Data Weighting for Tagging Data in Integrated Size-Structured Models

André E. Punt[1,2], Roy A. Deng[3], M.S.M. Siddeek[4], Rik. C. Buckworth[5], and Vicki Vanek[6]

[1]*School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020, USA*
[2]*CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, Tasmania, Australia*
[3] *CSIRO Oceans and Atmosphere, Queensland Bioscience Precinct, 306 Carmody Road, St Lucia, QLD, 4072, Australia*
[4]*Alaska Department of Fish and Game, Division of Commercial Fisheries, 1255 W. 8th Street, Juneau, Alaska 99811, USA*
[5]*CSIRO Oceans and Atmosphere, Tropical Ecosystems Research Centre, 564 Vanderlin Drive, Berrimah, NT, 0828 Australia*
[6]*Alaska Department of Fish and Game, Division of Commercial Fisheries, 351 Research Court, Kodiak, AK 99615, USA*

Abstract

Increasingly, stock assessments for hard-to-age species such as crabs, prawns, rock lobsters, and abalone are being based on integrated size-structured population dynamics models that are fit to a variety of data sources. These data sources include tagging data to inform growth. Diagnostic statistics and plots have been developed to explore how well integrated population models fit the data types typically used for assessment purposes (index data, size- and age-compositions, and conditional age-at-length data). However, such statistics and plots are not available for tagging data, when these data are used to estimate growth. This paper outlines two diagnostic statistics that can be used to evaluate fits to tagging data, and develops a method based on 'Francis weighting' for weighting tagging data in integrated models. For illustration, the methods are applied to Aleutian Islands golden king crab (*Lithodes aequispinus*) in Alaska, and tiger prawns (*Penaeus semisulcatus* and *P. esculentus*) in Australia's Northern Prawn Fishery. Some degree of growth model mis-specification was revealed for *P. semisulcatus*, and there were conflicts in the data for the tiger prawns. The standard errors for the estimates of mature male biomass for golden king crab were larger when the tagging data were downweighted based on the proposed weighting method. This serves to emphasise that assessments and their interpretations can be impacted by how tagging data are weighted.

Keywords: data weighting; size-structured stock assessment methods; tagging data

Corresponding author: A.E. Punt
a. *Email*: aepunt@uw.edu
b. *Phone*: 1-206-221-6319
c. *Fax* : 1-206-685-7471

Highlights

- Assessment results can be sensitive to the assumptions for data weighting.
- Tagging data may be overdispersed relative to the commonly assumed Bernoulli likelihood.
- 'Francis weighting' can be extended to apply to tagging data.

1. Introduction

There is an increasing trend towards the use of integrated size-structured stock assessments for species that are difficult to age (Punt et al., 2013). For example, assessments for crab stocks off Alaska are based on size-structured population dynamics models that often divide the population into new and old shell crab (i.e. crab that did and did not moult the previous season; e.g. snow crab *Chionoecetes opilio*; Turnock and Rugulo, 2014; and red king crab *Paralithodes camtschaticus*; Zheng and Siddeek, 2014) while the assessment of tiger prawns (*Penaeus semisulcatus* and *P. esculentus*) in Australia's Northern Prawn Fishery (NPF) is based on a sex- and size-structured population dynamics model (Punt et al., 2010; Buckworth et al., 2015).

Integrated size-structured stock assessment methods make use of several sources of data. For example, assessments of golden king crab *Lithodes aequispinus* in the Aleutian Islands, Alaska, include data on landings in numbers, the size-composition of the landings, the size-composition of observer records for all crab arriving on deck, catch-rate indices for the retained component of the catch, and tagging data (Siddeek et al., in press). In contrast, assessments of tiger prawns in the NPF are based on weekly catch and effort data, survey indices of abundance, survey and commercial size-composition data, and tagging data. However, it is not uncommon for information in data sources to be in conflict with each other to some extent (e.g. Richards, 1991). Thus, each data type (and each data point within each data type) included in a stock assessment needs to be assigned a weight. In principle, this weight should relate to the deviation between the data point and its expected value (Punt, this volume), although on occasion weights reflect a subjective evaluation of the reliability of the data type (e.g. ICCAT, 2013). However, it is not straightforward to objectively select weights, and history reveals that data weighting can be influential on assessment results (e.g., Richards, 1991).

The primary purpose of the tagging data in the assessments of Aleutian Islands golden king crab and of tiger prawns in the Northern Prawn Fishery is to allow growth (rather than fishing mortality) to be estimated. The component of the likelihood function for the tagging data (Punt et al., 2009) is:

$$L=\prod_i \left(S_{C_i}[\mathbf{X}^{T_i}]_{R_i,C_i} / \sum_j S_j[\mathbf{X}^{T_i}]_{R_i,j} \right) = \prod_i p(C_i \mid R_i, T_i) \tag{1}$$

where T_i is the time-at-liberty for the ith recapture, $\mathbf{X}$ is the size-transition matrix (which specifies the probability of growing from one size-class to each of the same or larger size-classes), R_i is size-class in which the ith recapture was when it was released, C_i is the size-class in which the ith recapture was when it was recaptured, and S_j is the selectivity of an animal in size-class *j* (logistic for the example applications reported here). The size-transition matrix can separate the processes of moulting from those of growth given moult (e.g., Zheng and Siddeek, 2014) or represent the combined effects of moulting and growth given moult within a single model (e.g., Punt and Kenndy, 1997; Haist et al., 2009)). The form of the size transition matrix for the case in which moulting is modelled explicitly is:

$$\mathbf{X}=\mathbf{X}'\mathbf{Q}+\mathbf{I}(\mathbf{I}-\mathbf{Q}) \tag{2}$$

where $\mathbf{Q}$ is a diagonal matrix with values given by the probability of moulting, $\mathbf{X}'$ is a matrix where each column is for a size before moult and each entry in each column is the probability of growing to that size given the size being represented by the column, and $\mathbf{I}$ is the identity matrix.

Equation 1 treats each recapture as a Bernoulli trial, i.e. each tagged animal is treated as a single data point, independent from all the others. However, there will be overdispersion if tagging is such that some of the tagged animals are pseudoreplicates. This can happen if groups of tagged animals are released together and hence may have moved together and hence been subject to the same environmental conditions and prey fields. Consequently, the growth and probability of recapturing an animal are not independent of those for some of the other tagged animals. To account for overdispersion, the size-composition data used in the current configuration of the assessment methods for the two example fisheries are downweighted, but this is not currently the case for the tagging data. Accounting for this overdispersion requires that the right hand side of Equation 1 is raised to a power (equivalent to multiplying the logarithm of the right hand side of Equation 1 by an overdispersion factor). Several approaches (e.g., McAllister and Ianelli, 1997; Francis, 2011; Punt, this volume) have been developed to estimate overdispersion factors for size-composition data, and these approaches have been used to weight the size-composition data for Aleutian Islands golden king crab and tiger prawns in the NPF. However, methods have not been developed to explore whether the growth model is mis-specified, whether there is overdispersion, and how tagging data used in size-structured stock assessment methods should be weighted.

This paper provides diagnostic statistics for evaluating the fits to tag-recapture data within size-structured integrated assessment models and for estimating an overdispersion factor for weighting tagging data. The approach follows the spirit of the approach of Francis (2011). The proposed diagnostics and weighting factors are illustrated using Aleutian Islands golden king crab and *P. semisulcatus* and *P. esculentus* in the NPF. These two cases were selected because although the assessments are both based on size-structured models, that for Aleutian Islands golden king crab is male-only, has an annual time step, and considers 5 mm size-classes. In contrast, the assessments for *P. semisulcatus* and *P. esculentus* are based on a sex-structured model that has a weekly time-step and 1 mm size-classes.

2. Material and Methods

2.1 Diagnostic statistics

Two diagnostic statistics are considered. Both statistics are computed by time-at-liberty. The first diagnostic statistic is a comparison of frequencies of observed numbers recaptured by size-class versus the model-predicted distribution for size-classes at recapture. The latter distribution is:

$$\hat{P}_j = \sum_i p(j \mid k_i) \qquad (3)$$

where $\hat{P}_j$ is the expected number of recaptures in size-class *j*, and $p(j \mid k_i)$ is the probability that the ith individual (which was released in size-class class *k*) was recaptured in size-class *j* (see Equation 1).

The second diagnostic statistic involves plotting the observed mean recapture size, $\bar{P}_L^{obs}$, versus release size-class *L*, along with the expected distributions of size-at-recapture, as a function of size-class-at-release, characterized by the expected (mean) size-at-recapture $\hat{\hat{P}}_L$ and the standard error of the observed mean size-at-recapture $SE[\hat{\hat{P}}_L]$, i.e.:

$$\hat{\hat{P}}_L = \sum_j \bar{L}_j p(j \mid L)\,; \qquad SE[\hat{\hat{P}}_L] = \sqrt{\sum_j (\bar{L}_j - \hat{\hat{P}}_L)^2 / N_L} \qquad (4)$$

where $\bar{L}_j$ is the mid-point of size-class *j*, and N_L is the number of releases of animals in size-class *L*.

2.2 Data weighting

Francis weighting (Francis, 2011) involves defining the overdispersion factor for catch size-composition data as the inverse of the variance of the standardized residuals for the mean size of the catch. By analogy, the weight *W* that should be assigned to the tagging composition of the likelihood is given by:

$$W^{-1} = \mathrm{var}[(\bar{P}_L^{obs} - \hat{\bar{P}}_L) / \mathrm{SE}[\hat{\bar{P}}_L]] \qquad (5)$$

In common with the diagnostic statistics, the weighting factors can be computed separately by time-at-liberty and by sex.

Data weighting would entail applying standard methods for weighting compositional data (e.g. Punt, this volume) and the above method for weighting the tagging data iteratively until convergence occurs. If the data are in conflict, it may be that this process will not converge and the weights for some subsets of the data will increase without limit while the weights for other subsets will be reduced to zero (Punt, this volume).

2.3 Applications

2.3.1 Aleutian Islands golden king crab

Siddeek et al. (in press) outline the stock assessment model used for golden king crab in the Aluetian Islands region. In relation to the tagging data, rectangular, king crab pots were used to capture crabs for tagging in all experiments, with the exception of the 1991 experiment where smaller, conical pots were used. Tagged animals were released during summer (July – September) before the fishery started. Location, date, and fishing depth were recorded for each pot retrieved. Upon pot retrieval, the carapace lengths (CL) of crabs were measured to the nearest millimetre and shell condition (old or new) recorded. Isthmus-loop (“spaghetti”) tags were used to tag crabs (Gray, 1965), and tagged crabs were released on or adjacent to the capture location. The majority of tag recaptures were obtained from the fishers during the commercial pot fishery. The tagging data used in the analyses comprised CL-at-release, CL-at-recapture, and time-at-liberty. The tagging records were restricted to male golden king crab releases in the size range 101 – 185 mm CL given the size-range included in the population dynamics model. There were 27,131 tagged crab releases in this size range, with a 6.33% overall return rate.

The assessment makes use of several data sources in addition to tagging data. Table 1 lists the specifications for all of the data sources included in the assessments.

2.3.2 Northern Prawn Fishery

Punt et al. (2010) outline the stock assessment model for the two tiger prawn species in the Northern Prawn Fishery. Tag-recapture data are available from experiments conducted in the northwestern Gulf of Carpentaria in 1981 (Kirkwood and Somers, 1984) and in 1983 and 1984 (Somers, 1987; Somers and Kirkwood, 1991). In these experiments, prawns were captured using chartered commercial prawn trawlers and tagged using numbered streamer tags (Floy tags in 1981 experiments, Hallprint tags for the later experiments). With date of release, species, sex, and size (mm CL), and presence/ absence of bopyrid parasites *Epipenaeon ingens* recorded, prawns were released adjacent to the trawl runs where they were captured. All recaptures were made by the commercial fleet.

In common with Kirkwood and Somers (1984), Somers and Kirkwood (1991) and Wang et al. (1995), the data used in the assessment were restricted to animals that were at liberty for at least two weeks and which were not infected (at release or recapture) by the bopyrid parasites. Only prawns for which species, sex, size-at-release, size-at-recapture, and time-at-liberty are known, were included in the analyses. Data were excluded for *P. semisulcatus* that were extreme outliers in the growth curve analyses of Somers and Kirkwood (1991). Those authors ascribed these extreme data to measurement error. The sample sizes by week (the time-step in the population model) are very small. Size results were thus pooled over time for the first diagnostic statistic and a 4-week period (“month”) for the second diagnostic statistic.

Table 1 lists the full set of information used when fitting the population models for the two tiger prawn species.

3. Results and discussion

3.1 Aleutian Islands golden king crab

Figure 1 shows the application of the first diagnostic plot (observed and model-predicted frequencies of recaptured crab by recapture size-class) when the tagging data are assumed to be Bernoulli trials (“unweighted” tagging data; solid lines). The model is able to mimic the observed distributions adequately for animals at liberty up to the third year-at-liberty, with the discrepancy between the observed and model-predicted distributions increasing with increasing time-at-liberty beyond the third year-at-liberty. This latter result is, however, not unexpected given sample sizes decline quickly with increasing time at liberty (Fig. 1).

The second diagnostic plot (Fig. 2) suggests that the model mimics the mean lengths-at-recapture very well for animals at liberty for the first year-at-liberty (note that the observed mean lengths-at-recapture are offset sightly to allow differences to be distinguished). The expected lengths-at-recapture in Fig. 2 (and the equivalent plot for tiger prawns) are not linear functions of the lengths-at-release because what is plotted is $\hat{\hat{P}}_L$ (Equation 4). This would be a linear function of length-at-release if allowance was made for the possibility of “shrinkage”, but that is not the case so $p(L\,|\,L)$ can be substantial for several size-classes, particularly the larger size-classes. For times-at-liberty greater than one year, $p(j\,|\,L)$ is a product of $\mathbf{X}$ over time which is not a linear transformation.

The discrepancies between the observed and model-predicted mean lengths-at-recapture increase with increasing time-at-liberty. However, sample sizes get smaller with increasing time-at-liberty (Fig. 1) so the predicted 95% confidence intervals for mean lengths-at-recapture also get wider (Fig. 2). As a result, differences between observed and model-predicted mean lengths-at-recapture appear less consequential when expressed as z-scores. Application of Equation 5 to the results in Fig. 2 leads to weighting factors of 0.5, 0.2, and 0.35 for times-at-liberty of 1, 2 and 3 years (rounded from 0.536, 0.185, and 0.352). The weighting factors computed using Equation 5 for times-at-liberty of 4, 5 and 6 years exceeded 1; “underdispersion” is unlikely so the weighting factors were set to 1 in subsequent analyses.

The fits to the frequencies of tag recaptures are essentially unchanged when the data are downweighted (Fig. 1, dashed lines). Figure 3 shows time-trajectories of mature male biomass (the population component used to provide management advice for crab stocks in the Bering Sea; NPFMC, 2008), with asymptotic 95% confidence intervals when each tag is assigned equal weight (i.e., default weighting) (upper panel in Fig. 3) and when the tagging data for animals at liberty for 1, 2 and 3 years are down weighted by factors of 0.5, 0.2 and 0.35 (lower panel in Fig 3). The estimates of mature male biomass are very similar, but the precision with which the estimates are obtained is lower (by between 0% and 14%) when the tagging data are downweighted.

3.2 Northern Prawn Fishery

The solid lines in Figure 4 show the application of the first diagnostic plot to the two prawn species (and by sex) when the tagging data are assumed to be Bernoulli events. This diagnostic plot is based on data aggregated over time given low sample sizes. The fits to the data for *P. semisulcatus* (upper panels in Fig. 4) are fairly poor, with evidence that the model overpredicts growth. In contrast, the fits to the data for *P. esculentus* (for which sample sizes are larger) (lower panels in Fig. 4) are much better.

Figure 5 shows the results of applying the second diagnostic approach by prawn species and sex. The over-prediction evident in Figure 4 is clear from the upper two sets of panels in Figure 5, with the model predictions consistently exceeding the observed mean lengths-at-recapture. The model mimics the observed mean lengths-at-recapture well for *P. esculentus*, as might be expected from Fig. 4.

The calculated weighting factors for all months, sexes and species are approximately 1 so there is no justifications for reducing the weights. However, the results in Fig. 4 are suggestive of conflicts in the data. Consequently, the assessment was conducted unrealistically increasing the weight assigned to the tagging data by a factor of 1,000 (Fig. 4, dashed lines). The model is now able to mimic the tagging data well. However, the fits to the length-frequency data (Fig. 6) deteriorate quite markedly for some species, sexes, and length-frequency types when the tagging data are upweighted. In addition, the selectivity pattern for *P. semisulcatus* on large individuals is reduced substantially when the tagging data are substantially upweighted.

3.3 General comments and thoughts

This paper has provided two diagnostic statistics that can be used to evaluate the ability of size-structured assessment models to replicate tagging data and also outlined a way to weight tagging data when these data are included in an assessment along with several other data sources. The results in Figures 4 and 6 show that the fits to length-frequency and tagging data can be in conflict. It is well known that models that fit to size-composition data can be very sensitive to assumptions about growth, especially when the parameters of the growth model are pre-specified (e.g. Aires-da-Silva, and Maunder, 2012). However, this is the first example that shows that the results of size-structured integrated stock assessments can also be sensitive to the weighting of tagging data.

While it would be ideal to report both plots when reporting assessment results, plots such as Figs 2 and 5 are more informative than plots such as Figs 1 and 4 because model mis-specification could be missed in the latter plots if, for example, growth of smaller animals was over-estimated while that of larger animals was under-estimated.

The statistics and plots outlined in this paper could be extended. For example, Figs 2 and 5 indicate measures of uncertainty as ±1.96 standard errors. However, the matrix $\mathbf{X}$ is usually lower triangular (because most crustaceans do not shrink following a moult). Consequently, for short times-at-liberty, the modal number of size-classes an animal (particularly a large individual) may grow will be 0 (i.e. it is most likely to remain in its size-class-at-release), which suggests a non-symmetric distribution for the size-increment and hence the mean size-at-recapture. Such a distribution could be computed using the quantiles of the distribution for the size-increment from the matrix $\mathbf{X}$. Figs 2 and 5 (and methods such as those of Francis (2011) and Punt (this volume)) focus on the mean age/length (for composition data) or mean length-at-recapture (for tagging data). However, similar statistics could be developed to explore whether the model is able to adequately mimic the variation in the length-at-recapture. A statistic to make this comparison has been developed for age/length data ("Andre Plots" in the r4ss package [Taylor et al., 2014] that compares the observed and model-predicted variances of the size- or-age-compositions under the assumption that the model-predicted variances are F-distributed). However, unless the number of recaptures is high (which is unlikely to be the case for tagging data), the power of such a statistic is likely to be low, and the statistic consequently uninformative.

In common with all diagnostics plots and statistics, the suggestions in this paper indicate potential problems, but cannot prove that the model is adequate if the diagnostics do not provide evidence for model mis-specification and/or overdispersion. Patterns of residuals (such as those for *P. semisulcatus* in Figs 4 and 5) could be suggestive of model mis-specification, i.e. in the case that the von Bertalanffy growth curve is an inadequate representation of growth. In the case of prawns, there is evidence that the growth rate varies cyclically during the year in some species (e.g., Somers, 1988; Xiao and McShane, 2000; Lloyd-Jones et al., 2012), but this is not accounted for in the assessment model for NPF tiger prawns. Whether differences in growth seasonally are sufficient to generate the types of patterns evident in Figs 4 and 5 is beyond the scope of the present paper. We note, however, that seasonality in growth was not found to be significant in blue endeavour prawns, *Metapenaeus endeavouri*, caught in the same fishery (Buckworth, 1992). Moreover, patterns in residuals may reflect issues with other aspects of the model and be unrelated to the tagging data and the model of growth.

The diagnostic statistics and the approach to weighting outlined in this paper are heuristics and based on analogy with approaches used to weight age- and size-composition data. Future work should consider the use of simulation studies to evaluate the performance of these statistics and the weighting method, particularly in cases in which the model is mis-specified (e.g., is based on a non-seasonal von Bertalanffy curve when this is not the case).

Finally, the assessments considered in this paper are based on the likelihood function given by Equation 1. However, tagging data can be included in other ways in integrated stock assessments. For example, the method of Haist et al. (2009) is used in New Zealand to conduct assessments of rock lobsters and abalone (e.g., Breen et al., 2012; Fu et al., 2010). This method assumes that the size-increment is governed by a truncated normal distribution and allows for measurement error when fitting the tag-recapture data. In principle, the approach of this paper could be applied to create diagnostic statistics to evaluate models implemented using the method of Haist et al. (2009).

4. Acknowledgements

We would like to thank two anonymous reviewers for their comments on an earlier draft of this paper. This work was partially funded by NOAA Award NA14NMF4370119, Australian Fisheries Management Authority (AFMA) project RR2015/0811 and the Joint Institute for the NOAA Ocean Acidification Program through the Joint Institute of the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative agreement No. NA10OAR4320148. This is JISAO Contribution No. xxx.

References

Aires-da-Silva, A.M., Maunder, M.N., 2012. Status of bigeye tuna in the eastern Pacific Ocean in 2011 and outlook for the future. Inter-American Tropical Tuna Commission (IATTC), Stock Assessment Report 13, pp. 18–29.

Breen, P.A., Haist, V., Starr, P J., Pomarede, M., 2012. The 2011 stock assessment and management procedure development for red rock lobsters (*Jasus edwardsii*) in CRA 4. New Zealand Fisheries Assessment Report 2012/09.

Buckworth, R.C., 1992. Movements and growth of tagged blue endeavour prawns *Metapenaeus endeavouri* (Schmitt 1926) in the western Gulf of Carpentaria, Australia. Aust. J. Mar. Freshw. Res. 43, 1283–1299.

Buckworth, R.C., Deng, R.A., Plaganyi, E., Punt, A., Upston, J., Pascoe, S., Miller, M., Hutton, T., Lawrence, E. B., Venables, B., 2015. Northern Prawn Fishery RAG Assessment 2013-15. AFMA Project R13/0005 170pp.

Francis, R.I.C.C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138.

Fu, D., McKenzie, A., 2010. The 2010 stock assessment of paua (*Haliotis iris*) for Milford, George, Central, and Dusky in PAU 5A. New Zealand Fisheries Assessment Report 2010/46.

Gray, G.W., 1965. Tags for marking king crabs. Progr. Fish-Cult. 27, 221-227.

Haist, V., Breen, P.A., Starr, P.J., 2009. A multi-stock, length-based assessment model for New Zealand rock lobster (*Jasus edwardsii*). NZ J. Mar. Freshw. Res. 43, 355-371.

ICCAT, 2013. Report of the 2012 white marlin stock assessment meeting. Coll. Vol.Sci. Pap. ICCAT 69, 1085–1183.

Kirkwood, G.P., Somers, I.F., 1984 Growth of two species of tiger prawn, *Penaeus esculentus* and *P. semisulcatus*, in the western Gulf of Carpentaria. Aust. J. Mar. Freshw. Res 35, 703-712.

Lloyd-Jones, L.R., Wang, Y-G, Courtney, A.J., Prosser, A.J., Montgomery, S.S., 2012. Latitudinal and seasonal effects on growth of the Australian eastern king prawn (*Melicertus plebejus*). Can. J. Fish. Aquat. Sci. 69, 1525–1538.

McAllister, M.K., Ianelli, J.N., 1997. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54, 284–300.

North Pacific Fishery Management Council (NPFMC). 2008. Amendment 24. Final Environmental Assessment for amendment 24 to the Fishery Management Plan for Bering Sea/Aleutian Islands King and Tanner Crabs to Revise Overfishing Definitions. North Pacific Fishery Management Council, 605 West 4th Ave, Anchorage, AK 99501.

Punt, A.E., This volume. Some insights into data weighting in integrated stock assessments. Fish. Res. 00, 00–00.

Punt, A.E., Kennedy, R.B., 1997. Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. Mar. Freshw. Res. 48, 967–980.

Punt, A.E., Buckworth, R.C., Dichmont, C.M., Ye, Y., 2009. Performance of methods for estimating size–transition matrices using tag–recapture data. Mar. Freshw. Res. 60, 168–182.

Punt, A.E., Deng, R.A, Dichmont, C.M. Kompas, T., Venables, W.N., Zhou, S., Pascoe, S., Hutton, T., Kenyon, R., van der Velde, T., Kienzle, M., 2010. Integrating size-structured assessment and bio-economic management advice in Australia's Northern Prawn Fishery. ICES J. Mar. Sci. 67, 1785–1801.

Punt, A.E., Haung, T-C., Maunder, M.N., 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. 70, 16–33.

Richards, L.J., 1991. Use of contradictory data sources in stock assessments. Fish. Res. 11, 225–238.

Siddeek, M.S.M., Zheng, J., Punt, A.E., Vanek, V., In press. Estimation of size-transition matrices with and without moult probability for Alaska golden king crab using tag–recapture data. Fish. Res. 00, 00–00.

Somers, I.F., 1987. Sediment type as a factor in the distribution of commercial prawn species in the western Gulf of Carpentaria, Australia. Aust. J. Mar. Freshw. Res. 38, 133–149.

Somers I.F., 1988. On a seasonally oscillating growth function. Fishbyte 6, 8–11

Somers, I.F., Kirkwood, G.P., 1991. Population ecology of the grooved tiger prawn, *Penaeus semisulcatus*, in the north-western Gulf of Carpentaria, Australia: growth, movement, age structure and infestation by the bopyrid parasite *Epipenaeon ingens*. Aust. J. Mar. Freshw. Res. 42, 349–267.

Taylor, I., Stewart, I., Hicks, A., Garrison, T., Punt, A., Wallace, J., Wetzel, C., Thorson, J., Takeuchi, Y., Monnahan, C. 2014. Package r4ss. https://github.com/r4ss

Turnock, B.,J., Rugolo, L.J., 2014. Stock assessment of Eastern Bering Sea snow crab. pp. 41-177. Stock Assessment and Fishery Evaluation Report for the king and Tanner crab fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, 605 W. 4th Avenue, #306, Anchorage, AK 9950.

Wang, Y-G., Thomas, M.R., Somers, I.F., 1995. A maximum likelihood approach for estimating growth from tag–recapture data. Can. J. Fish. Aquat. Sci. 52, 252–259.

Xiao, Y., McShane P., 2000. Use of age- and time-dependent seasonal growth models in analysis of tag/recapture data on the western king prawn Penaeus latisulcatus in the Gulf St. Vincent, Australia. Fish. Res. 49, 85–92.

Zheng, J., Siddeek, M.S.M. 2014. Bristol Bay red king crab stock assessment in fall 2014. pp. 178-323. Stock Assessment and Fishery Evaluation Report for the king and Tanner crab fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, 605 W. 4th Avenue, #306, Anchorage, AK 9950.

Table 1. Data available for the stocks considered

Data type	Year range
Aleutian islands golden king crab	
Retained catch in numbers in total	1981 - 1984
Retained catch in numbers by length-class	1985 - 2014
Total (retained + discarded) catch in numbers	1990 - 2014
CPUE (observer)	1995 - 2014
Groundfish bycatch in numbers by length-class	1989 - 2014
Tagging data (recaptures to 2012)	1991, 1997, 2000, 2003, 2006 releases
P. semisulcatus and *P. esculentus*	
Weekly catch (mass) and effort (vessel days fished)	1970 - 2014
Fishery length-frequency data	2002 - 2003 (sex-specific)
Spawning survey indices	2002 - 2014
Spawning survey length-frequencies	2002 - 2014 (sex-specific)
Recruitment survey indices	2003 - 2014
Recruitment survey length-frequencies	2003 - 2015 (sex-specific)
Tagging data	1981 - 1984

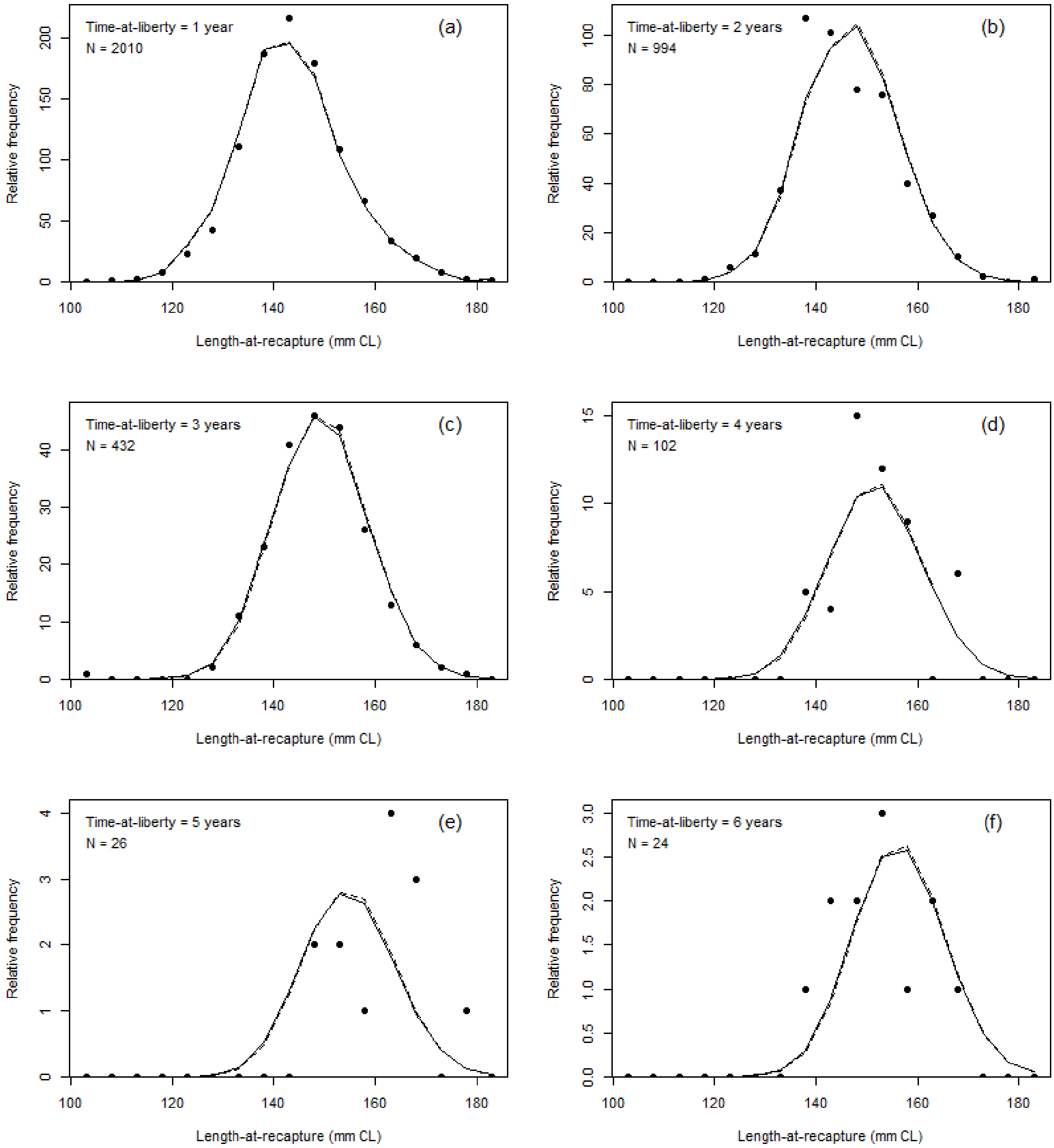

Figure 1. Observed tag recaptures (filled circles) versus model-predicted tag recaptures (solid and dashed lines) by size-class at recapture for male Aleutian Islands golden king crab. Results are shown for animals that were at liberty for between one and six years. N denotes the sample size. The solid lines show results for the original weighting and the dashed lines those when the tagging data are downweighted.

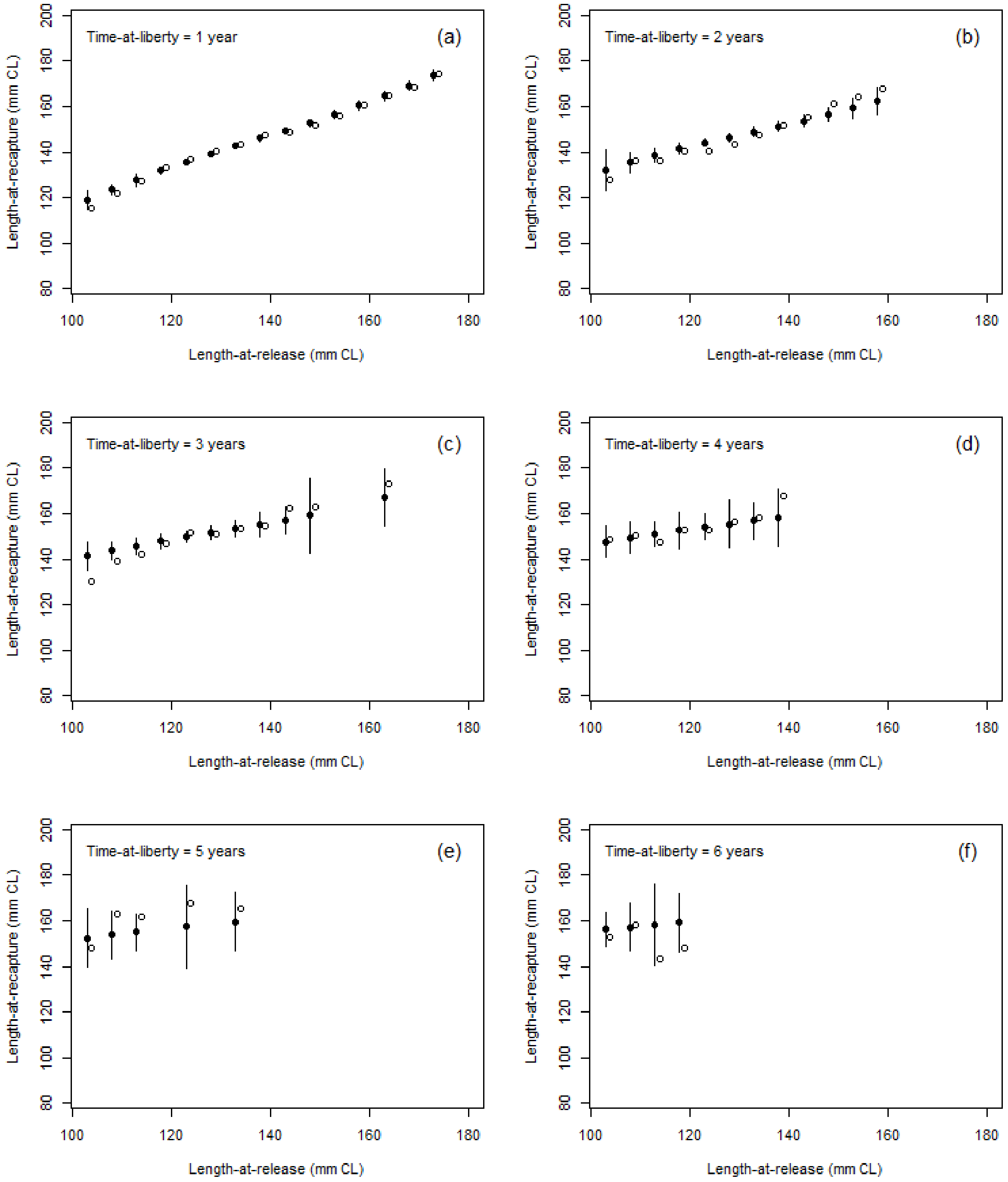

Figure 2. Observed (open circles) and model-predicted (filled circles) mean lengths-at-recapture versus release length for male Aleutian Islands golden king crab. The vertical lines indicate 95% confidence intervals for model-predicted mean lengths-at-recapture. Results are shown for animals that were at liberty for between one and six years. The model predictions correspond to the model configuration in which the tagging data are treated as the outcomes from Bernoulli experiments.

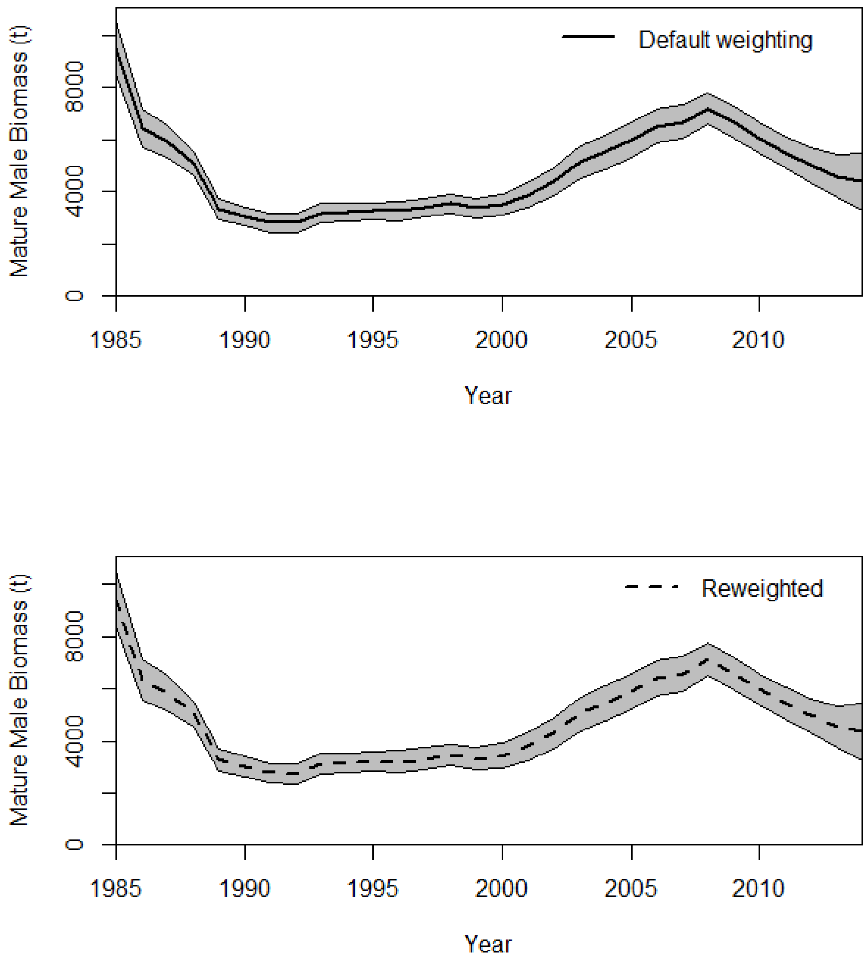

Figure 3. Time-trajectory of mature male biomass for Aleutian Islands golden king crab. Results are shown when the tagging data are taken to be the results of Bernoulli trials and when the tagging likelihood is downweighted using Equation 5.

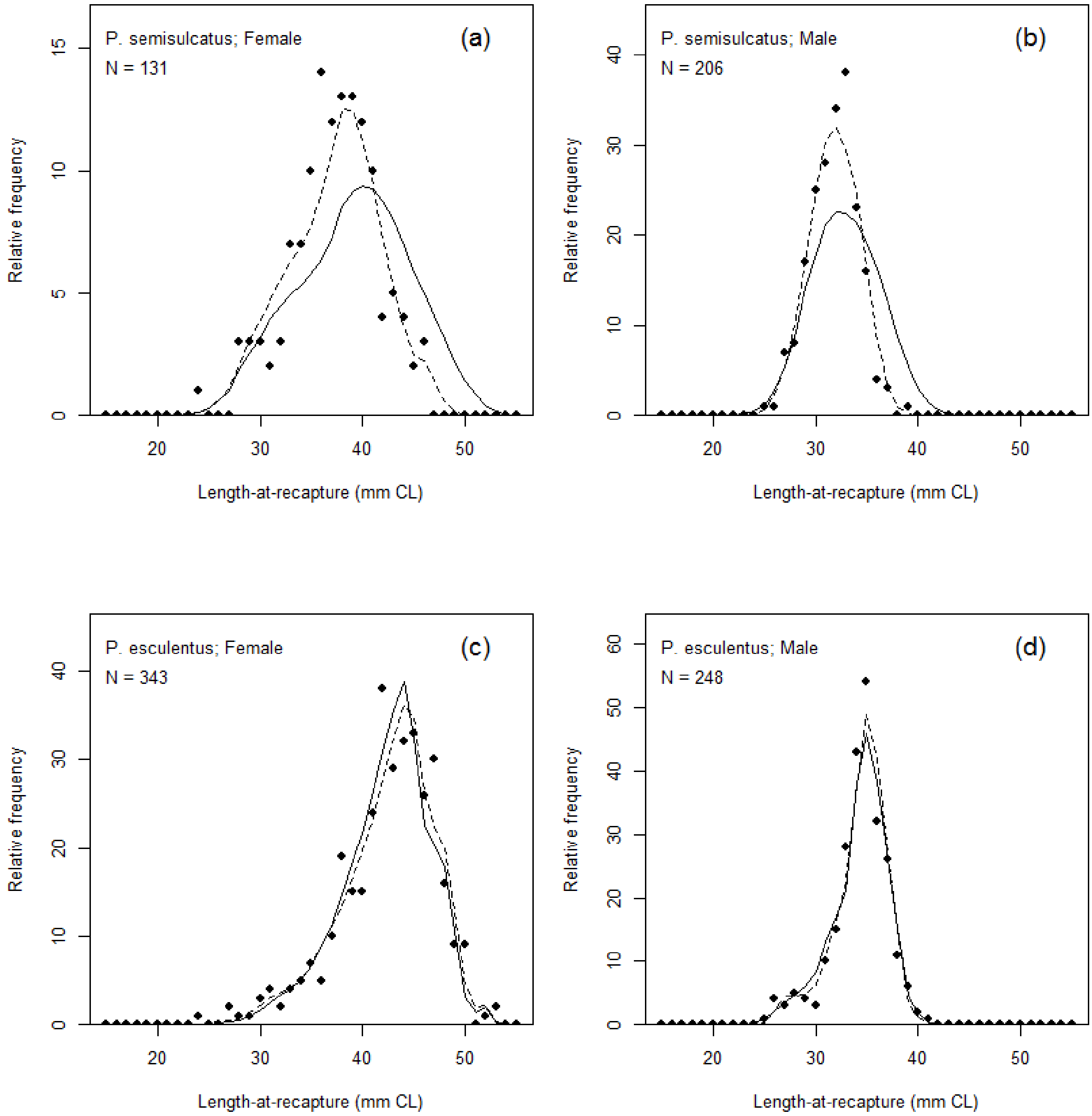

Figure 4. Observed tag recaptures (filled circles) versus model-predicted tag recaptures (solid and dashed lines) by size-class at recapture for two prawn species in Australia's northern prawn fishery by sex. Data are aggregated over time-at-liberty owing to small sample sizes. N denotes the sample size. The solid lines show results for the original weighting and the dashed lines those when the tagging data are upweighted 1000-fold.

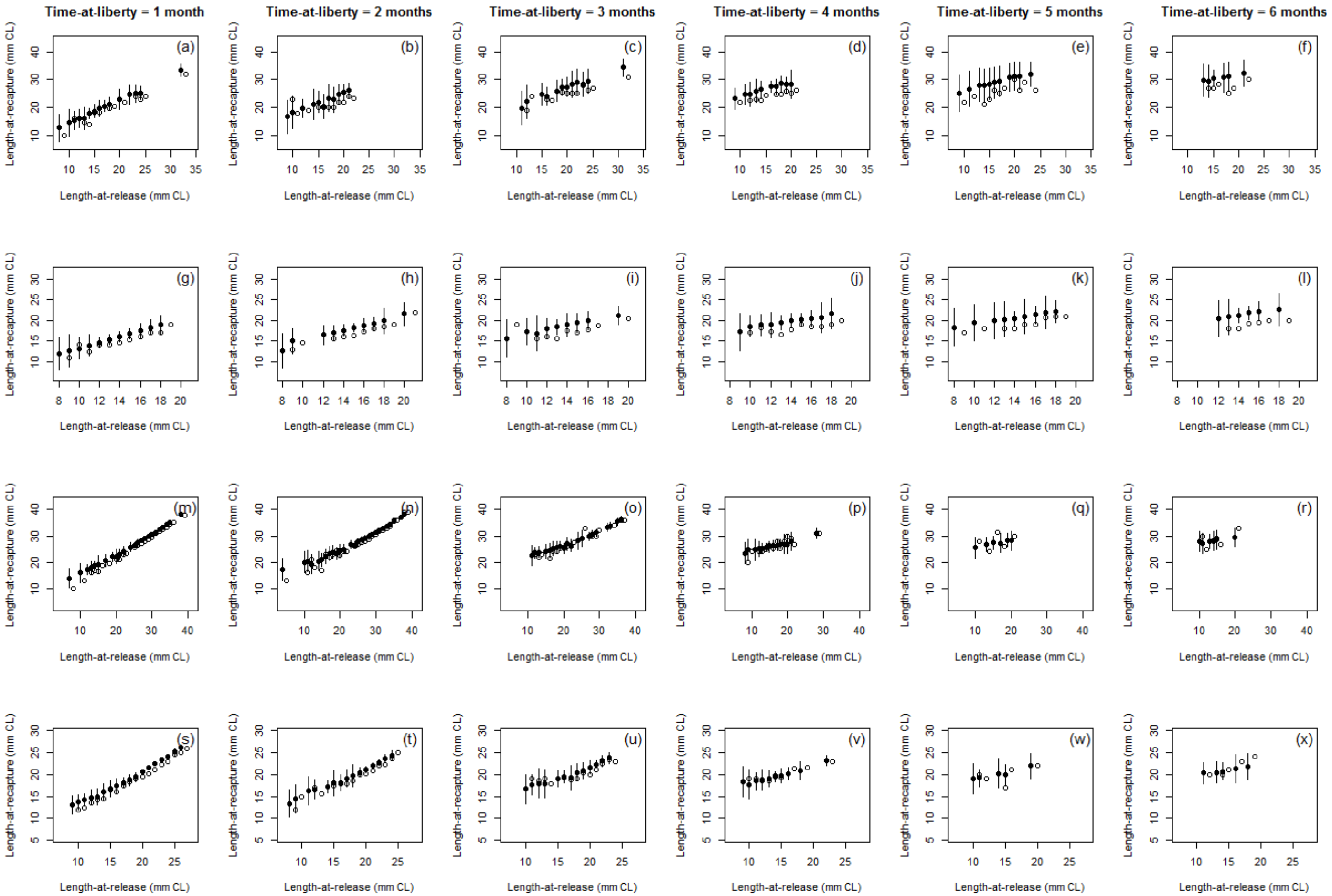

Figure 5. Observed (open circles) and model-predicted (filled circles) mean lengths at recapture versus release length for two prawn species in Australia's northern prawn fishery by sex. The vertical lines indicate 95% confidence intervals for model-predicted mean lengths-at-recapture. Results are shown for animals that were at liberty for between one and six months. Results for female *P. semisulcatus*, male *P. semisulcatus*, female *P. esculentus*, and male *P. esculentus* are respectively shown in each row (top to bottom). The model predictions correspond to the model configuration in which the tagging data are treated as the outcomes from Bernoulli experiments.

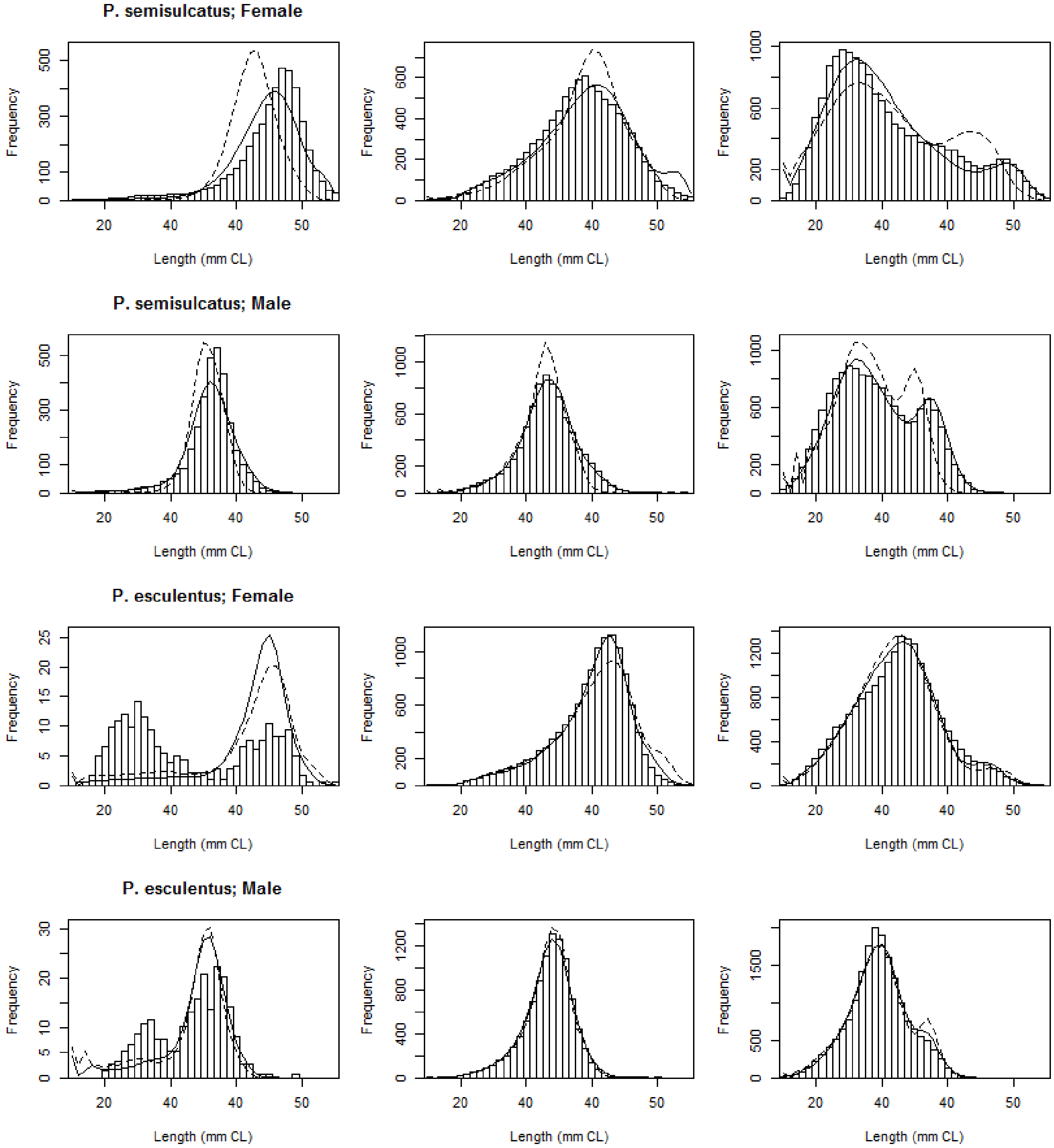

Figure 6. Observed (bars) model-predicted length-frequency distributions when the observations and model-predictions are aggregated over time (weighted by their assumed effective sample sizes). Results are shown for two species and two sexes (rows) and for three types of length-frequencies (catch, left; spawning, middle; recruitment, right). The solid lines show model predictions for the original weighting and the dashed lines those when the tagging data are upweighted 1000-fold.