

Estimating relative abundance from historic Antarctic whaling records.

WILLIAM K. de la MARE

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

ABSTRACT

Catch per unit effort data (CPUE) are often the only form of data available from historic fisheries that can be used to infer patterns of distribution and abundance of exploited populations. Information derived from CPUE underestimates variations in relative abundance when effort data is only measured in total operating days. Gross effort includes both searching time and handling time, but it is only the first of these times that is useful in deriving an index of relative abundance. A method is developed for improving the linearity of the relationship between relative abundance and CPUE by estimating the searching time. The searching time is found by subtracting an estimate of time lost due to handling from the gross effort. However, an additional correction is required if handling time can occur past the end of the operating day. An expectation maximisation (EM) algorithm is used to combine maximum likelihood estimates of the handling time with the expected additional operating time due to handling the last catch of each day. Simulation tests show that the method leads to estimates of catch per unit of searching time (C/CSW) that are much closer to proportionally related to local density than gross CPUE. However, the method does not produce unbiased estimates of handling time and some non-linearity can remain in the relationship between local density and catch per unit of searching time. Although developed to improve the analysis of historic whaling data, the methods may be useful for other fisheries where only historic gross catch and effort data are available and that have time budgets that involve both searching and handling.

KEYWORDS: ANTARCTICA, CPUE, SEARCHING TIME, WHALING, SIMULATION

INTRODUCTION

It is a general problem that much of the data from historic fisheries comprise only basic catch data and gross effort. It is well established that such data do not provide a linear index of abundance because of the effects of the time lost to searching due to the time spent handling the catch (Beddington, 1979; Cooke, 1985). Various methods have been proposed to correct for handling time, e.g. Sampson (1988). The method developed here is derived from a formal approach to estimating searching and handling time directly from a statistical model. Although applied here to whaling, similar problems exist in other fisheries where the operating time budget includes both searching and handling (Punsley, 1987; Maunder *et al.* 2006).

Catch per unit effort (CPUE) was an important data source in historic attempts to manage whaling. For example the first quantitative assessments of the state of Antarctic whale populations (e.g. IWC, 1964) relied heavily on such data. CPUE data was instrumental in protecting some of the most severely depleted whale populations. However, CPUE data often showed no trend and hence methods of estimation such as the de Lury method (Chapman, 1974) used at the time could not always be successfully applied. It was also recognised that the relationship between abundance and CPUE would be non-linear due to the effects of handling time. Beddington (1979) proposed that the relationship would take the form of Holling's "disc" equation (Holling, 1965):

$$\frac{C}{s} = \frac{C}{E' - hC} \quad (1)$$

Where C is the catch taken for a given amount of catcher searching time worked s (CSW). E' is the total operating and h is the time taken to handle each catch. However E' is not recorded in typical historical data, only the numbers of days and catcher- vessels, from which we can calculate the gross effort in terms of catcher-days. An **E-M** algorithm (Dempster *et al.*, 1976) is applied here to deal with the latent variable E' , by using its expected value. Another complication is that the data do not include the days when searching took place but no catches were taken, and so the maximum likelihood estimator used here is based on the distribution of searching time given the number of whales taken. The component of searching time not recorded because the data are censored for days with zero catch is also corrected for by using its expected value in the E-M algorithm.

Since the 1980s CPUE has been overtaken in whale population abundance estimation by the use of sightings surveys, with the population abundance over time being derived from population models. However, some species were substantially depleted by whaling and sightings are still too few to develop clear patterns of spatial and temporal distributions. So for these species the commercial catch data are of interest in determining the spatial distributions of whales prior to their substantial depletion, and as more information on current distribution accrues, how these distributions may have changed. For these reasons it is still worth trying to analyse whale catch records, even though we would now not use the data to form a time series index of relative abundance.

There are many difficulties in interpreting catch and effort data (Maunder *et al.*, 2006), and many reasons why C/CSW will not form a linear index of abundance (Cooke, 1985). C/CSW was usually assumed to be

proportional to local whale density, but the constant of proportionality is unlikely to be constant over time due to, for example, changes in whaling efficiency and effort being concentrating on preferred species. For example, in the 1930s blue whales were the preferred species and hence a low C/CSW for fin whales in a given place does not imply that fin whales were scarce on the whaling grounds, only that when fin whales were found there would be a low probability of chasing them if the whaling crews expected to encounter a blue whale in the near future. Thus, fin whale C/CSW will increase as the abundance of blue whales declines, but this does not mean that the abundance of fin whales increased, only the relative probability of catching them. However, the aim for developing the methods here is not to attempt to derive a time series index of relative abundance to infer population trends over long periods, but rather to improve the contrast in measures of spatial relative abundance within years or blocks of several years. These improved indices will help in identifying spatial distributions of species abundance and assist in identifying the nature of possible inter-species interactions. All analyses are implemented using R (R Development Core Team, 2011) and the scripts are available on request.

MODELS

Assume that within a small locality and short period of time that the probability of catching a whale for a given amount of searching time is constant and follows a Poisson process (i.e. whales are locally randomly and independently distributed and sufficiently abundant so that local depletion does not occur within the short period). The searching time is interspersed with episodes of handling time after each whale is taken and before searching resumes. Search time on day i is given by:

$$s_i = E'_i - hC_i \quad | hC_i \leq E'_i \quad (2)$$

where E'_i is the total operating time, h is the handling time for each whale taken and C_i is the total catch. It is assumed that h is the same for all species and whaling expeditions. The total operating time consists of a period in which searching is feasible, which is defined as the nominal operating day E_i . However, there may be some additional operating time (\tilde{E}_i) incurred if handling the last catch of the day continues past the end of the nominal operating day. Consequently;

$$E'_i = E_i + \tilde{E}_i \quad (3)$$

There are two classes of searching time; those which result in the capture of a whale (successful), and those which terminate at the end of the nominal operating day without a capture (unsuccessful). For each catcher, a day may consist of several successful searches followed by zero or one unsuccessful searches. There are zero unsuccessful searches when the handling time for the last catch of the day overlaps the end of the nominal operating day E , otherwise there is one unsuccessful search. The total searching time is given by:

$$s_i = x_i + y_i \quad (4)$$

where x_i and y_i are the searching times for successful and unsuccessful searches respectively. The estimator developed here treats a series of days ($i = 1 \dots n$) in the same locality as replicates. On each day a number of whales are taken, so that the calculated estimate of C/CSW (denoted λ below) for a given locality and time period is given by:

$$\lambda = \frac{\sum_i C_i}{\sum_i s_i} \quad (5)$$

However, the days on which searching occurs with zero catches are not recorded in the original data. If not accounted for, the omission of such days will lead to upward bias in the calculated density estimates.

The data available give only the total number of catchers in operation for each expedition. An expedition comprises a factory ship and a number of catcher-vessels. In the 1930s there were on average about 6-8 catchers per expedition. In the 1950s the number virtually doubled. There are two aspects to converting the recorded number of catcher days into catcher operating time. The first is allowing for the amount of daylight available for searching. To a first approximation this can be taken as the time between sunrise and sunset, which can be calculated for specified dates and latitudes using standard astronomical formulae. The second and more difficult aspect is to allow for the probability that the last catch of the day by each catcher vessel will entail some handling time after sunset. Not allowing for this possibility will bias the estimates of C/CSW because the amount of searching time will be too low if it is assumed that the last catch of the day will only occur at a time before sunset strictly no closer than one handling time, that is $\tilde{E}_i \equiv 0$. Such an assumption seems very unlikely. Since there is no detailed time budget data available for the early whaling operations the length of the operating day is replaced with its expected value when calculating C/CSW . The operating day used here is based on

adding to the nominal operating day the expected value of handling time that occurs after sunset. Generally, subject to the assumption that searches are independent and identically distributed events, for N catcher-vessels the expected amount of operating time per day is given by:

$$E' = N \left(\tau + \sum_{r=1}^R \int_T^{\tau} (x+h-\tau) d(x-rh; r, \dots) dx \right) \quad (6)$$

where $d(x; r, \dots)$ is a probability density function for the search time required to take r catches, τ is the length of the nominal operating day and:

$$R = \left\lceil \frac{\tau}{h} \right\rceil \quad (7)$$

and;

$$T = \sup \{ \tau - h, (r-1)h \} \quad (8)$$

This formulation assumes that if a whale were sighted just before sunset that it would be possible to finish chasing and harpooning the whale during twilight. This is not unreasonable because twilight is quite long in the Antarctic during the whaling season. However, if elsewhere this was known to be unlikely, then h could be reduced by a proportion in equation (6) and other equations as appropriate. In high latitudes in summer there will be continuous daylight. If catching operations were also continuous then the handling time after midnight should be subtracted from the following day's working hours. This adjustment is ignored here because with typical Antarctic whaling latitudes and dates the number of days with continuous daylight is small (less than 0.4% – see Fig. 1). Given the further assumption that catching is locally a Poisson process, it does not matter that a search that crosses midnight is partitioned across two days.

If it is assumed that catching is strictly a Poisson process, the density function for the amount of searching time to take r whales has a gamma distribution:

$$f(x; r, \lambda) = \frac{\lambda^r}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} \quad | x, r, \lambda > 0 \quad (9)$$

where λ is the catch rate per unit searching time (C/CSW) and x is the cumulative (successful) searching time expended in catching r whales. $\Gamma(\cdot)$ is the gamma function.

However, de la Mare (1986) showed that CPUE data are consistent with statistical distributions that have λ as a random variable. Weather, for example, has an important effect on daily catch rates (Ohsumi and Yamamura, 1978, Zahl, 1985). Any time trend in local abundance due to local depletion or whale movements will also contribute to the apparent variability in C/CSW. Thus this model has less restrictive assumptions than the gamma model.

A reasonable distribution for such cases is the compound gamma distribution (or gamma mixture of gammas) in which the searching time has a gamma distribution as in equation (9), but with λ also having a gamma distribution:

$$g(\lambda; a, b) = \frac{a(a\lambda)^{b-1} e^{-a\lambda}}{\Gamma(b)} \quad (10)$$

The mean value of λ is locally constant and given by:

$$E[\lambda] = \frac{b}{a} \quad (11)$$

The coefficient of variation for λ is given by:

$$\kappa_\lambda = \frac{1}{\sqrt{b}} \quad | b > 2 \quad (12)$$

The compound gamma distribution is a type 2 inverse beta distribution which has the following density function:

$$h(x; r, a, b) = \frac{x^{r-1} a^b}{B(r, b)(x+a)^{r+b}} \quad (13)$$

where $B(\dots)$ is the incomplete beta function. The distribution does not have a proper variance for $b \leq 2$.

The probability density functions considered further here for equation (6) are those given in equations (9) and (13). The integrals can be evaluated by numerical methods quite quickly by making use of standard numerical approximations for the cumulative distribution functions for the gamma or compound gamma distributions. The corrected operating time forms the expectation (**E**) step for the **E-M** algorithm. Fig. 2 shows the calculated and simulated values for the additional handling times for the compound gamma model. The points represent simulated values; the curves are calculated by solving (6) by numerical methods.

The data consist of a number of catches taken each day by each expedition. To develop a maximum likelihood estimator for any of the parameters requires a method to calculate the probabilities that the catches on given days have the observed values. The cumulative probability of taking less than a given number of whales (r) can be found from the probability density functions for the distributions of (successful) search time x to take a given catch r . Including the allowance for additional handling time, the amount of search time available to take r whales on a given day is:

$$s_r = N\tau - (r-1)h \quad (14)$$

The probability that the catch taken (c) is less than r is given by

$$p(c < r; N, \tau) = p(x_r > N\tau) = 1 - p(x_r \leq N\tau) \quad (15)$$

where x_r has a probability density function for the search time used to take r whales. For the gamma and compound gamma waiting time distributions the order of events is unimportant, so that generically, the cumulative probability for either model can be calculated as:

$$p(c < r; N, \tau, h) = 1 - \int_{(r-1)h}^{N\tau} d(x - (r-1)h; r \dots) dx \quad (16)$$

where the additional parameters are either λ or a and b for the gamma or compound models respectively. The integral is readily evaluated using standard numerical approximations for the cumulative distributions of the gamma or compound gamma distributions. The probability of taking exactly r whales is given by:

$$p(r; N, \tau, h) = p(c < r+1) - p(c < r) \quad (17)$$

Thus, the log likelihood function for a set of n observed daily catches is given by:

$$L(h, \dots | C_1 \dots C_n) = \sum_{i=1}^n \ln(p(C_i)) \quad (18)$$

The maximum likelihood estimators for h will form the maximisation (**M**) step in applying an **E-M** algorithm. In this paper no distinction between species is made when applying the **E-M** algorithm, and so the handling time is averaged over the species. The missing observations from days with zero catch are uninformative about the handling time. Essentially h is a nuisance parameter that needs to be estimated in order to calculate CSW. However, h is historically interesting because of what it shows about the development of the Antarctic whaling industry.

Equation (16) can also be used to calculate the probability that a zero catch will be taken for a given amount of search effort. This can be used for correcting for the unrecorded amount of searching time expended on days when no whales were caught. Each day of searching can be considered a Bernoulli trial with a probability $1 - p(0)$ of an expedition catching at least one whale (success). The expected number of days without a catch, m , to obtain k days with catches has a negative binomial distribution. With constant effort, the expected value of m is given by;

$$E[m] = \frac{k p(0)}{1 - p(0)} \quad (19)$$

However, the amount of effort in a given location can change each day, for example, when an additional expedition arrives or departs. Consequently $p(0)$ can vary each day. Thus, as part of the **E** step, an additional correction for the hypothetical days without catches can be added to the cumulative search time;

$$E'' = \frac{\tau}{n} \left[\sum_{i=1}^n N_i \right] \left[\frac{\bar{p}(0)}{1 - \bar{p}(0)} \right] \quad (20)$$

where;

$$\bar{p}(0) = \frac{1}{n} \sum_{i=1}^n p(0; N_i, \tau) \quad (21)$$

This formulation allows for the amount effort to vary each day by averaging over n trials of one day (and hence $k = 1$). With the typical effort in each expedition, the correction will be small for $\lambda > 1$. There may be some localities and time periods where searching occurred but no whales were taken; in such cases the correction for missing search days cannot be calculated. This is a source of bias if total search time is pooled for combinations of localities and time periods.

In principle, the likelihood functions could also be used to find the estimates for λ or a and b . However, simulation tests show that the **E-M** estimates have better properties for data with the very heterogeneous numbers of replicates at the different locations and time periods observed in historic Antarctic whaling. The calculated values of λ explicitly include the estimated searching time lost in unsuccessful searches on the days when at least one whale was caught and the EM algorithm also allows correction for the censoring of the zero catch records and the unrecorded additional operating time. Given the maximum likelihood estimate of handling time, the total search time can be calculated by means of equation (3) and the C/CSW of by species can also be obtained from equation (5) by substituting the catch of each species.

ESTIMATION METHOD

The method is applied by dividing the daily catch records into cells corresponding to unique latitude and longitude regions and specified time periods. The estimators are applied to find the maximum likelihood estimate of a common value of h across all the cells (**M** step) after finding for each cell the local estimate of λ , using the expected total operating time E' from (14) and including E'' from (20) (the **E** step). For the compound model the value of b is also found by maximum likelihood subject to $b > 2$ and using $a = b/\lambda$. The **E-M** steps are iterated until sufficient convergence is obtained; this usually takes less than 10 iterations to converge to an accuracy of 1 part in 10000.

This approach gives the marginal likelihood for the handling time, and so naturally leads to estimating the confidence limits from likelihood ratios or for forming Bayes' posterior distributions for handling time and density if required. In reality, it is likely that both handling time and daily catch rates are random variables, and they are probably correlated as well. However, it is very unlikely that the effects of variability in handling time could be disentangled from variability in daily catch rates. The approach here is to assign the all the variability to the latter. Handling time as estimated here may actually also include some operational time lost various other activities such as re-fuelling, stop-catch periods and relocation (stop-catch periods occur when the factories seek to avoid a glut of whales to process).

SIMULATION TESTS

The estimates of λ can be expected to be biased because a random variable (the estimated handling time) is used in the denominator. Simulations are used here to determine the likely nature of any bias. Simulated data are generated either according to the gamma or compound gamma models for various fixed levels of handling time. The simulated data are generated by direct simulation of the catching process by N catchers. Each catcher searches for the next whale with random searching time drawn from an exponential distribution (equation [9] with $r = 1$). In the gamma model λ is fixed for each locality, whereas in the compound model each day's λ is drawn from a gamma distribution (equation [10]) with CVs of 0.2 or 0.4. The search time for each whale is subtracted from the remaining nominal operating time (nominal operating time is from sunrise to sunset) for each catcher along with one instance of handling time. If any nominal operating time remains for a catcher after subtracting the searching and handling time for the most recent whale then a new search commences. The new search either leads to another capture or is truncated by the end of the nominal operating day. However, if a whale is sighted by a given catcher near the end of the operating day it can be taken even if the handling time will extend past the end of the nominal operating day. Consequently, additional operating time is incurred for the day. Thus, the simulated data conform to the assumptions of the estimation method.

In the trials presented here the data generated are consistent with the properties of whale catch and effort data derived from the Antarctic pelagic whaling industry (data are from IWC, 2010). The "true" levels of λ are derived from the observed daily catches from pelagic whaling in two periods (1930-1940) and (1950-1955). The first period had daily catches conforming roughly to a gamma distribution with mean = 1.63 and variance = 0.96. This is consistent with the observed property from the 1930s data that there is a peak in the daily catch distribution, but there is also a negligible probability of catching more than 6 whales per catcher per operating day. This upper limit of 6 whales per day suggests that handling time will be less than 4 hours (0.167 days) in the 1930s. For the 1950s the expected daily catch is simulated with a random number drawn from an exponential distribution with a mean = 0.586. This is consistent with the observed property that most of the observed daily catches are quite small, with no obvious peak away from the origin, and with a negligible probability of taking more than 10 whales per catcher per operating day. The observed values suggest that handling time in the 1950s was less than 2.4 hours (0.1 days).

The expected daily catch \bar{C}_i for cell i is converted into $\bar{\lambda}_i$ using:

$$\bar{\lambda}_i = \frac{\bar{C}_i}{\tilde{d}_i - \bar{C}_i h} \quad | \quad \tilde{d}_i > \bar{C}_i h \quad (22)$$

where \tilde{d}_i is a randomly generated day-length for cell i . The generated values for $\bar{\lambda}_i$ are truncated at 100.

Nominal operating day-lengths are drawn from random normal distributions with the same means and standard deviations as calculated for the respective early or late periods (early mean = 0.708, early std. dev. = 0.087; late mean = 0.706, late std. dev. = 0.095).

Partitioning the data into the cells results in them containing a heterogeneous number of days in which catches occurred. In the application of the methods to Antarctic pelagic whaling data the cells are defined by 1° of latitude and longitude and by dividing each month into three approximately ten-day periods. The most common number of operating days per cell is one; only a few cells have more than six observed catching days. This is simulated by setting the number of days with catch in each cell by rounding up a random number drawn from an exponential distribution with a mean = 0.4.

In the 1930's the average number of cells containing data is 1038 per year and in the period 1946-1955 data exist in an average of 912 cells per year. The simulation tests were based on 50 replicates of applying both the gamma and compound gamma methods to 1000 cells in which simulated catch and nominal effort data were generated. The simulated data are conditioned on the general properties of Antarctic whaling data described above.

Results of the trials are presented in Figs 3-10. To avoid obscuring the results from over-plotting, each density scatterplot shows a randomly selected (without replacement) subset of 10% of the 50000 estimates from each trial. The scatterplots are on log-log scales. The scatterplots are annotated with the residual coefficient of variation, which is estimated by fitting a linear model to logarithms of the simulated data.

Fig. 3 shows that the estimates of handling time from applying the gamma model estimator to the early simulated data. The gamma model is the most appropriate for the case when the CV in daily catch rate is zero, i.e. the left column. Clearly, the distributions of handling times exhibit systematic bias that depends on the handling time, being biased high for the low handling time (0.05 days) and low for the high handling time (0.15 days). The left column of the scatterplots in Fig 4 shows that the biases in handling time estimates cause some departure from the 1:1 line at higher densities, as would be expected given equation (1). The other two columns of Fig 3 also shows that the gamma model estimates are not robust to the effects of random variability in daily catch rates (λ); becoming attracted to either end of the interval used in the numerical search for the maximum likelihood. At the higher values of CV the estimates have failed completely for the lower handling times. The rightmost column of Fig 4 shows the estimates of λ to be biased low by up to an order of magnitude at high densities. Figs 7 and 8 show the results for trials conditioned on data from the early 1950s. Overall these results are consistent with those from the early period simulations. There are no results presented for a handling time = 0.15 days (3.6 hours) since this is not consistent with the observed catches per catcher day.

Figs 5, 6, 9 and 10 give the results from the compound gamma estimator using the same sets of early and late simulated data as was used with the gamma estimator. The left column of Fig 5 shows that in terms of bias, the compound estimator is not substantially worse than the gamma estimator, even though the latter is correct in this case. However, the distributions of estimates are wider. These results demonstrate that the compound estimator is reasonably robust to the failure of the assumption that daily C/CSW is stochastic. However the bias in the handling time estimates do lead to some bias in estimates of C/CSW at higher densities. For the cases where the C/CSW is stochastic, the compound estimator is less biased than the gamma estimator, and does not fail at the higher CV. Although some bias exists in the estimates of handling times, the scatterplots are reasonably linear, although the cases with the highest handling time do underestimate the higher densities. The results conditioned on the 1950s data (Fig 9) show less bias and less variability in the estimates of handling time than for the trials conditioned on the 1930s data.

Additional trials were carried out where both the daily lambda and handling time are (correlated) random variables, such that poorer searching conditions tends to lead to lower values of λ and longer handling times. Each day's λ is drawn from a gamma distribution with a CV=0.2 and each days handling time is generated from:

$$\tilde{h} = \bar{h} \left(0.3 + \frac{0.3}{1 + e^{-0.1(10 - \tilde{\lambda})}} \right) + \varepsilon \quad (23)$$

where \tilde{h} is the stochastic handling time for a particular day, \bar{h} is the mean value of handling time and $\tilde{\lambda}$ is the stochastic realisation of λ ; ε is a random number drawn from a gamma distribution with expected value $0.5\bar{h}$ and a CV = 0.2. This leads to an overall variability in the daily catch rate roughly equal to a CV of 0.3. Not surprisingly, Figs 11 and 13 show the distributions of estimates of handling times are broader. The median values of the estimates are lower, but the overall pattern of bias is similar to that from the fixed handling time

trials. The scatterplots of estimated versus true C/CSW (Figs 12 and 14) shows that the compound estimator is reasonably robust to the effects of random variability in daily handling times in the range tested.

The gamma based estimator is not used in the analyses presented below because it is not robust to stochasticity in daily C/CSW .

APPLICATION TO ANTARCTIC PELAGIC WHALING

In this application a single handling time is estimated for each year using the compound model. The data are divided spatially into 1° latitude by 1° longitude cells and into 3 time periods per month according to day 1-10, 11-20, >20. Consequently the method yields an estimate of λ and b for each $1^\circ \times 1^\circ$ cell and ten day period.

Fig. 15 shows the estimates for handling times estimated for each year from 1930 to 1986. Only data south of $50^\circ S$ are used and days when only sperm whales were caught (predominantly before the opening of the baleen whale season) are excluded. The estimate for 1973 appears to be an outlier. The low estimate for 1965 is also probably unreliable, and comes from the period of considerable turmoil in the industry after blue and humpback whales were protected by the IWC.

The estimates show a general declining trend in handling time consistent with improvements in catching efficiency. Interestingly the changes in the handling time are also consistent with changing conditions under which whaling was conducted. By 1950 a high proportion of catcher vessels were more powerful and equipped with both radar and "whale scaring" sonar as well as nylon rope for the harpoon line (Tonnesen and Johnsen, 1982). These innovations had substantial effects on handling times, as is clear from the estimates. In 1955 the "Sanctuary" in the Pacific sector was opened and handling times fell further due to the greater abundance of whales (particularly fin whales) in this region. Handling times increase after blue and humpback whales were protected and catching operations move northwards to concentrate on sei whales, although fin whales were still regularly encountered there as well. However, densities were lower for both these species in this region, which is the probable reason for the longer estimated handling time. After 1972 a growing proportion of the catches were minke whales, which can be found in high densities near the ice-edge. Consequently the handling time again falls for a few years. This was particularly so for early 1970s when up to 18 minke whales were sometimes taken in a day by a single catcher. The handling time increases through to 1986 when the IWC "moratorium" comes into effect. The increase in apparent handling time is most likely driven by the industry maximising production through selecting for larger minke whales and the regular occurrence of "stop catch periods" (Ohsumi, 1979).

Ohsumi (1979) gives direct estimates of handling time (including "stop catch" or "resting" time) for Japanese Minke operations in 1976 and 1977 of 1.9 hours per whale. The largest component of the time budget is resting time (1.0 hours), of which an unspecified proportion occurs at night. Yamamura and Ohsumi (1981) give direct estimates for handling time from Japanese and USSR data for Antarctic minke whaling for an unspecified set of years in the 1970s; Japan = 0.7 hours, USSR = 2.0 hours, average = 1.35 hours. The combined compound gamma estimates obtained here for the Japanese and USSR operations averaged over 1974 to 1979 is 1.50 hours. The results are thus consistent with the direct time budget data; although the direct estimates in Yamamura and Ohsumi (1981) do not include stop-catch and resting time.

The estimated daily variability (CV) in C/CSW , shown in Fig. 16, also appears to be related to changes in the abundance of whales (the estimated CV in 1969 was virtually zero and is omitted as not reliable because the data records include only 39 whales taken south of $50^\circ S$ in that year). The CV tends to increase as the abundance of whales declines through to the 1960s, but decreases as the industry switches to more abundant species thereafter. This is consistent with the distributions of whales becoming patchier with declining abundance. The CV declines further throughout the minke whaling period as industrial strategies become focussed on reducing the variability in daily production.

Figs 17 and 18 show plots of the apparent densities of fin whales estimated as the catch/catcher day (Fig 17) and the C/CSW estimated with the compound gamma estimator (Fig 18) by applying equation (6). The colours represent the apparent densities plotted on common scale. Comparing the figures shows that the estimation of the handling and searching time has been successful in enhancing the contrast in the data, such that patterns of distribution difficult to discern in Fig 17 are much clearer in Fig 18.

CONCLUSION AND DISCUSSION

The aim in developing these methods was to improve the contrast in relative abundance data. The simulation tests show that the gamma model based method is not robust to the effects of stochastic variation in daily catch rates. The compound gamma estimator performs better under these realistic circumstances and is therefore to be preferred. Estimates of handling time and density are not unbiased and the direction of bias is different for high and low handling times. However the effects of the biases in estimated handling times are not important at low densities. Higher density estimates may be biased either high or low depending on the bias in handling time. Given, for example, that the whaling CPUE saturates at six to ten whales per day regardless of the true density, the simulations allow for the number of whales that would be caught per day in the absence of handling time to exceed the maximum CPUE by an order of magnitude. The method of estimation is shown in these

circumstances capable of producing an index of abundance that is reasonably linearly related to true density over 2-3 orders of magnitude.

There are many complications in the real operations that cannot be accounted for by purely statistical modelling based on crude catch statistics. These include:

- whaling preferences for the larger and more profitable whales will change the apparent abundance of the different species over time
- searching speed and efficiency increases over time
- inaccurate reporting
- change in the frequency of failed chases
- different species will have different handling times
- cooperative catching – groups of whales will be reported to other catchers
- use of scouting vessels
- changes in production priorities that lead to “stop catch” periods

Consequently, it is highly unlikely that any method exists that would transform the rudimentary available data into a fully linear index of abundance, particularly over decades. However, the tests of the methods developed here demonstrate that they are at least capable of attaining the more modest aim of improving on the catch per catcher day as a measure of relative local abundance. The intent is to apply these methods over restricted time periods to make inferences about the spatial and within season distributions of species in the Antarctic. These studies will be reported elsewhere.

REFERENCES

- Beddington, J. R. 1979. On some problems of estimating population abundance from catch data. *Rep. int. Whal. Commn* 29:149-54.
- Chapman, D. G. 1974. Estimation of population size and sustainable yield of sei whales in the Antarctic. *Rep. int. Whal. Commn* 24:82-90.
- Cooke, J. G. 1985. On the relationship between catch per unit effort and whale abundance. *Rep. int. Whal. Commn* 35:511-9.
- de la Mare, W. K. 1986. Further consideration of the statistical properties of catch and effort data, with particular reference to fitting population models to indices of relative abundance. *Rep. int. Whal. Commn* 36:419-23.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B* 39(1):1-38.
- Holling, C. S. 1966. The functional response of invertebrate predators to prey density. *Mem. Entomological Soc. Canada* 98:(48) 1-86.
- IWC, 1964. Special Committee of Three Scientists Final Report. *Rep. int. Whal. Commn* 14:40-92.
- IWC, 2010. The IWC catch database. International Whaling Commission, Cambridge.
- Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P. and Harley, S. J. 2006. Interpreting catch per unit effort to assess the status of individual stocks and communities. *ICES J. Mar Sci.* 63:1373-85.
- Ohsumi, S. and Yamamura, K. 1978. Catcher's hours' work and its correction as a measure of fishing effort for sei whales in the Antarctic. *Rep. int. Whal. Commn* 28:459-67.
- Ohsumi, S. 1979. Population assessment of the Antarctic minke whale. *Rep. int. Whal. Commn* 29:407-20.
- Punsly, R. 1987. Estimation of the relative annual abundance of yellowfin tuna, *Thunnus albacares*, in the eastern Pacific Ocean during 1970-1985. *Inter-American Tropical Tuna Commission Bulletin*, 19: 263-306.
- Sampson, D. B. 1988. Fish capture as a stochastic process. *J. Cons. int. Explor. Mer* 45:39-60.
- Tønnesen, J. N. and Johnsen, A. O. 1982. *History of modern whaling*. Hurst and Co. London. 798pp.
- Yamamura, K. and Ohsumi, S. 1981. Comparison of the yearly changes in CPUE for minke whales in the Antarctic based on Japanese and Soviet effort data. *Rep. int. Whal. Commn* 31:327-32.
- Zahl, S. 1985. Revised model for adjustment of Antarctic Japanese minke whale data 1973-1982. *Rep. int. Whal. Commn* 35:223-6.

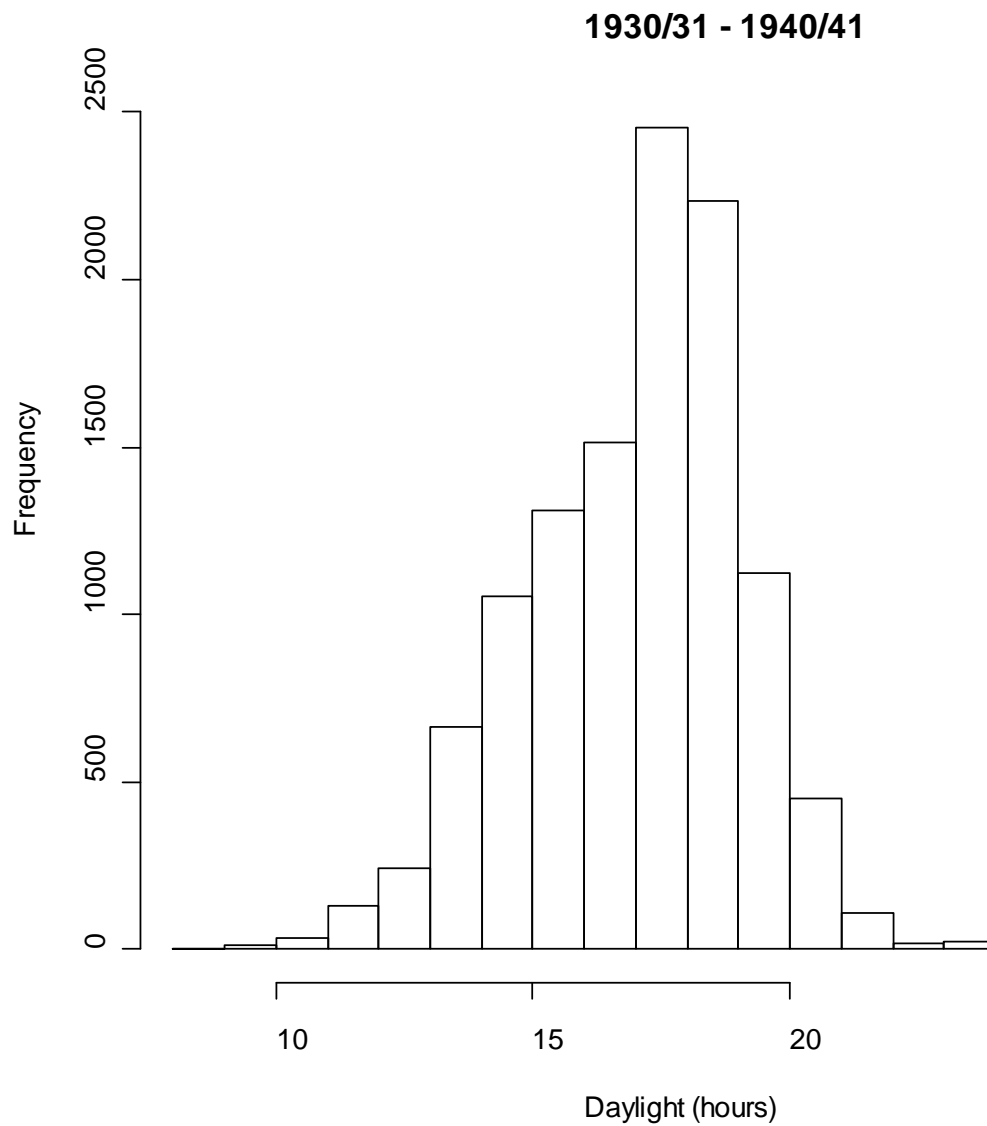


Fig. 1. Distribution of day lengths calculated from the dates and latitudes of pelagic whaling in the 1930s.

Additional handling time (compound

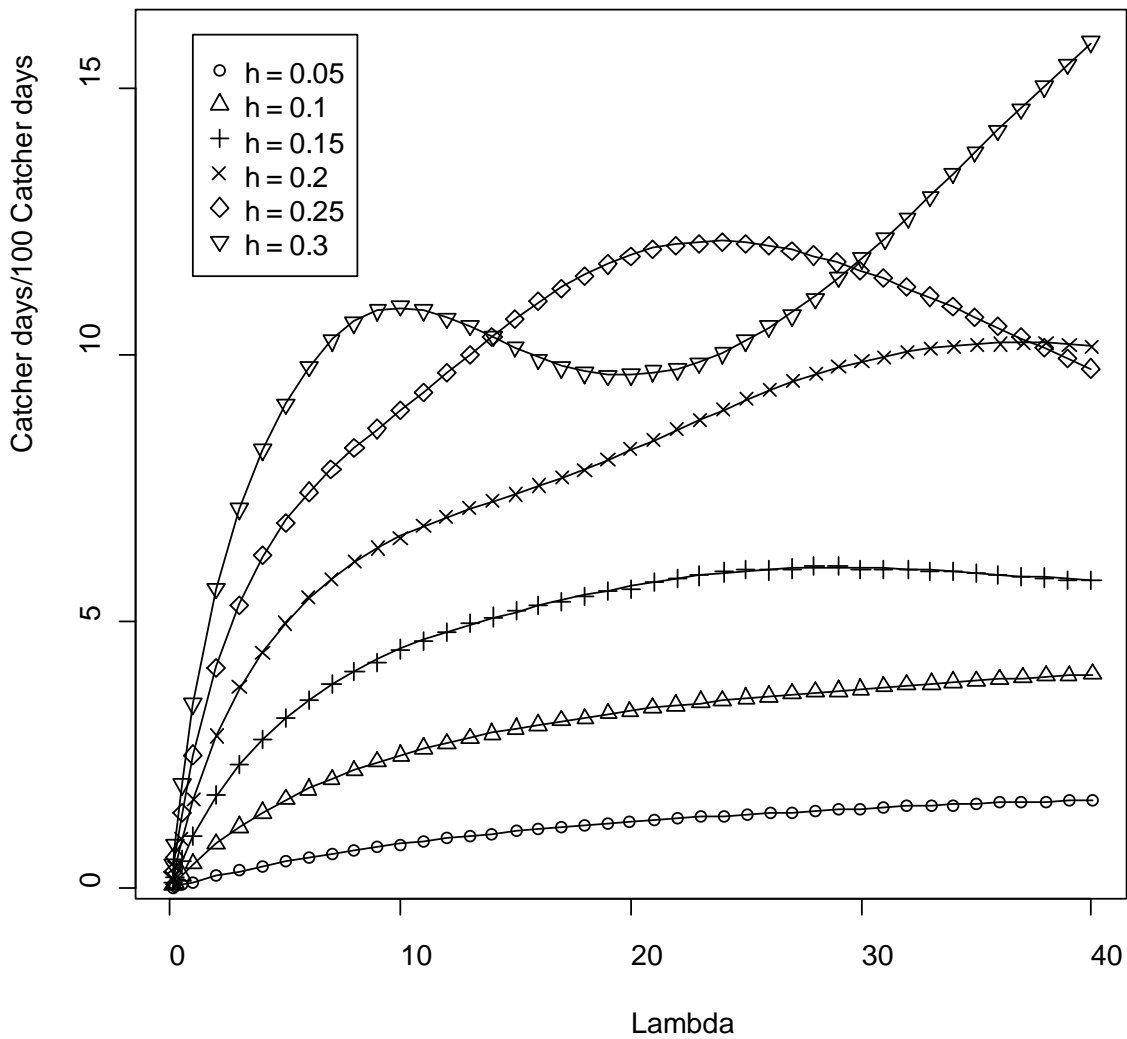


Fig. 2. Expected values of additional handling time versus whale density (λ , in terms of whales per day) for the compound gamma model at various handling time values (h , measured in days). The points are values generated by direct simulation; the curves are calculated by solving equation (6) by numerical methods. The corresponding curves for the gamma model are very similar.

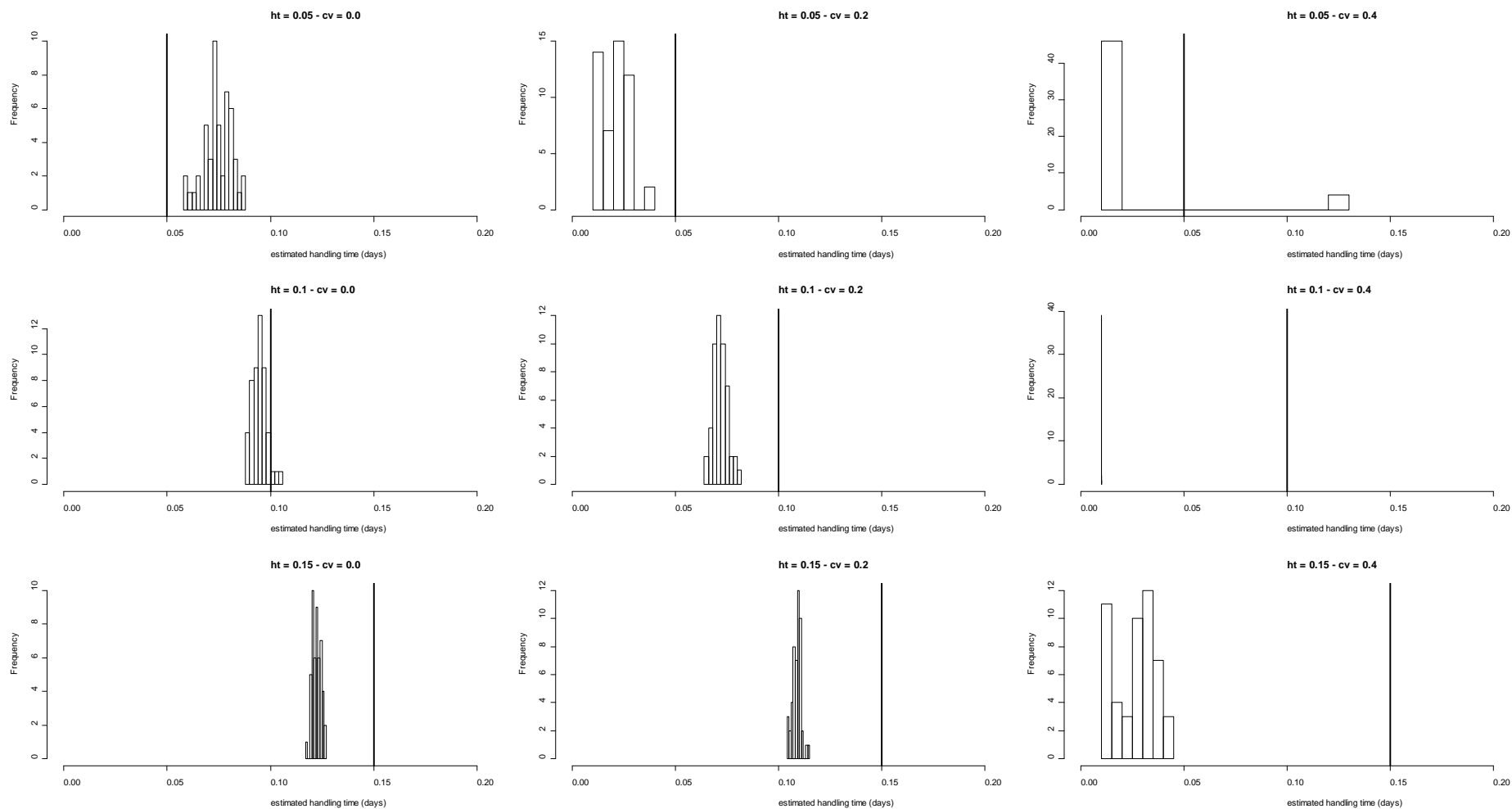


Fig. 3. Estimated handling times from 50 stochastic replicates of the simulations where the fitted model is a gamma (simple Poisson) model. The true handling time (ht) is shown by the thick vertical line in each histogram. The data are generated using the gamma model (CV = 0.0) or the compound gamma model for variable daily catch rates (CVs > 0.0). The generated data are conditioned on the observed catch distributions and day lengths of the 1930s.

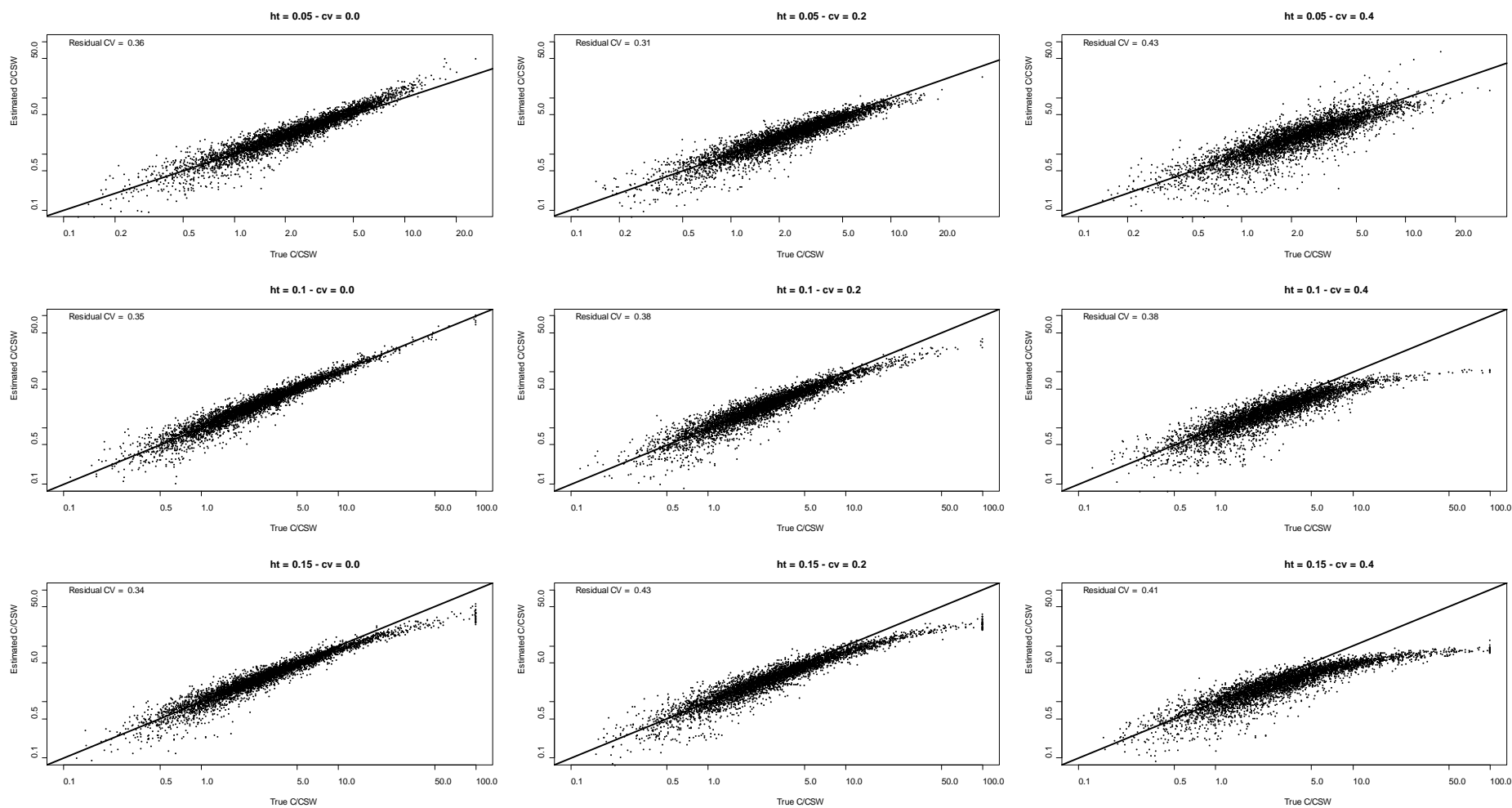


Fig. 4. Estimated versus true catch per catcher search time (C/CSW) for a range of handling times and variability in daily catch rates estimated using the gamma (simple Poisson) model, conditioned on observed catch distributions of the 1930s and day lengths. The plot shows 5000 points randomly selected from the 50000 estimates generated in each trial.

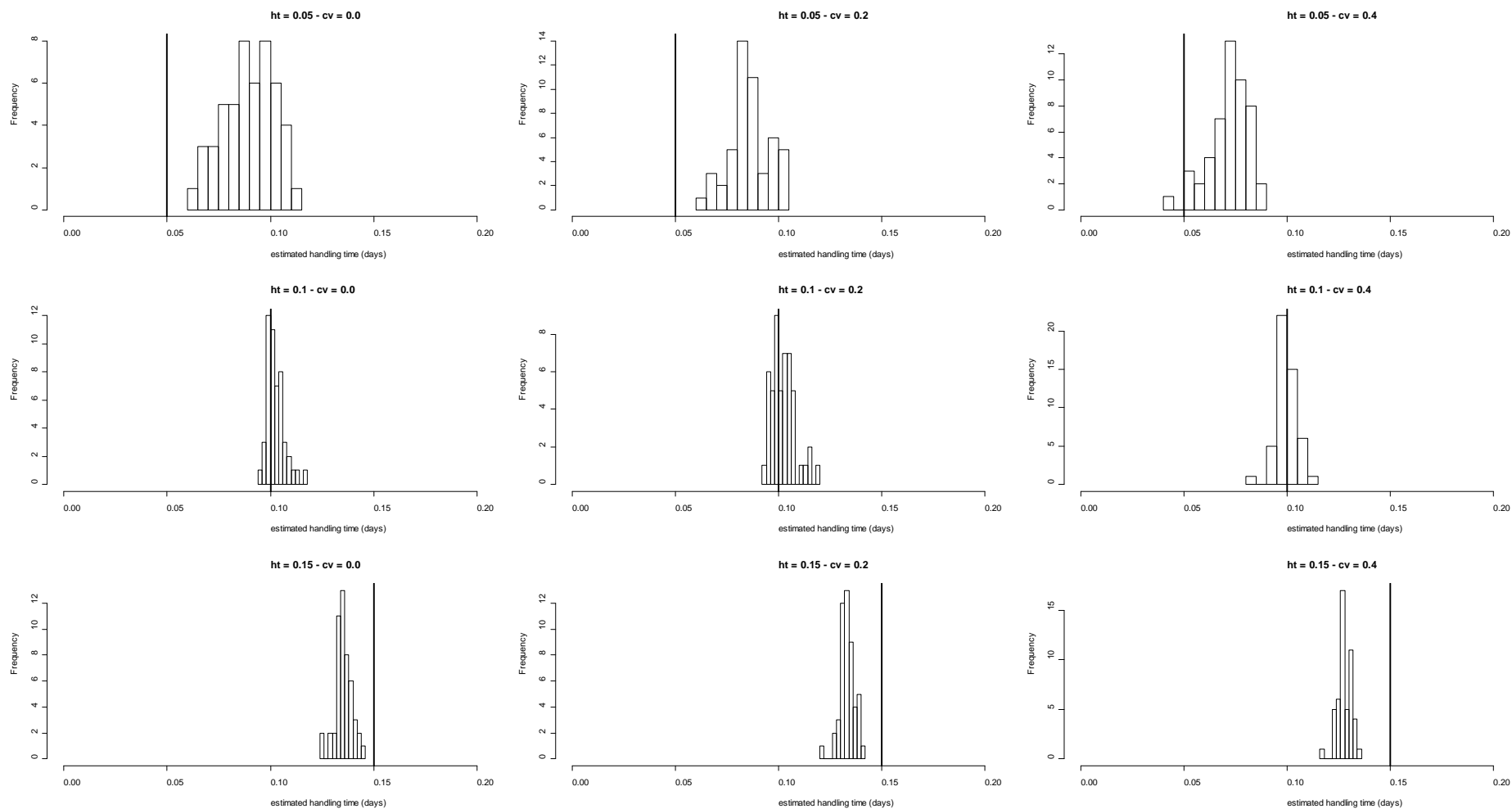


Fig. 5. Estimated handling times from 50 stochastic replicates of the simulations where the fitted model is a compound gamma model. The true handling time (ht) is shown by the thick vertical line in each histogram. The data are generated using the gamma model (CV = 0.0) or the compound gamma model for variable daily catch rates (CVs > 0.0). The generated data are conditioned on the observed catch distributions and day lengths of the 1930s.

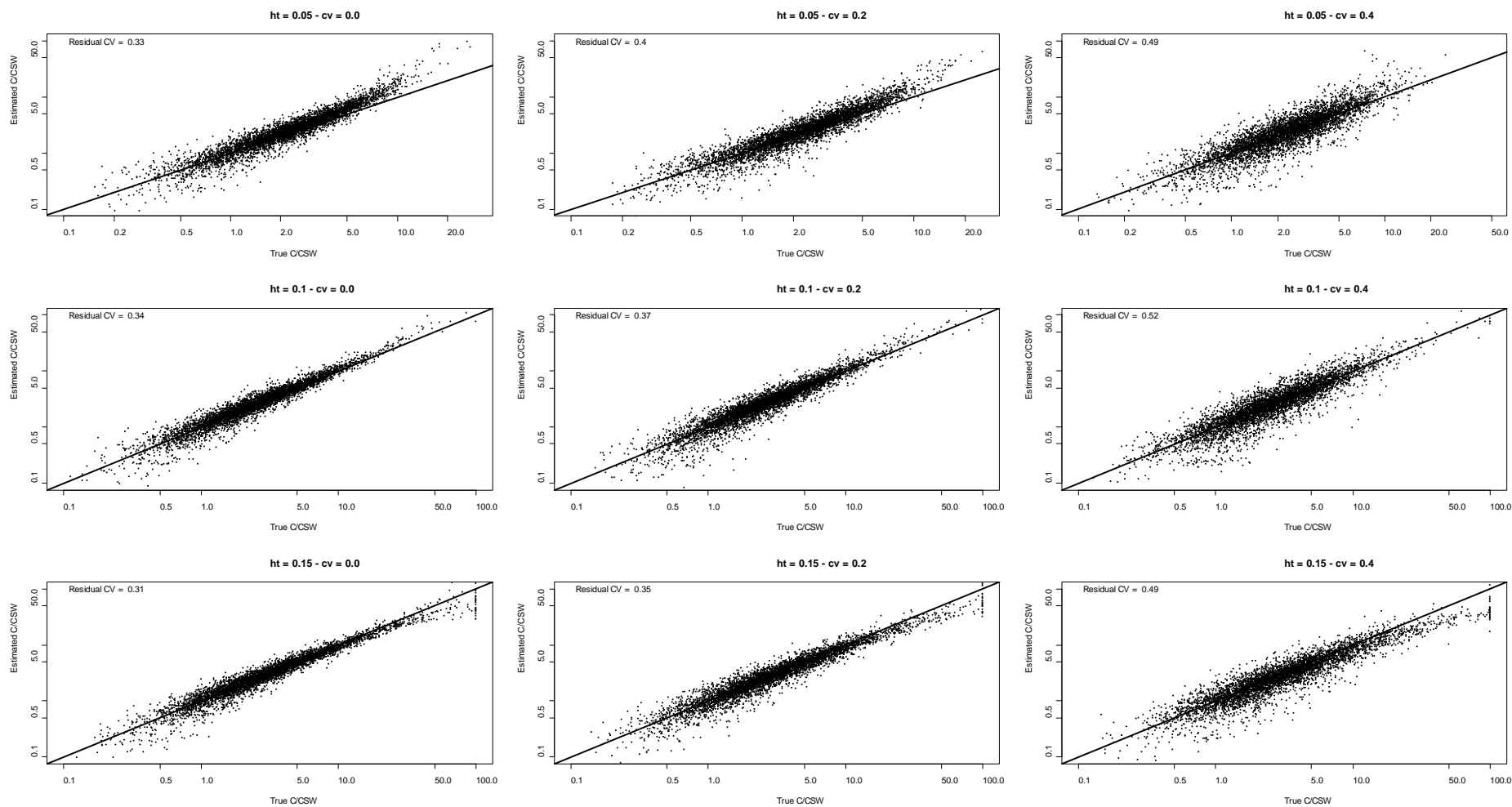


Fig. 6 Estimated versus true catch per catcher search time (C/CSW) for a range of handling times and variability in daily catch rates estimated using the compound gamma model, conditioned on observed catch distributions of the 1930s.

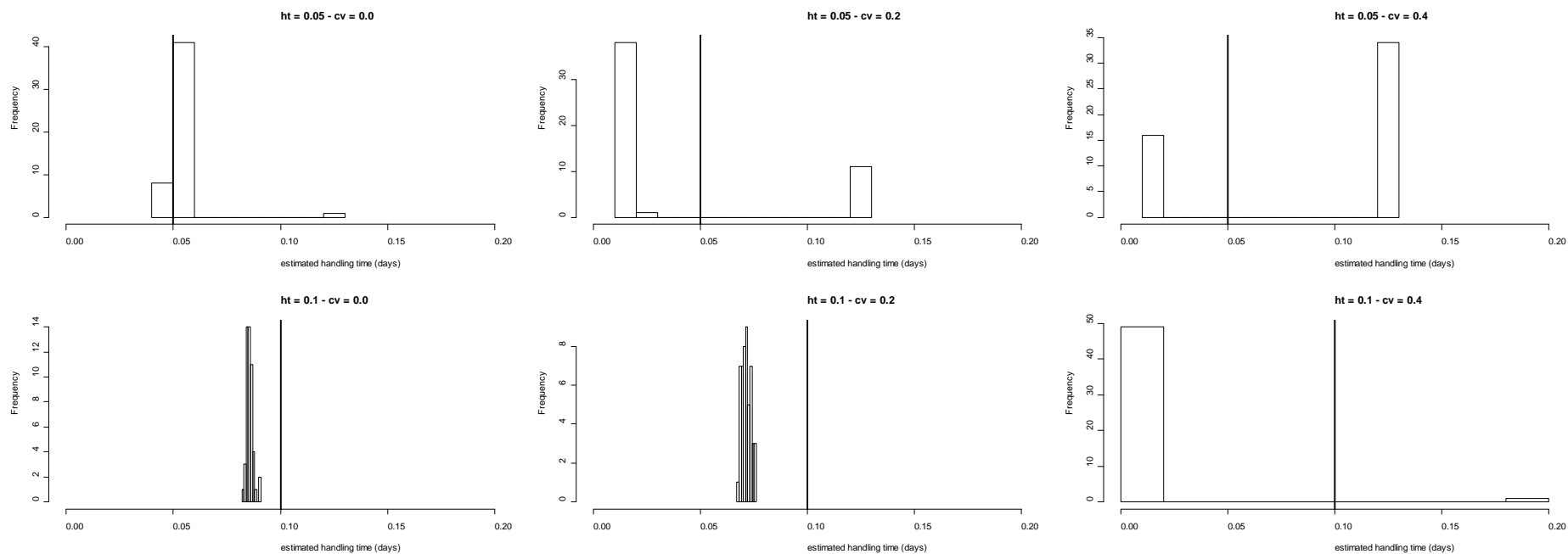


Fig. 7. Estimated handling times from 50 stochastic replicates of the simulations where the fitted model is a gamma (simple Poisson) model. The true handling time (ht) is shown by the thick vertical line in each histogram. The data are generated using the gamma model (CV = 0.0) or the compound gamma model for variable daily catch rates (CVs > 0.0). The generated data are conditioned on the observed catch distributions and day lengths of the early 1950s.

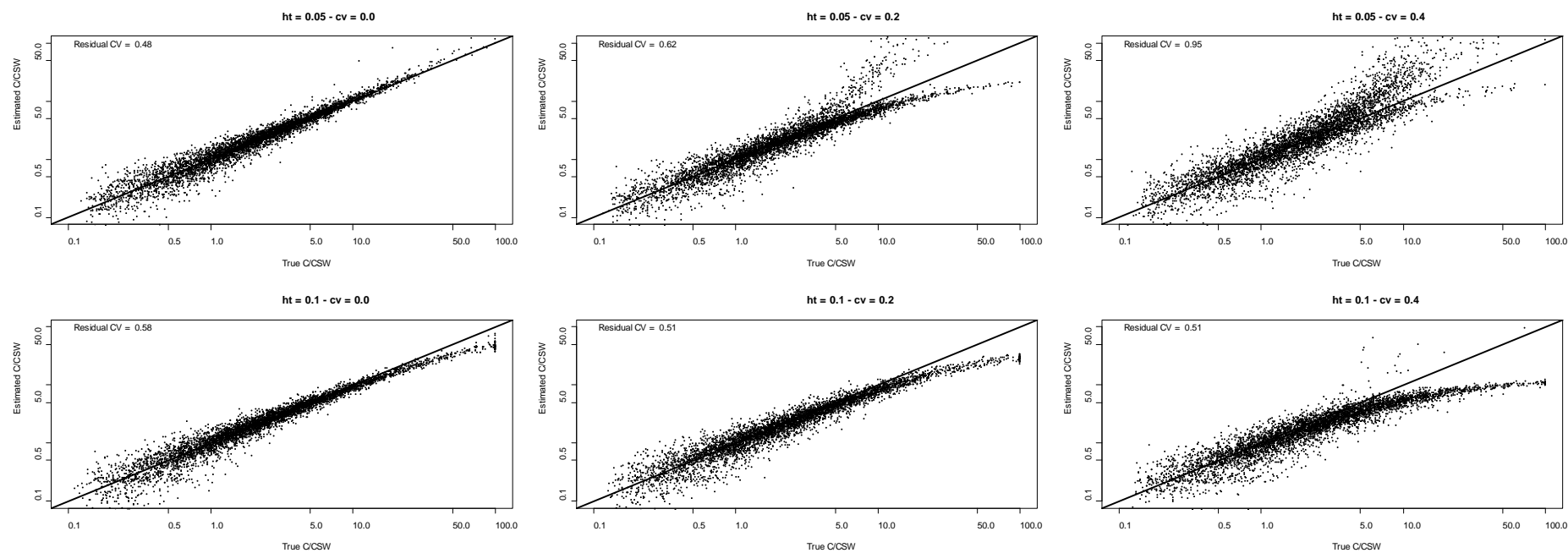


Fig. 8 Estimated versus true catch per catcher search time (C/CSW) for a range of handling times and variability in daily catch rates estimated using the gamma (simple Poisson) model, conditioned on observed catch distributions of the early 1950s.

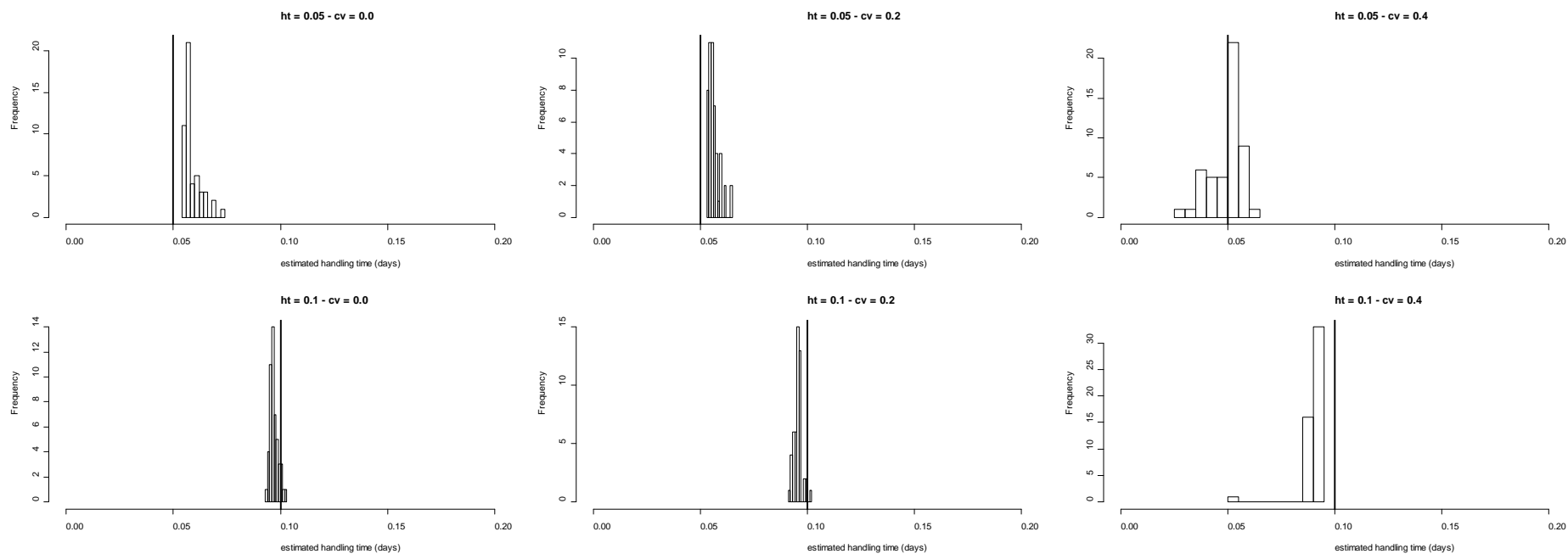


Fig. 9. Estimated handling times from 50 stochastic replicates of the simulations where the fitted model is the compound gamma model. The true handling time (ht) is shown by the thick vertical line in each histogram. The data are generated using the gamma model (CV = 0.0) or the compound gamma model for variable daily catch rates (CVs > 0.0). The generated data are conditioned on the observed catch distributions and day lengths of the early1950s.

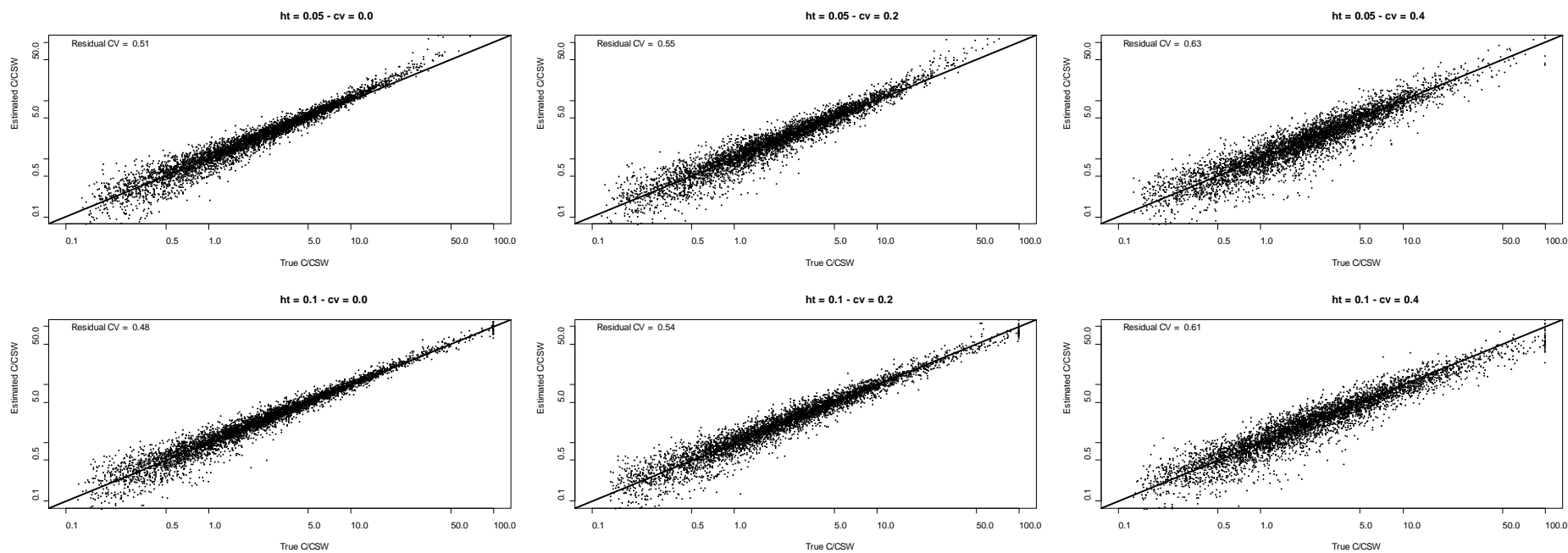


Fig. 10 Estimated versus true catch per catcher search time (C/CSW) for a range of handling times and variability in daily catch rates estimated using the compound gamma model, conditioned on observed catch distributions of the early 1950s.

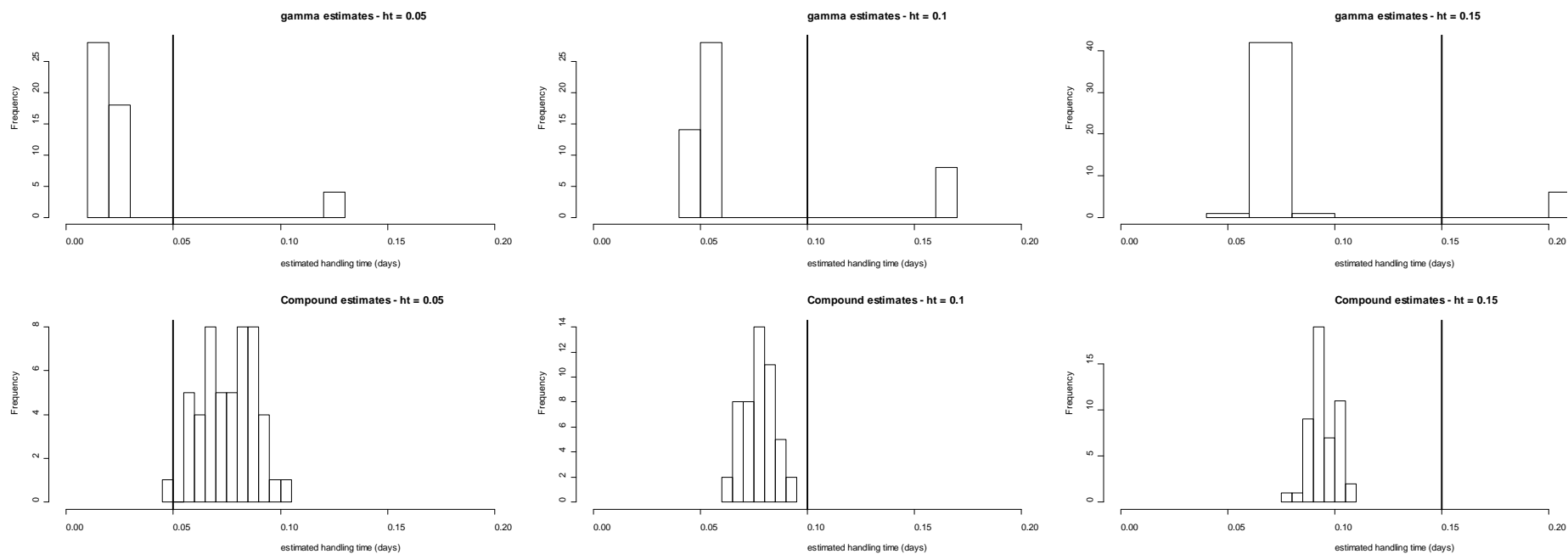


Fig. 11. Estimated handling times from 50 stochastic replicates of the simulations where the fitted model is either the gamma or compound gamma models. The data are generated using a random handling time correlated with variable daily catch rates from the compound gamma model (realised CVs ≈ 0.3). The true mean handling time (ht) is shown by the thick vertical line in each histogram. The generated data are conditioned on the observed catch distributions and day lengths of the early1930s.

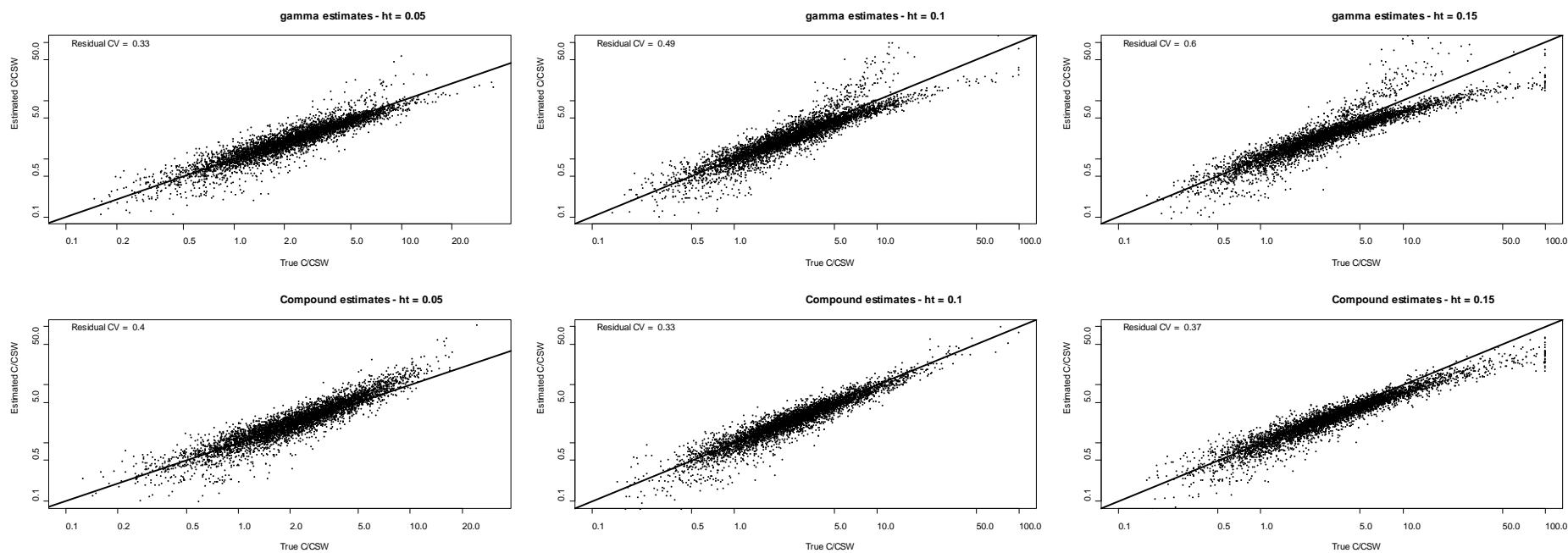


Fig. 12 Estimated versus true catch per catcher search time (C/CSW) for a range of handling times and variability in daily catch rates estimated using the gamma or compound gamma models, conditioned on observed catch distributions of the early 1930s. The generated data include the effects of random variability in handling time.

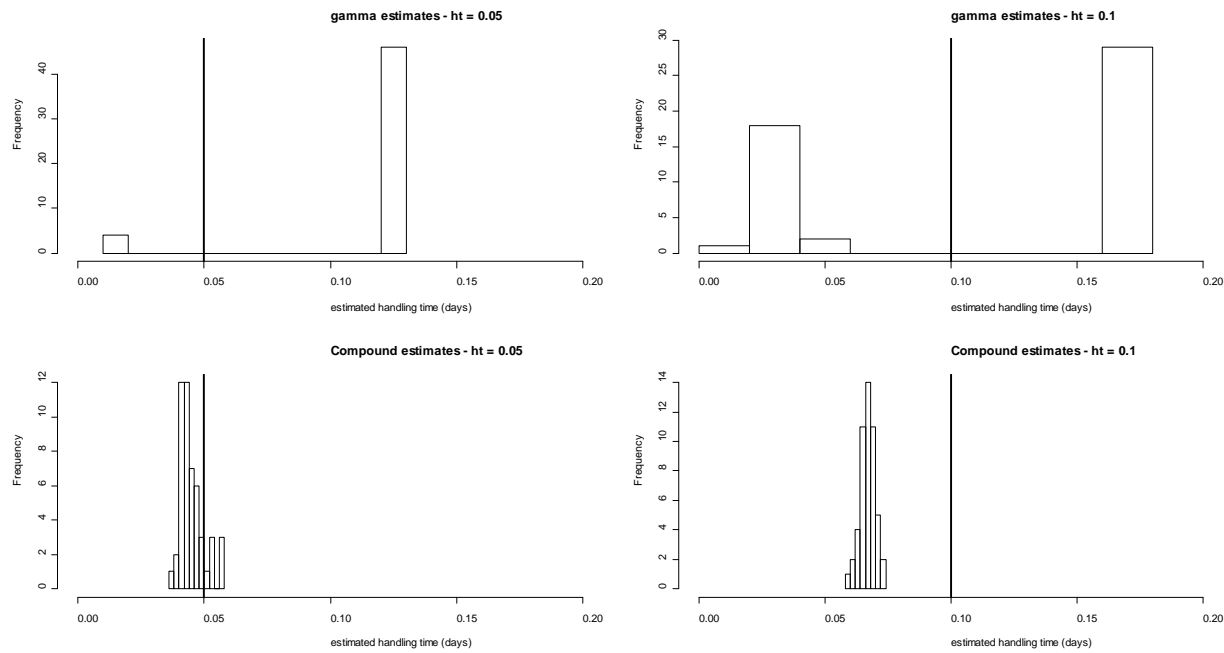


Fig. 13. Estimated handling times from 50 stochastic replicates of the simulations where the fitted model is either the gamma or compound gamma models. The data are generated using a random handling time correlated with variable daily catch rates from the compound gamma model (realised CVs ≈ 0.3). The true mean handling time (ht) is shown by the thick vertical line in each histogram. The generated data are conditioned on the observed catch distributions and day lengths of the early1950s.

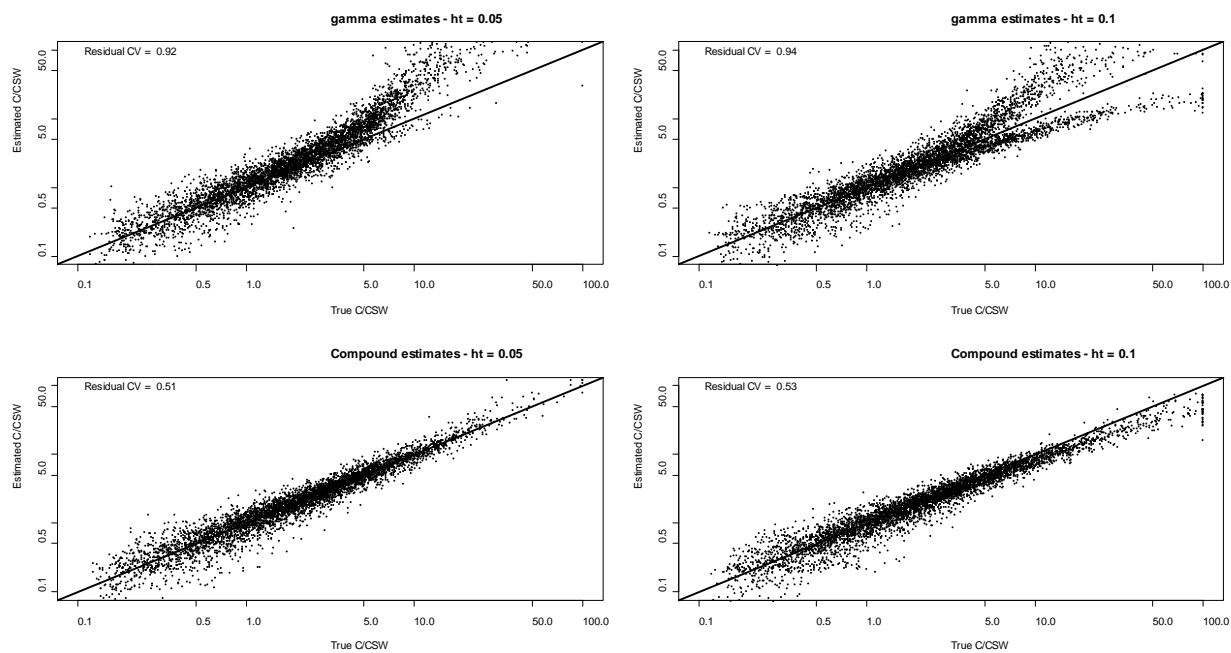


Fig. 14. Estimated versus true catch per catcher search time (C/CSW) for a range of handling times and variability in daily catch rates estimated using the gamma or compound gamma models, conditioned on observed catch distributions of the early 1950s. The generated data include the effects of random variability in handling time.

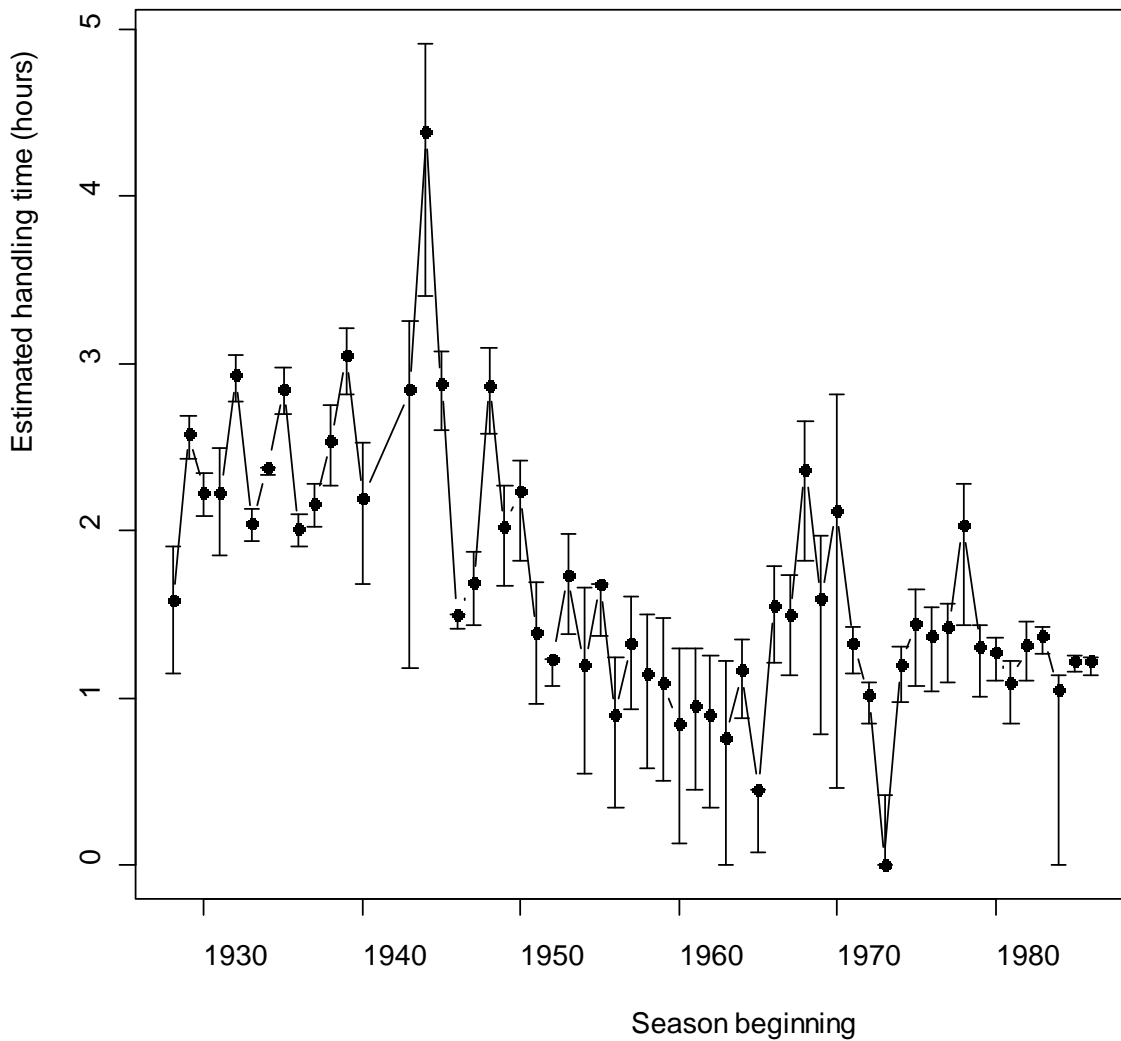


Fig. 15. Estimated handling times estimated from Antarctic pelagic whaling date each year 1930- 1986 using the compound gamma model. Error bars show the 95% confidence intervals. There was no Antarctic pelagic whaling in 1941 or 1942. In 1943 and 1944 only one whaling expedition operated. 1964 was the year the IWC protected Blue and Humpback whales and from then until 1972 whaling concentrated on sei and fin whales north of the South Polar Front. After 1972 minke whales became an increasing part of the catch and were the only species of baleen whale taken from 1978 onwards.

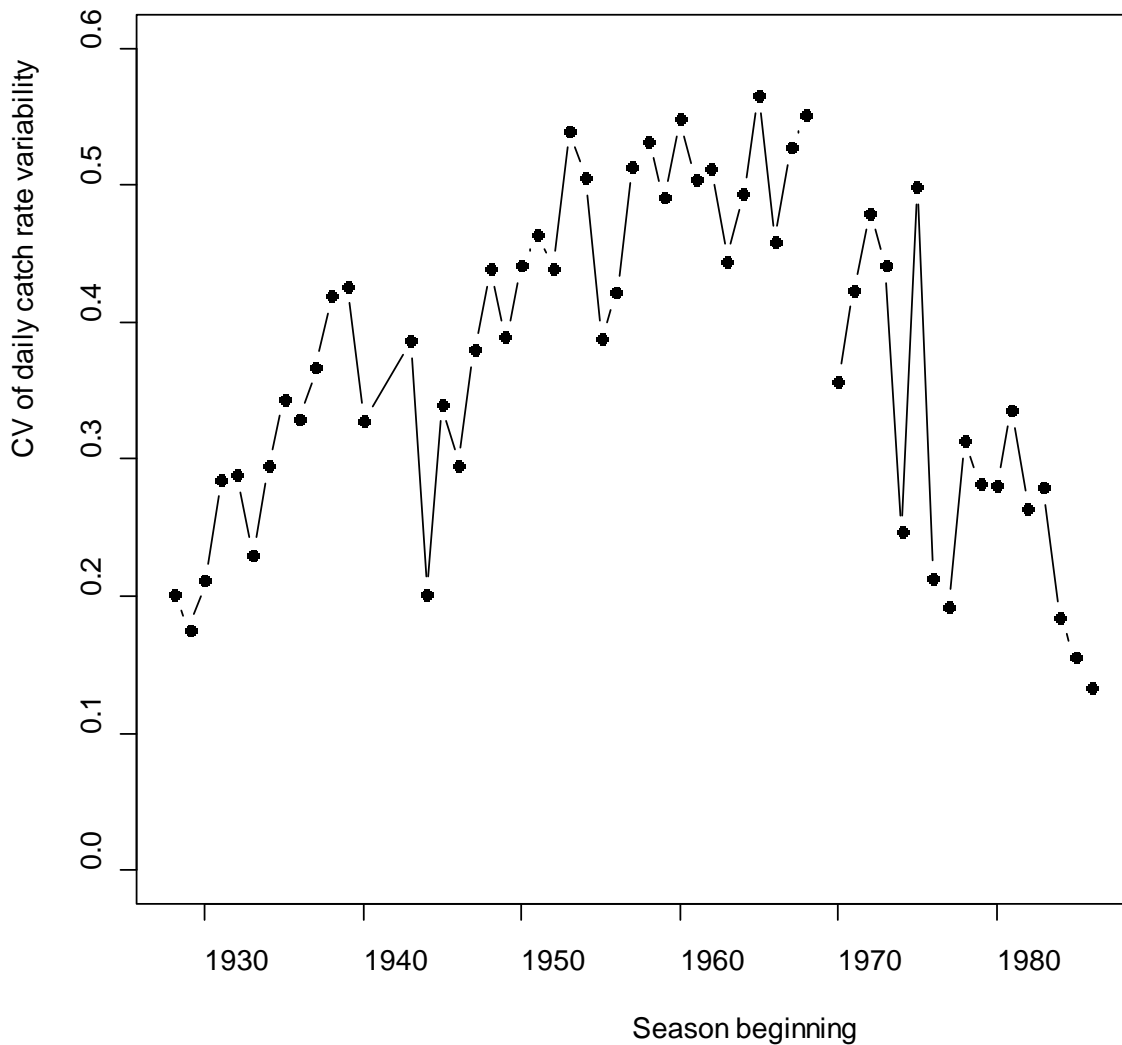


Fig. 16 Estimated coefficients of variation for the day to day variability in Antarctic pelagic catch rates estimated using the compound gamma model. The CV's are calculated from those cells where there are at least three days of recorded catch and averaged. In 1969 only one cell had three or more days with catches and so the CV for this year is unreliable.

Fin C/GCD - 1935/36 to 1940/41

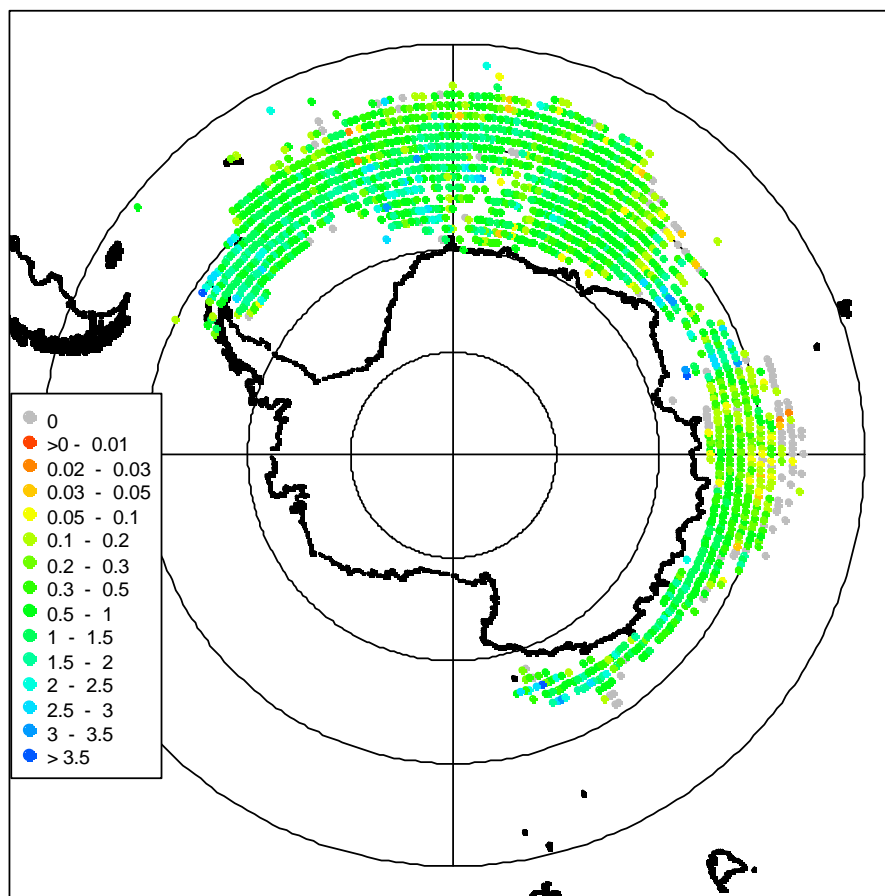


Fig. 17 Chart showing the apparent density of fin whales from whaling operations in the period 1935-1940 based on catch/catcher day.

Fin C/CSW - 1935/36 to 1940/41

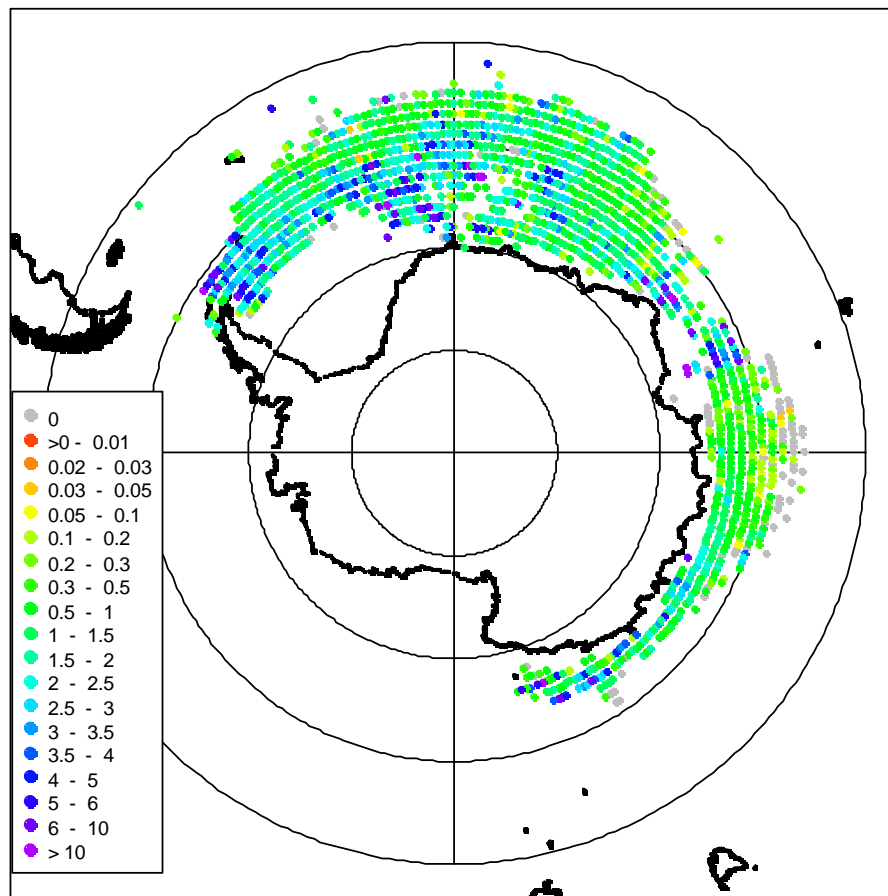


Fig. 18 Chart showing the apparent density of fin whales from whaling operations in the period 1935-1940 based on catch/CSW obtained from the compound gamma estimates of handling time and searching times.

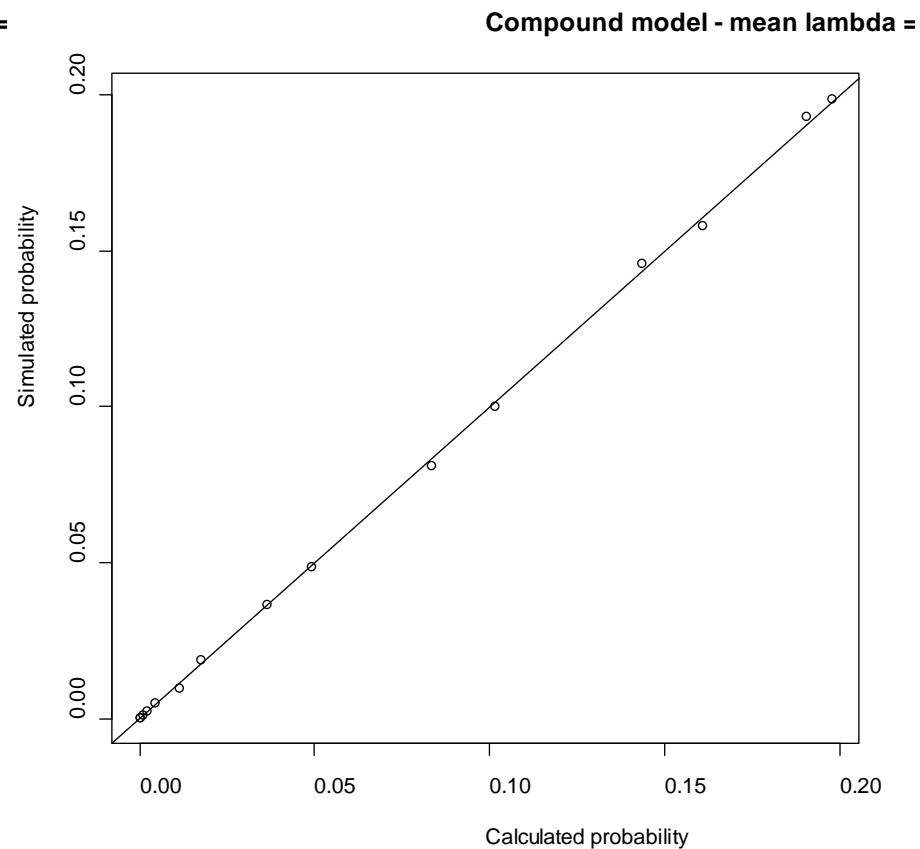
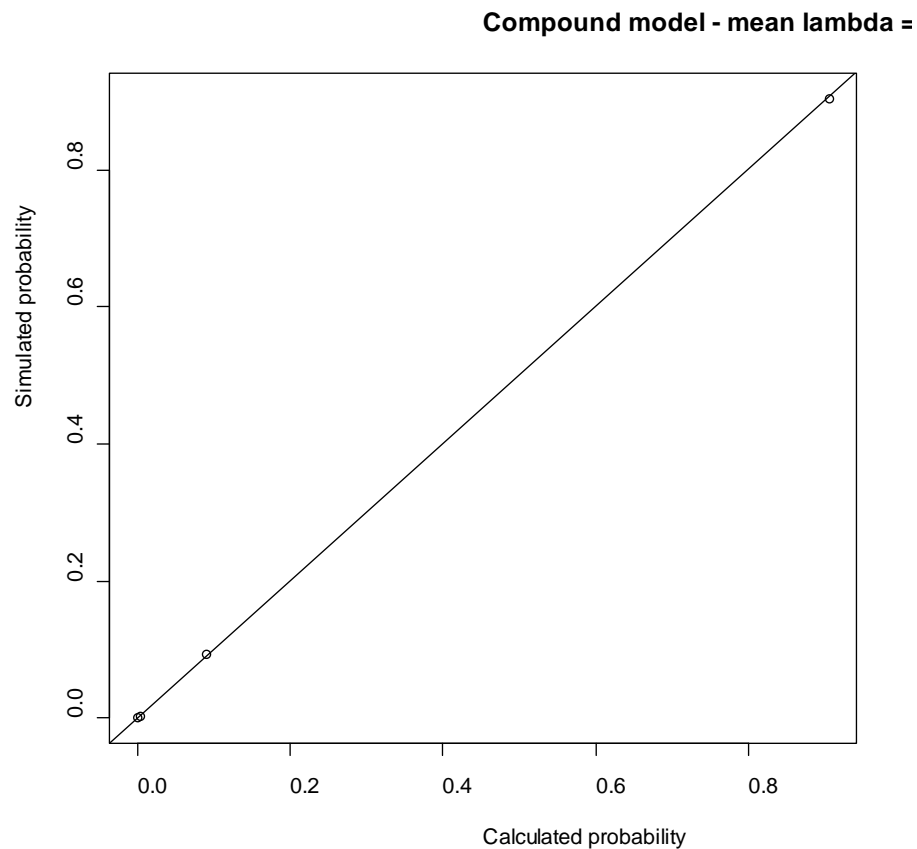


Fig ? Two widely spaced examples of probabilities of catch frequencies calculated according to equation (17) compared with the corresponding probabilities estimated by direct simulation of the catching process.

OPTIONAL FIGURE CONFIRMING THAT METHOD FOR CALCULATING CATCH PROBABILITIES HAS REQUIRED PROPERTIES