Modified hierarchical Bayesian biomass dynamics models for assessment of short-lived invertebrates: a comparison for tropical tiger prawns

Shijie Zhou[A,E], *André E. Punt*[A,B], *Roy Deng*[A], *Catherine M. Dichmont*[A], *Yimin Ye*[A,C] *and Janet Bishop*[D]

[A]CSIRO Marine and Atmospheric Research, PO Box 120, Cleveland, Qld 4163, Australia.

[B]School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195-5020, USA.

[C]Present address: Fishery Management and Conservation Service, Food & Agriculture Organisation of the United Nations, Viele delle Terme di Caracalla, 00153 Rome, Italy.

[D]Present address: 20 Tooth Street, Nobby, Qld 4360, Australia.

[E]Corresponding author. Email: shijie.zhou@csiro.au

Abstract. Conventional biomass dynamics models express next year's biomass as this year's biomass plus surplus production less catch. These models are typically applied to species with several age-classes but it is unclear how well they perform for short-lived species with low survival and high recruitment variation. Two alternative versions of the standard biomass dynamics model (Standard) were constructed for short-lived species by ignoring the 'old biomass' term (Annual), and assuming that the biomass at the start of the next year depends on density-dependent processes that are a function of that biomass (Stock-recruit). These models were fitted to catch and effort data for the grooved tiger prawn *Penaeus semisulcatus* using a hierarchical Bayesian technique. The results from the biomass dynamics models were compared with those from more complicated weekly delay-difference models. The analyses show that: the Standard model is flexible for short-lived species; the Stock-recruit model provides the most parsimonious fit; simple biomass dynamics models can provide virtually identical results to data-demanding models; and spatial variability in key population dynamics parameters exists for *P. semisulcatus*. The method outlined in this paper provides a means to conduct quantitative population assessments for data-limited short-lived species.

Additional keywords: maximum likelihood, observation error, process error, squid, state-space, surplus production.

Introduction

Biomass dynamics (or production) models remain one of the most popular tools for analysing both finfish and shellfish population dynamics (Smith and Addison 2003). They are typically used when information on the age–structure of the catch is unavailable and hence when more sophisticated methods of stock assessment such as Virtual Population Analysis or statistical catch-at-age analysis cannot be applied, and when information on the size of the biomass of the population alone is adequate for management purposes. Biomass dynamics models use difference equations in which new biomass equals old biomass plus growth less catch, and 'growth' includes recruitment, somatic growth and natural mortality (Punt 2003; Chaloupka and Balazs 2007).

A variety of formulations of the biomass dynamics model have been developed and examined (reviewed in Quinn and Deriso 1999). An implicit assumption of most biomass dynamics models is that natural mortality is not very high so that a fairly large proportion of the biomass at the start of the next (annual) time-step consists of the biomass at the start of the

current time-step. However, the suitability of these models and their assumptions have rarely been examined for short-lived species such as tropical prawns and squids that exhibit high annual recruitment variation and for which the catch comprises only a single age class. Rather, researchers have used alternative approaches for assessment of short-lived species. For example, Roel and Butterworth (2000) and Isoda *et al.* (2005) adopted different recruitment functions for different stock sizes when assessing short-lived squids using biomass dynamics models. Bellido *et al.* (2001) used generalised additive models for modelling variation in abundance of squid rather than applying population dynamics models.

A single-stock, single-fleet biomass dynamics model based on the assumption that the dynamics are deterministic has been applied to species-aggregated data for two tiger prawn species (*Penaeus semisulcatus* and *P. esculentus*) in Australia's Northern Prawn Fishery (NPF), but the results were unrealistic (Dichmont *et al.* 2005). It was not clear whether this was because a biomass dynamics model is not suitable for a short-lived species or because the method used to fit the model was inappropriate.

The assumption that all of the error is in the observation process and that the dynamics are deterministic (the assumptions of an observation-error only estimator) is standard when applying biomass dynamics models (Punt and Hilborn 1996; Mueter and Megrey 2006). This assumption is often made because simulation studies have shown this method of fitting biomass dynamics models is more robust than the conventional alternative of assuming that the observations are made without error, but the dynamics are stochastic (Polacheck *et al*. 1993). In contrast, the dynamics of short-lived species are subject to considerable process error (annual recruitment constitutes a large proportion of future biomass) as well as observation error.

The concerns with the use of an observation-error estimator may be overcome using a Bayesian state-space formulation of the biomass dynamics model (Meyer and Millar 1999). Moreover, concern that it is inappropriate to apply a standard biomass dynamics model to short-lived species may be resolved by reformulating the biomass dynamics model. This paper therefore outlines two alternative formulations of the standard biomass dynamics model, which better account for the high rate of natural mortality for short-lived species, and applies them to the grooved tiger prawn (*P. semisulcatus*). These formulations account for the multi-species nature of the fishery for *P. semisulcatus* by fitting the model to catch-rate data from a fleet that targets *P. semisulcatus* and from another fleet that targets another prawn species in the Australia's NPF, *P. esculentus*, and has a bycatch of *P. semisulcatus*.

There is evidence that tiger prawns in the NPF constitute multiple 'stocks' (Dichmont *et al*. 2005). The analyses of this paper therefore analyse the data for *P. semisulcatus* using a hierarchical formulation of the Bayesian state-space method and hence impose priors on the between-stock variation in some key population dynamic parameters. This avoids the need to specify prior distributions for the parameters of the model using (non-existent) auxiliary information and imposes the assumption that the values for the parameters should not differ markedly among stock areas. Finally, the results from the alternative models are compared with those from two other models that have been applied to data for *P. semisulcatus*: (1) a model that aggregates data spatially, assumes a single fleet and estimates parameter values using an observation-error estimator (Dichmont *et al*. 2005; also see this fishery used as an example in Haddon 2001; Chapter 10) and (2) a weekly delay-difference model that incorporates additional parameters such as recruitment pattern, catchability, availability, growth and natural mortality, and estimates annual recruitment (Dichmont *et al*. 2003).

Methods

Alternative biomass dynamics models

In this section, the population dynamics models are assumed to be deterministic and any dependence on 'stock' is omitted for ease of presentation. The standard (or conventional) formulation of the biomass dynamics model is (Polacheck *et al*. 1993; Punt and Hilborn 1996):

$$B_y = B_{y-1} + rB_{y-1}\left(1 - \frac{B_{y-1}}{K}\right) - C_{y-1}, \tag{1}$$

where B_y is the biomass at the start of year y, r is the intrinsic growth rate, K is the carrying capacity and C_y is the total catch during year y. For short-lived species whose catchable biomass is made up entirely of new recruitment, last year's biomass contributes little to the biomass this year so Eqn 1 can be simplified to:

$$B_y = rB_{y-1}\left(1 - \frac{B_{y-1}}{K}\right) - C_{y-1}. \tag{2}$$

Eqn 2 is the popular logistic model for annual terrestrial organisms such as insects and plants (Gillman and Hails 1997) when the catch term C_y is omitted. Eqn 2 can be extended further with two alternative assumptions: (1) most of the catch occurs before spawning and (2) density-dependence is more likely to depend on this year's biomass (B_y) rather than last year's biomass (B_{y-1}), i.e.:

$$B_y = r(B_{y-1} - C_{y-1})\left(1 - \frac{B_y}{K}\right). \tag{3}$$

Rearranging Eqn 3 leads to:

$$B_y = \frac{r(B_{y-1} - C_{y-1})}{1 + \frac{r}{K}(B_{y-1} - C_{y-1})}. \tag{4}$$

Eqn 4 has the appearance of a classical Beverton-Holt stock-recruitment model when the mean weight is the same over time. The biomass remaining after fishing ($B_{y-1} - C_{y-1}$) represents the spawning biomass in the Beverton-Holt model, and r is the maximum recruits-per-spawner at low stock size.

The parameter K in Eqns 2 and 4 cannot be interpreted as carrying capacity (unlike in Eqn 1). Here, we defined the carrying capacity B_∞ as the equilibrium population size in the absence of fishing. Table 1 lists the equations for B_∞, the population growth rate λ (slope at origin), the biomass at which *MSY* is achieved and the *MSY* for each of models 1, 2 and 4. Table 1 also lists the relationships between the parameters r and K for each model. Eqn 1 will be referred to as the 'Standard' model, Eqn 2 as the 'Annual' model and Eqn 4 as the 'Stock-recruit' model. These three models can be extended to consider multiple stocks by substituting r_s for r, K_s for K and $B_{s,y}$ for B_y where s denotes stock.

Parameter estimation

The values for the parameters of the three models were estimated by fitting them to data on catch per unit effort (CPUE). For a multi-stock, multi-fleet fishery where some fleets target the species of interest and other fleets take it as bycatch, the model-estimate corresponding to the catch-rate for stock s, fleet f and year y, $\hat{U}_{s,f,y}$ is:

$$\hat{U}_{s,f,y} = q_{s,f}P_y B_{s,y}, \tag{5}$$

where $q_{s,f}$ is the catchability coefficient for stock s and fleet f and P_y is the relative fishing power during year y. The observed catch-rate was assumed to be log-normally distributed about its expected value in common with most applications of biomass dynamics models (Polacheck *et al*. 1993; Meyer and Millar 1999):

$$U_{s,f,y} \sim \text{log-normal}\{\ell n(E[\hat{U}_{s,f,y}], \tau_{U,s,f})\}, \tag{6}$$

Table 1.   Three alternative biomass dynamics models, the relationship between the parameters *r* and *K* for each model and the equations defining the population growth rate λ, carrying capacity B_∞, B_{MSY} and *MSY*

Note: the relationships between $r^{Std} \sim r^{SR}$ and $K^{Std} \sim K^{SR}$ are true only when $B_{MSY}^{Std} = B_{MSY}^{SR}$

	Standard	Annual	Stock-recruit
Model	$B_y = B_{y-1} + r^{Std} B_{y-1}\left(1 - \dfrac{B_{y-1}}{K^{Std}}\right) - C_{y-1}$	$B_y = r^{Anl} B_{y-1}\left(1 - \dfrac{B_{y-1}}{K^{Anl}}\right) - C_y$	$B_y = \dfrac{r^{SR}(B_{y-1} - C_{y-1})}{1 + \frac{r^{SR}}{K^{SR}}(B_{y-1} - C_{y-1})}$
r relationship	r^{Std}	$r^{Anl} = 1 + r^{Std}$	$r^{SR} = \left(\dfrac{2}{2 - r^{Std}}\right)^2$
K relationship	K^{Std}	$K^{Anl} = \dfrac{(1 + r^{Std})K^{Std}}{r^{Std}}$	$K^{SR} = \dfrac{K^{Std}}{r^{Std}}$
λ	$1 + r^{Std}$	r^{Anl}	r^{SR}
B_∞	K^{Std}	$\dfrac{(r^{Anl} - 1)K^{Anl}}{r^{Anl}}$	$\dfrac{(r^{SR} - 1)K^{SR}}{r^{SR}}$
B_{MSY}	$0.5K^{Std}$	$\dfrac{(r^{Anl} - 1)K^{Anl}}{2r^{Anl}}$	$K^{SR}\left(1 - \dfrac{1}{\sqrt{r^{SR}}}\right)$
MSY	$\dfrac{r^{Std}K^{Std}}{4}$	$\dfrac{(r^{Anl} - 1)^2 K^{Anl}}{4r^{Anl}}$	$K^{SR}\left(1 - \dfrac{1}{\sqrt{r^{SR}}}\right)^2$

where $\tau_{U,s,f}$ is the precision (the inverse of the variance) of the observation error for the catch-rate data for stock *s* and fleet *f*. $\tau_{U,s,f}$ is allowed to differ among fleets because it would not be expected that fleets that target a species and which take it as bycatch would lead to indices of abundance with the same precision.

Eqns 1, 2 and 4 are deterministic. However, it is necessary to hypothesise how realised biomass relates to the expectation based on Eqns 1, 2 and 4 to account for process error in the population dynamics (and hence formulate the biomass dynamics models as state-space models). For the purposes of this paper, we assumed that deviations about the expected biomass are log-normally distributed (Meyer and Millar 1999; Chaloupka and Balazs 2007), i.e.:

$$B_{s,y} \sim \text{log-normal}\{\ell n(E[B_{s,y}]), \tau_{B,s}\}, \qquad (7)$$

where $\tau_{B,s}$ is the precision of the process error for stock *s*. The prior for the biomass at the start of the first year of the modelled period is assumed to be the same as for the carrying capacity for stock *s*.

It is necessary to specify prior distributions for all of the parameters of the model to implement each of the three state-space models within a hierarchical Bayesian framework. Under the assumption that the key parameters are unlikely to differ substantially among areas, it was assumed that *r*, *K* and *q* were log-normally distributed about a common mean, i.e. *r*, *K* and *q* for each stock are random effects about a common mean, and:

$$\begin{aligned} K_s &\sim \text{log-normal}(\mu_K, \tau_K), \\ r_s &\sim \text{log-normal}(\mu_r, \tau_r), \qquad (8) \\ q_{s,f} &\sim \text{log-normal}(\mu_{q,f}, \tau_{q,f}), \end{aligned}$$

where μ_K, μ_r and $\mu_{q,f}$ are the prior means for *K*, *r* and fleet-specific catchability, respectively, and τ_K, τ_r and $\tau_{q,f}$ are the corresponding prior precisions. Collectively, these parameters are known as hyper-parameters (Harley and Myers 2001; Su *et al.*

2001). We assumed a normal distribution, $N(M_\theta, T_\theta)$, for μ_θ, where θ is either *K*, *r*, or *q*. Bayesian hierarchical models have the advantages that there is no need to set the values for the parameters of the priors, but only those of the hyper-parameters, and that the results of models are less sensitive to the values for the parameters of the hyper-prior than those of the prior. We specified values for the means (M_θ) of these hyper-priors (McAllister *et al.* 2004; Askey *et al.* 2007) by considering results of other studies and set the values for T_θ to large values so that the hyper-priors were relatively non-informative, but still proper (Gelman 2006). We tested a wide range of values for M_θ and found that the results were not sensitive to them. For example, setting $M_K = 8.2$ or 9.6 had little impact on the results. Two alternative approaches for setting the values for the hyper-parameter T_θ were considered: (1) set so that the coefficient of variation (CV) of the hyper-prior is 150% and so that the hyper-prior is relatively non-informative (McAllister *et al.* 2004) and (2) a half-Cauchy distribution. The Cauchy distribution was obtained as the ratio of a normal and the square root of a χ^2 distribution with one degree of freedom (Gelman 2006; Van Dongen 2006). The results for the two methods for setting T_θ were similar so results are only shown for the half-Cauchy hyper-prior.

The hyper-priors for the τ_θ, as well as the priors for the observation precisions, $\tau_{U,s,f}$, and the process precisions, $\tau_{B,s}$, were set to proper, but reasonably non-informative gamma distributions with mean 1 and variance 1000, i.e. G(0.001, 0.001).

In summary, the hierarchical structure of the alternative biomass dynamics models contains the following levels:

1. Hyper-priors: $\mu_\theta \sim N(M_\theta, T_\theta)$, $\tau_\theta \sim G(0.001, 0.001)$
2. Hyper-parameters: $\mu_K, \mu_r, \mu_{q,f}, \tau_K, \tau_r, \tau_{q,f}$
3. Priors: $\log(K_s) \sim N(\mu_K, \tau_K)$, $\log(B_{s,1}) \sim N(\mu_K, \tau_K)$, $\log(r_s) \sim N(\mu_r, \tau_r)$, $\log(q_{s,f}) \sim N(\mu_{q,f}, \tau_{q,f})$, $\tau_{U,s,f} \sim G(0.001, 0.001)$, $\tau_{B,s} \sim G(0.001, 0.001)$
4. Parameters: $K_s, r_s, q_{s,f}, B_{s,1}, \tau_{U,s,f}, \tau_{B,s}$
5. Data: $U_{s,f,y}$

Given the assumptions regarding the nature of the state-space model, the priors for the parameters and those for hyper-priors, the posterior distribution is proportional to:

$$
\begin{aligned}
&p(\mu_K)p(\tau_K)p(\mu_r)p(\tau_r)p(\underline{\mu}_{q,f})p(\underline{\tau}_{q,f}), \\
&p(\underline{K}_s|\mu_K,\tau_K)p(\underline{B}_{1970,s}|\mu_K,\tau_K)p(\underline{r}_s|\mu_r,\tau_r) \\
&\quad p(\underline{q}_{s,f}|\underline{\mu}_{q,f},\underline{\tau}_{q,f})p(\underline{\tau}_{B,s})p(\underline{\tau}_{U,s,f}), \\
&\prod_{s,y}\Big(p(B_{s,y}|B_{s,y-1},K_s,r_s,C_y,\tau_{B,s}) \\
&\prod_f p(U_{s,f,y}|B_{s,y},q_{s,f},P_y,\tau_{U,s,f})\Big),
\end{aligned}
\tag{9}
$$

where the underlined parameters denote a vector or matrix over stock s, fleet f and/or year y.

The Gibbs sampler, a Markov chain Monte Carlo (MCMC) technique, implemented using the WinBUGS package (http://www.mrc-bsu.cam.ac.uk/bugs, verified 1 November 2009) was used to sample parameter vectors from the posterior distribution (Eqn 9). Three Markov chains were constructed based on dispersed initial values and the results of the first 4000 cycles of each chain were taken as the burn-in period. The results of an additional 60 000 cycles from the three chains were saved, which formed the basis for further analysis. Whether the burn-in period was sufficient and the MCMC algorithm converged adequately to the posterior were evaluated by visually examining the three chains for each parameter in Eqn 9 and using the CODA package (Best *et al*. 1996).

Model diagnostics and selection

The fit of the model to the data was evaluated using the following criteria: (1) graphical assessment of the 95% prediction credibility intervals, (2) χ^2 goodness-of-fit statistics, (3) posterior predictive *p*-values and (4) Kolmogorov–Smirnov (KS) two-sample tests (Sheskin 1997). We calculated these statistics from posterior predictive distributions for the time-series of CPUE. For each observed catch-rate, this distribution was obtained by sampling parameters from the posterior distribution (Eqn 9) and then, conditional on those samples, sampling catch rates from the log-normal distribution assumed to capture observation error (Eqn 6). The posterior predictive distribution of catch rate for each fleet, stock and year, $u^{pred}_{s,f,y}$ is:

$$
p\big(u^{pred}_{s,f,y}|\mathbf{U}\big)=\int P\big(u^{pred}_{s,f,y}|\underline{\theta}\big)P(\underline{\theta}|\mathbf{U})d\underline{\theta}.
\tag{10}
$$

In this equation, $\underline{\theta}$ denotes all parameters, including model parameters and hyper-parameters. The 2.5th, 50th and 97.5th percentiles from the posterior predictive distribution were plotted together with the observed catch rates $U_{s,f,y}$.

The second criterion for model checking based on the posterior predictive distribution involved comparing the realised discrepancy χ^2_{rel} (between the observed catch rates and the posterior expected catch rates) and the posterior predictive χ^2_{pred} discrepancy (between catch rates from the posterior predictive distribution and the posterior expected catch rate) (Gelman *et al*.

1996). That is, for each fleet f, this χ^2 discrepancy was:

$$
\chi^2_f(u_f|\underline{\theta})=\sum_s\sum_y[\ln(u_{s,f,y})-\ln(\hat{U}_{s,f,y})]^2\tau_{U,s,f},
\tag{11}
$$

where u_f is either the observed CPUE or the predicted CPUE from Eqn 10 for fleet f.

We also calculated the posterior predictive *p*-value for the χ^2 discrepancy as:

$$
p\big(u^{pred}_f\big)=\int P\big[\chi^2_n\geq\chi^2_{pred,f}\big(u^{pred}_f|\underline{\theta}\big)\big]P(\underline{\theta}|\mathbf{U})d\underline{\theta},
\tag{12}
$$

where χ^2_n is the standard χ^2 distribution and n is the number of data points for each fleet.

The non-parametric KS two-sample test was used to test the hypothesis that the predicted catch rates for each replicated sample and the observed catch rates were from the same distribution. The *P*-value of this test is displayed using histograms. Furthermore, the proportion of replicates in which the null hypothesis is rejected at $\alpha=0.05$ is defined as the overall KS test *P*-value:

$$
p^{KS}_f=\frac{1}{n}\sum_s\sum_y I(p_{s,y,f}<0.05),
\tag{13}
$$

where n is the total number of data points for fleet f, I is the indicator function that takes the value of 1 when its argument is true and zero otherwise, and $p_{s,y,f}$ is the probability value from the KS for each species, year and fleet.

We used two criteria to compare alternative models: the Deviance Information Criterion (DIC) (Spiegelhalter *et al*. 2002) and the mean square predictive error loss function (MSPE) on the log-scale (Ghosh and Norris 2005; Webster *et al*. 2008). The latter was defined as:

$$
MSPE=\frac{1}{n}\sum_s\sum_f\sum_y\big[\ln\big(u^{pred}_{s,f,y}\big)-\ln(U_{s,f,y})\big]^2,
\tag{14}
$$

where $u^{pred}_{f,s,y}$ is sampled from Eqn 10. These equations were coded directly in the WinBUGS program, except for the KS test and Eqn 13, which were implemented using R. The best model was the one that had the smallest values for $E[MSPE]$ and DIC.

Application to P. semisulcatus

The grooved tiger prawn *P. semisulcatus* is a tropical species with a typical life span of less than 18 months and an assumed natural mortality rate of 0.045 week^{-1} (Dichmont *et al*. 2003). Sex-specific length–frequency data from scientific surveys show that the catch just before the start of fishing season is largely composed of a single cohort (Ye *et al*. 2007). Given the high mortality typically associated with prawns, this implies that few animals will survive an entire year. Catch-rate data for each stock were available for two 'fleets': one that targets *P. semisulcatus* and another that catches this species as bycatch when targeting another commercially-valuable prawn species, *P. esculentus*. Fishing power in Eqn 5 is expressed relative to that at the start of 1993 (Dichmont *et al*. 2003) so $q_{s,f}$ is the catchability coefficient at the start of 1993. There is a need to model changes over
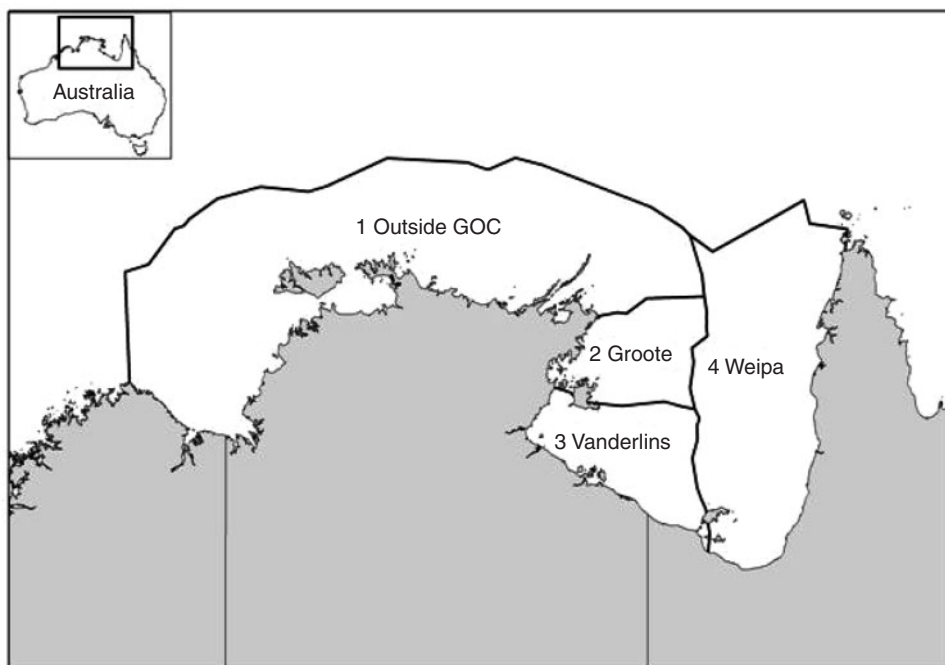
Fig. 1.   The four stock regions in the Northern Prawn Fishery for *Penaeus semisulcatus*.

time in fishing power because of improvements in technology and fishing skill in the NPF (Bishop 2006).

Compulsory commercial logbook data form the primary basis for the assessment of prawn species in the NPF, including that for *P. semisulcatus*. These data can be divided into fishing days, which targeted one of the two tiger prawns or other species (in particular, the common banana prawn, *P. merguiensis*) based on the probability for each fishing day of catching banana or tiger prawns (Venables *et al.* 2006). In addition, although catch and effort data are recorded by species group rather than species (e.g. *P. semisulcatus* and *P. esculentus* combined rather than individually), information on, for example, the date and location of shots can be used to split the species-combined catches to those of individual species (Venables and Dichmont 2004; Dichmont *et al.* 2005). Furthermore, the catches of tiger prawns by day are assigned to one of the two 'fleets' based on whichever tiger species had the highest relative probability of being caught on that day given where the fishing occurred (Venables and Dichmont 2004).

The two species of tiger prawns in the NPF have each been divided into seven putative 'stocks' based on geographic and biological information. These have been combined into four stocks for assessment purposes (Fig. 1) primarily because the abundance in some of these putative stocks is so low that data are uninformative, precluding the application of assessment models (Dichmont *et al.* 2005). For simplicity, we refer to these four stocks as follows: Outside Gulf of Carpentaria (GoC) (Stock 1), Groote (Stock 2), Vanderlins (Stock 3) and Weipa (Stock 4). The application to *P. semisulcatus* is based on 291 catch-rate data points (i.e. 2 fleets, 4 stocks and 38 years with a few 0-effort data points).

Results

Model diagnostics and selection

The convergence diagnostics generally do not exhibit evidence for non-convergence after ~2000 cycles of the MCMC algorithm (e.g. a value for Gelman-Rubin statistic around 1.0), suggesting that the length of the burn-in and the number of subsequent cycles is sufficient for the results to form the basis for inference. The fits of the three models to the catch-rate data are visually very similar, and suggest that the models mimic the data well apart from the catch rates for the bycatch fleet for the Outside GoC and Weipa stocks (Stocks 1 and 4; Fig. 2). Consequently, detailed results are only shown for one of these models (Stock-recruit) for the model diagnostics.

The posterior predictive distributions for the catch rates for the target fleet mimic the observed catch rate data and are relatively narrow (e.g. Fig. 3*a*). In contrast, and as expected from Fig. 2, the posterior predictive distributions for catch rates for the bycatch fleet are much broader, especially for Stock 1 (Outside GoC) and Stock 4 (Weipa) (Fig. 3*b*).

The realised discrepancy χ^2_{rel} and the predictive discrepancy χ^2_{pred} do not indicate problems of model fit. For example, the proportion of points above the 45° line, which is the *P*-value for this χ^2-test, is close to 0.5. The predictive *P*-values are similar between models and fleets: 0.533, 0.481 and 0.518 for the target fleet for the Standard, Annual and Stock-recruit models, and 0.515, 0.523 and 0.507 for the bycatch fleet for these three models, respectively.

The results of the KS two-sample test are somewhat different from those of the χ^2 goodness-of-fit test. Although the target fleet has high KS *P*-values (the overall *P*-value = 1), the distribution of *P*-values for the bycatch fleet is relatively uniform (Fig. 4).
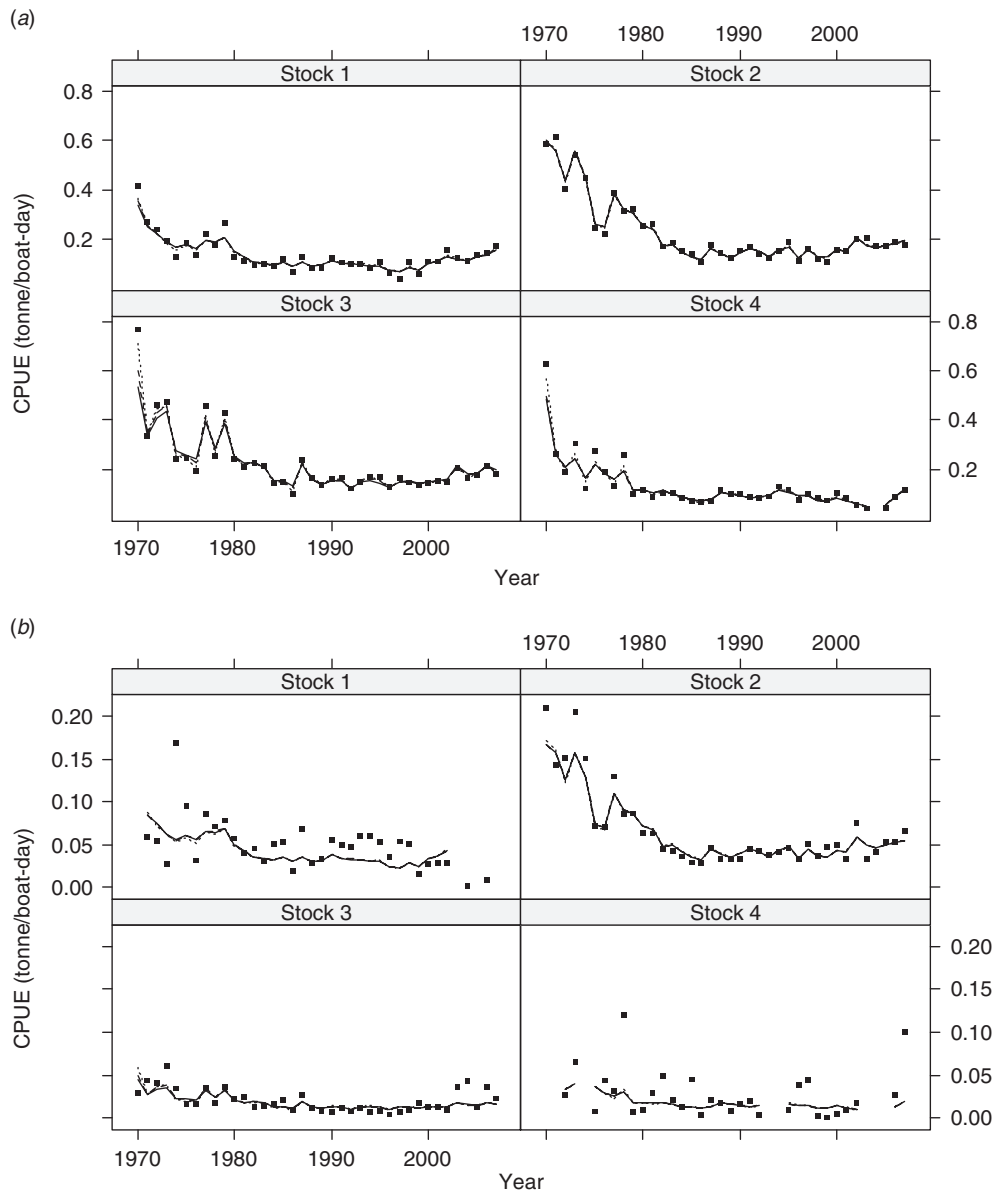
Fig. 2. Observed catch rates for the (*a*) targeted and (*b*) bycatch fleets (dots), and the posterior median time-trajectories of predicted catch-rate from three alternative models: Standard (solid line), Annual (dashed line) and Stock-recruit (dotted line). Stock 1, Outside Gulf of Carpentaria; Stock 2, Groote; Stock 3, Vanderlins; Stock 4, Weipa.

The overall *P*-value (Eqn 13) is 0.969, meaning that the null hypothesis that the predicted and the observed data are from the same distribution was rejected for nearly 3% of the replicates. This KS test indicates that the model fits the catch rate data for the target fleet better than the catch rate data for the bycatch fleet. It also indicates that the KS test is more sensitive than the overall χ^2-test.

The model selection method based on DIC and the mean square predicted loss selected the Stock-recruit model as 'best' and the Standard model as 'worst'. The Standard and Annual models had, respectively, DICs 56.85 and 47.59 greater than that for the Stock-recruit model. The extent of difference in DIC between the Stock-recruit model and other two models is 'definitive' (DIC difference >10; Spiegelhalter *et al*. 2002), whereas the difference between the Annual and Standard models is 'substantial' (DIC difference between 5 and 10). The mean MSPEs were consistent with the inferences based on DIC; 0.425, 0.422 and 0.419 for the Standard, Annual and Stock-recruit models, respectively.

Quantities of management interest

The posterior distributions for B_{MSY} and *MSY* do not differ substantially among models. The posterior medians for *MSY* (summed across stocks) were 1927, 1921 and 2001 tonnes for
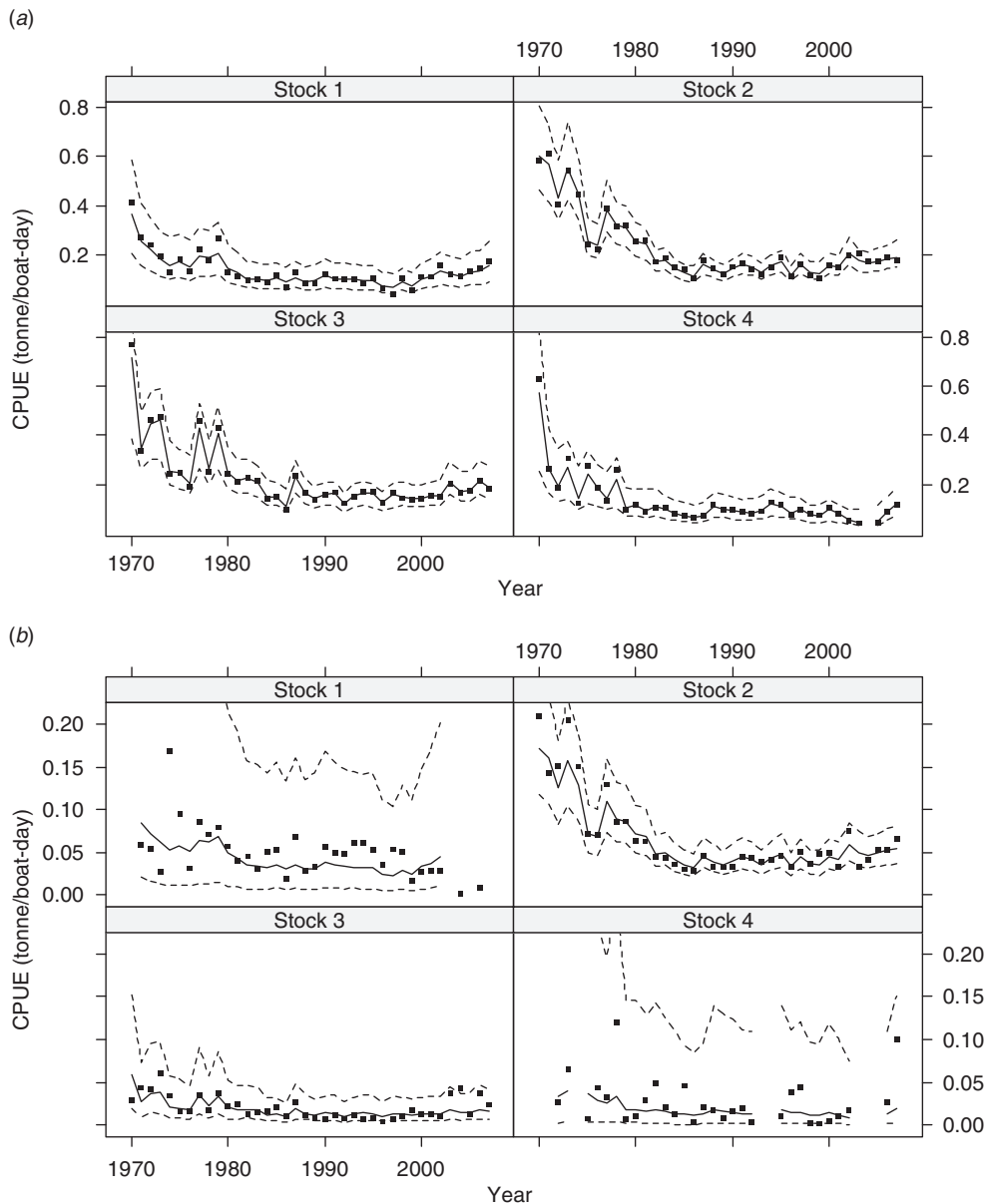
Fig. 3. Observed catch rates (dots) for the (*a*) targeted and (*b*) bycatch fleets, and the posterior predictive distributions (medians and 95% credibility intervals) for catch rates based on the Stock-recruit model.

the Standard, Annual and Stock-recruit models, respectively. These values are slightly higher than the estimate of *MSY* from the weekly delay-difference model currently used to provide management advice (Dichmont *et al.* 2003) ($M\hat{S}Y = 1768$ t), but fall within the 95% confidence intervals for this estimate (1517–2043 t).

The time-trajectories of biomass relative to B_{MSY} (a key management indicator for the NPF) from the three biomass dynamics models are similar for each individual stock (Fig. 5) and when the data for all stocks are aggregated (Fig. 6), although the Stock-recruit model tends to produce a higher estimate of B/B_{MSY} than the other two models. All of the analyses suggest that the stocks have been reduced in abundance since the start of fishing in

1970, dropped below B_{MSY} during early the 1980s and increased in abundance in recent years. The posterior median values for B_{2007}/B_{MSY} exceeded 1 for three of the four stocks. The exception is the Groote stock (Stock 2), for which the posterior median for B_{2007}/B_{MSY} is 0.86, 0.87 and 0.95 for the Standard, Annual and Stock-recruit models, respectively.

When aggregated over stocks, the ratio of current biomass to B_{MSY} exceeds 1 and the time-trajectory of B/B_{MSY} is remarkably similar to that from the weekly delay-different model even though the latter is substantially more complicated than a biomass dynamics model (Fig. 6). In contrast, the results from a Schaefer biomass dynamics model implemented as a maximum-likelihood observation-error estimator and fitted by

means of maximum likelihood (Dichmont *et al.* 2005) differ markedly from those of the Bayesian state-space models and the weekly delay-difference model even though it uses the same basic data and makes the same assumptions about changes over time in fishing power as the other biomass dynamics model (Fig. 6).

The posterior distribution provides a convenient way to examine parameter uncertainty. The coefficients of variation for K,
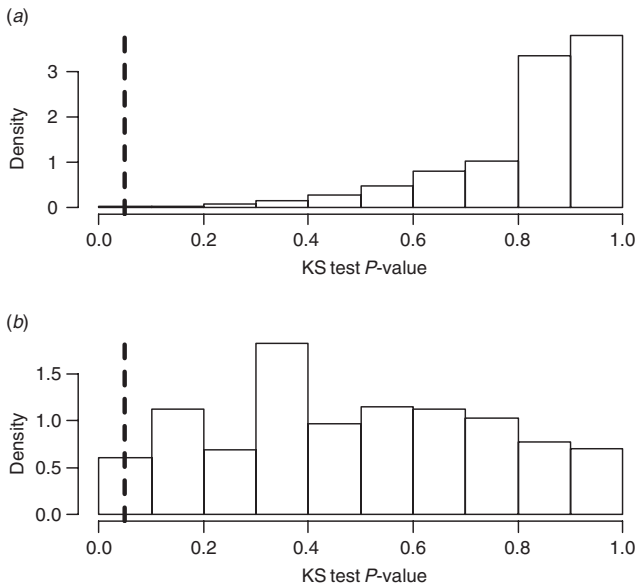
B_{MSY}, MSY and growth rate r are fairly small (generally below 20% for each stock and when results are aggregated spatially). Catchability q for the target fleet is also precise: a CV of 18%, 15% and 15% for the Standard, Annual and Stock-recruit models, respectively. However, the CV of q for the bycatch fleet is high: 52%, 55% and 54% for the Standard, Annual and Stock-recruit models, respectively. The posteriors for the catchability of the target fleet are similar among stocks but those for the catchability of the bycatch fleet vary among stocks (Fig. 7). The process and observation error variances are similar among the three models. However, the observation error variances differ substantially between the target and bycatch fleets. The observation error variances for the bycatch fleet also differ substantially among the four stocks.

Discussion

This study demonstrates that biomass dynamics models are appropriate for short-lived species when both process and observation error are taken into account. The biomass dynamics model that assumes that density-dependence is governed by current year biomass (the Stock-recruit model) appears to be particularly effective for short-lived species. While the methods developed in this paper have clear advantages, some caveats should be taken into consideration.

Advantages of hierarchical Bayesian state-space models

Comparisons between hierarchical Bayesian biomass dynamics models, the weekly delay-difference model and a standard observation-error estimator indicate the former has clear advantages. The estimates of the ratio of biomass to B_{MSY} from the hierarchical Bayesian biomass dynamics models are virtually identical to those from a more sophisticated weekly delay-difference model. In contrast, the estimates of this ratio from
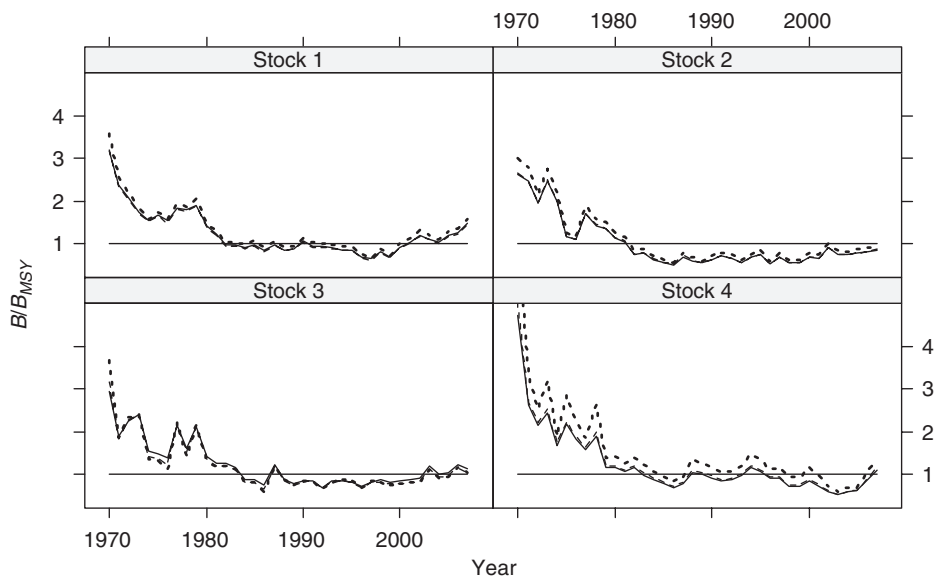

Fig. 4. Distribution of Kolmogorov–Smirnov (KS) test *P*-values comparing posterior predictive catch per unit effort (CPUE) from the Stock-recruit model and the observed CPUE. The vertical dashed line is where $P = 0.05$. (*a*) Target fleet; (*b*) bycatch fleet.

Fig. 5. Posterior median time-trajectories for B/B_{MSY} by stock (Stock 1, Outside Gulf of Carpentaria; Stock 2, Groote; Stock 3, Vanderlins; Stock 4, Weipa) and model (Standard, solid line; Annual, dashed line; Stock-recruit, thick dotted line).

the standard observation-error estimator are markedly different. This can be attributed to making allowance for process error and hence capturing the dynamics of the resource better.

Past attempts to assess even data-rich prawn species by stock have led to unreliable or unrealistic results (Dichmont *et al*. 2005). The use of a Bayesian estimation framework, which imposes hyper-priors on the key parameters of the model, clearly improved the stability of the model by allowing the assessment for the more data-poor stocks to 'borrow strength' from those for the more data-rich stocks. The benefits of hierarchical Bayesian techniques in this respect have been identified for several applications in the past (Rivot and Prevost 2002; McAllister *et al*. 2004).

The results of a stock-specific assessment reveal spatial differences in both parameter values and stock status (although perhaps less than would have been the case had the assessment not imposed priors on the parameters). In particular, although biomass of *P. semisulcatus* is assessed to be above B_{MSY} when results are aggregated over stocks in the NPF, the stock-specific results indicate that at least one stock (Groote) has not recovered to the extent that the other stocks have and remains below B_{MSY}.

Potential violation of model assumptions

The Stock-recruit model is selected as 'best' using DIC and the mean square predicted loss even though the fits to the data were
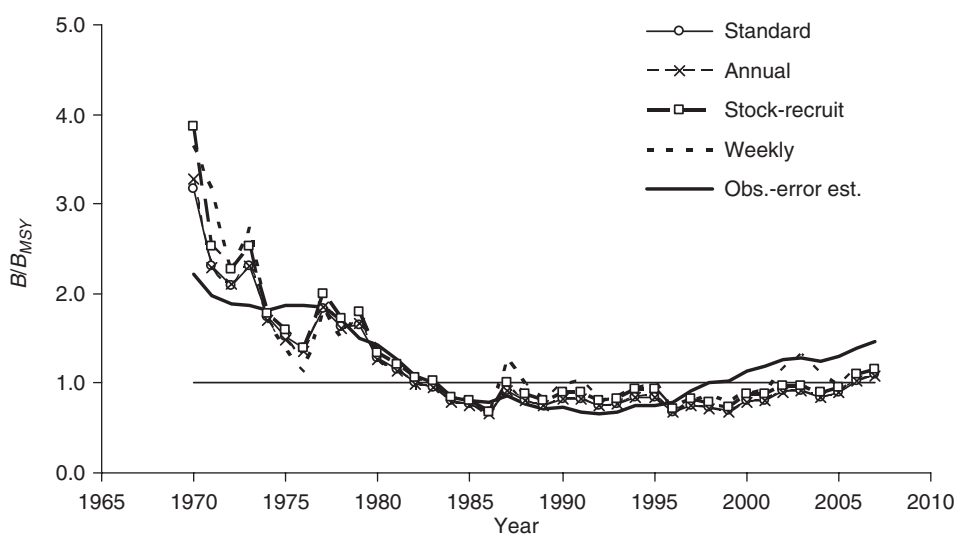

Fig. 6. Posterior median time-trajectories for B/B_{MSY} for the three alternative biomass dynamics models aggregated over stock, the weekly delay-difference model, and a maximum likelihood observation-error estimator.
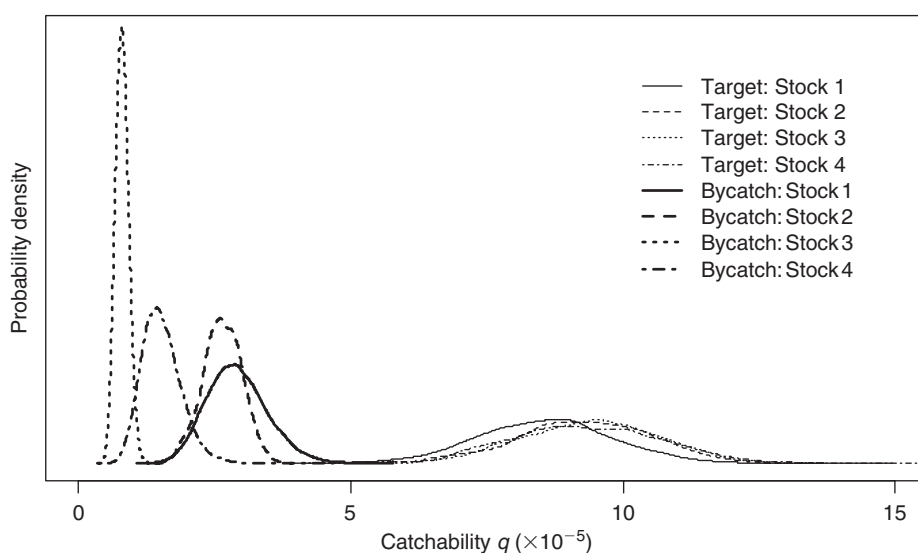

Fig. 7. Posterior distributions for catchability q ($\times 10^{-5}$) by fleet (target and bycatch) and stock from the Stock-recruit model.

visually very similar to those of the other models. This model assumes that very few prawns survive a year, and that the density-dependence is a function of current rather than past biomass. However, these assumptions will be violated to some extent for *P. semisulcatus* because at least some animals survive an entire year. Moreover, spawning occurs over an extensive period indicating that a discrete formulation for the biomass dynamics will always be an approximation irrespective of assumptions regarding density-dependence and survival.

Effects of observation error between target and bycatch fleets

We presented results for the three alternative biomass dynamics models where the precision parameter $\tau_{U,s,f}$ (the inverse of the variance) of the observation error for the catch-rate data varies among stocks and fleets. We compared two alternative assumptions regarding the variance of the observation error: (1) it is the same across stocks and fleets and (2) it differs among stocks, but is the same for each fleet. The time-trajectories of B/B_{MSY} from these models are much smoother than those shown in Figs 5 and 6. However, the models fit the data poorly. For example, ΔDIC is 513 for the variant of the Stock-recruit model in which it is assumed that the observation error variance is the same among stocks and fleets, and 201 when the observation error variance varies among stocks, but not between fleets.

The results of poor fits to the bycatch fleet data for two stocks (Outside GoC and Weipa) also suggest that assuming a constant observation error variance across stocks and fleets is inappropriate. These poor fits are mainly due to very limited catch and effort data. For example, only 2% and <1% of total effort by the bycatch fleet occurred on the Outside GoC and Weipa stocks, respectively.

Application of the method to other short-lived invertebrates

The hierarchical Bayesian biomass dynamics models developed in this paper could be applied to other short-lived invertebrate species for which only catch and effort data are available. The data for *P. semisulcatus* are adequate to apply fairly complicated stock assessment methods. However, this is not the case generally for species in the NPF for which information on stock status is needed. These species lack information on recruitment pattern, catchability, availability, growth and natural mortality, which precludes application of, for example, the method of Dichmont *et al.* (2003) to the data for these species. The similarity of results between the biomass dynamics models implemented in the state-space framework and those of the weekly delay-different models provides some confidence that the biomass dynamics models outlined in this paper may be applied to data for species such as blue and red endeavour prawns (*Metapenaeus endeavouri* and *M. ensis*), red-legged banana prawns (*Fenneropenaeus indicus* – formerly *Penaeus indicus*) and king prawns (*Melicertus latisulcatus* and *M. longistylus*), which are of commercial value and for which data on catch and effort are available, but for which data on biological parameters such as growth and natural mortality are either absent or considered unreliable. Of course, model diagnostics and examination are needed when one applies this method to other species because *P. semisulcatus* is perhaps unusual among tropical prawns because recruitment appears to be functionally related to spawning stock size, and among-year fluctuations in recruitment are relatively small.

References

Askey, P. J., Post, J. R., Parkinson, E. A., Rivot, E., Paul, A. J., and Biro, P. A. (2007). Estimation of gillnet efficiency and selectivity across multiple sampling units: a hierarchical Bayesian analysis using mark-recapture data. *Fisheries Research* **83**, 162–174. doi:10.1016/J.FISHRES.2006.09.009

Bellido, J. M., Pierce, G. J., and Wang, J. (2001). Modelling intra-annual variation in abundance of squid *Loligo forbesi* in Scottish waters using generalised additive models. *Fisheries Research* **52**, 23–39. doi:10.1016/S0165-7836(01)00228-4

Best, N., Cowles, M. K., and Vines, K. (1996). 'CODA Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output.' (MRC Biostatistics Unit: Cambridge, UK.)

Bishop, J. (2006). Standardizing fishery-dependent catch and effort in a complex fishery where technology changed. *Reviews in Fish Biology and Fisheries* **16**, 21–38. doi:10.1007/S11160-006-0004-9

Chaloupka, M., and Balazs, G. (2007). Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock. *Ecological Modelling* **205**, 93–109. doi:10.1016/J.ECOLMODEL.2007.02.010

Dichmont, C. M., Punt, A. E., Deng, A., Dell, Q., and Venables, W. (2003). Application of a weekly delay-difference model to commercial catch and effort data for tiger prawns in Australian's Northern Prawn Fishery. *Fisheries Research* **65**, 335–350. doi:10.1016/J.FISHRES.2003.09.024

Dichmont, C. M., Deng, A. R., Venables, W. N., Punt, A. E., Haddon, M., *et al.* (2005). A new approach to assessment in the NPF: spatial models in a management strategy environment that includes uncertainty. Fisheries Research and Development Corporation Report 2001/002, Cleveland, Australia.

Van Dongen, S. (2006). Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology* **242**, 90–100. doi:10.1016/J.JTBI.2006.02.002

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis (Online)* **1**, 515–534. doi:10.1214/06-BA117A

Gelman, A., Meng, X., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**, 733–807.

Gillman, M., and Hails, R. (1997). 'An Introduction to Ecological Modelling—Putting Practice into Theory.' (Blackwell Science: Oxford, UK.)

Ghosh, S. K., and Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioural effects. *Journal of Agricultural Biological & Environmental Statistics* **10**, 35–49. doi:10.1198/108571105X28651

Haddon, M. (2001). 'Modelling and Quantitative Methods in Fisheries.' (Chapman and Hall/CRC: Boca Raton, FL.)

Harley, S. J., and Myers, R. A. (2001). Hierarchical Bayesian models of length-specific catchability of research trawl surveys. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1569–1584. doi:10.1139/CJFAS-58-8-1569

Isoda, Y., Bower, J. R., and Hasegawa, S. (2005). Assessing environmental effects on recruitment of Japanese common squid (*Todarodes pacificus*) in the Japan Sea using a biomass dynamics model. *Bulletin of Fisheries Science Hokkaido University* **56**, 19–31.

McAllister, M. K., Hill, S. L., Agnew, D. J., Kirkwood, G. P., and Beddington, J. R. (2004). A Bayesian hierarchical formulation of the DeLury stock assessment model for abundance estimation of Falkland Islands' squid (*Loligo gahi*). *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 1048–1059. doi:10.1139/F04-084

Meyer, R., and Millar, R. B. (1999). BUGS in Bayesian stock assessments. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 1078–1087. doi:10.1139/CJFAS-56-6-1078

Mueter, F. J., and Megrey, B. A. (2006). Using multi-species surplus production models to estimate ecosystem-level maximum sustainable yields. *Fisheries Research* **81**, 189–201. doi:10.1016/J.FISHRES.2006.07.010

Polacheck, T., Hilborn, R., and Punt, A. E. (1993). Fitting surplus production models: comparing methods and measuring uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences* **50**, 2597–2607. doi:10.1139/F93-284

Punt, A. E. (2003). Extending production models to include process error in the population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences* **60**, 1217–1228. doi:10.1139/F03-105

Punt, A. E., and Hilborn, R. (1996). 'Biomass Dynamic Models. User's Manual.' FAO Computerized Information Series (Fisheries). (Rome: FAO.)

Quinn, T. J., and Deriso, R. B. (1999). 'Quantitative Fish Dynamics.' (Oxford University Press: New York.)

Rivot, E., and Prevost, E. (2002). Hierarchical Bayesian analysis of capture-mark-recapture data. *Canadian Journal of Fisheries and Aquatic Sciences* **59**, 1768–1784. doi:10.1139/F02-145

Roel, B. A., and Butterworth, D. S. (2000). Assessment of the South African chokka squid *Loligo vulgaris reynaudii*: Is disturbance of aggregations by the recent jig fishery having a negative impact on recruitment? *Fisheries Research* **48**, 213–228. doi:10.1016/S0165-7836(00)00186-7

Sheskin, D. (1997). 'Handbook of Parametric and Nonparametric Statistical Procedures.' (CRC Press: Boca Raton, FL.)

Smith, M. T., and Addison, J. T. (2003). Method for stock assessment of crustacean fisheries. *Fisheries Research* **65**, 231–256. doi:10.1016/J.FISHRES.2003.09.017

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B. Methodological* **64**, 583–639. doi:10.1111/1467-9868.00353

Su, Z., Adkison, M. D., and van Alen, B. W. (2001). A hierarchical Bayesian model for estimating historical salmon escapement and escapement timing. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1648–1662. doi:10.1139/CJFAS-58-8-1648

Venables, W., and Dichmont, C. M. (2004). A generalized linear model for catch allocation: an example from Australia's Northern Prawn Fishery. *Fisheries Research* **70**, 409–426. doi:10.1016/J.FISHRES.2004.08.017

Venables, W. N., Kenyon, R. A., Bishop, J. F. B., Dichmont, C. M., *et al.* (2006). Species distribution and catch allocation: data and methods for the NPF, 2002–2004. Final report. AFMA Project No. R01/1149. Australian Fisheries Management Authority, Canberra.

Webster, R. A., Pollock, K. H., Ghosh, S. K., and Hankin, D. G. (2008). Bayesian spatial modelling of data from unit-count surveys of fish in streams. *Transactions of the American Fisheries Society* **137**, 438–453. doi:10.1577/T06-138.1

Ye, Y., Kenyon, R. A., Burridge, C., Dichmont, C. M., Pendrey, R., *et al.* (2007). An integrated monitoring program for the Northern Prawn Fishery 2005/06. Project Australian Fisheries Management Authority R05/0599, Australia.