# Guyana/Suriname Seabob Stock Assessment: Method Summary

Paul A H Medley

paulahmedley@yahoo.co.uk

21<sup>st</sup> September 2019

## **1** Introduction

The following report provides a technical summary of the seabob stock assessments for Suriname and Guyana. It does not include results, but only describes the modelling approach and methodology.

The stock assessment is an integrated assessment that uses all available information in a single model to estimate the past stock dynamics and current status for seabob within the national boundaries of Suriname and Guyana. The assessment consists of a population model that describes the dynamics, an observation model that calculates what would be the expected observed values of total catch, catch and effort and size composition derived from the population model and a likelihood model that link the observation model to the data.

All assessment work, apart from the initial data preparation to produce raw data (total catch, catcheffort and average count-per-pound by commercial size category, and the sample counts by sex/size weight bin) have been carried out in RMarkdown scripts (notebooks). All analyses are therefore fully documented and reproducible. The specific work carried out in each case is described in these documents. All results are also presented in these, not in this document.

The original data for catch effort and random size/sex sampling is held in MS Access databases. Data have been extracted from the databases using SQL queries. Total catch data was obtained from the respective governments in spreadsheets.

The model is implemented in Stan (mc-stan.org). All important model calculations were defined as functions, which could be exposed in R and tested separately. This ensured model code was testable and correct.

#### 1.1 Summary of Changes from 2013 Stock Assessment

The stock assessment model has broadly remained the same as previously developed in ADMB (Medley 2013), with the following key changes:

- The stock assessment has been applied to both Guyana and Suriname fisheries with the same priors and model structure, but independent data. This allows direct comparison between the assessments and should be informative on differences between the fisheries.
- The stock assessment is fully implemented in a Bayesian MCMC using Stan (mc-stan.org).

- Growth is modelled using a size transition matrix, which avoids maintaining the population in age categories and the excessive smoothing that occurs in converting weight to age or *vice versa*.
- Selectivity is now modelled as a flexible spline function.
- Recruitment is modelled properly as a log-normal latent variable ("random effect").
- Commercial size grading is modelled directly as a process using multiple normal distributions.
- Catches have been corrected for the presence of white belly shrimp.
- The random size composition now incorporates a model of allocation between mature females, males and immature males/females.
- The likelihood for the random size composition now accounts for overdispersion by using a Dirichlet multinomial.
- The entire analysis from (almost) raw data to the final fit has been implemented in RMarkdown. This includes detailed documentation of the entire analysis. For example, the stepwise decision making in processing the data to a form suitable for the model is provided, with supporting graphs and equations in a single document. This also means that any changes made to the input data, decisions or analysis can be rapidly incorporated and the full analysis re-run relatively easily.

## 2 Data Preparation

#### 2.1 Total Catch

Total landings are reported to governments by each processor. Information reported has not always been consistent, but has improved over the years. There are initiatives to improve data reporting in Guyana so that it is more timely and accurate.

Monthly landings were available from all processors back to January 2002. Before this, monthly data were not consistently available, but annual landings were reported. Landings are reported as total tail weight in pounds or kilograms by commercial size categories. Annual landings are available as gross weight to the start of the fishery. Discards are assumed to be negligible. Landings of "sour" shrimp are reported in the landings, although they would not be used.

In contrast to 2013, all catches have been corrected for the presence of white belly shrimp (*Nematopalaemon schmitti*). The average quantity of white belly was estimated from the catch sampling and removed from the smallest size category, which includes the "broken" category. Although the percentage of white belly compared to total catch is small, the white belly formed a slightly more significant proportion of the smallest shrimp with which it would be combined.

## 2.2 Catch and Effort

Catch and effort data were obtained from spreadsheet forms used by Noble House Seafoods and Heiploeg Suriname to record landings and processing operations. The spreadsheets are used for internal monitoring of their business. Data were extracted from these forms and held in a database for further manipulation. Using the database, it was possible to match trip information (trip dates of departure and return) with processed landings weight, fuel used, and commercial size grades produced.

The landed catch is recorded as pounds or kilos of processed shrimp, representing about 43% of the live weight. Effort might be measured in several ways, but here days-at-sea are used as this is consistently measured.

Plots of <u>trip</u>effort against catch reveal an asymptotic relationship with landings (Figure 1). This has been found previously (Medley 2013) and examined using generalized linear models. Alternate measures were also examined in this case, and it was found number of trips was a better indicator of catch than days-at-sea. Therefore, a logistic model was fitted to the data to adjust the raw effort and linearize the relationship within trips between catch and effort (Figure 2). This resulted in a relationship between landings (catch) and effort which was acceptable (Figure 3).

The reasons why trip length has an asymptotic relationship with landings is not clear, but it is apparent that trip length does not measure trawling effort. Clearly within a trip time <u>is</u> being spent on other activities, such as travelling to the fishing ground. For more recent data, the trip activity should be cross referenced with the VMS and other information.



Figure 1 Box plot of landings (kg) plotted against the raw measure of effort: days-at-sea for Suriname, showing asymptotic relationship.



Figure 2 Landings (kg) plotted against the days-at-sea with the logistic standardisation model.



Figure 3 Landings (kg) plotted against the standardised measure of effort showing a linear relationship.

#### 2.3 Commercial Size Category

The processing facilities routinely collect average count data from the commercial categories. This should monitor the average size within each category. This information should be useful within the

stock assessment model to fit to changes in mean size within the category if such changes are significant. One processing facility provided average counts recorded by the quality control staff.

The change of shrimp size in the population and changes in selectivity will cause not only changes in the landings recorded as change in the amounts of each commercial categories, but may also change size within categories over time. A simple analysis of variance (Medley 2013) estimating the average count data dependent on the Year term as a factor suggested that Year has a significant effect on within-category size (Table 1). This would indicate that average count data should be included in some form in the stock assessment model.

Table 1 Analysis of variance	for average counts in comm	ercial category for a stand	dard log-linear model	(Medley 2013)
------------------------------	----------------------------	-----------------------------	-----------------------	---------------

	Residual degrees of freedom	Residual Deviance	Degrees of freedom	Deviance	Pr(>Chi)
AvgCount ~ Category	105404	99057335			
$AvgCount \sim Category + Year$	105394	97897240	10	1160095	< 2.2e-16
AvgCount ~ Category * Year	105312	90699966	82	7197274	< 2.2e-16



Commercial Category

Figure 4 Box and whisker plot for average count data in commercial size categories, showing the median, 25<sup>th</sup> and 75<sup>th</sup> percentiles and 1.5 times the interquartile range for the average counts in each category. The solid lines represent the minimum and maximum count for each category (based on the name) and the dotted line is the mid-point.



## Figure 5 The same data as Figure 4 plotted as peeled tail weight. These data would be used in the stock assessment model.

In contrast to the 2013 stock assessment, a grading model was used to allocate catch size composition among grades. This avoided the need to allocate specific tail weights to commercial grades. The average count was still useful to help define parameters in the grading model.

The expected catch within each commercial size category is calculated through a sequential selection process that models as far as possible the actual grading applied in the processing facilities (Figure 6). The grading parameters are based on the target size category and adjusted to some extent to align with the average counts. The parameters (mean and standard deviation for the weight selectivity normal curve) cannot be fitted freely as they are aliased with other parameters.



Figure 6 Estimated grading selectivity applied in processing facilities. The numbers in each grade name refer to the count range.

## 2.4 Sex and Maturity

A significant improvement in model fit compared to 2013 assessment was achieved by modelling not only maturity, but sex allocation as applied in the random sampling. The basis for this was that immature males are easily confused with immature females as external parts may not be present (e.g. lost through damage or not yet formed through a moult). The maturity model is therefore included in the stock assessment model and data are provided in three categories: immature males/females, mature females and males.

There is now a considerable data set linking female size (tail weight in grams) to maturity (presence of a "green vein") in females. This allows the maturity ogive to be estimated, which can be used to estimate spawning stock biomass within the stock assessment model. For the 2013 assessment, this was done externally to the stock assessment model. It is now done within it because it was found that the size composition data could be best explained if some (estimated) proportion of immature females were actually males. In this context, maturity also increased the correct sex identification so need to be estimated within the model.

When calculating the SSB, the maturity ogive based on the "green vein" observations were found to result in a very small proportion of females reaching maturity. This was considered probably unrealistic and reflected the method used to identify maturity. Alternative studies based on detailed observations on reproductive organs (Table 2) indicated much smaller sizes at 50% maturity compared the "green vein" data (Figure 7). Castilho et al. (2015) was chosen as a better estimate of 50% maturity based on a 3 year study and being in the middle range of the various studies. This was used to calculate spawning stock biomass for stock-recruitment, reference points and to determine relative depletion. It was not used for other purposes. The decision to use these published estimates was taken by the review meeting in 2019.

CL 50% mm	Reference
13.2	Almeida et al., 2012
24.0	Campos et al., 2009
15.5	Castilho et al, 2015

Table 2 Maturity ogive logistic parameter based on other studies on seabob.



Figure 7 Alternative maturity models based on carapace length (mm). Willems and RS were based on "green vein" observations in Suriname and Guyana. Campos, Castillo and Almeida were published maturity studies from elsewhere on seabob (Table 2).

#### 2.5 Tail Weight: Random Samples

The random samples needed to be converted from unpeeled tail weight to processed tail weight to be used in the assessment. The tail weights were multiplied by 0.78 to adjust for peeling based on morphometric data collected in 2007 (CRFM 2009, Table 5 p.115). Unpeeled tails are measured on electronic scales to within 0.01 of a gram. Within the database, these are held as whole numbers (integers) and compiled into 0.2g class frequencies. These data are provided in the three sex and maturity categories described above.

However, for Suriname in 2015 the accurate scales failed and scales measuring to the nearest gram were used instead. This degraded the precision of the measurements considerably. These data are still used, but treated differently in the model.

## **3** Population Model

#### 3.1 Overview

The model used in this assessment was a statistical integrated model, implemented with the Stan software (<u>mc-stan.org</u>) in R. In essence, a statistical catch-at-age model simulates population dynamics in time including biological and fishing processes and fits these to all available observations.

The model is based on a standard forward-projection design, but applies a transition matrix for growth. Therefore, rather than modelling age explicitly, animals are modelled based on their size. Standard models (e.g. Stock Synthesis III) was not used so that the assessment was able to use the growth model with the weight data and grading information correctly. Age data were not available. Where possible the observations and model are kept distinct. The model is adjusted to fit a sufficient data set. In some cases, exact fits can be obtained because there are enough parameters to allow the model to closely follow the data. This applies to the total catch. For other data, where observations are presumed to be significant, the model may not fit the observations closely,





#### 3.2 Monthly Catches

and some error is acceptable.

The basic population model time step is one month, which was considered appropriate for this species. Separate models are run for males and females. When monthly catch and size composition data are available, a simple approach can be used to model the population, clearly separating the model and data. For each sex and weight, the numbers at the beginning of each month are calculated based on mortality parameters and standard negative exponential model:

$$N_{w t+1} = N_{wt} e^{-M - F_t S_w}$$

where  $F_t$  = fishing mortality in month t and  $S_w$  = selectivity for weight w. Catches in each weight bin were calculated using the standard catch equation. Multiplying the numbers by the mid-point weight bin provides the catch weight.

Growth between weights was estimated in a transition matrix based on the von Bertalanffy growth function (see Equation 1 below) for the mean increment, and another parameter for the variation around the mean.

Selectivity was modelled primarily as a cubic spline function based on tail weight, which allowed selectivity to be very flexible. Because of this flexibility and for simplicity, selectivity was linked directly to weight rather than length. Selectivity can also be modelled as a exponential logistic curve, however the spline curve did not necessarily indicate that this would be a good choice. The cubic spline function was transformed through a logistic function to ensure it was mapped to values between 0.0 and 1.0.

From 1997-2002, only catches are available. The model is run during this period to obtain a level of depletion for the start of the main data series using catches allocated on a proportional basis among the weight bins.

## 3.3 Growth

Most age-related data are available as tail weight. This includes both commercial size category data and scientific sampling (see section 2). The mean growth of seabob is assumed to follow the von Bertalanffy (vB) growth curve. The general from of the growth model is:

1)

$$L_t = L_{\infty} (1 \quad e^{-Kt})$$

Parameter  $L_{\infty}$  is estimated within the stock assessment for each sex. The growth transition matrix is based on the normal distribution, with mean growth defined by the vB equation:

$$\mu_{i} = m_{i} + (L_{\infty} \quad m_{i})(1 \quad e^{-K})$$
 2)

and transition probabilities from weight bin *i* to *j*:

$$G_{ij} = N(\mathbf{h}_j \mid \boldsymbol{\mu}_i, \sigma) \quad N(\mathbf{l}_j \mid \boldsymbol{\mu}_i, \sigma) \quad \mathbf{i} < \mathbf{j}$$

$$G_{\mathbf{i}\mathbf{i}} = N(\mathbf{i} \mid \boldsymbol{\mu}_\mathbf{i}, \sigma) \quad \mathbf{3}$$

$$G_{\mathbf{i}\mathbf{j}} = 0 \quad \mathbf{i} > \mathbf{j}$$

Where N()=cumulative normal distribution,  $m_i$  = mid-point length for weight bin i, and  $h_i$  is the upper length bound and  $l_i$  is the lower bound for the bin. Notice that  $\sigma$  contributes to the growth rate and negative mean growth increments are allowed, but negative does not occur, the shrimp are more likely to remain in the current bin. Lengths for each weight bin were calculated using the lengthweight conversion.

The length-weight conversion was not directly used in this stock assessment due to the use of transition matrices which allowed greater flexibility in the growth. The length-weight relationship was estimated using a log-linear model from the morphometric data collected in 2007/8 (CRFM 2009).

### 3.4 Sex and Maturity

Unfortunately, relatively little is known about the biology of seabob, so interpretation of the observations is difficult.

The primary problem with the previous 2013 assessment model was the under estimation of females in the smallest size category (Figure 9). This indicated a poor fit with the standard selectivity which was not resolved.

For the 2019 model, residual patterns were resolved to some extent using sex-specific cubic spline selectivity curves. However, there was no justification for the resulting selectivity curves shapes as they would imply separation of sexes by location or other factor. Furthermore, selectivity explained observations using nuisance parameters with no meaning, making the data uninformative on parameters of interest. For example, selectivity was confounded with sex differences in growth, which is the most likely explanation for observed size differences.



Figure 9 Previous random sampling model female (top) and male (bottom) plot of residuals by weight class.

While it has not been confirmed, the most likely explanation is misidentification of immature males in the smallest sizes. Males are only identified by positive identification of the petasma, which may not be present or identifiable in immature shrimp. This would mean that the numbers of immature females would be over-observed in the data and would help explain the patterns in residuals. The interpretation of the data would therefore be positively identified mature females and males are correct observations, whereas immature females could be either female or male, with unknown proportion allocated to each. This is supported from the observations. Of the total number of 1094946 shrimp observed, 59% were identified as female, 35% male and 6% unknown. The unknown proportion is small and probably underestimates misidentification, but does indicate uncertainty in sex designation particularly for smaller shrimp. The 6% unknown sex have been added to the immature group and another "maturity" curve included that allocates males in the population model to the immature females. Hence the model fits the population males and females to each of three groups: mature females, positively identified (probably mature) males and immature shrimp. Immature shrimp are all shrimp not positive identified as males or mature females.

Mature females are estimated with a standard logistic curve:

$$p_{fi} = \frac{r_1}{1 + e^{-f_5(w_i - f_{50})}}$$

Positively identified males include a logistic representing maturity as well as a fixed proportion (m<sub>1</sub>) applying across all sizes:

4)

$$p_{\rm mi} = m_1 + \frac{(1-m_1)}{1+e^{-m_{\rm s}(w_i-m_{\rm 50})}}$$
5)

The third category, immature "females" are all other animals not allocated to these groups (i.e. the complements to these functions: see Figure 10).



Figure 10 Ilustration of the logistic models allocating shrimp male and female shrimp to the three categories.

#### 3.5 Recruitment

The stock recruitment model used was the Beverton-Holt model:

$$R = \frac{R_0 S}{R_b + S} \tag{6}$$

where R=Expected recruitment, S=spawning stock biomass from the previous month or earlier depending on the length of the larval stage,  $R_0$  = the maximum recruitment when the spawning stock is very large (i.e. unexploited) and  $R_b$  the spawning stock biomass when recruitment is 50% of the maximum value.

For comparative purposes the "steepness" parameter was calculated separately:

$$R = \frac{4hR_0S}{R_0S_R(1-h)+S(5h-1)}$$
7)

Where  $S_0$  =spawning biomass per recruit when the stock is unexploited and h =steepness parameter (0.2< h <1.0).

The recruitment was modelled as a log-normal, with equation 6 the log-normal mean, and individual deviations fitted as random effect parameters from 2002-2017. Before 1997-2002, when only catches are available (model burn-in), no deviations from equation 6 are fitted.

### 3.6 Likelihood

#### 3.6.1 Overview

The log-likelihood was calculated for each data component based on the Dirichlet-multinomial or normal log-likelihoods.

- The negative log-likelihood for the size composition by size and sex in the random samples is calculated from the Dirichlet multinomial. This is the equivalent to beta binomial, but for multiple category data. For this likelihood, there is an additional dispersion parameter as well as standard multinomial parameters.
- The likelihood for the total catch and catch- effort data were based on the normal. Assuming a Poisson probability function for the catch, the scale parameter was assumed to be the square root of the predicted catch multiplied by an estimated constant σ, so the negative log-likelihood would be:

$$LL = \sum_{k} \frac{(o_k - e_k)^2}{\sigma e_k} + \ln(\sigma e_k)$$
<sup>8)</sup>

Where  $o_k$  = observed catch,  $e_k$  = predicted catch for a particular month k. The predicted total catch weight is predicted from the model fishing mortality and selectivity, with catches summed over all sizes and sex. The predicted catch weight for a given level of effort is derived from the estimated total catch and fishing mortality:

9)

$$e_k = qfC/F$$

Where q = catchability parameter, f = observed effort, C = estimated total catch (in numbers) and F = estimated fishing mortality in each month.

- The catches and catch and effort within commercial size categories is based on integrating over possible catch allocations among categories. This was necessary because commercial categories overlap and are incomplete. The details are given in section 3.6.3 below.
- The average count per pound data was assumed to follow a Poisson and therefore the loglikelihood Equation 8 was used. In this case, the predicted count was calculated from the predicted size composition using the size grade selectivity. Most categories had a standard deviation for the observed counts taken in each month which was used as the count standard deviation.
- A recruitment deviations were modelled directly as a log-normal random effect, which included estimating the recruitment scale parameter (σ<sub>R</sub>).

In Stan, the normal log-likelihood is provided as a function, which was used to avoid coding errors and having to deal with numerical exceptions. The Dirichlet-multinomial likelihood was not provided, so a function was written to calculate the log-likelihood in this case. In addition, another bespoke function was used to calculate the commercial size category likelihoods.

#### 3.6.2 Random Sampling

The log-likelihood for the random sampling data was a Dirichlet-multinomial, which includes an additional parameter to account for overdispersion in the multinomial distribution. The distribution includes a conjugate Dirichlet prior, which is the origin for the overdispersion parameter, and makes the mathematics more tractable.

The random sampling data are defined in 3 categories (mature females, identified males and "immature females") of 30 weight bins each. The population and observation models calculate the proportion expected in each bin category, so this can be compared with the observed proportions.

The relative weight of the multinomials compared to other sources of information is controlled by the dispersion parameter.

It should be noted that there are limitations to this approach. Further analyses of these data could yield an improved likelihood accounting for within trip correlations for example. Therefore, although it is more flexible than assuming a strict multinomial, further improvements in the interpretation of the random size composition data is still likely to be possible.

#### 3.6.3 Log-Likelihood for Commericial Size Composition

The log-likelihood used for categories is the same as that usewd for the previous 2013 stock assessment. An explanation is presented here. However, with the new approach to modelling the grades, the number of overlapping categories was greatly reduced. As noted below, where a single non-overlapping category within a month was identified, the catch relative to the total catch within that category used a normal distribution as an approximation to the Poisson distribution.

For each size composition, it is possible to estimate the number of seabob within it. This is the sum of seabob over the size category from smallest to largest. For example, the 90-110 count per pound size category would contain sizes varying from 5.04 (1000/(2.20462\*90)) down to 4.12 (1000/(2.20462\*110)) grams weight. The expected number of seabob in each size category can be obtained from the population model based on the fishing mortality for each size category, population abundance in each category and the fishing selectivity.

All commercial categories can be defined as a subset of a larger category which contains it. In the simplest case, the category contains all the catch, so no larger category is required. Therefore in this case, the "larger" category and the category are the same and the log-likelihood is simply based on the expected landings directly from the model. In this simple case, the likelihood for numbers in a particular size category would be Poisson:

$$L = \frac{\mu_a^{A}}{x_a!} e^{-\mu_a}$$
 10)

v

where  $x_a$ =numbers observed in the category A and  $\mu_a$ =expected numbers in category A. The expected numbers can be calculated by simply summing over all sizes from the model within the category. Suitable alternatives to the Poisson can be used to account for over-dispersion and/or to simplify the calculations. Taking advantage of the situation where  $x_a$  and  $\mu_a$  are very large, as in this case, the normal likelihood or log-normal could be used.

Unfortunately, this likelihood cannot be applied in this simple form unless the data are manipulated to allocate all catches to non-overlapping well-defined categories. This should be avoided if possible, since the model would not be fitted to raw data and such manipulations can introduce unknown bias in the result. Instead, it was considered preferable to develop a likelihood which captures what size information there is in the data rather than impose such information by manipulating the data.

In all cases, a significant proportion of the catch will be undifferentiated by size. Any catch allocated to a particular size range can therefore always be defined within the context of a larger size category which is complete. The known catch in the smaller category represents a minimum catch within this range, where other catches within the larger category might also be in the smaller one. The likelihood becomes the sum of likelihoods across possible allocations of catch between the two categories.

To illustrate the basic calculation, we consider categories A and B covering separate size categories (Figure 11) for each of which the statistical model can estimate the expected catch in a particular month. The category B may envelop A (B<sub>1</sub> and B<sub>u</sub>) or extend it only (B<sub>1</sub> or B<sub>u</sub>), but it should always be possible to calculate the expected catch for both A and B.



The data however is only available partially for A and B, and otherwise the total catch is made up in C, where C landings have been ungraded among A and B. We need to sum the likelihood over possible allocations of landings in C between A and B, so the likelihood for the joint Poisson likelihood becomes:

$$L = \sum_{x_a = x_a}^{X_c + X_a} \frac{\mu_a^{x_a} \quad \mu_b^{(X_c + X_a + X_b - x_a)}}{x_a! (X_c + X_a + X_b - x_a)!} e^{-\mu_a - \mu_b}$$
11)

where  $X_a$ ,  $X_b$  and  $X_c$  are the observed landings in A, B and C (unallocated A+B), and  $\mu_a$  and  $\mu_b$  are the expected catches in A and B which are estimated from the model.

This can be simplified to some extent by reformulating to create a Poisson term for the total catch in A+B and a sum of binomial terms for the proportion in category A:

$$L = e^{-\mu_a - \mu_b} \frac{(\mu_a + \mu_b)^{(X_c + X_a + X_b)}}{(X_c + X_a + X_b)!} \sum_{x_a = X_a}^{X_c + X_a} \frac{(X_c + X_a + X_b)!}{x_a! (X_c + X_a + X_b - x_a)!} \left(\frac{\mu_a}{(\mu_a + \mu_b)}\right)^{x_a} \left(\frac{\mu_b}{(\mu_a + \mu_b)}\right)^{X_c + X_a + X_b - x_a}$$
12)

Similarly, the likelihood for several categories within a larger category can be described using a multinomial.

For large catches it is not possible to sum over possible catches and Eq. 12 can only be simplified by closely approximating the binomial with a normal probability. The binomial term then becomes:

$$B\left(p = \frac{\mu_{a}}{(\mu_{a} + \mu_{b})}, n = (X_{c} + X_{a} + X_{b})\right) \approx N\left(\frac{(X_{c} + X_{a} + X_{b})\mu_{a}}{(\mu_{a} + \mu_{b})}, \sqrt{\frac{(X_{c} + X_{a} + X_{b})\mu_{a}\mu_{b}}{(\mu_{a} + \mu_{b})^{2}}}\right)$$
13)

Similarly, the total catch likelihood can be approximated with a normal density:

$$L \approx N\left((X_{c}+X_{a}+X_{b}); (\mu_{a}+\mu_{b}), \sqrt{(\mu_{a}+\mu_{b})}\right) \int_{x_{a}=X_{a}}^{(X_{c}+X_{a})} N\left(\frac{(X_{c}+X_{a}+X_{b})\mu_{a}}{(\mu_{a}+\mu_{b})}, \sqrt{\frac{(X_{c}+X_{a}+X_{b})\mu_{a}\mu_{b}}{(\mu_{a}+\mu_{b})^{2}}}\right) dx_{a}$$
 14)

The cumulative normal can be well approximated numerically (West, 2004), so the likelihood can be calculated reasonably easily for each datum.

The binomial part of the likelihood is only informative on landings below those expected in category A. As the expected landings fall below the observed landings in category A, the log-likelihood declines. Clearly, as higher landings have been observed than those estimated, the estimated landings become less likely. Conversely, since any ungraded landings in C could be allocated to A, there is no information on higher estimated landings in A as all are equally possible. Therefore, the

log-likelihood asymptotically approaches 1.0 as the expected catch increases. As the expected landings exceed the total landings observed (A+C), then the likelihood begins to decline again (Figure 12). In this case, the estimated landings in A exceed the possible observed landings in A (A+C), and the estimate becomes less likely. The result is a flat-topped likelihood, where the flat top covers the likely range of the landings within the category. Because the likelihood will include a term for the total catch as well, the likelihood should have a mode for the full model, but additional information is likely to be required to be able to estimate parameters defining stock size composition.



Figure 12 Example conditional log-likelihood of expected number in a category (parameter  $\mu_a$ ) where the observed catch in categories A and C are 2000 and 500, so 500 may or may not belong to category A.

The process of analysing size category data was somewhat simplified by using the size grade selectivity model, which explicitly accounts for error in shrimp grade allocation, as well as external analysis of the grading data. The main overlaps between categories apply to the smaller size grades, which may also relate to some extent as labelling for quality control rather than related to shrimp size. Therefore, as much simplification was carried out of the categories as possible and as a result only three clearly overlapping categories were identified in all cases.

#### 3.7 Priors

The choice of priors for the various parameters generally followed the Stan advice (see <a href="https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations">https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations</a>). The following table outlines the choices made for the priors. It is important to note that informative priors are important in achieving a tractable MCMC fit, but their influence has been kept to a minimum.

	Parameter	Prior	
stc	Total catch scale	Fixed = 1.0	For the total catches, the scale was fixed at 1.0, which was the default for the Poisson, effectively fitting the total catch exactly.
sca	Size grade catch scale	Cauchy(0, 10)	For the normal scaling parameters a half-Cauchy distribution was used. This is a weak prior that slightly discourages higher scaling
sce	Catch effort	Cauchy(0, 10)	values.
dmbeta	Dirichlet- multinomial dispersion parameter	G(0.2*WN, 0.1)	A weak gamma distribution was used, with parameters scaled to the number of weight bins (WN).
FLinf	Female asymptotic growth	N(31.41, 5)	Weak normal for carapace length (mm) based on largest commercial size grade category
dMLinf	Male-Female asymptotic growth Difference	N(-1.99, 2)	Weak normal "empirical Bayes" for carapace length based on the mean difference overall between males and females.
К	Growth rate	Fixed = 0.2	Confounded with Gsig below, but evaluations suggested 0.2 /month appropriate (Ribeiro De Campos et al. 2011, Soomai et al. 2012, review report).
Gsig	Growth sigma	N(2.0, 1)	Weak normal allowing a reasonable carapace length growth increments in a month.
M50	50% maturity	N(4.32, 0.5)	Applies to males and females, normal prior based on "empirical
Msp	Steepness maturity	N(1.58, 1.0)	Bayes" ML estimate of female maturity estimated externally from all random sampling data combined.
рММ	Proportion male	beta(9.5, 0.5)	Weak prior for 95% male identification rate.
Sy	Cubic spline selectivity	N(0, 5)	Weak normal prior on cubic spline parameters to prevent arbitrary small or large values.
lq_t	Log- catchability	N(12.5, 1.0)	Normal prior for log-q relative to $R_0$ , primarily to help convergence in MCMC at reasonable values.
IRO	BH log unexploited recruitment	N(19, 3)	Weak normal prior based on sufficient recruitment to support fishery catch, but prevents arbitrarily high biomass.
IRb0	BH log 50% recruitment	Fixed: h=0.8	Early attempts to estimate were successful, but fixed based on steepness=0.8 to aid MCMC convergence.
IRs	Recruitment log-normal sigma	N(0.5, 0.5)	Weak normal prior covering likely range of reasonable recruitment variance.
lRac	Recruitment autocorrelation	N(0, 0.5)	Weak normal prior on recruitment autocorrelation.
М	Natural Mortality	Fixed = 0.2	Hernández et al. 2003 (cited in Soomai et al. 2012) suggested 0.183 /month based on growth.
IFO	Log mortality for initial depletion level	N(-1.61, 1.0)	Weak normal prior on initial fishing mortality to account for catches before time series start.
leFd	Log monthly fishing mortality difference	N(0, 0.5)	Prior on fishing mortality change from month to month – acts as a penalty on very large changes. This was necessary to ensure MCMC convergence. It presupposes that large changes in fishing effort from month to month are unlikely due to limits in fishing capacity. Closures with zero catch are excluded from fitting.
CatMW	Grade mid- weights	Fixed	Based on mid-point weights for the count range, but adjusted slightly downward to account for bias.
CatSD	Grade standard deviations	Fixed	Set at 16% of the grade widths. Because the grading model is sequential, the grading SD should be small.

## 4 Model Configuration

The basic structure of the model was designed to minimise problems with aliasing and other issues preventing the MCMC algorithm exploring the full posterior. Three simple linear transforms of the parameters were found to help. These consisted of scaling the log-q (catchability) parameter relative to the recruitment (R<sub>0</sub>), defining male maximum average size relative to females and defining fishing mortality as the difference between succeeding values scaled on observed total catch (the first value being a free point estimate). These changes do not affect the underlying model and are only there to support the MCMC algorithm convergence.

It is worth noting an important change in estimating fishing mortality. It was difficult for the MCMC to converge on fishing mortality estimates and, because they changed so much year to year with large changes in catches, a single simple prior was unable to help encourage convergence. To resolve this, the log fishing mortality parameter estimates were adjusted based on the change in catches year to year. The fitted parameters are the difference between sequential log-fishing mortalities  $(ldF_t)$ . The first log fishing mortality  $(lF_1)$  in the series needs to be fitted as a free parameter and thereafter the parameter is fitted condition on the log catches:

$$\begin{split} & lF_t \approx lC_t \quad lB_t \\ lF_t \quad lF_{t-1} \approx lC_t \quad lB_t \quad lC_{t-1} + lB_{t-1} \\ ldF_t = lF_t \quad lF_{t-1} \quad lC_t + lC_{t-1} \sim N(0,\sigma) \end{split}$$

If the log biomass does not have large changes month to month, a normal PDF around zero provides a good prior for these parameters. This is an example of a non-centred parameterization type solution.

The initial base case was determined from exploring various configurations for the models using the maximum posterior density estimates (MPD). MCMC simulations were only carried out for selected cases to estimate uncertainty. It should be noted, however, that with random effects the MPD is not necessarily reliable so all management advice and HCR evaluation should be based on MCMC results. The MCMC however takes a long time to run, so decisions on model configuration need to be based on the MPD results.

For consistency, the models and priors were kept the same between Suriname and Guyana. No evidence was found to suggest a different model configuration or structure. Keeping the models and configurations the same means that the results are comparible.

There is no known reason why the sex ratio of recruits will be other than 50:50. The sex ratio in the random sampling favours females, but it is not clear that sexes have been correctly identified in immature shrimp. Accounting for males within the "immature females" category based on expected maturity levels produced much improved residuals, suggesting that this is the most likely cause of problems fitting to the random sampling data.

The size grade category average count standard deviation was not fitted, although the average count itself was. In practice it is not clear that standard deviation is useful in this way. In theory, the standard deviation would measure the size variation within the size category, which contains small amount of information on landings size composition. However, given that grading may be adjusted in response to these counts, any useful information may be lost. Including the standard deviations in the model did not appear to make any difference to the model fit, suggesting there was no

information on parameters of interest, so the standard deviations were not fitted. They were used instead as determining the precision of the average count data.

Cubic spline selectivity was fitted, with the same selectivity for both sexes. Sex differences in random samples would be used to inform on difference in growth. The selectivity function is logit-transformed to map values from -/+infinity 0-1. The cubic spline function was used to avoid bias effects from the choice of a parametric selectivity function. The cubic spline "knots" were chosen based on initial exploration to allow smooth curves in a variety of forms that explained the data in both fisheries. The same knots were used, but the parameters to estimate the selectivity at each knot were fitted independently.

A break in catchability was introduced for both Suriname and Guyana to improve fit to the catcheffort data. This removed a pattern in the residuals suggesting that catchability had fallen in January 2012 in Guyana. This has not been fully justified and may need to be examined. The inclusion of the break in catchability does not affect the final results of the stock assessments.

The parameter profiling using the optimizer did not work well. Profiles over the MCMC also did not work. The posterior mode is most likely not well defined and therefore the optimizer is unable to adjust after parameters are fixed. The profiles do demonstrate the importance of the higher precision data as this appears to be influential on important parameters. The most effective method would be to use MCMC to do the profiling, but this would clearly take too long as each MCMC simulation takes a few days to run. Improved ways to profile parameters needs to found.

## 5 Checks and Diagnostics

The model code was checked by comparing exposed Stan function results against equivalent R functions.

The model was run with simulated data to check its theoretical performance. As well as a check that the model fits appropriately, the outputs illustrate the "best possible" residual plots for comparison.

Plots of observed-expected values, standardised residuals-expected values, standardised residualstime, standardised residuals-month (seasonality) and, where appropriate, standardised residualssize category, were produced for each data source.

Key parameters were profiled based on the data source. This was done by fixing the parameter being profiled around the MPD and fitting remaining parameters at different levels. This is indicative, but not strictly speaking a marginal PDF.

A retrospective analysis was conducted by sequentially removing the last month's data to detect any bias in final stock status estimates.

All checks and diagnostics are contained within RMarkdown and can be re-run easily. The MCMC themselves are slow to converge and would take some time to complete.

## 6 Further Work

The following tasks were identified as requiring attention during the review workshop:

1. The stock status compared to unexploited status appears very low. It would be worth considering whether this is realistic. This was resolved by noting that the maturity based on

recording eggs in females was shifted well to the right compared to studies based on observations on reproductive organs themselves, which should be much more accurate. Published estimates for 50% maturity were used in this assessment, but further work on the local maturity is recommended.

- 2. Check whether males make up a significant proportion of the "immature females" category, and, if possible, estimate this proportion. It may also be possible to review and improve the sampling protocol.
- 3. Check whether the break in catchability in January 2012 was justified. This has not been yet been justified based on any external information and may need to be examined. Candidates for an effect include sargassum episodes and changes in the inner line that may affect catches. It would be preferable to include explicit effects of this sort than arbitrary changes. The inclusion of the current break in catchability does not affect the final results of the stock assessments.
- 4. Analyses of the size composition data with respect to their error structure could improve the estimation of the underlying size composition of the catches, accounting for, among other things, the sizes of the trip catches and correlation within trips.
- 5. The reported trip activity (days-at-sea) for more recent data should be cross referenced with the VMS data to improve effort measure. This would be particularly useful in helping to understand why trip length is so asymptotic with respect to catches.
- 6. There is a significant difference in catch rates between Suriname and Guyana. This is not well understood, and current treatment is precautionary rather than scientifically justified. Although catch rates are much higher in Suriname, the model does not estimate that the stock is more abundant or more productive. There are at least two testable hypotheses which would bear further investigation:
  - The Suriname stock is smaller than Guyana but also more densely distributed. Higher shrimp density would raise the catch rates. This is the current model hypothesis. This could be tested by looking at the area trawled footprint. If this hypothesis is true, the Suriname trawl area should be significantly lower even taking into account the smaller fleet size.
  - 2. There is significant density dependent effects on the shrimp adult population. This is not addressed in the model. If density dependent growth was present, then growth parameters in Suriname should be lower. Although this has not been carefully examined, growth rates (K) may be a little lower in Suriname, but maximum lengths appear to be similar, and overall support for this is not strong. There is no assessment natural mortality, but if there was higher natural mortality through density dependent effects in the Suriname stock, this might explain the model fit and the overall mortality rate estimated for this stock. This could be explored further with alternative model runs, but estimating the density dependent effect without raising fishing mortality above the current F<sub>MSY</sub> would be difficult.