CENTER FOR INDEPENDENT EXPERTS (CIE) INDEPENDENT PEER REVIEW REPORT

SEDAR 35 STOCK ASSESSMENT REPORT: U.S. CARIBBEAN RED HIND

Natalie Anne Dowling, Ph.D.

OCTOBER 2014

CONTENTS:

EXECUTIVE SUMMARY

This report comprises an impartial and independent peer review of the SEDAR 35 stock assessment undertaken for Caribbean Red Hind, a grouper species. The assessment focuses on three locations: Puerto Rico, and St Thomas/St John and St Croix in the U.S. Virgin Islands.

The Assessment Workshop Report is the third of three sections of the SEDAR 35 report. The other two sections comprise an Introduction and the Data Workshop Report. There was minimal cross-referencing or linkage between the report sections. The management history of the fishery provided in the Introduction is barely mentioned in the subsequent sections.

There is no mention of previous stock assessments having been undertaken for this species, whether in these or other regions, and as such, no notion of past precedence.

There was no attempt to define the assumed stock structure for purposes of the assessment. At a minimum, a consideration of the area covered by the analyses, versus the potential of the species for movement and mixing (per RD09), would have been useful. The lack of specification of stock structure assumptions also brings into question the reliability of spawner-per-recruit based reference points.

There is a range of historical fishery dependent and independent data available for the fishery, including several time series of catch rates, for which two CPUE standardisations were attempted in order to obtain proxy abundance indices. These were considered to be inappropriate for use in a formal stock assessment, and were considered only briefly and in a qualitative sense in the Discussion section of the report. Most of the other data was short-term or temporally patchy, and/or associated with low sample sizes.

There was little attempt to compare or reconcile data from different gear/fleets, different regions and sectors (i.e. commercial, recreational, survey), nor to consider time series in the context of the management history (i.e. significant changes in management that may have affected observed trends).

- It would have been highly useful to have seen the extent to which trends were consistent across these categories.
- At a minimum, it would have been useful to have seen time series of mean length presented for all available length information, as this is what was used in the assessment. Ideally, such comparisons should have been made for the available catch, CPUE and survey abundance information, also. As it stands, it is very difficult to grasp the overall picture: Are things different in a different area and/or under a different gear/survey protocol? Or a different sector (commercial/recreational/independent survey)? Or for a different data set in the same area?

The Caribbean Red Hind Assessment Workshop focused on using length frequency data obtained from a port-sampling data collection program, termed the Trip Interview Program, to obtain estimates of total mortality via a mean-length estimator. A simple mean-length estimator was fitted to the most temporally consistent source of length data to obtain estimates of total mortality. Fishing mortalities were obtained by subtracting estimated natural mortality, and compared to overfishing reference points derived from standard spawner-biomass-per-recruit estimates. There were two Fmsy proxy reference points, and therefore six separate probabilities of overfishing (two for each of three regions) were calculated.

These methodologies appear sound and consistent with the best available assessment approaches based on length data. The mean-length estimator is a current acknowledged approach for estimating total mortality from length distribution data. It does not assume equilibrium population dynamics. It appears to have been properly configured and used consistent with standard practices (e.g. Edwards et al. 2012; Erisman et al. 2014).

The following should be noted:

- The assumptions of constant recruitment and mortality associated with the mean-length estimator, are not given consideration. The history of the fishery should be acknowledged in this context.
- Knife-edged vulnerability was assumed for the per-recruit analyses, but not considered in terms of the data.
- The definition of size-at-recruitment to the fishery, Lc, was vague.
- Very little detail was provided on the derivation of natural mortality estimates.
- Sources of uncertainty in the assessment methods were not investigated quantitatively. Temporal inconsistencies in terms of sampling size were mentioned throughout the report, together with their possible impact on length distributions, but there was no direct suggestion as to how to reconcile these.
- It is mentioned that the size of fish in the USVI may be market driven for plate size, suggesting that the selectivity may be dome shaped, which is in violation of the assumption of knife-edged selectivity in the mean-length estimator. Expanding the mean-length estimator to accommodate other selectivity patterns should be undertaken.

A sensitivity analysis was undertaken by fitting mortalities under a range of combinations of life history parameters and length of first capture. Beyond the suite of sensitivity scenarios, and providing resultant probabilities of overfishing, the report did not critically review or evaluate its approach, or consider its findings in a managerial context (as per Assessment Workshop TOR 7).

The technical conclusions are three sets of probabilities of overfishing. These have been presented with no implications of uncertainty, or (with the exception of recommending studies on basic life history) recommendations regarding narrowing the range of mortality estimates and so increasing the certainty around the probability of overfishing. Specifically, the implications of:

- the breadth of the ranges of Z (and hence the uncertainty in stock status) in terms of the efficacy of the overfishing probabilities in a management context
- having separate sets of probabilities for each of the three regions, without specifying what is assumed about stock structure

- the equal consideration/contribution of all sensitivity runs in calculating the probabilities of overfishing
- the lack of recommendation regarding the preferred choice of overfishing reference point

are not explicitly considered.

Against the current assessment approach, the outcomes for each gear/fleet should be reconciled within regions (i.e. what is the extent of overlap of the fishable sizes targeted/captured by the different gear/fleet types? To what extent can the data from each be combined?), so that the probabilities of overfishing are useful in a management context. Currently each gear/fleet is considered to contribute equally to the frequency distributions used to determine the probabilities. Ideally, there should be a better quantitative articulation of the extent to which the information from each gear/fleet contribute to the overall frequencies on which the probabilities of overfishing are based. Additionally, the assessment results showed variation in Z estimates and temporal trends between gear/fleet types at the same location. The obvious response would be to consider whether this is due to differences in selectivity, whereby the different fleets/gears are targeting different sizes of the stock.

It appears that the Workshop participants undertook the assessment on the basis that the approach was appropriate for the best available data source, and simply reported on this and its outcomes. While the assessment approach may be the best scientific information available, the report provides little more than the technical details and a brief consideration of the outcomes.

My remaining concerns with the Assessment Workshop Report are summarized as follows:

- The presentation of available data across the three sections of the SEDAR-35 report was difficult to navigate, both within and between reports. A simple summary table of the available information (source, location, type of data, fishery in/dependent, time series, and indicating what was provided to the Assessment Group, and what was actually used) allowing for ready comparison would have been highly useful.
- Prior to circulating for review, reports should be cross-checked to ensure that report reference lists are complete and that key papers are included as background reading (or at least links provided).
- Generally, the data decisions appear sound and robust, but, particularly where data have been excluded, these could have been better justified. I also recommend not being overly hasty in discarding data. Even if time series of CPUE, for example, are not considered of adequate quality to enable a formal stock assessment, the data may be useful in informing simpler, more empirical assessments (e.g. Froese 2004; Dowling et al. 2008; Prince et al. 2012; Edwards et al. 2012; Erisman et al. 2014). At the very least, they give some notion of historical high catches, and size-based catch rates.
    - With the exception of the direct comparison of TIP length data and SEAMAP-C length data, there was little effort made to reconcile i) the TIP data across the different locations and gear/fleet types, let alone across ii) the various sources of length data, or iii) the other available data. As a minimum, some

qualitative consideration of how the TIP estimates of mean length and Lc vary between gear types and location, and what this may mean for the interpretation of stock status, would have been highly useful.

  - Alternatively, a clear justification of why stock status estimates will differ according to gear type and location, and why this is acceptable in the context of using these for management purposes, could have been provided.
  - Reviewing the TIP length data for representativeness (temporally, spatially, of the size structure of the total fished population, of the total fishing effort) is important. This should include a careful consideration of the TIP length data in the context of the other available length frequency information.

- No alternative assessment approaches were reviewed or considered. I recommend attempting to reconcile length-based methods with production models and/or other assessment approaches, either via a review of approaches, attempting to fit a production model to existing CPUE time series, and/or or better justifying the current approach. Even acknowledging the supposed problems with the two standardized CPUE indices (AW01; DW04), and the St Thomas spawning aggregation density time series, it still would have been of interest to have at least attempted to have fitted a simple production model to these.
- There was no estimated reference point corresponding to a threshold for an overfished state of the stock (however "stock" is defined), nor justification for the lack of this. Limit reference points often correspond to 20% of unfished biomass, B_0, so a crude approximation could be made that the probability of being overfished equates to the probability that $F_{cur} > F_{SPR20\%}$.
- The probabilities of overfishing were determined from frequency distributions that embraced a large range of possible fishing mortalities. Whether by improving the understanding of life history, and/or by reviewing the available information and parameters, weighting or narrowing the range of mortality estimates and so increasing the certainty around the probability of overfishing should be a key priority for future assessments.
- The same concerns regarding the large range of values of Z apply to the large range in the estimates of $F_{30\%}$ and $F_{40\%}$ (per Figures 39-41). There should be a recommended preferred reference point (that which is more precautionary, in the absence of other information).
- The report concluded that it was unlikely that overfishing was occurring for Caribbean red hind in Puerto Rico, despite probabilities of 25% and 40%. This conclusion is highly subjective and lacks direct justification.
- The probabilities of overfishing according to the $F_{30\%}$ and $F_{40\%}$ proxies for F_{msy}, are, I believe, high enough to be of concern that overfishing is indeed occurring, particularly for St Thomas/St John and St Croix, where the lowest probability is 42%. There is no discussion of a minimum reference probability above which some decision rule is invoked or further investigation is required, but even the lowest probability of overfishing of 25% for Puerto Rico should warrant some attention.

- The interpretation of the results relative to the other sources of available input data (as per the Assessment Workshop Report's Discussion section) is questionable in various ways.
- It was stated that the mean length estimator approach does not include projections of stock dynamics, and that these were therefore not conducted. There exists adequate life history and catch information that a simulation-based management strategy evaluation (MSE) model could have been developed, incorporating a length-based assessment. A simulation-based MSE could be developed, to facilitate projections and undertake a risk analysis by applying decision rules according to the probabilities of overfishing, and thus determine the relative importance of resolving the range of mortalities contributing to these probabilities.
- The Assessment Workshop argues that the mean length estimator approach precludes the setting of ACLs. However, direct estimates of abundance and sustainable yield are not required to set a recommended ACL. I strongly encourage that more effort be dedicated to considering approaches to developing ACLs. Catch time series, triggers (as reference point proxies) or reference points, and empirical decision/harvest control rules can all be used to set an interim ACL.

BACKGROUND

This report comprises an impartial and independent peer review of the SEDAR 35 stock assessment undertaken for Caribbean Red Hind, a grouper species. The assessment focuses on three locations: Puerto Rico, and St Thomas/St John and St Croix in the U.S. Virgin Islands.

The Assessment Workshop Report is the third of three sections of the SEDAR 35 report; the other two sections comprise an Introduction and the Data Workshop Report. There was minimal cross-referencing or linkage between the report sections. The management history of the fishery provided in the Introduction is scarcely mentioned in the subsequent sections.

There is no mention of previous stock assessments having been undertaken for this species, whether in these or other regions, and as such, no notion of past precedence.

There is also no indication of assumptions regarding stock structure.

There is a range of historical fishery dependent and independent data available for the fishery, including several time series of catch rates, for which two CPUE standardisations were attempted in order to obtain proxy abundance indices. These were considered to be inappropriate for use in a formal stock assessment, and were considered only briefly and in a qualitative sense in the Discussion section of the report. Most of the other data was short-term or temporally patchy, and/or associated with low sample sizes.

There was little attempt to compare or reconcile data from different gear/fleet sectors and different regions, nor to consider time series in the context of the management history (i.e. significant changes in management that may have affected observed trends).

The assessment approach focused on using length frequency data obtained from a port-sampling data collection program, termed the Trip Interview Program, to obtain estimates of total mortality via a mean-length estimator. Fishing mortalities were obtained by subtracting estimated natural mortality, and compared to overfishing reference points derived from spawner-biomass-per-recruit estimates. A sensitivity analysis was undertaken by fitting mortalities under a range of combinations of life history parameters and length of first capture.

These methodologies appear sound and consistent with the best available assessment approaches based on length data.

No alternative assessment approaches were reviewed or considered.

The outcomes equate to six values for the probability of overfishing occurring, with two values (each corresponding to an alternative threshold reference point) for each of the three locations. There was no estimate corresponding to whether the stock (however this is defined) is overfished. There was no attempt to reconcile the outcomes between areas and between gear/fleet types – the latter are assumed to contribute equally to the overall frequency distribution from which the overfishing probability was derived.

The probabilities were determined from frequency distributions that embraced a large range of possible fishing mortalities. Narrowing the range of mortality estimates and so increasing the certainty around the probability of overfishing should be a key priority for future assessments.

The report concluded that it was unlikely that overfishing was occurring for Caribbean red hind in Puerto Rico, despite probabilities of 25% and 40%. This conclusion was apparently based upon the fact that a proxy abundance index time series (DW04) supposedly contradicts the mean length results in suggesting an increase in F for the vertical line fleet, and upon the fact that the results from mean length analyses suggested conflicting dynamics between the fleets, in terms of temporal trends in predicted Z. Neither of these justify a "no overfishing" conclusion.

The probabilities of overfishing according to the $F_{30\%}$ and $F_{40\%}$ proxies for F_{msy}, are, I believe, high enough to be of concern that overfishing is indeed occurring, particularly for St Thomas/St John and St Croix, where the lowest probability is 42%. There is no discussion of a minimum reference probability above which some decision rule is invoked or further investigation is required, but even the lowest probability of overfishing of 25% for Puerto Rico should warrant some attention. This is regardless of the temporal trends in Z predicted by the mean length estimators. While low sample sizes were a problem, particularly for the USVI regions, this should be even more reason to respond in a precautionary manner, to probabilities of overfishing that are already high.

It was stated that the mean length estimator approach does not include projections of stock dynamics, and that these were therefore not conducted. There exists, however, adequate life history information to have developed a simulation-based management strategy evaluation model that would have enabled stock projections to have been modelled. This would also have provided a framework via which to evaluate and establish target reference points and ACLs.

DESCRIPTION OF THE INDIVIDUAL REVIEWER'S ROLE IN THE REVIEW ACTITIVIES

I was approached by the Center for Independent Experts (CIE), following a recommendation from colleague, Dr. Ana Parma from the Centro Nacional Patagónico of Argentina.

My background is in management strategy evaluation (MSE) and the development of formal harvest strategies (monitoring, assessment, decision/harvest control rules) for data-poor species and fisheries. I was project officer on the development of the Australian Commonwealth Harvest Strategy Guidelines and led the development, finalisation and implementation of harvest strategies for all Commonwealth data-poor stocks. I subsequently secured and led two contracts for the FAO developing guidelines for harvest strategy development in data-poor contexts.

While the majority of my work is not focused on stock assessments per se, via my extensive experience with data poor harvest strategies, I have a strong awareness of the ranges of assessment approaches and options available in a data poor context.

I have never been actively involved in assessments of red hind or other grouper species either in the Caribbean or elsewhere.

SUMMARY OF FINDINGS FOR EACH TERM OF REFERENCE

Annex 2: Terms of Reference for the Peer Review

SEDAR 35 Caribbean Red Hind Assessment Desk Review

NB lack of page numbering in the assessment report will make referencing difficult.

1. **Evaluate the data used in the assessment, addressing the following:**
   a) **Are data decisions made by the Assessment Workshop sound and robust?**

Generally, the data decisions do appear sounds and robust, but, particularly where data have been excluded, could have been better justified. There were numerous available time series of standardized catch rates, fishery independent survey abundances, and catch, which were not used directly in the formal assessment, yet the rationale behind this decision was not discussed in detail.

A typical example is where the fishery independent St Thomas spawning aggregation data have been described, and maximum densities used to obtain an index of spawning abundance. The report states, "given the caveats…that have been highlighted in this section, the resulting density estimates should be used as ancillary qualitative information, rather than be incorporated into a quantitative stock assessment model". Yet, the caveats mostly pertained to the use of mean rather than maximum density.

Similarly, the SEAMAP-C time series of standardized CPUE proxy abundance indices (per SEDAR35-DW-04) were referred to briefly, but the main focus of the SEAMAP-C data discussion pertained to the length data. The rationale for the standardized CPUE time series not being uptaken was not provided.

Better justification was provided for the decision not to use the fishery-dependent CPUE-based abundance indices developed for Puerto Rico (SEDAR35-AW-01) ("…based on self-reported data of unknown accuracy"), but the argument that the indices are based on landings only and therefore should be excluded partially on this basis does not seem to make sense: all CPUE indices are based on landings. I also disagree with the conclusion that these indices showed "no overall directional trends in CPUE" – I feel that the lack of detectable trends is a result of the scale of the y-axis that was chosen to include the broad confidence interval values.

The rationale for excluding the St Croix spawning aggregation data and the Mona Island and Abrir la Sierra, Puerto Rico spawning aggregation data are sound, but the potential of the latter should have been noted explicitly.

Moreover, even acknowledging the supposed problems with the two standardized CPUE indices, and the St Thomas spawning aggregation density time series, it still would have been of interest to have at least attempted to have fitted a simple production model to these. This is of especial relevance given that the mean length methods used do not enable projections of stock dynamics. Even if the abundance indices are unreliable and/or there was inadequate contrast to enable a production model to be fitted, the attempt to have done so would have flagged more strongly the potential value of collecting robust CPUE and/or density information.

The updating of the several data sources as per point 1 of the Assessment Workshop Terms of References appears appropriate and thorough.

More generally, I found the presentation of available data across the three sections of the SEDAR-35 report highly difficult to navigate, both within and between reports. A simple summary table of the available information (source, location, type of data, fishery in/dependent, time series, and indicating what was provided to the Assessment Group, and what was actually used) allowing for ready comparison would have been highly useful (in fact, in order to help try to grasp the data landscape, I undertook this task myself).

It would also have been helpful had the order and groupings of the presentation of data been consistent between the Data Workshop Report and the Assessment Workshop report. Links to working papers and reference documents were not always made explicitly (e.g. the "St Croix trap study" cites Bryan et al. (2013) which is presumably RD01 (yet RD01 states that it describes a protocol for belt visual surveys, so the reader is left confused – particularly as this reference [along with many others] is not listed in the reference list of the report).

Of greater concern was that there was very little effort made across both reports to attempt to reconcile in any detail the data trends and inferences between gear/effort types, location and sector (i.e. commercial, recreational, survey). This applies both within and between data types. It would have been highly useful to have seen the extent to which trends were consistent across these categories.

At a minimum, it would have been useful to have seen time series of mean length presented for all available length information, as this is what was used in the assessment. Figure 3.5.8 in the Data Workshop Report shows this, but only for the TIP length frequency information, and Figures 11 and 12 in the Assessment Workshop Report compare the TIP and SEAMAP-C length-frequency information. It would have been useful to have seen this for all sources of length information (including the SEAMAP-C surveys, the NOS Biogeography Visual Survey Database, and that from DW03, RD02, RD03 and RD04). Ideally, such comparisons should have been made for the available catch, CPUE and survey abundance information, also. As it stands, it is very difficult to grasp the overall picture: are things different in a different area and/or under a different gear/survey protocol? Or a different sector (commercial/recreational/independent survey)? Or for a different data set in the same area?

Aside from a cursory attempt to reconcile the Fmsy-based probabilities of overfishing with recent trends in CPUE, there was no attempt made to compare the different types of data, to ascertain whether they were reinforcing or apparently contradicting each other in terms of what they suggested regarding the status of the stock.

It would also have been useful to have understood the difference between minimum/mean length of capture from the length frequency data, and the length at maturity, in terms of age. That is, how many spawning seasons (if any) do red hind experience before becoming vulnerable to fishing gear?

I do agree that it was appropriate to choose a mean-length estimator approach on the basis that length frequency data are currently the most temporally consistent source of species-specific information.

The Assessment Workshop report states "Annual length-frequency plots were constructed for each stratum that the Panel agreed has sufficient sample sizes". While the issues around sample sizes were touched on in the Data Workshop report, the criteria for a "sufficient" sample size are not explicitly presented.

Finally, there was no attempt to define what constituted a stock, for purposes of the assessment, nor were any assumptions specified regarding stock structure. This would have provided a context for how results could or should be reconciled across regions and/or gear/fleet types. At a minimum, a consideration of the area covered by the analyses, versus the potential of the species for movement and mixing (per RD09, which indicated red hind were capable of spawning migrations of up to 33km), would have been useful.

b) Are data uncertainties acknowledged, reported, and within normal or expected levels?

Uncertainties in:

- the von Bertalanffy growth parameters
- the values of Lc
- the length-weight relationship
- maximum age (used to derive M)

are acknowledged and considered in sensitivity tests around the estimates of total mortality and the years in which this changes. Von Bertalanffy, maximum ages and length-weight parameters were based on those reported in the literature.

Ideally, the suite of sensitivity analyses undertaken could have been summarized clearly in a table illustrating the total number of combinations/scenarios.

I am not convinced of the value of fitting a linear model to the growth parameters reported in the literature, so as to obtain an additional nine parameter pairs. The relationship between the pairs of reported parameters does not seem significantly linear (Figure 17, r^2 = 0.5608; significance value not reported). Moreover, the addition nine parameter pairs are interpolations within the range of those reported in the literature. As such, the inference adds more uncertainty while not providing new values outside the range of those from the literature. It would have been simpler and more sensible to have only used the four reported sets of values found in the literature.

The definition of size-at-recruitment to the fishery, Lc, was vague and this was a significant concern. Page 15 of the Data Workshop Report states, "visual inspection of length-frequency distributions can be used to determine the size-at-recruitment (Lc) to the fishery where the mode of a well-defined distribution can represent Lc".

- It seems strange to me that the mode, as opposed to some lower percentile, is assumed to represent Lc. The mode would surely represent more the prime size of capture rather than the first size of capture. According to Gedamke and Hoenig

(2006), on whose paper and equation the mean length model was based (noting also that this citation did not appear in the reference list of the report, nor as a reference document), Lc is the "smallest size at which animals are fully vulnerable to the fishery and the sampling gear". The Assessment Report, under "Assumptions", refers to Lc only as "the length at full recruitment". I still do not see how the mode of sampled lengths from the catch equates to Lc as per Gedamke and Hoenig's (2006) definition.

- The same page also states, "The mode of each annual distribution was determined as well as the overall mode for each gear type". Was the overall mode determined by averaging the annual modes, or combining the data across years and taking the mode of this combined distribution? The former gives equal weighting to each year; the latter down-weights years with lower sample sizes.

The Assessment Workshop report states, "Lc was selected visually from the annual length-frequency distributions while considering the annual sample size" – how was the sample size considered? Also, "The highest Lc value over the time series was chosen…(this) avoids violating model assumptions and the confounding of selectivity and mortality in the calculation of annual mean lengths" – I don't follow this. Why/how? Then, "mean lengths were calculated from lengths larger than Lc" – this makes no sense to me. If Lc is the modal length, as per the Data Workshop report, then this would only leave the upper percentiles of sampled lengths from which to obtain a mean – which seems contradictory to the definition of the latter. I then refer to Gedamke and Hoenig (2006) who state that "mean length" is in fact the mean length of animals larger than Lc – in which case the definition provided makes sense (assuming Lc is indeed the modal length in the sampled catch). However, it is not clear if this is the mean length of animals larger than Lc in the population, or in the fishable population. These definitions need to be carefully clarified and explained.

Further, in terms of the sensitivity analysis, the Assessment Workshop report states, "The sensitivity of the mean length estimator to the selection of Lc was explored using two alternative assumptions, the value chosen by visual inspection and used in the initial analysis, and the average mode of the annual length-frequency distributions for each stratum". But does not the "visual inspection" equate to the identification of the mode? Without clarity on how these modes were combined across years to obtain the Lc based on "visual inspection", it is difficult to understand how this differs from taking "an average mode of the annual length frequency distributions". This needs to be clarified.

Moreover, it would seem more appropriate to test sensitivity to a range of Lc values derived from the alternative available sources of length-frequency data.

c) Are data applied properly within the assessment model?

See the above concerns regarding the definition and derivation of Lc and mean length. These need to be clarified. Incidentally, the reference to Thorson and Prager (2011) against the statement "Lc was selected visually from the annual length-frequency distribution" doesn't add anything – this paper, among other things, is about how the use of logistic catch curves relaxes the assumption of knife-edge selectivity, and eliminates the need to select this age visually from observed catch data. The citation doesn't add confidence regarding, or provide clarification of, how Lc was selected visually.

The assumptions of constant recruitment and mortality associated with the mean length estimator, are not given consideration. The history of the fishery should be acknowledged in this context: there have been many management changes pertaining to spatial and seasonal closures, and to gear. There have also been temporal changes in the length-frequency distribution/mean length for several gear/area strata. Whether these have corresponded to management changes has not been investigated. However, given the management changes and the changes in length, some consideration should have been given regarding the extent to which recruitment and mortality could be assumed to have been constant.

Knife-edged vulnerability was assumed for the per-recruit analyses– but is this suggested by the data? Moreover, if this is considered to be the case, then the mean length estimator would be more consistent with the per-recruit analyses if a lower percentile of the length distribution was used to calculate Lc – i.e. if a fish of a certain minimum size has been caught by the gear, regardless of in what relative proportion, this size is assumed to be that at which the species is fully recruited to the fishery.

Otherwise, the data appear to have been applied properly within the mean length estimator and the per recruit analyses.

d) Are input data series reliable and sufficient to support the assessment approach and findings?

I question the representativeness of the TIP length information (the only source used for the assessment), in the following ways:

- Under the "SEAMAP-C" section of the Assessment workshop report, it is stated, "selectivity of the SEAMAP-C survey resulted in a smaller modal length than the selectivity of the vertical line fleet". While the TIP information is derived from a fishery-dependent port-based interview program, how certain is it that the selectivities that correspond to each of the TIP gear types may be considered to be representative?
- There was no attempt to standardize for the variability in sample sizes between years. If mean length and Lc estimates are time invariant, this is not particularly important, but if estimates of Lc in particular vary between years, and so does sample size (and/or external factors such as management measures pertaining to gear changes or spatial closures), then this should be considered when combining the time series of data within each stratum to obtain an overall estimate of Lc. (There was an acknowledgement of the possible effect of the Hind Bank (St Thomas) closure on mean length, but the influence of spawning season on mean length was not apparent). The issue of inter-annual sample size variability persists across the three locations (Puerto Rico, St Thomas, St Croix), but taking pages 14-15 of the Data Workshop Report describe the Puerto Rico TIP data as an example:
    - The sample sizes were very low for all gears other than the vertical line.
    - The increases in annual mean length since 1983 for the pot and trap, and vertical line fleets: do these correspond to management measures pertaining to gear/selectivity? What is the impact, if any, on Lc?

- There appears to have been some kind of regime shift c.1998 for the pot and trap, and vertical line fleets corresponding to shifts in modal mean length. Possible reasons for this were not investigated in detail. Various (over all, pre-1988, post-1998) estimates of Lc were made as a result, but neither correspond to those used in the sensitivity analyses in the Assessment Report.

With the exception of the direct comparison of TIP length data and SEAMAP-C length data, there was little effort made to reconcile i) the TIP data across the different locations and gear/fleet types, let alone across ii) the various sources of length data, or iii) the other available data. As a minimum, some qualitative consideration of how the TIP estimates of mean length and Lc vary between gear types and location, and what this may mean for the interpretation of stock status, would have been highly useful. This would have considered the length statistics in the context of the gear selectivity and stock structure assumptions (noting that there is no discussion of stock structure across any of the reports). Alternatively, a clear justification of why stock status estimates will differ according to gear type and location, and why this is acceptable in the context of using these for management purposes, could have been provided. For example, if the probabilities of overfishing at one location are greater for one gear type than another in the same location, different control rules can be applied to each gear type accordingly. However, there remains the question of how these probabilities are reconciled to form an overall estimate of the status of the stock (presuming the gears are targeting the same stock). Conversely, if the probabilities of overfishing for the same gear type are different between areas, area-specific control rules can be imposed, but this will again depend on what is assumed about stock structure.

2. Evaluate the methods used to assess the stock, taking into account the available data.

a) Are methods scientifically sound and robust?

I have no problem with the mean length estimator method. I like that Gedamke and Hoenig (2006) get around the assumption of equilibrium population dynamics. This appears to the standard current acknowledged approach for estimating total mortality from length distribution data, accounting to non-equilibrium conditions (e.g. Edwards et al. 2012; Erisman et al. 2014).

The yield- and spawner-biomass-per-recruit calculations were standard.

As such, both the mean length estimator, and the per recruit analysis appear to me to be sound and robust, with the assumptions of each clearly stated.

However, while citations are made in the text, there are generally no full references given for key papers pertaining to the Methods. Specifically, I refer to

- Gedamke and Hoenig (2006) (extension of Beverton-Holt length-based mortality estimator)
- Thorson and Prager (2011) (visual selection of Lc from length-frequency distributions)
- Hewitt and Hoenig (2006); Hoenig (1983) (regression approach to estimate M)

- Botsford (1981); Walters and Martell (2004) (equilibrium vulnerable biomass-per-recruit)
- Menza et al. (2008) (fishery independent data)

I appreciate that the Assessment Working Group had to complete their report within a small window of time. However, it is ironic that these references, which were arguably the most critical in terms of underpinning the chosen methodology, were omitted from the reference list. I find it also somewhat bewildering that a 200 page report of survey protocols that collected data that was not used in the assessment (RD01) and a similarly large PhD thesis focusing largely on larval development (RD06) were included as reference documents, yet core references such as those listed did not even make it into the reference list of the report, let alone be included as reference documents.

Very little detail was provided on the derivation of natural mortality estimates. The previous studies on which the calculation was based are cited but do not appear in the reference list, nor in the list of background documents. These equations should appear (particularly given that the relatively well-known von Bertanlanffy growth equations are presented), if only in an appendix. Kenchington (2013) provides natural mortality estimators for information-limited fisheries; it would have been useful to have better understood the method chosen and why this was chosen over other alternatives.

b) Are assessment models configured properly and used consistent with standard practices?

The mean length estimator appears to have been properly configured and used consistent with standard practices (e.g. Edwards et al. 2012; Erisman et al. 2014).

There are some issues with the notation pertaining to the yield- and spawner-biomass-per-recruit calculations:

- K should be uppercase in equation 4.
- Equation 8: a_{mat}, not t_{mat}, or change a_{mat} to t_{mat} in the text
- Equations 9 and 10: s_t and s_{ot} seem to have been replaced with x_t and x_{ot} in equations 13 and 14 – this needs to be consistent
- Equation 11: use single letters for subscripts, i.e., not "*vb*"
- Equations 11,13,14: What is *l*?
- Equation 14: "*o*" needs to be subscripted to *x* (or *s*, as the case may be)

Otherwise these appear to have been properly configured and used consistent with standard practices.

c) **Are the methods appropriate for the available data?**

The methods appear to be appropriate for the available data, given that length-frequency data are currently the most temporally consistent source of species-specific information.

However, the Methods section of the Assessment Workshop report jumps straight from data evaluations to a description of the modelling approach. More context would have been appreciated, vis-à-vis

- A concluding summary of the available data, leading into a consideration of
- Available assessment options given the data and its quality, presented as a brief review. This should include both methods that pertain to the range of available data (particularly as the chosen length-based methods excludes stock projections, but acknowledging that other data were rejected for various reasons), and those that pertain to length data only.
- Background on (if)/how the species has been assessed in the past.
- Hence, greater justification for the chosen approach.

As it stands, the report reads as, "The length data were the most temporally consistent source of species-specific information, so a mean length estimator was used". The reader is left to wonder whether other approaches were even considered, and whether perhaps there is a historical precedent of this type of analysis having been undertaken, and, as such, an attitude of complacency. Perhaps it is merely a case of knowledge of the fishery and its history and assessment approaches being assumed known – but if so, there is a lot of knowledge that is assumed.

No explicit consideration was given regarding alternative assessment approaches. Erisman et al. (2014) provide various options for assessing stocks (for species forming spawning aggregations) using simple metrics based on catch length composition.

- Other methods that could have been used to evaluate stock status include
    - Assessment from Froese's (2004) indicators of sustainability. Froese (2004) suggested that management could be based on three size-based indicators and their target reference points: (i) percentage of mature fish in the catch, with 100% as target (Pmat); (ii) percent of fish of optimum length (i.e. the length where the number of fish in a given unfished year class, multiplied by their mean individual weight, is maximum and where thus the maximum yield and revenue can be obtained.) in the catch, with 100% as target (Popt); and (iii) percentage of large fish in the catch (Pmega), with 0% as target and 30–40% being acceptable if there is no upper size limit for the fishery.
    - Spawning potential ratio (SPR) analyses: SPR is the ratio of the total fecundity of the fished population under a given exploitation rate, to the total fecundity of the unfished population.
    - (possibly) SAFE: A Sustainability Assessment for Fishing Effects (SAFE): a fishing mortality based method that can quantify the effects of fishing on sustainability for large numbers of species with limited data (Zhou *et al.* 2009)
    - (possibly) production models – as stated above: even acknowledging the supposed problems with the two standardized CPUE indices, and the St Thomas spawning aggregation density time series, it still would have been of interest to have at least attempted to have fitted a simple production model to these. This is of especial relevance given that the mean length methods

used do not enable projections of stock dynamics. Even if the abundance indices are unreliable and/or there was inadequate contrast to enable a production model to be fitted, the attempt to have done so would have flagged more strongly the potential value of collecting robust CPUE and/or density information.

- The Assessment Workshop report seems to skirt around the issue of setting Annual Catch Limits (ACLs), arguing that the mean length estimator approach precludes this. However, one arguably does not need direct estimates of abundance and sustainable yield to set a recommended ACL (although obviously the ACL becomes more defensible the more that is understood about the stock). Methods that could have been used to set ACLs include, but are not limited to:
    - Froese's (2004) size-based sustainability indicators
    - Simple, empirical catch/CPUE- time series-based regression methods – e.g. RD04 has a good catch time series
    - and/or traffic light, CUMSUM control indices, or hierarchical decision trees – indicators could include changes in catch composition, landings and size of landings, as per RD04. Also, RD04 has a historical (1988-2001) time series corresponding to a heavily fished period, which could be used to inform reference points for indicators.

3. Evaluate the assessment findings with respect to the following:

a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?

The assessment findings are limited to estimates of fishing mortality and spawner biomass per recruit.

In the absence of any information regarding stock structure (or assumptions regarding same), the first question is whether the results should be interpreted as each of the three areas/regions equating to a separate stock. On the basis of the manner in which the results are presented, I will assume this to be the case, but this is an issue that should be clarified in the report.

Total mortality estimates are provided by location, and, correspondingly, probabilities of overfishing (across all crosses of sensitivity analyses [13x,von Bertalanffy pairs, 2x Lc, 3x length-weight relationship, 2x M] and gear/fleet types) against each of two F_{MSY} proxies: $F_{30\%}$ and $F_{40\%.}$

It would have been useful if something similar to the following summary table had been provided:

Location and gear/fleet type	Z estimates: most strongly supported by AIC	Lc	Z ranges	Pr(overfishing)
Puerto Rico: diving	0.312	370mm 323mm	0.68-1.0 0.3-0.8	Using $F_{30\%}$: 25% Using $F_{40\%}$: 40%
Puerto Rico: pot and trap	34% decline in 1995 from 0.444 to 0.292	283mm 310mm	Avg. increase of 32% (from what to what?) Avg. increase of 36%(from what to what?)	
Puerto Rico: vertical line	61% decline in 1987 from 0.379 to 0.235	298mm 340mm	Avg. increase of 31% (range 0.19-0.58) Avg increase of 38% (range 0.25-0.68)	
St Thomas/ St John: pot and trap	44% increase in 1983 from 0.270 to 0.390	302mm 320mm 340mm	Avg increase of 38% (range 0.34-0.93) Avg increase of 64% (range 0.45-1.0) Avg increase of 57% (range 0.35-0.83)	Using $F_{30\%}$: 42% Using $F_{40\%}$: 57%
St Croix: diving	0.476	264mm 296mm	Similar for both Lc values	Using $F_{30\%}$: 54% Using $F_{40\%}$: 66%
St Croix: pot and trap	0.295	277mm 340mm	Avg increase of 86% (no range given) Avg increase of 79% (range 0.27-0.65)	
St Croix: vertical line	75% increase from 0.286 to 0.502	287mm 350mm	Constant Z; no values given Avg. increase of 64%	

Are these estimates reliable? On the basis of the consistently large ranges of possible mortality values (and the large range in the $F_{30\%}$, $F_{40\%}$ values to which they were compared), probably not. The fact that some of the Z estimates include values of 1.0 is also of concern.

I don't know if there is some manner in which the suite of possible mortality values arising from the sensitivity analyses can somehow be weighted, or reduced – e.g., are there combinations of input values for Lc, von Bertalanffy parameters, natural mortality and length-weight parameters that are more biologically plausible/consistent than others? (e.g. growth parameters with a lower L_{∞} may be more likely to be associated with a lower Lc value). (Also, less weighting could be given to runs where von the Bertalanffy parameters have been inferred from linear regression). The combinations that are weighted more highly could perhaps have a greater relative representation in the frequency distributions that are used to determine the probabilities of overfishing. Alternatively, those combinations that are considered less plausible could be omitted from the pool of values.

I assume that the numbers of lines on each of the panels of Figures 39-41, and the summed frequencies in Figures 42-44 equate to 13 von Bertalaffy parameter pairs x 2 Lc values x 3 length-weight relationships x 2 natural mortality values x (1 or 3) fleet types = 176 (St Thomas) or 528 (Puerto Rico and St Croix). If so, this should be explicitly stated. (I assume that "like was compared with like" in calculating the ratios for F_{cur} (from the mean length estimator) relative to $F_{30\%}$ and $F_{40\%}$ (from the YPR, SPR analyses), in terms of the population and Lc parameters).

Even if we assume that the values of F are reliable for each area and gear/fleet stratum, there remains:

i) the issue of some contradictions in trends in Z for different sensitivity runs, and/or differences in timing for changes in Z

ii) the issue of the variation in Z estimates and temporal trends between gear/fleet types at the same location.

Beside the above suggestion of assigning relative plausibilities to the different sensitivity combinations, and acknowledging the issue of frequently low sample sizes, I have no suggestions as to how to reconcile i).

Regarding ii), the obvious response would be to consider whether this is due to differences in selectivity, whereby the different fleets are targeting different sizes of the stock. Certainly for Puerto Rico, and, to a lesser extent, St Croix, there is little overlap in the Lc values for the diving fleet and those for the pot and trap, and vertical line fleets, which may help to explain the contradictory trends in Z. Given that M is invariant of the gear/fleet type, the differences in Z are solely due to fishing mortality. Where these differ between gear/fleets, emphasis should be put on those gears/fleets with the highest relative effort, and/or those that are targeting the larger, more fecund individuals (or else very small individuals < L_{mat}, but this does not seem to have occurred for red hind). For Puerto Rico, the diving sector appears to target larger individuals than the other gear types, but for St Croix, the diving sector appears to target smaller individuals.

Due to the large number of combinations of scenarios considered, it is difficult to determine whether estimates of total mortality are consistent with input data and population biology characteristics.

- There is a frequently observed negative correlation between estimated current total mortality and asymptotic length, and positive correlation between estimated current total mortality and the von Bertalanffy growth coefficient. These trends should be the inverse of each other: a larger L_∞ goes along with a lower K, corresponding to a slow-growing, less productive stock that ultimately reaches a larger maximum size. However, for a given mean length and Lc, I would have thought a slow-growing, less productive stock would experience a higher total mortality than that for a fast-growing, more productive stock (i.e. one with a lower L_∞ and higher K) – and indeed this is suggested by equation (1) (the equilibrium Beverton-Holt estimator). The relationship between total mortality and the von Bertalanffy parameters was sometimes dome-shaped, which suggests that total mortality is lower at the extremes of von Bertalanffy parameter combinations.
- There was no consistent trend of estimated Z being higher or lower for higher or lower values of Lc.
- There was often an interaction between Lc and the von Bertlanffy growth parameters in terms of the nature of the relationship between the estimated Z and each of L_∞ and K.

Then, there are issues associated with the interpretation of the results relative to the other sources of available input data (as per the Assessment Workshop Report's Discussion section):

- For Puerto Rico, the estimated temporal changes in mortality for the vertical line fleet were compared to the CPUE standardization undertaken using SEAMAP-C data (DW04) for the same fleet. The latter suggests abundance is declining, and this is assumed to equate to an increase in fishing mortality. This may not be the case: fishing mortality could in fact be declining in response to low abundance. Moreover, the conclusion cannot be drawn that "This is contradictory to the mean length estimator result for the vertical line fleet, which suggests mortality declined". It does not follow that low abundance equals high fishing mortality in the same time period. There is typically a lag between fishing mortality and the response in terms of abundance. Low abundance indices can follow a period of high fishing mortality, and in the time during which abundance is low, fishing mortality may subsequently ease in response to this, such that the population then recovers.
- Again for Puerto Rico, it is stated that the fishery dependent relative abundance indices (per AW01) were "flat, suggesting that abundance has not changed". I disagree; I believe the mean trend appears flat due to the scales of the y-axes that are accommodating the confidence intervals.
- Again for Puerto Rico, it is stated that effort has declined for the pot and line and the vertical trap fleets, corresponding to a modelled decline in Z and F, while diving effort for red hind has increased, corresponding to no change or an increase in Z. These correlations do make sense.
- For St Thomas, it is concluded that, because reported landings and effort have declined since 2008, this equates to a decline in fishing mortality, and an eventual increase in fish size. This is interpreted to be contradictory to the result from the mean length estimator, which suggests that mortality has increased due to a

reduction in mean length. This supposed contradiction is shaky: there is no consideration of time frame for recovery of mean length. Moreover, there is a temporal mismatch in the comparisons: the modelled decline in mean length is based on a time series from ~1983 to ~2012, during which mean length declined from the mid-1980s, but remained relatively constant since ~1995. Thus the modelled increase in mortality corresponds to a time long prior to the declines in landings and effort in 2008.

b) Is the stock overfished? What information helps you reach this conclusion?

There was no reference point corresponding to a threshold for an overfished state.

I don't particularly follow the rationale of "The discussion about whether to use $F_{0.1}$ and $F_{SPR30\%}$ (or $F_{SPR40\%}$) was centered on biological considerations and acceptable risk. The Panel agreed that the risk of recruitment overfishing outweighed the risk of growth overfishing, and given the seasonal and spatial closures for red hind, $F_{30\%}$ and $F_{40\%}$ were reasonable F_{MSY} proxies". This doesn't make much sense unless it is put in the context of the sentence from the "Modelling approach" section: "F_{SPR} based metrics are most often considered when there is a concern that recruitment overfishing is possible since SPR is a function of not only mortality and weight, but also maturity". The two sections describing the choice of reference points (in the "Modelling approach" and "Results" sections) should be combined.

Second, the rationale for including both $F_{30\%}$ and $F_{40\%}$ as F_{MSY} proxies, was not made clear – was the intention that F< $F_{40\%}$ would correspond to growth overfishing, and F< $F_{30\%}$ to recruitment overfishing? Presumably $F_{0.1}$ was excluded because it was obtained from yield per recruit analysis and so did not embrace recruitment overfishing. Thirdly, none of the "Per recruit analysis" section of the Results section justifies the lack of choice of reference point for an overfished stock status.

In a broad-brush sense, limit reference points often correspond to 20% of unfished biomass, B_0, so a crude approximation could be made that the probability of being overfished equates to the probability that $F_{cur} > F_{SPR20\%}$.

c) Is the stock undergoing overfishing? What information helps you reach this conclusion?

The probabilities of overfishing according to the $F_{30\%}$ and $F_{40\%}$ proxies for F_{msy}, are, I believe, high enough to be of concern that overfishing is indeed occurring, particularly for St Thomas/St John and St Croix, where the lowest probability is 42%. There is no discussion of a minimum reference probability above which some decision rule is invoked or further investigation is required, but even the lowest probability of overfishing of 25% for Puerto Rico should warrant some attention. This is regardless of the temporal trends in Z predicted by the mean length estimators. While low sample sizes were a problem, particularly for the USVI regions, this should be even more reason to respond in a precautionary manner, to probabilities of overfishing that are already high.

Point evaluations of overfishing/no overfishing for each of the reference case scenarios would have been useful.

The report conclusion that, for Puerto Rico, "The results indicate that there is a low probability that Red Hind are experiencing overfishing" is highly subjective and lacks direct justification. It appears to be based upon the fact that a proxy abundance index time series (DW04) supposedly contradicts the mean length results in suggesting an increase in F for the vertical line fleet, and upon the fact that the results from mean length analyses suggested conflicting dynamics between the fleets, in terms of temporal trends in predicted Z. Neither of these justify a "no overfishing" conclusion. Indeed, it was the diving sector that suggested Z was constant or increasing with time, and it is this sector that appears to target larger fish for Puerto Rico. An increase in Z on a larger (more fecund) sector of the population is cause for concern.

In terms of temporal changes, it should be reiterated that values for M are fixed temporally such that any predicted changes in Z are directly attributed to changes in F.

d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?

There is no stock-recruitment relationship provided, nor estimates of steepness.

e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

While $F_{30\%}$ and $F_{40\%}$ are proxies for F_{msy}, there no reference point or proxy that corresponded to a threshold for an overfished state.

The same concerns regarding the large range of values of Z apply to the large range in the estimates of $F_{30\%}$ and $F_{40\%}$ (per Figures 39-41): it is difficult to know which (if any) among these are the more plausible. In the "Modelling Approach" section of the report, the sentence "The probability of overfishing integrated across all modelled sources of uncertainty was then determined" is vague and does not specify precisely how this integration was performed. I assume all sensitivity combinations contributed to the frequency distributions presented in Figures 42-44.

Again, I assume that "like was compared with like" in calculating the ratios for F_{cur} (from the mean length estimator) relative to $F_{30\%}$ and $F_{40\%}$ (from the YPR, SPR analyses), in terms of the population and Lc parameters. However, aside from the ambiguous sentence, "…the ratio between $F_{current}$ and F_{MSY} proxies….was obtained to determine overfishing status for a given sensitivity run", this is not clarified in the text: it is possible that only the mean or median $F_{30\%}$ and $F_{40\%}$ were used in the ratio calculations.

The lack of specification of stock structure assumptions also brings into question the reliability of spawner-per-recruit based reference points. If the fish captured by the various gear/fleets are generally above the size of maturity, then presumably they collective contribute to the spawner biomass. Yet separate spawner-per-recruit estimates are presented for each gear/fleet combination. Better justification for this needs to be provided.

As stated earlier, even acknowledging the supposed problems with the two standardized CPUE indices, and the St Thomas spawning aggregation density time series, it still would have been of interest to have at least attempted to have fitted a simple production model to these. This is of especial relevance given that the mean length methods used do not enable projections of stock dynamics. Even if the abundance indices are unreliable and/or there was inadequate contrast to enable a production model to be fitted, the attempt to have done so would have flagged more strongly the potential value of collecting robust CPUE and/or density information. If production models were able to be fitted, estimates of biomass-based reference points can be inferred from these.

Moreover, there exists adequate life history and catch information that a simulation-based management strategy evaluation (MSE) model could have been developed, incorporating a length-based assessment. The simulation could then have been used to have selected an appropriate $F_{x\%}$, as per Edwards et al. (2012) (and associated references under the "reference points" section of this paper). As it stands, there is no attempt made to recommend the use of $F_{30\%}$ over $F_{40\%}$, or vice-versa.

I acknowledge that the time and/or required to develop an MSE may have been exceeded that available, but it would be of value to have flagged this as a future recommendation.

4. **Evaluate the stock projections, addressing the following:**
   a) **Are the methods consistent with accepted practices and available data?**
   b) **Are the methods appropriate for the assessment model and outputs?**
   c) **Are the results informative and robust, and useful to support inferences of probable future conditions?**
   d) **Are key uncertainties acknowledged, discussed, and reflected in the projection results?**

Against their terms of reference, the Assessment Workshop surmises, "Due to the limited data available, a data poor methodology was attempted that does not include projections of stock dynamics. Therefore, projections were not conducted for this assessment".

As stated immediately above, I believe there exists enough life history information and catch data that a simulation-based operating model could have been built and projections conducted in the context of a Management Strategy Evaluation (MSE) framework. This could have been a simple model, but it would at least have provided a platform for undertaking projections (thereby inferring future conditions), and for selecting appropriate reference points.

Even for data-poor cases, Australian examples (e.g., Dichmont and Brown, 2010; Dowling, 2011; Haddon, 2011; Klaer and Wayte, 2011; Plaganyi et al., 2013) indicate that a formal MSE enables objective performance evaluation, robustness testing, and detecting responses that cannot be intuitively anticipated.

5. **Consider how uncertainties in the assessment, and their potential consequences, are addressed.**

- **Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods**

The methods used to evaluate uncertainty are limited to undertaking sensitivity runs embracing population (i.e. life history parameter) uncertainty. Specifically, uncertainties in:

- the length at full recruitment, Lc (up to 2 values per area and gear/fleet stratum)
- von Bertalanffy growth parameters (13 sets of values)
- the length-weight relationship (3 sets of values)
- natural mortality (2 values)

were considered.

While these sensitivity runs certainly embrace the population–based uncertainty relevant to the required inputs for the assessment approach, they resulted in a large range of Z-values from the mean-length estimator and large ranges of $F_{\%SPR}$ based reference points, conferring uncertainty to the resultant probabilities of overfishing.

However, short of recommending studies on basic life history, there were not suggestions around how to narrow the range of estimated mortalities, or what subset of the range was considered most plausible.

As stated against TOR 3a) above, I don't know if there is some manner in which the suite of possible values arising from the sensitivity analyses could somehow be weighted, or reduced – e.g., are there combinations of input values for Lc, von Bertalanffy parameters, natural mortality and length-weight parameters that are more biologically plausible/consistent than others? (e.g. growth parameters with a lower L_∞ may be more likely to be associated with a lower Lc value). (Also, less weighting could be given to runs where the von Bertalanffy parameters have been inferred from linear regression, rather than empirically derived). The combinations that are weighted more highly could perhaps have a greater relative representation in the frequency distributions that are used to determine the probabilities of overfishing. Alternatively, those combinations that are considered less plausible could be omitted from the pool of values considered.

In the Discussion section for Puerto Rico, it was mentioned that "Further complicating this (per-recruit) analysis is the potential unknown component of mortality associated with the regulatory discards during the closed season for red hind….it is important to better understand whether the magnitude of incidental catch of red hind during the seasonal closure is a negligible component of mortality." There was no attempt to incorporate some hypothetical estimate of discard mortality within the per-recruit analyses.

As stated against TOR 1b) above, I was confused about the following statement: "The sensitivity of the mean length estimator to the selection of Lc was explored by using two alternative assumptions, the value chosen by visual inspection and used in the initial analysis, and the average mode of the annual length-frequency distributions for each stratum". How do these differ? Surely the value chosen by visual inspection should equate to the mode?

No specification or discussion regarding stock structure, or assumptions around the same, was provided.

There was little evaluation of uncertainty in the data sources or the assessment methods themselves. In terms of data sources, as stated against TOR 1d) above,

- there was no attempt to standardize for the variability in sample sizes between years within the TIP data.
- with the exception of the direct comparison of TIP length data and SEAMAP-C length data, there was little effort made to reconcile i) the TIP data across the different locations and gear/fleet types, let alone ii) the various sources of length data, or iii) the other available data. As a minimum, some qualitative consideration of how the TIP estimates of mean length and Lc vary between gear types and location, and what this may mean for the interpretation of stock status, would have been highly useful.

The Discussion section mentioned the following issues pertaining to data sources, but the report did not attempt to address these:

- Re: Puerto Rico:
    - "In the per-recruit analyses to develop overfishing probabilities, the fleets were assumed to be equally representative of the population. Without spatially explicit data with respect to area and depth it is difficult (to) ascertain whether this assumption is being met". Relative effort would also have been useful in this context.
    - Discards in Puerto Rico during closed season possibly affecting mortality (as mentioned above).
- Re: St. Thomas/St. John:
    - The Hind Bank spawning aggregation data index was considered as a source of ancillary data because of many years with low sample sizes of red hind length from St. Thomas. This index lacked a clear temporal trend and was characterized by inter-annual variability, but the report stated that it was not possible to disaggregate the degree to which the variability was explained by environmental covariates or sampling variability.
- Re: St. Croix
    - Low sample size was a major concern with the length-frequency data.
- Re: St. Thomas/St. John and St. Croix
    - Possibly market-driven demands affecting size distribution and hence selectivity. As market-driven selectivity is generally dome-shaped, this would violate the assumption of knife-edged selectivity in the per recruit analysis.

In terms of the assessment methods themselves, sources of uncertainty were not investigated quantitatively. The "Research Recommendations" section states that "the ability to use the mean-length estimator is contingent upon having length-frequency data that are temporally consistent and representative of the population, and upon having reliable estimates of life history parameters". Temporal inconsistencies in terms of sampling size were mentioned throughout the report, together with their possible impact on length distributions, but there was no direct suggestion as to how to reconcile these.

It is also mentioned that the size of fish in the USVI may be market driven for plate size, suggesting that the selectivity may be dome shaped, which is in violation of the assumption of knife-edged selectivity in the mean-length estimator. Expanding the mean-length estimator to accommodate other selectivity patterns was suggested as future research, rather than attempted within the current workshop.

- **Ensure that the implications of uncertainty in technical conclusions are clearly stated.**

The technical conclusions are three sets of probabilities of overfishing. These have been presented with no implications of uncertainty, or (with the exception of recommending studies on basic life history) recommendations regarding narrowing the range of mortality estimates and so increasing the certainty around the probability of overfishing. Specifically, the implications of:

- The breadth of the ranges of Z (and hence the uncertainty in stock status) in terms of the efficacy of the overfishing probabilities in a management context
- Having separate sets of probabilities for each of the three regions, without specifying what is assumed about stock structure
- The equal consideration/contribution of all sensitivity runs in calculating the probabilities of overfishing
- The lack of recommendation regarding the preferred choice of overfishing reference point

are not explicitly considered.

As a secondary point, I have some issues the CPUE standardisations undertaken in DW04 and AW01, and the manner in which these were undertaken. I have made some brief points in my appended notes against each of the working papers/reference documents (Appendix 3).

6. Consider the research recommendations provided by the Assessment workshop and make any additional recommendations or prioritizations warranted.

The research recommendations provided by the Assessment Workshop are:

- (top priority) Undertake studies on basic life history (e.g. age-growth relationships, length/age at maturity).
    - Agree, in so much as these should reduce existing uncertainty – but are these realistic given the existing capacity? Why are not previous studies considered representative? Are there existing studies for the same species elsewhere that may be helpful?
- Review the current TIP sampling structure to ensure sampling is representative.
    - Agree - but “representative” in what sense? Temporally, spatially, of the size structure of the total fished population, of the total fishing effort?
    - I think this should rate as a higher priority than undertaking fishery-independent surveys. The priorities should be immediately focused on improving the input to, and outcomes, the existing assessment approach.

- (top priority) undertake fishery-independent surveys that enable the development of abundance indices, and that collect age, length, weight and reproductive data.
    - Fair enough – but again, are these realistic given the existing capacity?
    - Moreover, this recommendation should be made in the context of the evaluations of the existing fishery independent data and/or survey protocols (e.g. the Mona Island and Abrir la Sierra (DW03) protocols and data had potential had the time series been longer).
- To expand the mean-length estimator to accommodate other selectivity patterns.
    - I think this is an excellent recommendation.
- To quantify the selectivity patterns for the different gear types.
    - I agree that this needs to be resolved, especially given the assumption of knife-edged selectivity underpinning the per-recruit analyses.
- To attempt to quantify discard and discard mortality rates.
    - Agree that this would be useful, but how could this be achieved? Quantifying discarding is notoriously difficult.
- To continue to improve the data collection of fishery-dependent catch and effort statistics so that traditional biomass-based assessment approaches can be employed (and hence annual catch limits determined and monitored).
    - I agree that this is a key priority.
    - However, there is presumably no way to improve the quality of the historical catch and effort statistics, so the issue is also one of how best to work with the existing data.
    - While it may be ideal to develop ACLs from estimates of abundance and sustainable yield, these are often unavailable. This does not preclude ACLs from being set. ACLs may be determined using simple empirical approaches, while acknowledging the increased risk associated with less information and certainty.

- **Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments**.

There are two arenas for research and monitoring. One is around improving the reliability and usefulness of the current assessment approach. This equates to narrowing the range of mortality estimates and so increasing the certainty around the probability of overfishing (and to introducing an overfished reference point threshold). The other is around improving the quality of information available into the future such that alternative approaches may be permitted.

Against the current assessment approach, I agree that

- Reviewing the TIP length data for representativeness (temporally, spatially, of the size structure of the total fished population, of the total fishing effort) is important. This should include a careful consideration of the TIP length data in the context of the other available length frequency information.

- Whether by improving the understanding of life history, and/or by reviewing the available information and parameters and weighting or narrowing these to a more plausible subset, working to reduce the range of mortality estimates is also important.

In addition, monitoring and research to resolve uncertainties around:

- Stock structure in the context of the three regions considered (whether by genetic testing (expensive), tagging studies (expensive), or considering spawning migration (per RD09) and larval transport data (per RD06), and/or studies of dispersal, mixing and stock structure from similar species elsewhere) (i.e. are we treating the three regions as three separate stocks, to which different stock statuses and hence difference management apply?) and
- Gear/fleet reconciliation within regions (i.e. what is the extent of overlap of the fishable sizes targeted/captured by the different gear/fleet types? To what extent can the data from each be combined?)

should be prioritized, so that the probabilities of overfishing are useful in a management context. Currently there are six sets of probabilities (3 regions x 2 overfishing threshold reference points), with each gear/fleet considered to contribute equally to the frequency distributions used to determine the probabilities. Ideally, there should be a recommended preferred reference point (that which is more precautionary, in the absence of other information), a better quantitative articulation of the extent to which the information from each gear/fleet contribute to the overall frequencies on which the probabilities of overfishing are based, and an increased confidence of how these should be applied in the context of what is understood about stock structure.

Additional work to determine:

- A suitable reference point corresponding to an overfished stock status
- A target reference point that could underpin management decision/control rules

is also recommended.

Against improving the quality of information available into the future, such that alternative approaches may be permitted

- There is no clear indication given as to whether formal logbook reporting is possible (or exists), but it seems that the best means of obtaining uniform catch and effort data on which alternative assessments may be based (e.g. simple production models).
- A reconciliation of the fishery independent approaches should be undertaken with a view to recommending a monitoring program that will optimize the utility of the information obtained. (To what extent would the protocol described in RD01 be sufficient?) Recommendations should be pragmatic given the available resources and capacity.

Meanwhile, I encourage

- Avoiding tossing out data for the sake of being overly Puritan. Even if time series of CPUE, for example, are not considered of adequate quality to enable a formal stock assessment, the data may be useful in informing simpler, more empirical assessments (e.g. Froese 2004; Dowling et al. 2008; Prince et al. 2012; Edwards et al. 2012; Erisman et al. 2014). At the very least, they give some notion of historical high catches, and size-based catch rates.

- At least attempting to fit a production model to the two sets of standardized CPUE (AW01; DW04). Even if there proves to be inadequate contrast in the data, the attempt to use it in an assessment would still place emphasis on what is needed from future data collection protocols.

Finally, I strongly encourage that more effort be dedicated to considering approaches to developing ACLs. The Assessment Workshop avoided developing ACLs because the mean length estimator "does not provide these metrics" and "in an ideal scenario, ACLs would be developed from estimates of abundance and sustainable yield". However, the yield-per-recruit/spawner-per-recruit analyses provide F_{MSY} target reference point proxies that could be used in determining an ACL via simulated projections. Second, fisheries arguably do not require "traditional biomass based assessment approaches" in order to set ACLs. Catch time series, triggers (as reference point proxies) or reference points, and empirical decision/harvest control rules can all be used to set an interim ACL (e.g. Dowling et al. 2008; Prince et al. 2012; Dowling et al. 2014).

- **Provide recommendations on possible ways to improve the SEDAR process.**

On the basis of the documents provided, I have the following recommendations:

- Link the Data Workshop Report more closely to the Assessment Workshop Report, so that
    - Data are consistently summarized between each report, preferably via a commonly presented summary table
    - There is improved clarity on how and whether data are used in the assessment. There is minimal detail regarding data in the Assessment Workshop Report. It would have been useful had the Data Workshop Report indicated whether and how each type/set of available data was used in the assessment, both as a summary sentence at the time of its presentation, and in an overall data summary table.

- Prior to circulating for review, cross-check reports to ensure that report reference lists are complete and that key papers are included as background reading (or at least links provided). Perhaps allowing slightly more time for completion of reports may assist with this.

- As a required part of the Assessment Workshop Report, provide historical context and past precedence for assessments previously undertaken. This was not provided in the current reports. If not previous precedent exists, this should be explicitly stated.

7. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

As stated above, in terms of the available data provided to the assessment workshop via the Data Workshop Report:

- The data were presented in such a way that it was difficult to navigate. A clear summary table of available data including their attributes, whether fishery dependent/independent, available time series, associated references, and how/whether used in the assessment, would have helped enormously.

In terms of the presentation of data/information in the Assessment Workshop Report:

- A clearer presentation of input data, as per the above-suggested summary table, would have been highly useful.
- Sensitivity combinations should be presented clearly in a table format that is easy to follow, and such that the number of crosses in explicitly stated.

Regarding the efficacy of the modelling approaches, and as stated previously, I recommend considering:

- Establishing a reference point threshold corresponding to an overfished state.
- Establishing a target reference point such that a ACL can be set using the current assessment approach, via empirical decision/control rules (e.g. slope-to-target) (Prince et al. 2012; Dowling et al. 2014).
- Undertaking a simulation-based MSE, to facilitate projections and undertake a risk analysis by applying decision rules according to the probabilities of overfishing, and thus determine the relative importance of resolving the range of mortalities contributing to these probabilities.
- Reconciling length-based methods with production models or other assessment approaches, either via a review of approaches, attempting to fit a production model to existing CPUE time series, and/or or better justifying the current approach. There is no indication as to whether assessments have been undertaken previously for this species, and if so, whether there is a precedent for the use of the mean-length estimator. While the mean-length estimator approach used appears sound, this is used as an excuse to avoid projections (when these could have been undertaken using simulations), the setting of ACLs and of target and overfished reference points.
- The application of results in a management context – how useful and informative are the outcomes? What are possible decision rules that could applied?

CONCLUSIONS AND RECOMMENDATIONS

The Caribbean Red Hind Assessment Workshop applied a simple mean-length estimator to the most temporally consistent source of length data to obtain estimates of total mortality. Fishing mortalities were obtained by subtracting estimated natural mortality, and compared to overfishing reference points derived from standard spawner-biomass-per-recruit estimates. There were two Fmsy proxy reference points, and therefore six separate probabilities of overfishing (two for each of three regions) were calculated.

The mean-length estimator is a current acknowledged approach for estimating total mortality from length distribution data, and does not assume equilibrium population dynamics. It appears to have been properly configured and used consistent with standard practices (e.g. Edwards et al. 2012; Erisman et al. 2014).

The following should be noted, however:

- The assumptions of constant recruitment and mortality associated with the mean length estimator, are not given consideration. The history of the fishery should be acknowledged in this context: there have been many management changes pertaining to spatial and seasonal closures, and to gear. There have also been temporal changes in the length-frequency distribution/mean length for several gear/area strata. Whether these have corresponded to management changes has not been investigated. However, given the management changes and the changes in length, some consideration should have been given regarding the extent to which recruitment and mortality could be assumed to have been constant.
- Knife-edged vulnerability was assumed for the per-recruit analyses– but is this suggested by the data?
- The definition of size-at-recruitment to the fishery, Lc, was vague.
- Very little detail was provided on the derivation of natural mortality estimates.
- Sources of uncertainty in the assessment methods were not investigated quantitatively. The "Research Recommendations" section states that "the ability to use the mean-length estimator is contingent upon having length-frequency data that are temporally consistent and representative of the population…". Temporal inconsistencies in terms of sampling size were mentioned throughout the report, together with their possible impact on length distributions, but there was no direct suggestion as to how to reconcile these.
- It is mentioned that the size of fish in the USVI may be market driven for plate size, suggesting that the selectivity may be dome shaped, which is in violation of the assumption of knife-edged selectivity in the mean-length estimator. I agree that expanding the mean-length estimator to accommodate other selectivity patterns should be undertaken.

Beyond undertaking a suite of sensitivity scenarios, and providing probabilities of overfishing, the report did not critically review or evaluate its approach, or consider its findings in a managerial context (as per Assessment Workshop TOR 7).

The technical conclusions are three sets of probabilities of overfishing. These have been presented with no implications of uncertainty, or (with the exception of recommending studies on basic life history) recommendations regarding narrowing the range of mortality estimates and so increasing the certainty around the probability of overfishing. Specifically, the implications of:

- The breadth of the ranges of Z (and hence the uncertainty in stock status) in terms of the efficacy of the overfishing probabilities in a management context,
- Having separate sets of probabilities for each of the three regions, without specifying what is assumed about stock structure,
- The equal consideration/contribution of all sensitivity runs in calculating the probabilities of overfishing,
- The lack of recommendation regarding the preferred choice of overfishing reference point

are not explicitly considered.

Against the current assessment approach, the outcomes for each gear/fleet should be reconciled within regions (i.e., what is the extent of overlap of the fishable sizes targeted/captured by the different gear/fleet types? To what extent can the data from each be combined?), so that the probabilities of overfishing are useful in a management context. Currently each gear/fleet is considered to contribute equally to the frequency distributions used to determine the probabilities. Ideally, there should be a better quantitative articulation of the extent to which the information from each gear/fleet contribute to the overall frequencies on which the probabilities of overfishing are based. Additionally, the assessment results showed variation in Z estimates and temporal trends between gear/fleet types at the same location. The obvious response would be to consider whether this is due to differences in selectivity, whereby the different fleets/gears are targeting different sizes of the stock.

It appears that the Workshop participants undertook the assessment on the basis that the approach was appropriate for the best available data source, and simply reported on this and its outcomes. While the assessment approach may be the best scientific information available, the report provides little more than the technical details and a brief consideration of the outcomes.

My remaining concerns with the Assessment Workshop Report are summarized (and reiterated from the above section) as follows, with recommendations as appropriate:

- General presentation
    - The presentation of available data across the three sections of the SEDAR-35 report was difficult to navigate, both within and between reports. A simple summary table of the available information (source, location, type of data, fishery in/dependent, time series, and indicating what was provided to the Assessment Group, and what was actually used) allowing for ready comparison would have been highly useful.
    - The Data Workshop Report should be linked more closely to the Assessment Workshop Report, so that there is improved clarity on how and whether data are used in the assessment.

  - Prior to circulating for review, reports should be cross-checked to ensure that report reference lists are complete and that key papers are included as background reading (or at least links provided). While citations were made in the text, there were generally no full references given for key papers pertaining to the assessment methods.
- The report lacks context in terms of data decisions and historical precedence regarding previous assessments.
  - The Methods section of the Assessment Workshop report jumps straight from data evaluations to a description of the modelling approach. More context would have been appreciated in order to have provided greater justification for the chosen approach.
  - As a required part of the Assessment Workshop Report, historical context and assessments previously undertaken should be explicitly described.
  - Generally, the data decisions do appear sounds and robust, but, particularly where data have been excluded, these could have been better justified. I also recommend not being hasty to discard data for the sake of being overly Puritan. Even if time series of CPUE, for example, are not considered of adequate quality to enable a formal stock assessment, the data may be useful in informing simpler, more empirical assessments (e.g. Froese 2004; Dowling et al. 2008; Prince et al. 2012; Edwards et al. 2012; Erisman et al. 2014). At the very least, they give some notion of historical high catches, and size-based catch rates.
  - With the exception of the direct comparison of TIP length data and SEAMAP-C length data, there was little effort made to reconcile i) the TIP data across the different locations and gear/fleet types, let alone across ii) the various sources of length data, or iii) the other available data. As a minimum, some qualitative consideration of how the TIP estimates of mean length and Lc vary between gear types and location, and what this may mean for the interpretation of stock status, would have been highly useful.
  - Alternatively, a clear justification of why stock status estimates will differ according to gear type and location, and why this is acceptable in the context of using these for management purposes, could have been provided.
  - Reviewing the TIP length data for representativeness (temporally, spatially, of the size structure of the total fished population, of the total fishing effort) is important. This should include a careful consideration of the TIP length data in the context of the other available length frequency information.
- There was no attempt to define what constituted a stock for purposes of the assessment, nor were any assumptions specified regarding stock structure.
  - Are we to treat the three regions as three separate stocks, to which different stock statuses and hence difference management apply?
  - At a minimum, a consideration of the area covered by the analyses, versus the potential of the species for movement and mixing (per RD09), would have

been useful. A priority for monitoring and research should be to resolve the issues of stock structure across the three regions considered (whether by genetic testing (expensive), tagging studies (expensive), or considering spawning migration (per RD09) and larval transport data (per RD06), and/or studies of dispersal, mixing and stock structure from similar species elsewhere).

  - The lack of specification of stock structure assumptions brings into question the reliability of spawner-per-recruit based reference points. If the fish captured by the various gear/fleets are generally above the size of maturity, then presumably they collective contribute to the spawner biomass. Yet separate spawner-per-recruit estimates are presented for each gear/fleet combination. Better justification for this needs to be provided.

- Both within and between data types, and for assessment outcomes, there was little attempt to reconcile the data trends and/or inferences between gear/effort types, location and sector (i.e. commercial, recreational, survey).
  - It would have been highly useful to have seen the extent to which trends were consistent across these categories.
  - At a minimum, it would have been useful to have seen time series of mean length presented for all available length information, as this is what was used in the assessment. Ideally, such comparisons should have been made for the available catch, CPUE and survey abundance information, also. As it stands, it is very difficult to grasp the overall picture: are things different in a different area and/or under a different gear/survey protocol? Or a different sector (commercial/recreational/independent survey)? Or for a different data set in the same area?
  - Aside from a cursory attempt to reconcile the Fmsy-based probabilities of overfishing with recent trends in CPUE, there was no attempt made to compare the different types of data, to ascertain whether they were reinforcing or apparently contradicting each other in terms of what they suggested regarding the status of the stock.
- No alternative assessment approaches were reviewed or considered. Erisman et al. (2014) provide various options for assessing stocks (for species forming spawning aggregations) using simple metrics based on catch length composition.
  - Other methods that could have been used to evaluate stock status include
    - Assessment from Froese's (2004) size-based indicators of sustainability
    - Spawning potential ratio (SPR) analyses
    - (possibly Sustainability Assessment for Fishing Effects (SAFE)(Zhou *et al.* 2009)
    - (possibly) production models
  - I recommend attempting to reconcile length-based methods with production models or other assessment approaches, either via a review of approaches,

attempting to fit a production model to existing CPUE time series, and/or or better justifying the current approach. There is no indication as to whether assessments have been undertaken previously for this species, and if so, whether there is a precedent for the use of the mean-length estimator. While the mean-length estimator approach used appears sound, this is used as an excuse to avoid projections (when these could have been undertaken using simulations) and the setting of ACLs and of target and overfished reference points.

  - Even acknowledging the supposed problems with the two standardized CPUE indices (AW01; DW04), and the St Thomas spawning aggregation density time series, it still would have been of interest to have at least attempted to have fitted a simple production model to these. This is of especial relevance given that the mean length methods used do not enable projections of stock dynamics. Even if the abundance indices are unreliable and/or there was inadequate contrast to enable a production model to be fitted, the attempt to have done so would have flagged more strongly the potential value of collecting robust CPUE and/or density information.
  - Against improving the quality of information available into the future, such that alternative approaches may be permitted:
    - There is no clear indication given as to whether formal logbook reporting is possible (or exists), but it seems that the best means of obtaining uniform catch and effort data on which alternative assessments may be based (e.g. simple production models);
    - A reconciliation of the fishery independent approaches should be undertaken with a view to recommending a monitoring program that will optimize the utility of the information obtained. (To what extent would the protocol described in RD01 be sufficient?) Recommendations should be pragmatic given the available resources and capacity.

- There was no estimated reference point corresponding to a threshold for an overfished state of the stock (however "stock" is defined), nor justification for the lack of this.
  - Limit reference points often correspond to 20% of unfished biomass, B_0, so a crude approximation could be made that the probability of being overfished equates to the probability that $F_{cur} > F_{SPR20\%}$.
- The probabilities of overfishing were determined from frequency distributions that embraced a large range of possible fishing mortalities. Whether by improving the understanding of life history, and/or by reviewing the available information and parameters, weighting or narrowing the range of mortality estimates and so increasing the certainty around the probability of overfishing should be a key priority for future assessments.
  - Short of recommending studies on basic life history, there were no suggestions of how to narrow the range of estimated mortalities, or what subset of the range was considered most plausible.

  - Point evaluations of overfishing/no overfishing for each of the reference case scenarios would have been useful.
  - Are there combinations of input values for Lc, von Bertalanffy parameters, natural mortality and length-weight parameters that are more biologically plausible/consistent than others? The combinations that are weighted more highly could perhaps have a greater relative representation in the frequency distributions that are used to determine the probabilities of overfishing. Alternatively, those combinations that are considered less plausible could be omitted from the pool of values.
- The same concerns regarding the large range of values of Z apply to the large range in the estimates of $F_{30\%}$ and $F_{40\%}$ (per Figures 39-41):
  - It is difficult to know which (if any) among these are the more plausible.
  - In the "Modelling Approach" section of the report, the sentence "The probability of overfishing integrated across all modelled sources of uncertainty was then determined" is vague and does not specify precisely how this integration was performed. I assume all sensitivity combinations contributed to the frequency distributions presented in Figures 42-44.
  - There should be a recommended preferred reference point (F_{msy} proxy) (that which is more precautionary, in the absence of other information).
- The report concluded that it was unlikely that overfishing was occurring for Caribbean red hind in Puerto Rico, despite probabilities of 25% and 40%. This conclusion is highly subjective and lacks direct justification. It was apparently based upon the fact that a proxy abundance index time series (DW04) supposedly contradicts the mean length results in suggesting an increase in F for the vertical line fleet, and upon the fact that the results from mean length analyses suggested conflicting dynamics between the fleets, in terms of temporal trends in predicted Z. Neither of these justify a "no overfishing" conclusion.
  - Indeed, it was the diving sector that suggested Z was constant or increasing with time, and it is this sector that appears to target larger fish for Puerto Rico. An increase in Z on a larger (more fecund) sector of the population is cause for concern.
- The probabilities of overfishing according to the $F_{30\%}$ and $F_{40\%}$ proxies for F_{msy}, are, I believe, high enough to be of concern that overfishing is indeed occurring, particularly for St Thomas/St John and St Croix, where the lowest probability is 42%. There is no discussion of a minimum reference probability above which some decision rule is invoked or further investigation is required, but even the lowest probability of overfishing of 25% for Puerto Rico should warrant some attention. This is regardless of the temporal trends in Z predicted by the mean length estimators. While low sample sizes were a problem, particularly for the USVI regions, this should be even more reason to respond in a precautionary manner, to probabilities of overfishing that are already high.

- The interpretation of the results relative to the other sources of available input data (as per the Assessment Workshop Report's Discussion section) is questionable in the following ways:
    - For Puerto Rico, the estimated temporal changes in mortality for the vertical line fleet were compared to the CPUE standardization undertaken using SEAMAP-C data (DW04) for the same fleet. The latter suggests abundance is declining, and this is assumed to equate to an increase in fishing mortality. This may not be the case: fishing mortality could in fact be declining in response to low abundance. Moreover, the conclusion cannot be drawn that "This is contradictory to the mean length estimator result for the vertical line fleet, which suggests mortality declined". It does not follow that low abundance equals high fishing mortality in the same time period. There is typically a lag between fishing mortality and the response in terms of abundance. Low abundance indices can follow a period of high fishing mortality, and in the time during which abundance is low, fishing mortality may subsequently ease in response to this, such that the population then recovers.
    - It is stated that the fishery dependent relative abundance indices for Puerto Rico (per AW01) were "flat, suggesting that abundance has not changed". I disagree; I believe the mean trend appears flat due to the scales of the y-axes that are accommodating the confidence intervals.
    - For St Thomas, it is concluded that, because reported landings and effort have declined since 2008, this equates to a decline in fishing mortality, and an eventual increase in fish size. This is interpreted to be contradictory to the result from the mean length estimator, which suggests that mortality has increased due to a reduction in mean length. This supposed contradiction is shaky: there is no consideration of time frame for recovery of mean length. Moreover, there is a temporal mismatch in the comparisons: the modelled decline in mean length is based on a time series from ~1983 to ~2012, during which mean length declined from the mid-1980s, but remained relatively constant since ~1995. Thus the modelled increase in mortality corresponds to a time long prior to the declines in landings and effort in 2008
- It was stated that the mean length estimator approach does not include projections of stock dynamics, and that these were therefore not conducted.
    - There exists adequate life history and catch information that a simulation-based management strategy evaluation (MSE) model could have been developed, incorporating a length-based assessment. A simulation-based MSE could be developed, to facilitate projections and undertake a risk analysis by applying decision rules according to the probabilities of overfishing, and thus determine the relative importance of resolving the range of mortalities contributing to these probabilities. This could be a simple model, but it would provide a platform for undertaking projections (thereby inferring future conditions), and for establishing and evaluating overfished and target reference points and ACLs. It could also be used to

select an appropriate $F_{x\%}$: there is no attempt made to recommend the use of $F_{30\%}$ over $F_{40\%}$, or vice-versa.

- The Assessment Workshop argues that the mean length estimator approach precludes the setting of ACLs. However, direct estimates of abundance and sustainable yield are not required to set a recommended ACL.
    - I strongly encourage that more effort be dedicated to considering approaches to developing ACLs. The Assessment Workshop avoided developing ACLs because the mean length estimator "does not provide these metrics", and "in an ideal scenario, ACLs would be developed from estimates of abundance and sustainable yield". However, the yield-per-recruit/spawner-per-recruit analyses provide F_{MSY} target reference point proxies that could be used in determining an ACL via simulated projections. Fisheries arguably do not require "traditional biomass based assessment approaches" in order to set ACLs. Catch time series, triggers (as reference point proxies) or reference points, and empirical decision/harvest control rules can all be used to set an interim ACL (e.g. Dowling et al. 2008; Prince et al. 2012; Dowling et al. 2014).
    - Methods that could have been used to set ACLs include, but are not limited to:
        - Froese's (2004) size-based sustainability indicators
        - Simple, empirical catch/CPUE- time series-based regression methods
        - and/or traffic light, CUMSUM control indices, or hierarchical decision trees.

REFERENCES

Dichmont, C., Brown, I. 2010. A case study in successful management of a data-poor fishery using simple decision rules: the Queensland Spanner Crab Fishery. Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science 2, 1–13.

Dowling, N.A., Smith, D.C., Knuckey, I., Smith, A.D.M., Domaschenz, P., Patterson, H.M., and Whitelaw, W. 2008. Developing harvest strategies for low-value and data-poor fisheries: Case studies from three Australian fisheries. Fisheries Research 94: 380-390.

Dowling, N., 2011. Management Strategy Evaluation testing of the Management Strategies used with North West Slope Trawl Fisheries. CSIRO Marine and Atmospheric Research, Hobart. 86 p.

Dowling, N.A., Dichmont, C.M., Haddon, M., Smith, D.C., Smith, A.D.M. and Sainsbury, K. 2014. Guidelines for developing formal harvest strategies for data-poor species and fisheries. Fisheries Research in review.

Edwards, C.T.T., Hillary, R.M., Levontin, P., Blanchard, J.L. and Lorenzen, P. (2012) Fisheries Assessment and Management: A Synthesis of Common Approaches with Special Reference to Deepwater and Data-Poor Stocks, Reviews in Fisheries Science 20:3, 136-153, DOI: 10.1080/10641262.2012.683210

Erisman, B.E., Apel, A.M., MacCall, A.D., Romna, M.J., Fujita, R. 2014. The influence of gear selectivity and spawning behaviour on a data-poor assessment of a spawning aggregation fishery. Fisheries Research 159: 75-89 DOI: 10.1016/j.fishres.2014.05.013

Froese R. 2004. Keep it simple: three indicators to deal with overfishing. Fish and Fisheries 5: 86–91.

Haddon, M., 2011. Management Strategy Evaluation testing of the Management Strategies used with South-Eastern Scallop Fisheries. CSIRO, Marine and Atmospheric Research, Hobart. 98 p.

Kenchington, T. 2013. Natural mortality estimators for information-limited fisheries. Fish., htpp://dx.doi.org/10.1111/faf.12027

Klaer, N., Wayte, S., 2011. Demersal MSE for trawl fish in the Southern and Eastern Scalefish and Shark Fishery and other like-species. CSIRO Marine and Atmospheric Research, Hobart. 67 p.

Plaganyi, E.E., Skewes, T.D., Dowling, N.A., Haddon, M., 2013. Risk management tools for sustainable fisheries management under changing climate: a sea cucumber example. Climatic Change 119(1), 181 – 197, 10.1007/s10584-012-0596-0)

Prince, J.D., Dowling, N.A., Davies, C.R., Campbell, R.A. and Kolody, D.S. 2011. A simple cost-effective and scale-less empirical approach to harvest strategies. ICES Journal of Marine Science 68: 947-960

Zhou, S.J., Griffiths, S.P., and Miller, M. 2009. Sustainability assessment for fishing effects (SAFE) on highly diverse and data-limited fish bycatch in a tropical prawn trawl fishery. Marine and Freshwater Research 60: 563-570.

APPENDIX 1: BIBLIOGRAPHY OF MATERIALS PROVIDED FOR REVIEW

SEDAR 35

Caribbean Red Hind

Workshop Document List

Document #	Title	Authors	Date Submitted
Documents Prepared for the Data Workshop			
SEDAR35-DW-01	Monitoring of Commercially Exploited Fisheries Resources in Puerto Rico	Aida Rosario Jimenez	20 Sept 2013
SEDAR35-DW-02	Reef Fish Monitoring	Aida Rosario Jiménez, Verónica Seda Matos, and Noemí Peña-Alvarado	20 Sept 2013
SEDAR35-DW-03	Red hind data from Puerto Rico	Michelle Scharer, Michael Nemeth and Daniel Matos	3 March 2014
SEDAR35-DW-04	Abundance Indices of Red Hind Collected in Caribbean SEAMAP Surveys from Southwest Puerto Rico	G. Walter Ingram, Jr.	13 May 2014
Documents Prepared for the Assessment Process			
SEDAR35-AW-01	Standardized Catch Rates for Red Hind from the Commercial Diving, Trap, and Vertical Line Fisheries in Puerto Rico	Adyan Rios	8 August 2014
Final Stock Assessment Reports			
SEDAR35-SAR1	Caribbean Red Hind		
Reference Documents			
SEDAR35-RD01	A Cooperative Multiagency Reef Fish Monitoring Protocol for the U.S. Virgin	David R. Bryan, Andrea J. Atkinson, Jerald S. Ault, Marilyn E. Brandt,	

	Islands Coral Reef Ecosystem, v. 1.00	James A. Bohnsack, Michael W. Feeley, Matt E. Patterson, Ben I. Ruttenberg, Steven G. Smith, Brian D. Witcher
SEDAR35-RD02	Fishery independent survey of commercially exploited fish and shellfish populations from mesophotic reefs within the Puerto Rican EEZ	Jorge R. García-Sais, Jorge Sabater-Clavell, Rene Esteves, Milton Carlo
SEDAR35-RD03	Portrait of the commercial fishery of red hind, *Epinephelus guttatus*, in Puerto Rico during 1992-1999	Daniel Matos-Caraballo
SEDAR35-RD04	Portrait of the commercial fishery of red hind, *Epinephelus guttatus*, in Puerto Rico during 1988-2001	Daniel Matos-Caraballo, Milagros Cartagena-Haddock, and Noemi Pena-Alvarado
SEDAR35-RD05	Evaluation of seasonal closures of red hind, *Epinephelus guttatus* (Pisces: Serranidae), spawning aggregations to fishing off the west coast of Puerto Rico, using fishery-dependent and independent time series data	Anthony Robert Marshak
SEDAR35-RD06	Description of larval development of the red hind *Epinephelus guttatus*, and the spatio-temporal distributions of ichthyoplankton during a red hind spawning aggregations off La Parguera, Puerto Rico	Edgardo Ojeda Serrano
SEDAR35-RD07	Brief Summary of SEAMAP Data Collected in the Caribbean Sea from 1975 to 2002	G. Walter Ingram, Jr.
SEDAR35-RD08	Population characteristics of a recovering US Virgin Islands red hind spawning aggregation following protection	Richard S. Nemeth
SEDAR35-RD09	Spatial and temporal patterns of movement and migration at spawning aggregations of red hind, *Epinephelus guttatus*, in the U.S. Virgin Islands	Richard S. Nemeth, Jeremiah Blondeau, Steve Herzlieb, and Elizabeth Kadison

APPENDIX 2: COPY OF THE CIE STATEMENT OF WORK

Attachment A: Statement of Work for Dr. Natalie Dowling

External Independent Peer Review by the Center for Independent Experts

SEDAR 35 Caribbean Red Hind Assessment Desk Review

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: SEDAR 35 will be a compilation of data, benchmark assessments of the stocks, and an assessment review conducted for Caribbean red hind. The review is responsible for ensuring that the best possible assessment is provided through the SEDAR process and will provide guidance to the SEFSC to aid in their review and determination of best available science, and when determining if the assessment is useful for management. The stocks assessed through SEDAR 35 are within the jurisdiction of the Caribbean Fishery Management Council and the territorial waters of Puerto Rico and the U.S. Virgin Islands. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of The CIE reviewers shall have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the tasks of the scientific peer-review described herein. Experience with data-limited assessment methods is desirable. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

Statement of Tasks: Each CIE reviewer shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, and other pertinent information. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer revie

Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs cannot be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.

1) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).
2) No later than September 12, 2014, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

4 August 2014	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact

18 August 2014	NMFS Project Contact sends the CIE Reviewers the report and background documents
25 August through 12 September 2014	Each reviewer conducts an independent peer review as a desk review
12 September 2014	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
26 September 2014	CIE submits the CIE independent peer review reports to the COTR
30 September 2014	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov)

Support Personnel:

Allen Shimada
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Allen Shimada@noaa.gov Phone: 301-427-8174

William Michaels
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
mshivlani@ntvifederal.com Phone: 305-968-7136

Key Personnel:

NMFS Project Contact:

Julie A. Neer, SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
Julie.Neer@safmc.net Phone: 843-571-4366

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review

   Appendix 2: A copy of the CIE Statement of Work

Annex 2: Terms of Reference for the Peer Review

SEDAR 35 Caribbean Red Hind Assessment Desk Review

1. Evaluate the data used in the assessment, addressing the following:
   e) Are data decisions made by the Assessment Workshop sound and robust?
   f) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   g) Are data applied properly within the assessment model?
   h) Are input data series reliable and sufficient to support the assessment approach and findings?
2. Evaluate the methods used to assess the stock, taking into account the available data.
   d) Are methods scientifically sound and robust?
   e) Are assessment models configured properly and used consistent with standard practices?
   f) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
   f) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   g) Is the stock overfished? What information helps you reach this conclusion?
   h) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   i) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   j) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. Evaluate the stock projections, addressing the following:
   e) Are the methods consistent with accepted practices and available data?
   f) Are the methods appropriate for the assessment model and outputs?
   g) Are the results informative and robust, and useful to support inferences of probable future conditions?
   h) Are key uncertainties acknowledged, discussed, and reflected in the projection results ?

5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
    - Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods
    - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
6. Consider the research recommendations provided by the Assessment workshop and make any additional recommendations or prioritizations warranted.
    - Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
    - Provide recommendations on possible ways to improve the SEDAR process.
7. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

APPENDIX 3: ADDITIONAL NOTES

Summary table of working papers and reference documents

Document	Author	Location	Timeframe	Nature of data	Notes
DW01	Jiminez	Puerto Rico (S&E Coast)	1987-88	fishery dependent	19.9% of #s sampled = red hind
DW02	Jiminez	Puerto Rico	2009	Fishery dependent	15.6% of weight sampled = red hind Size composition by station varied by area & depth Fisher experience influenced CPUE - Both of these should thus be considered in CPUE standardisation!
DW03	Scharer-Umpierre	Puerto Rico Mona Island Abrir la Sierra (size distn only)	2005 2010 2005-2012:	Fishery independent – diver visual census Passive acoustic monitoring + dive surveys	Transect & roving survey snapshots: low abundance Adult red hind at Mona Island most commonly observed over low coral cover and medium structural relief Effect of reserves wrt density Size distribution from 2 spawning sites 2005 spatial distn & habitat utilisation 2010: temporal changes in abundance at Mona Island 2005-2012: length and density info from uw visual surveys from 2spawning aggregation sites
DW04	Walter-Ingram, Jr	SW Puerto Rico	1991-2011 with gaps 2002-04 inclusive, 2007-09, 2012-	SEAMAP survey Handline CPUE only used	CPUE standardisation by year, spawning season, spawning aggregation area, depth BUT - No interactions - No fisher experience (relevant if data collected by sub-contracting skippers) - Need finer area term

					- RD08 suggest size and density and therefore biomass increase by >60% following permanent closure in 1999 – not accounted for - Effect of depth not removed in standardised index
AW01	Rios	Puerto Rico	1990-2012 (species-specific data only since mid 2001)	Commercial – self-reported fisher logbooks	Separate CPUE standardisation for each gear type by year, coast, season with 2-way interactions - No depth, fisher experience
RD01	Bryan et al	USVI	2010 snapshot	Reef belt visual surveys	Cooperative multiagency reef fish monitoring protocol for USVI - To what data does this link? Referred to as "StCroix trap study" in assessment report
RD02	Garcia-Sais et al	Puerto Rico (3 sites) -Abris la Sierra - Isla Desecheo - Bajo de Sico	2011-2012	Fishery independent surveys of exploited fish and shellfish	Length-frequency (p 53, 76, 78, 80, 83, 85) Density, population size (p 26, 28, 37, 39, 42, 43) NB page 1SEAMAP C since 1992 – notes narrow spatial and temporal coverage P53 "Red hind presented modal size (30cm) just > size at 1st reproduction (25cm – though see RD04 below))which may be indicative of high fishing pressure on this species" (may also reflect size at which not cryptic/targeting or spawning aggregations; will also depend on growth rate and age – how long does it take to grow that extra 5cm so how many spawnings?) P2 low abundance, migratory behaviour and large home range imply large geographical scale survey required for population stock assessment
RD03	Matos-Carabello	Puerto Rico	1992-1999	Commercial data: Commonwealth of Fisheries Statistics Project (program with Puerto Rico's Department of Natural & Environmental Resources)	CPUE, length-frequency Figure 3 CPUE – why not standardise this? Could input to a production model

RD04	Matos-Carabello	Puerto Rico	1988-2001	Commercial data: Commonwealth of Fisheries Statistics Project (program with Puerto Rico's Department of Natural & Environmental Resources)	P354 cites Sadovy and Figuerola as 215mm min length maturity Before 1987, red hind reported in the grouper category Overfishing symptoms = catch composition changes, decrease in landings, decrease in size l/f distributions 1988-2011 (at 1st glance mode length looks similar (~280-320mm) to that of later years) Figure 1 = decent catch time series 1988-2000 High fishing pressure during this period 1995 closures: increased landings of larger fish 1995-98 relative to 1992-94
RD05	Marshak	Puerto Rico	1988-2006	SEAMAP C And fishery dependent data	Evaluation of 1996 seasonal closures - CPUE initially thought to increase within aggregations - Subsequent increase in effort led to decrease in CPUE - Increase in average length in both data types, but due to limited recruitment - Closure initially effective but changes in fishing strategy overrode recovery SEAMAP C CPUE Figs 5,6,7,8,17,18,21
RD06	Ojeda Serrano	Puerto Rico		Biological	Larval development and spatio-temporal distributions of ichthyoplankton – retention could be local or up to 60km downstream
RD07	Walter-Ingram	Puerto Rico and USVI	1975-2002	2-pager overview of SEAMAP	1975-84: most was longline sampling, but couldn't standardise because of lack of knowledge of LL hooks per set (but could assume common number of hooks per set) Catch rates for each area-gear combination (1991-2002??) – fish trap and handline. Only latter used in CPUE standardisation of DW04

RD08	Nemeth	St Thomas, USVI	1999-2004 (closed seasonally in 1990; permanently in 1999)	Tag and release fishing and fish survey transects	Population response of a spawning aggregation - Should take this closure as a regime shift in GLM - Also tag studies inform movement - Recovery with respect to mean length gives idea of lag between fishing impact and mean length and could therefore help inform assessment in reconciling trend in CPUE indices and mean length trends (Fig 5-7) - Could density estimates be used in some kind of assessment for spawning biomass? Suggests size and density and thus biomass increased by <60% following permanent closure in 1999 – this should somehow be accounted for. Cited page 74 Data Workshop Report as source of l/f info
RD09	Nemeth et al	USVI St Thomas St Croix	1999-2006	Fishery independent scuba surveys, draft fishing and fish traps to determine arrival/departure from spawning sites; then focused density estimates using belt transects. Tag and release to determine migration.	STT 6-33km spawning migrations from area of 500km^2; STC 5-18km from area of 90km^2 Implies there is an important interaction effect between time of year (spawning season) and area (if latter is a spawning area) Striking similarities between STT and STX in timing, movement and migrations re: spawning aggregations Risk of hyperstability Cited page 74 Data Workshop Report as source of l/f info

GENERAL NOTES AGAINST EACH WORKING PAPER/REFERENCE DOCUMENT/REPORT SECTION

DW01:

- Should be stated upfront that this is a fishery dependent monitoring program
- How was relative abundance inferred from catch?
- Line 3: population fluctuations *in availability*
- Last line: was the species composition the same for both coasts?

DW02:

- Background?? How many sampling periods? What fishery/gears? (hook and line – stated later) Define "stations" and other terms

DW03:

- P1: whether data is fishery independent or dependent should be explicitly stated.
- P1 visual estimation of fish size – is this reliable?
- P2 What is "landscape composition"?
- P3 "frequency of occurrence" relative to what? 3.9% - from both belt transects and roving surveys? 25 individuals across 613 belt transects
- P4 this paragraph should lead the Results; 1st sentence is not a sentence
- P4 why weren't data collected at the spawning aggregation site?
- P4 "latter data" = 2010, 2011??
- P4 "Supposed protection" = 1st mention of reserves
- P6 are the belt transects and roving surveys same as those in part 1? Or different?
- P6 "NTZ" – what is this? 1st use
- P6 Methodology – what years?
- P7 "a potential reserve effect" of what?
- P8 why is it important to obtain data yearly? This won't improve any single abundance estimate.
- P9 by "fieldwork" – is this referring to the dive surveys?
- P10 That the sampling times were pooled needs to be in the Methods
- P10 where do the length estimates come from? I thought this section pertained to acoustic monitoring
- P11 p=0.00 – not really, surely?!

- P11 last paragraph: “Data collected from 2005 to 2012” – does this mean acoustic survey data?

DW04:

- P1 why does higher variability in weight data matter?
- Were skippers sub-contracted to undertake cruises? If so, their experience will be an important factor and should be a GLM explanatory variable.
- P1 notation: “c” also used for CPUE
- Equation 3 reduces to c/(1+c) How is this a vector of the presence/absence?
- P2 How large are the MPAs? Should a finer scale area term be used?
- P2 ***** In generating the standardised CPUE abundance indices, the effect of depth should be removed as a confounding variable. Standardised CPUE abundance indices should only be functions of time and area.
- I would have expected an interaction effect between spawning season and aggregation area, especially given RD09’s finding re: spawning migration
- RD08 suggests size and density (and therefore biomass) increases by over 60% on spawning grounds following permanent closure in 1999. This should somehow be accounted for.
- Figure 2 (Std CPUE indices) – some context ito history of fishery – when did it commence? Closures implemented (show with vertical line)

AW01

- Not used in assessment; species-specific data only available after mid-2001
- Are the effort units appropriate for the gear?
- To identify trips that targeted red hind, could have also used Principal Component Analysis or Cluster Analysis and identified those clusters with an appreciable red hind catch composition.
- Additional confounding factors? What is being filtered out from standardised signal? Not much different from nominal.
- DISAGREE that standardised indices show no overall directional trends – this is a function of the scale of the y-axis.
- Comparison by gears?
- One overall index with gear as a factor explanatory variable?
- Why would the GLM with the interaction term between year and coast not converge? (Table 6)
- Caption for Figures 1-3c is vague

SECTION I: Introduction

Up front should be species name and description

What is area of fishery/relevance to the assessment? What is the presumed stock structure? That is, what are we dealing with in terms of the assessment?

A map of the entire area of relevance would have been good.

Some of the management actions should be incorporated in GLMs or at least flagged temporally on plots. There have been many knife-edged regime changes.

P7 Where is the socio-economic information?

NB P8 overfished, overfishing definitions

P10 1993 prohibited juvenile harvest – should check l/f information per DW04 (available since 1991) to evaluate impact

P13 1999 first full closure

P14 spawning season closure

P14-15 management reference points (few comments but nothing noteworthy)

P2-17 HAPCs?? Not in glossary

P2-17 Lists confirmed spawning locations

P2-19 overfishing limit "maximum rate of fishing a stock can withstand while still providing MSY on a continuing basis" – I don't get this. Still needs to provide MSY, but above this cannot provide MSY – by definition! Since MSY is itself a proxy, this is not precautionary.

P22 reducing length of fishing season does not necessarily reduce catch, as effort can be condensed across a shorter season. The decision rule as to when to resume the original season length is not specified.

SECTION II: Data Workshop Report

Main concern is the lack of attempt to compare/reconcile data from different gear sectors and regions – i.e. what is the overall picture? Are things different in a different area and/or under a different gear? Or a different sector (commercial/recreational/independent survey)?

Summary tables indicating what was provided to the Assessment Group, and what was actually used, would have been helpful.

Summary of ALL available length-frequency information – from where, what years – as per Fig 3.5.8 but for the different data sources, would have been helpful.

General – would be good to have summary table of ALL data sources, types and what is used in assessment – latter also flagged at end of each section

P8 What is the assumed stock structure? Single or multiple?

P8 natural mortality 0.18-0.68 is a huge range

P8 avg size female maturity 250mm or 215mm (3years) TL, mean size at sexual transition 380mm TL – how does this compare to mean length? (see also RD02 "red hind presented a modal size just above the size at first reproduction")

P9 ?? stock-recruitment relationship?? ??steepness??

P9 movement/spawning migrations as per RD09 (18-33km) plus high site fidelity to a home reef. Upshot for stock structure assumption?

P9 large range in length-weight relationship parameters

PUERTO RICO

P13 Puerto Rico commercial red hind landings from 1986

P13 how is expansion factor determined?

P13-14 Trip Interview Program (TIP) described – identifying Lc (but never specific on exactly how – the very minimum size, or some lower percentile, or what?)

P14 TIP <1500 indiv red hind lengths pa on average

P14 Figure 3.5.6, 3.5.7 show very low numbers

P14 TIP is the increase in mean length significant? Confounded with decline in sample size, but may it also correspond to management measures re: gear/selectivity?

P15 TIP shift in median and mean length for pot and trap, vertical line gears ~1998 suggests some kind of regime shift- fig 3.5.8 p 35. Gear-type specific (pot-trap and vertical line); suggests could be to do with gear controls, also fig 3.5.9 p36

P15 why would mean length be affected during spawning season relative to rest of year? May be more heavily fished in spawning season but in same relative length proportion, especially if Lc> spawning size. And what about spatial closures in spawning areas?

P15 recreational landings estimated how? From intercepts?

P15 both commercial and recreational showed peak in 2005 landings followed by large drop in 2006.

P16 what is a “wave”?

P16 reliability of recreational discard info?

P16 how well do rec fisher participants in phone surveys recall the number of trips they undertook?

P16 how does the recreational data compare with the commercial data? Doesn’t seem to be big changes over time

For me, the recreational fishing sampling is so low as to be negligible, so for purposes of assessment not that useful, especially not l/f data UNLESS this is markedly different from the commercial data. Recreational catch could possibly be included, e.g. in an operating model MSE context.

P26 only 19 of 2786 recreational intecepts had red hind landings.

P31 fig 3.5.4 significant changes in gear over time. Depending on time series used in assessment, beware standardised CPUE if from one gear only (which it is)

ST THOMAS & ST JOHN

P41 St Thomas and St John (previous was Puerto Rico) How does the commercial catch, effort compare to that of Puerto Rico?

P42 how small was the small sample size from 1991-2006?

P42 figures should have a vertical line showing length at maturity

P42 again, how does mean length compare to that in Puerto Rico?

P42 as per Puerto Rico, why should mean length be affected by fishing activity in spawning season?

P43 Should I bother checking the ftp site to look at the St Thomas Mother’s Day Tournament info?

P52 fig 4.5.4 etc. what are the red vertical lines?

Pg 54 fig 4.5.6 difficult to compare vertical line, pot and trap l/f info with different x-axis scales

ST CROIX

P56 St Croix – again, what % of Puerto Rico’s catch does this represent?

P57 fig 5.4.8 should be 5.4.7

P57 influence in early 1980s and 90s on observed mean lengths from vertical line fleet difficult to see

P72 Fishery independent data – how does this compare to the fishery-dependent data. Bottom line: how representative do we believe each data source to be?

P72 Biogeography visual surveys – only used for l/f? abundance? Say so up front.

P72 Menza et al (2008) not in reading materials – how/was this used?

P72 why were Puerto Rico visual surveys undertaken in August?

P72 proportion of red hind relative to what? Number of surveyed sites? Is this an appropriate metric?

P73 visual survey size ditributions based on <10 individuals. Inference re: skewed toward smaller red hind is useless

P73 why October sampling in St Croix compared to August for Puerto Rico?

P73 Size range 2.5-41cm. How binned?

P73 St John & St Thomas visual survey July – different timing again to St Croix and Puerto Rico.

P74 Proportion of red hind in St John/St Thomas visual surveys much higher than other 2 regions.

Summary table across all 3 regions and sub-sites would have been good.

P74 DW03 summarised – 2005, and 2005-2012 data made available to assessment analysts

P74 RD08 and RD09 cited here in context of available source of l/f info. BUT I have concerns wrt what these 2 papers show wrt increase in biomass post-closure and re: spawning migrations, and factoring these into GLM standardisations of CPUE.

P75 mentions RD02 (big survey protocol) 2011, 2012 fishery indep visual survey – but total number of observed red hind low. Since only one year data with relatively few observations of red hind, not used.

P75 Why isn't cryptic nature of the species mentioned earlier?!

P75 summaries SEAMAP-C per DW04 (& RD07) – although from both Puerto Rico and USVI, and for multiple gear types, emphasises only handline CPUE (in numbers) off SW Puerto Rico only, used in standardisations.

SECTION III ASSESSMENT REPORT

Commerical landings: Puerto Rico – concern about expansion factors BUT still have a time series of catch

Recreational data Puerto Rico: reporting coverage as % would have been preferable (NB none in USVI)

AW01: fishery dependent CPUE indices (from self-reported logbooks) – only Puerto Rico; species specific data for USVI only available since 2011. Separate indices by gear (diving, trap, vertical line). States shows "wide confidence intervals and no overall directional trend" disagree with latter; artefact of scale due to CIs. "Recommended as a qualitative supplement to the quantitative meanL analysis" – in what way?! Why not attempt a production model while acknowledging limitations?

Fishery-independent data

- Spawning aggregation data
    - DW03 Puerto Rico (3 sets)
        - Short temporal scale – not recommended for use in formal assessment: agree, but note potential of each
    - RD08, RD09 St Thomas, St Croix
    - Tried to develop relative abundance indices for spawning population
        - St Thomas – used annual max density. Resultant time series showed inter-annual variability and again "given caveats highlighted in this section" (?) have used as ancillary qualitative info – why?!? Could have again tried production model. No attempt to reconcile this to CPUE trends.
        - St Croix – too large a gap in time series: agree
- SEAMAP-C
    - Why were relative abundance indices not discussed, or justified why not used to fit production model?
    - Discussed length data and compared these to TIP data. Post 1999 decline in number of annual lengths and shift to larger red hind (NB corresponded with first permanent closures – p13 Section 1 - why not mention this?) Examined changes in sampling methods but not changes affecting commercial fishing.
    - Also, suggests mean depth of survey stations relatively constant over time (Figure 9) but I disagree – decrease then increase
    - Selectivity of SEAMAP-C resulted in smaller modal length than vertical line fleet
    - Did not use SEAMAP-C data for mean length estimator as most data in 1990s and lacking from recent years; could not be combined with TIP data because of differences in selectivity would have violated model assumptions – agree
- St Croix trap study – per RD01 document
    - Says was 2010 pilot trap study but RD01 says visual belt survey - ??
    - 89 red hind captured; range of lengths similar to fishery –dependent length data

Table 6 – best to plot these

Figures 16 is missing its secondary axis

Figure 21: Circles and numbers are what? Does the line equate to the modelled values and the circles to the data, and the numbers the sample size?