

Stock assessment of short-lived invertebrates using hierarchical Bayesian models

Shijie Zhou¹, Andre E. Punt^{1,2}, Roy Deng¹, Catherine M. Dichmont¹

¹CSIRO Marine and Atmospheric Research, PO Box 120, Cleveland, QLD 4163, Australia

²School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020, USA

Email: Shijie.zhou@csiro.au

Abstract: Fisheries for short-lived invertebrates typically have the following characteristics: (1) annual catch is comprised of a single age-class; (2) recruitments vary substantial inter-annually and the population dynamics are hence subject to large amounts of process error; (3) data contain large observation errors; (4) the resource is composed of several stocks; (5) several fleets may target or by-catch a species and stock; and (6) only catch and effort data are available. Thus, stock assessment for such fisheries is very difficult. Biomass dynamics (or production) models are a popular tool for stock assessment of finfish and shellfish fisheries that only have catch and effort data. 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 items (1-2) for short-lived species may violate this assumption. Moreover the standard biomass dynamics model assumes that all of the error is in the observation process and that the dynamics are deterministic. This model has been shown inefficient for short-lived invertebrates, either due to its inappropriate assumptions or the traditional estimation methods.

We construct two alternative versions of the standard biomass dynamics model (Standard) 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 (SR).

We test these models by fitting them to catch and effort data for the grooved tiger prawn *Penaeus semisulcatus* using a hierarchical Bayesian technique. The results from these models are compared to those from a far more complicated weekly delay-difference model. The analyses show that: the Standard model is flexible for short-lived species; the SR model provides the most parsimonious fit; and simple biomass dynamics models can provide virtually identical results to data-demanding models.

Keywords: *surplus production, process error, observation error, squid, state-space, maximum likelihood*

1. INTRODUCTION

It is difficult to conduct quantitative population dynamic analysis for short-lived invertebrates because recruitments exhibit large inter-annual variation and the population is hence subject to considerable process error, data contain large observation errors, and available data are limited to catch and effort only. The biomass dynamics model is the simplest stock assessment model that requires catch and effort data only (Smith and Addison 2003). Traditionally, the standard biomass dynamics model assumes that all of the error is in the observation process when estimating model parameters (Punt and Hilborn 1996; Mueter and Megrey 2006). This model has been shown inefficient for short-lived invertebrates, either due to its inappropriate assumptions or because of the use of the traditional estimation methods.

In this paper we present two alternative formulations of the standard biomass dynamics model which better account for the high variation in recruitment and high rate of natural mortality for short-lived species. We implement these models using hierarchical Bayesian techniques to capture both process and observation error. We test the method by applying it to the catch and effort data for the grooved tiger prawn (*P. semisulcatus*) in Australia's Northern Prawn Fishery (NPF). Data are available from a fleet which targets *P. semisulcatus* and from another fleet which targets another prawn species in the NPF, *P. esculentus*, and has a by-catch of *P. semisulcatus*.

The Bayesian state-space method imposes priors on the between-stock variation in some key population dynamic parameters. This allows a more "objective" basis for inference by estimating parameters of prior distributions from data rather than by using subjective information (Gelman 2006). The results from the standard biomass dynamics model and the two alternative models are compared to those from a weekly delay-difference model that incorporates additional parameters such as recruitment pattern, catchability, availability, growth, natural mortality, and estimates annual recruitment (Dichmont et al. 2003).

2. METHOD

2.1. Alternative biomass dynamics models

The standard biomass dynamics model is formulated as (Polacheck et al. 1993; Punt and Hilborn 1996):

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

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 entirely made up of new recruitments, the last year's biomass directly contributes little to the current year's biomass so Eqn 1 can be simplified as:

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

We further assume that major catch occurs before spawning so that density is more likely to depend on this year's biomass (B_y):

$$B_y = r(B_{y-1} - C_{y-1}) \left(1 - \frac{B_y}{K} \right), \text{ or equivalently, } B_y = \frac{r(B_{y-1} - C_{y-1})}{1 + \frac{r}{K}(B_{y-1} - C_{y-1})} \quad (3)$$

Eqn 1 will be referred to as the 'Standard' model, Eqn 2 as the 'Annual' model, and Eqn 3 as the 'SR model'. When considering multiple stocks, these three models can be extended to consider an individual stock by substituting r_s for r , K_s for K , and $B_{s,y}$ for B_y where s denotes stock.

2.2. Hierarchical Bayesian modelling

Catch and effort data can be used to calculate catch-rate, which is used for parameter estimation. The model-estimate corresponding to the catch-rate for stock s , fleet f , and year y for a multi-stock, multi-fleet, fishery is $\hat{U}_{s,f,y} = q_{s,f} P_y B_{s,y}$, 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 .

We assume observation error, process error, and the parameters each follow log-normal distributions, with expected values μ_θ and precision parameters τ_θ . Hyper-parameters μ_θ are assigned a normal distribution, and τ_θ a inverse gamma distribution. Bayesian hierarchical models have the advantage that one need only specify the values for hyper-parameters rather than the parameters of the priors, and that the results of analysis are less sensitive to the values for the parameters of the hyper-prior than those of the prior (Gelman 2006). For $\mu_\theta \sim N(M_\theta, T_\theta)$, we specified values for the means (M_θ) of these hyper-priors (Rivot and Prevost 2002; McAllister et al. 2004; Askey et al 2007) by considering results from other studies. We tested a wide range of values for M_θ and found that the results were not sensitive to those values. We set assigned the hyper-parameter T_θ as a half-Cauchy distribution (Dongen 2006; Gelman 2006). The hyper-priors for the τ_θ were set to proper, but reasonably non-informative gamma distributions with mean 1 and variance 1000, i.e., $gamma(0.001, 0.001)$, which is appropriate as a prior and leads to stable estimates.

The hierarchical structure of the alternative biomass dynamics models contains the following levels and each level conditions on the previous one:

- Hyper-priors: $\mu_\theta \sim N(M_\theta, T_\theta)$, $\tau_\theta \sim$ half-Cauchy distribution;
- Hyper-parameters: $\mu_\theta(\mu_K, \mu_r, \mu_{q,f})$, $\tau_\theta(\tau_K, \tau_r, \tau_{q,f})$;
- Priors: $\log(P) \sim N(\mu_\theta, \tau_\theta)$, $\tau_{U,s,f}$ and $\tau_{B,s} \sim G(0.001, 0.001)$;
- Parameters P: $K_s, r_s, q_{s,f}, B_{s,l}, \tau_{U,s,f}$ (observation errors), and $\tau_{B,s}$ (process errors);
- Data: $U_{s,f,y}$.

Given these assumptions regarding the nature of the state-space model, the priors for the parameters and those for hyper-priors, the posterior distribution for the parameters is proportional to the joint distribution of the likelihood of the data given the parameters, the priors for all of the parameters and corresponding hyper-priors. The MCMC sampler, implemented using the WinBUGS package was used to sample parameter vectors from the posterior distribution. Three Markov chains were conducted based on dispersed initial values, and the first 4,000 simulations from 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 MCMC algorithm converged adequately to the posterior was evaluated by visually examining the three chains for each parameter and using the Gelman-Rubin diagnostic statistic (Best et al. 1996).

2.3. Model diagnostics and selection

The fit of the model to the data is evaluated using the following criteria: (a) graphical assessment of the 95% prediction credibility intervals, (b) χ^2 goodness of fit statistics, (c) posterior predictive p values for statistics in (b), and (d) the Kolmogorov-Smirnov goodness of fit statistic (Sheskin 1997). We calculated these statistics from posterior predictive distributions for the time-series of catch-rate. For each observed catch-rate, this distribution is obtained by sampling parameters from the posterior distribution (Eqn 4) corresponding to MCMC simulations, and then, conditional on those samples, sampling catch-rates from the log-normal distribution assumed to capture observation error. The posterior predictive distribution of catch rate for each fleet, stock and year, $u_{s,f,y}^{pred}$ is:

$$P(u_{s,f,y}^{pred} | \mathbf{U}) = \int P(u_{s,f,y}^{pred} | \underline{\theta}) P(\underline{\theta} | \mathbf{U}) d\underline{\theta}. \quad (4)$$

In this equation, $\underline{\theta}$ denotes all parameters, including model parameters and hyper-parameters.

The second criterion compared the realized discrepancy χ_{rel}^2 (between the observed catch rates and the posterior expected catch rates) and the posterior predictive χ_{pred}^2 discrepancy (between catch rates from the posterior predictive distribution and the posterior expected catch rate) (Gelman et al. 1996). We also calculated $p(u_f^{pred}) = \int P(\chi_n^2 \geq \chi_{pred,f}^2(u_f^{pred} | \underline{\theta})) P(\underline{\theta} | \mathbf{U}) d\underline{\theta}$, the posterior predictive p -value for the χ^2 discrepancy, where χ_n^2 is a chi-square distribution, and n is the number of data points for each fleet.

The nonparametric Kolmogorov-Smirnov (KS) two-sample test is used to test the null hypothesis that the predicted catch rates for each replicated sample and the observed catch rates were from the same distribution (Sheskin 1997). Further, the proportion of replicates in which the null hypothesis was

rejected at $\alpha = 0.05$ defined the overall KS p -value, $p_f^{KS} = \frac{1}{n} \sum_s \sum_y I(p_{s,y,f} < 0.05)$, 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 test for each species, year and fleet.

We use 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 is defined as:

$$MSPE = \frac{1}{n} \sum_s \sum_f \sum_y [\ln(u_{s,f,y}^{pred}) - \ln(U_{s,f,y})]^2 \quad (5)$$

where n is again the total number of data points and $u_{s,f,y}^{pred}$ is sampled from Eqn 4. Here we select the best model according to the smallest values for MSPE and DIC.

2.4. Application to *P. semisulcatus*

The grooved tiger prawn *P. semisulcatus* is a tropical prawn 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. Given the high mortality typically associated with prawns, this implies that few animals will survive an entire year. Catch-rate data for each stock are available for two “fleets”: one that targets *P. semisulcatus* and another that catches this species as by-catch when targeting another commercially valuable prawn species, *P. esculentus*.

The two species of tiger prawns in the NPF have each been divided into four putative “stocks” based on geographic and biological information (Dichmont et al. 2005). For simplicity, we refer to these four stocks as: Outside Gulf of Carpentaria as Stock 1, Groote region as Stock 2, Vanderlins region as Stock 3, and Weipa region as Stock 4.

3. RESULTS

3.1. Model diagnostics and selection

The three Markov chains converge well after about 2,000 cycles of the MCMC algorithm and the value for Gelman-Rubin statistic stays around 1.0. The fits of the three alternative biomass dynamics models to the catch-rate data are visually very similar. Figure 1 shows the result of the SR model for the target and bycatch fleets. The posterior predictive distributions for the target fleet mimic the observed catch rate data well but the posterior predictive distributions for the bycatch fleet are not as good, exhibiting much lower precision, particularly for Stock 1 and Stock 4.

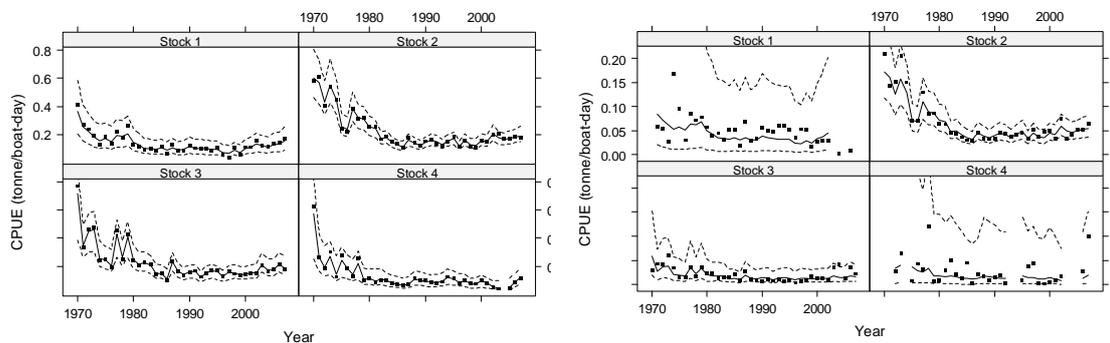


Figure 1. Observed catch-rates (dots) and the posterior predictive distribution medians (solid line) and 95% credibility intervals (dashed lines) for catch-rate based on the SR model for the target fleet (left panel) and the by-catch fleet (right panel).

The realized discrepancy χ_{rel}^2 and the predictive discrepancy χ_{pred}^2 do not indicate problems of fit to the catch rates. 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 SR models, and 0.515, 0.523, and 0.507 for the bycatch fleet for these three models, respectively. The target fleet has higher KS p -values, but the distribution of p -values for the bycatch fleet is relatively flat (Figure 2). The overall p -value is 0.969, meaning that nearly 3% of the replicates reject the null hypothesis that the predicted and the observed catch-rate data are from the same distribution. This KS test also indicates that the model fits the target catch rate data better than the bycatch catch rate data. Figure 2 also indicates that the KS test is more sensitive than the overall chi-square test.

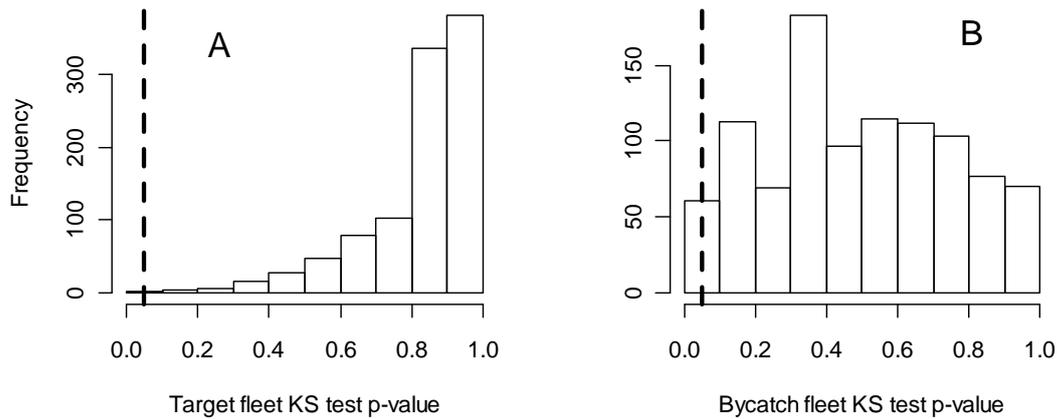


Figure 2. Distribution of Kolmogorov Smirnov (KS) test p values comparing posterior predictive catch rate from SR model and the observed CPUE. The vertical dashed line is where $p = 0.05$. A: target fleet, B: bycatch fleet.

The DIC for the Standard model is 56.85 deviance points greater than that for the SR model, and the DIC for the Annual model is 47.59 deviance points greater than that for the SR model. Model selection based on the mean square predicted loss also selects the SR model as “best”, and the Standard model as “worst”.

3.2. Quantities of management interest

Interestingly, the posterior distributions for the maximum sustainable yield MSY and biomass that supports MSY , B_{MSY} , do not differ substantially among models. The posterior medians for MSY (summed across stocks) are 1,927; 1,921; and 2,001 tonnes for the Standard, Annual, and SR models, respectively. They are slightly higher than the estimate of MSY from the weekly delay-difference model using maximum likelihood estimation ($M\hat{S}Y = 1,768t$). However, these medians fall within the 95% confidence intervals for MSY from the weekly delay difference model (1,517-2,043t). The 95% credibility intervals (e.g., 1,487-2,656 for the SR model) are wider than the 95% confidence interval based on the delay-difference model.

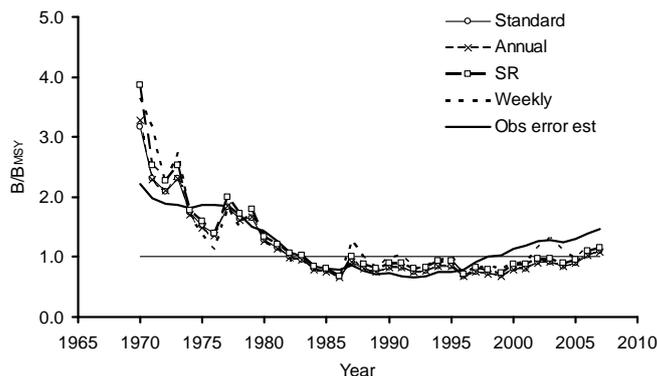


Figure 3. Posterior median time-trajectories for B/B_{MSY} for the three alternative biomass dynamics models (Standard, Annual, and SR) aggregated over stocks, the weekly delay-difference model (dashed), and a maximum likelihood observation-error estimator (solid).

The time-trajectories of biomass relative to B_{MSY} from the three biomass dynamics models are similar (Figure 3), although the SR model tends to produce a higher estimates of B/B_{MSY} than the other two models. These time-trajectories of B/B_{MSY} are also similar to that from the weekly delay-difference model even though the latter is substantially more complicated than a biomass dynamics model. In contrast, the results from a standard biomass dynamics model implemented as a maximum-likelihood observation-error estimator differ markedly from those of the Bayesian state-space models and the weekly delay-difference model (Figure 3).

The posterior distribution provides a convenient way to examine parameter uncertainty. Although not shown here, the coefficients of variation (cv) for K , B_{msy} , MSY , and growth rate r are fairly small (generally below 20% for each stock and when results are aggregated spatially over entire NPF area). Catchability q for the target fleet is also precise, with a cv of 18%, 15%, and 15% for the Standard, Annual and SR models respectively. However, the cv of q for the bycatch fleet is high: 52%, 55%, and 54% for the Standard, Annual and SR models respectively. The results also show that the extents of process and observation error are similar among the three models, but 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.

4. DISCUSSION AND CONCLUSION

A variety of formulations of the biomass dynamics model have been developed and examined. However, few studies have considered the suitability of these models for assessment of short-lived species. Our study demonstrates that biomass dynamics models are flexible for short-lived species when both process and observation error are taken into account. In particular, the alternative biomass dynamics model that assumes density-dependence is governed by current year biomass (the SR model) may be particularly effective for short-lived species. 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 will 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.

The estimates of the ratio of biomass to B_{MSY} from the biomass dynamics models are virtually identical to those from a more sophisticated weekly delay-difference model which requires more information on biological processes than a biomass dynamics model. In contrast, the estimates of this ratio from a standard observation-error estimator are markedly different, suggesting that allowing for process error permits the biomass dynamics models to capture 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 a hierarchical Bayesian techniques in this respect has been identified for several applications in the past (Su et al. 2001; Rivot and Prevost 2002; McAllister et al. 2004).

The similarity of results between the alternative Bayesian biomass dynamics models implemented and those of the weekly delay-different models provide some confidence that the biomass dynamics models outlined in this paper can be applied to species for which biological data such as growth, natural mortality, and recruitment are not available but only catch and effort information is available.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for their comments on an earlier version of this paper. This work is supported by the Australian Fisheries Research and Development Corporation.

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.
- Best, N., Cowles, M.K., and Vines, K. (1996). 'CODA convergence diagnosis and output analysis software for Gibbs sampling output.' (MRC Biostatistics Unit, Cambridge.)
- 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**, 333-350.
- Dichmont, C.M., Deng, A.R., Venables, W.N., Punt, A.E., Haddon, M., and Tattersall, K. (2005). A new approach to assessment in the NPF: spatial models in a management strategy environment that includes uncertainty. FRDC Report 2001/002.
- Dongen, S.V. (2006). Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology* **242**, 90-100.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515-534.
- Gelman, A., Meng, X., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**, 733-807.
- Ghosh, S.K., and Norris, J.L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioural effects. *Journal of Agriculture, Biological and Environmental Statistics* **10**, 35-49.
- 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.
- 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.
- 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.
- Punt, A.E. and Hilborn, R. (1996). 'Biomass dynamic models. User's manual.' FAO Computerized Information Series (Fisheries). (Rome: FAO.)
- Rivot, E., Prevost, E., (2002). Hierarchical Bayesian analysis of capture-mark-recapture data. *Canadian Journal of Fisheries and Aquatic Sciences* **59**, 1768-1784.
- Sheskin, D. (1997). 'Handbook of parametric and nonparametric statistical procedures.' (CRC Press, Boca Raton, Florida.)
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of Royal Statistical Society Series B* **64**, 583-616.
- 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.
- 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.