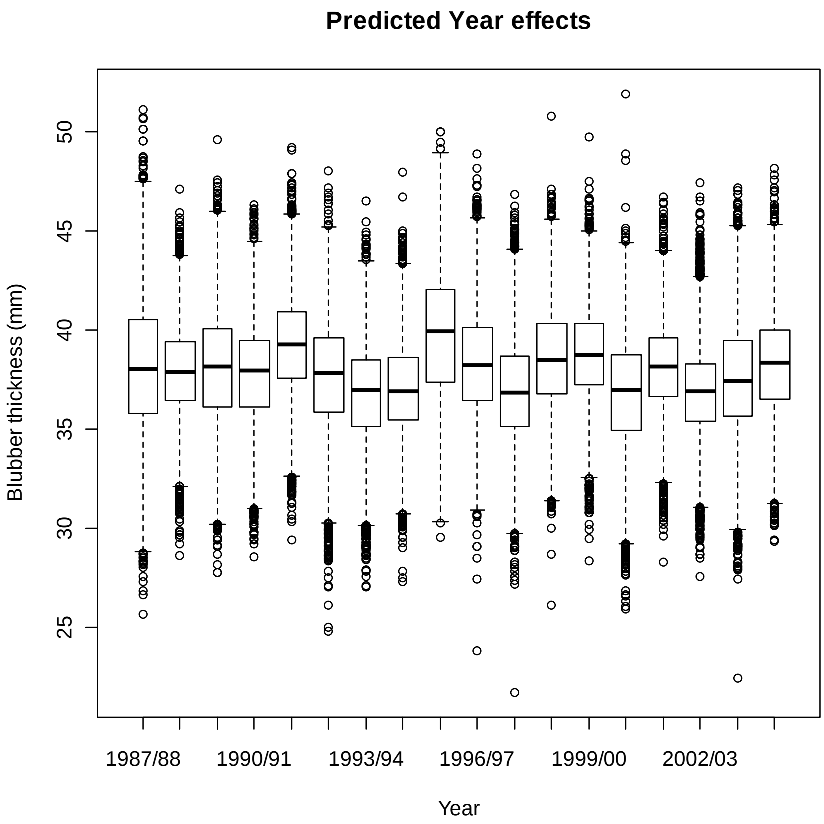

Date mean
Date variance
Area/Stratum
IVEN
IVES
VEN
VES
IVWN
IVWS
IVPB
VWS
VWN
IIIE
VIW

Fig. 22.



Predicted Year effects
Blubber thickness (mm)
25
30
35
40
45
50
1987/88
1990/91
1993/94
1996/97
1999/00
2002/03
Year

Fig. 23.