**Author(s):** Chrysafi, Anna; Cope, Jason M.; Kuparinen, Anna

**Title:** Eliciting expert knowledge to inform stock status for data-limited stock assessments

**Year:** 2019

**Version:** Accepted version (Final draft)



**Please cite the original version:**

Chrysafi, A., Cope, J. M., & Kuparinen, A. (2019). Eliciting expert knowledge to inform stock status for data-limited stock assessments. Marine Policy, 101, 167-176. https://doi.org/10.1016/j.marpol.2017.11.012

## Abstract

Data-limited fisheries are a major challenge for stock assessment analysts, as many traditional data-rich models cannot be implemented. Approaches based on stock reduction analysis offer simple ways to handle low data availability, but are particularly sensitive to assumptions on relative stock status (i.e., current biomass compared to unperturbed biomass). For the vast majority of data-limited stocks, stock status is unmeasured. The present study presents a method to elicit expert knowledge to inform stock status and a novel, user-friendly on-line application for expert elicitation. Expert opinions are compared to stock status derived from data-rich models. Here, it is evaluated how experts with different levels of experience in stock assessment performed relative to each other and with different qualities of data. Both "true" stock status and expert experience level were identified as significant factors accounting for the error in stock status elicitation. Relative stock status was the main driver of imprecision in the stock status prior (e.g., lower stock status had more imprecision in elicited stock status). Data availability and life-history information were not identified to be significant variables explaining imprecision in elicited stock status. All experts, regardless of their experience level, appeared to be risk neutral in the central tendency of stock status. Given the sensitivity to stock status misspecification for some popular data-limited methods, stock status can be usefully elicited from experts, but expert subjectivity and experience should be taken under consideration when applying those values.

Keywords: expert elicitation; data-limited; stock status; stock-assessment; fisheries management

## 1. Introduction

Concerns for the future of ocean health and the sustainable exploitation of fish stocks are rising and a large portion (estimate to be about 80% in Costello et al. 2012 [1]) of global fish catches still come from unassessed fisheries. Fisheries policies around the globe (e.g., The Law of the Sea [2], the Code of Conduct of the UN [3], the Magnuson-Stevens Act in the US [4], and the EU Common Fisheries Policy [5] all mandate for responsible fisheries and sustainable harvesting via biomass based target and limit reference points. The capacity to estimate biomass via population dynamics modelling has greatly advanced [6,7] and is able to provide scientifically-derived information for management action. The drawback of these models is that they typically require extensive and high-quality/quantity data. Data collection requires time, investment, and appropriate scientific personnel, all in perpetuity, which are typically lacking in developing countries [8,9]. Small-scale fisheries— often multi-species and multi-gear in nature— epitomize the challenges of stock assessment in data-limited situations [10]. Low contribution to Gross Domestic Product (GDP) is a major reason behind the limited quantitative information available for small-scale fisheries [8]. This difference in data-collection investment, directly causes such fisheries to require alternative methods to traditional approaches in quantitative stock assessment [11].

The desire to assess the status and sustainability of any fish stock has resulted in the recent and rapid development of data-limited assessment methods [12,13]. A variety of approaches are now available, some of the simplest of which use only life history information, while others use simple population models and attempt to compensate for the data limitations through model re-parameterization and/or explicit parameter distributions. When data are not available on exploitation history, approaches that use Beverton-Holt invariants and other life-history information [14] are used instead. Collecting fishery length compositions is rather inexpensive and therefore many approaches that use such data have emerged [15,16], and can be combined with life-history information [17,18] in order to assess the status of a fishery. Other approaches that utilize only catch data [19] can give an indication of the status of a fishery but need to be used with caution as catch data do not always reflect abundance [20]. Catch time-series can also be coupled with prior information on important biological and exploitation information in order to predict future yields required in management [21,22]. Additional overviews of the variety of data-limited methods are found in [23,24].

Some of the data-limited approaches are based on modified versions of classical stock reduction analysis (SRA; [25,26]), which require prior information on the relative stock status, expressed in the form of $B_y/B_0$ [27,28,29], where B is biomass (usually spawning biomass), y is year and 0 represents the initial condition. In these methods, the stock status input is specified as a distribution, from which values are resampled. The resampled status and the given catches, selectivity and biology of the stock solve for a particular biomass time-series. Model performance evaluations [12,30] have shown that under- or over-estimation of stock status leads to under- or over- estimation of potential yields respectively. Stock status however, is typically a derived stock assessment output, so usually not available unless a stock assessment has been carried out. Alternative ways of deriving this important stock metric would be a valuable contribution to management. Cope et al. [31] developed a method that estimates a stock status prior using Productivity-Susceptibility Analysis (PSA) vulnerability scores, being the first step towards solutions for defining stock status in data-limited situations. Other possible alternatives for defining stock status, when facing data limitations, include the utilization of information from species with similar exploitation/biology [32].

Another potential approach is using expert opinion. Expert opinions are highly valued in different research fields [33,34,35] and particularly within a Bayesian framework [36]. When it comes to applied questions, not all the elements of required knowledge can be covered by data or meta-analyses. Thus, expert subjective opinions can provide the required information for an unknown parameter in the form of a probability distribution. The use of probability is very appropriate for the representation and quantification of all forms of uncertainty but one should always acknowledge the existence of bias and imprecision in expert elicited probability distributions due to rules of cognitive heuristics [37]. In traditional elicitation approaches, the analyst has the task of eliciting the expert opinions and to interpret them in the way he/she considers appropriate based on the given task. Results can therefore be a mixture of expert views and analyst views [36]. During the elicitation process, experts are provided with a combination of analysts' assistance and sufficient background information around the topic from which their knowledge and experience is utilized to make inference for an unknown quantity [38]. Commonly, expert elicitation is carried out via questionnaires or interviews where the experts provide qualitative answers to the analyst that then he/she translates into probability distributions [38]. It is also possible that experts are asked to draw a probability distribution for the unknown quantity by assigning percentages to different outcomes, stacking probability chips (e.g. 5% probability chip), or sliding bars of a given probability distribution in order to depict their beliefs and construct the probability density [33,36,39].

The aim of this study was to imitate a real case of data-limited stock assessment, where the assessor/expert is challenged with the task of defining stock status based on limited data in order to parameterize a data-limited method. To achieve this, fisheries experts' knowledge on stock status was elicited, by providing them a realistic amount of data (catch time-series and fishery length compositions) that could be available in the case that they would conduct the assessment for a given stock. This paper presents a novel approach for eliciting expert knowledge aimed to be user-friendly, easy to access and use, and can be customized for different scenarios. This work presents how experts with different levels of experience in stock assessment perceived the given data to inform stock status and ground-truth expert opinion to the "best available science" from complex stock assessments. It also explores whether additional data could help experts form a better stock status prior and any possible connection between expert performance and life history type. The ultimate goal was to investigate the extent to which subjectivity, experience in stock assessment and potential interactions between expert performance and life history can be quantified, into a stock status prior, and how or whether these priors may improve current applications of data-limited methods to inform management decisions.

## 2. Material and methods

### Data collection

Stock assessments from the US. Pacific Fishery Management Council (PFMC) and Alaska Fisheries Science Center (AFSC) were used as the basis for data of this analysis, where 18 stocks with analytical assessments using Stock Synthesis (SS; [7]) or similar statistical catch-at-age models that cover a variety of life history types were selected. For the PFMC stocks, SS files were obtained in order to extract the necessary catch and length composition information. For the AFSC stocks, data for fishery length compositions were obtained from the domestic observer program. Catch data were taken from the AFSC assessment reports and if they were unavailable in the report, they were provided by the corresponding assessment authors. Additional information on life history and management of all stocks was extracted from the assessment reports or cited references. If information on maximum observed age was unavailable, the last age group (i.e., the plus group) in the assessment was used instead. When a fishery consisted of multiple fleets, length compositions were combined in order to create a single length composition for the entire fishery in the following manner:

1. Standardize length compositions of each fleet to proportions.
2. Multiply those proportions by the total catch of the fleet in the specific year.
3. Sum across fleets.
4. Turn those numbers into proportions by length bins.

The species selected for the study comprised the families Sebastidae and Pleurocentidae, but varied widely in life history values. Further details about the specific stocks and life history information is provided in Figure 1 and Table 1.

[FIGURE 1 ABOUT HERE]

[TABLE 1 ABOUT HERE]

## Expert group

For the elicitation process, six experts with different levels of experience in stock assessment were selected to quantify bias and imprecision in stock status determination. Experts were separated in three groups of two: i) Level 1 (Experienced) with high experience in stock assessment, being formally employed as stock assessors (Expert 1 and Expert 2), ii) Level 2 (Novice) being under training in stock assessment with limited direct assessment experience (Expert 3 and Expert 4) and iii) Level 3 (Inexperienced) with a background in fisheries ecology (Expert 5 and Expert 6). All experts represent USA west coast academic and federal research institutions and are familiar with the biology of west coast groundfishes. Experts from different backgrounds, e.g. fishers or NGOs, were  not considered eligible, as the aim of this work was to imitate a real case of data-limited stock assessment conducted by fisheries experts without any external input. This way we demonstrate how fisheries expert opinions could be gathered, evaluated, and analyzed under such a scenario. However, obtaining a combined stock status prior based on opinions of various backgrounds, is one way of conducting expert elicitation but not the one considered here. Furthermore, as stock assessments are commonly carried out by a few experts (and most likely only one)we were thus interested in individual expert performance and not cumulative performance. Therefore, the number of experts participating in the elicitation process was small on purpose and should not affect the results but be indicative of individual experts within their expertise level.

## Simulated stocks

In addition to the assessed stocks, two simulated stocks that represent the two main families (Sebastidae and Pleuronectidae) were used for expert calibration purposes. Because expert opinions were compared with model outputs, the two simulated stocks were considered in order to compare expert judgment to known stock status values. The DLM tool developed by Carruthers et al. [12] for management strategy evaluation (MSE) purposes was used as the platform for the simulated stocks. Within the tool, the user can select different options to create a user-specified operational model for the MSE. Here, the built-in Rockfish and Sole stocks, combined with the Generic fleet and Imprecise-Biased Observation model (OM) were used to created simulated data. The catch time-series, fishery length compositions and life history information provided to the experts for the elicitation process were taken from a 50-years simulation with randomly selected parameter combinations. All experts received the same data for both simulated stocks. This duration of random simulation is sufficiently long to develop a range of exploitation patterns similar to industrial fishing in US waters [12].

## Data combinations

### Case-study fisheries

Within a data-limited context, there are many potential data combinations that can be tested, but for the purpose of this work, we selected a sub-set of catch time-series and fishery length compositions. This sub-set was considered appropriate, as it is not uncommon for a data-limited fishery to have limited fishery-based length compositions and/or a reconstructed catch history. For the elicitation process, experts were provided with four different data sets for each stock, constituting different combinations of the selected data types, representing different degrees of data richness. Figure 2 gives examples of these combinations; full data sets for all stocks are provided in the supplementary material (Table S1). Catch sets were either a full or "short" time-series. To construct the short time-series data sets, a rule of 30% of the total years available for each species was selected. Length compositions were defined for the full time-series data sets as the oldest available fishery length compositions (in the cases of multi-fleets the earliest year where length compositions for all fleets was used) and one for the most recent year. The short time-series data sets provided the length compositions from the year midway between the first and final available years, and the most recent year. Furthermore, no additional noise was added to the stock assessment data given it is real data, and thus already sufficiently noisy. In addition, life history (e.g., growth, natural mortality and longevity) and management information were provided for all four data sets. Species names were not given to ensure that experts would judge based on the given data and not on prior knowledge about the specific stocks.

[FIGURE 2 ABOUT HERE]

Sophisticated stock assessment models account for sex-specific differences in life histories, and sex-specific variation in growth is common [40,41,7]. On the other hand, given that data-limited assessment tools are often more biologically simplified, they routinely use only single inputs for life-history information (e.g natural mortality M, maximum length $L_{max}$, von Bertalanffy K, maximum age $A_{max}$) to represent the entire stock across sexes [27,29]. To account for this methods' characteristic, and to make sure that the information that experts can extract from length compositions and life history parameters match each other, only females data were provided with the assumption that they are sufficient to make inference for the entire stock. Given that spawning stock biomass is typically used to describe stock status, this is not a poor assumption. However, there were a few cases where sex-specific length compositions were not available for some fleets, so sex-combined compositions were used. These mix-sex data always came from fleets with low contribution to the total catch and the effect on the final constructed length compositions was negligible.

### Simulated stocks

The same rule of 30% of total available years n catch time-series was also applied here and therefore, the short time-series data sets included catches from the last 15 years of simulation and length compositions from years 40 and 50. For the full time-series data sets, the entire 50 years of exploitation catch time-series and the length compositions of years 30 and 50 were provided to experts, with the assumption that it is unlikely that length compositions are available from the beginning of the exploitation of a stock (Table S1).

### Expert groups

The experts were separated in two groups (A=Expert 1, Expert 3 and Expert 5 and B=Expert 2, Expert 4, Expert 6) and the first group received the simulated rockfish as first species and the simulated sole as last species in the elicitation process, and vice versa for the second group. The 18 stocks were given in the same order. Data sets allowed one to ask a) if and how additional data (i.e., longer time series) would help experts to improve their belief about stock status and b) if during the elicitation process the experts would gain experience and confidence that would lead in improved distribution estimates of stock status.

### Elicitation tool

The "Shiny" R package [42] was used to develop a web-based application for expert elicitation. The two versions of the web application, corresponding to the two groups of experts mentioned above, can be found in https://chrysafi1.shinyapps.io/Beta_shiny_1/ and https://chrysafi1.shinyapps.io/Beta_shiny_2/ respectively. Step-by-step instructions on how to use the application are provided within the tool. Experts were able to select the stock and data set, which produced the respective graphs of catch and length compositions and background information. To construct a stock status prior, the tool provided a beta distribution probability density function (PDF) and scroll bars for the mean and standard deviation that users could freely move to capture the density that best fit their perception of stock status. In addition, a beta cumulative distribution function (CDF) with the median and interquartile range was provided in order to help experts visualize the distribution. The beta distribution is commonly used when describing relative stock status priors [27] and is bounded from 0 to 1, similar to value for expected stock status. The application also included a table, self-populated with the output of the beta distribution values once the expert completed each data set. After all stocks and data sets were completed, the final table of elicited stock status priors could be downloaded as a comma-separated-values (csv) file and submitted. All experts have or were working towards a PhD and basic knowledge in probability theory was assumed, though the visualization tool was developed to make the recommended prior distribution clear. A detailed description of the project and the aims were provided, and experts were given time to look thoroughly through the application to cover any arising questions before conducting the elicitation experiment.

## Performance measures

Relative error was used to quantify individual expert performance in comparison to model- and simulated-based "true" stock values and defined as

$$\mathrm{RE} = \frac{S_e - S_m}{S_m} \qquad (1)$$

Where RE is relative error, $S_e$ is the expert stock status distribution, and $S_m$ is the true stock status for the stock. The median RE was used to describe bias and the interquartile range (IQR) of the RE was used to describe imprecision in expert performance [43].

Linear and linear mixed models were used to identify which explanatory variables best described expert bias and imprecision. RE data where log transformed (log(log(median RE+2)) and log(log(IQR+1)) respectively) to fit the normal distribution. The variables of expert level, data set (being each of the four data combinations), "true" stock status and life history (LH) were treated as categorical explanatory variables (Table 2). LH categorization was derived using silhouette-based cluster validity analysis [44,45] of the M/k ratio for each stock and LH was explored as either a fixed or random effect. Backward stepwise model selection and the Akaike information criterion (AIC) and AIC with a correction for finite sample sizes (AICc) were used for final model selection [46]. The final models were checked for homogeneity and normality of residuals. All analyses were carried out using R [47] and the "lme4" package [48]. Data from all 20 stocks (18 real and 2 simulated) were used for the analysis.

[TABLE 2 ABOUT HERE]

## 3. Results

### Explanatory variables for expert performance

Life history traits were first treated as random effects, but this treatment did not explain any variability in the data, so linear fixed-effects models were used instead for all subsequent exploration. "Data set" and LH were not significant explanatory variables for either median RE nor IQR. For both cases, the linear model with interactions between expert level and "true" stock status as explanatory variables yielded the lowest AICc values (Table 3). A graphical inspection of the transformed median RE and IQR, for all explored variables, verify the results of the linear models (Figure 3). For bias (median RE), "true" stock status (F-test, DF=3, $p<0.0001$) explained 60.5% of the variability and the interactions between true stock status and expert level (F-test, DF=6, $p<0.0001$) explained 2.5% of the variability (adj. $R^2=0.6257$). Expert level (F-test, DF=2, $p<0.0001$) alone explained only 0.3% of the variability in expert bias. For imprecision (IQR), "true" stock status (F-test, DF=3, $p<0.0001$) explained 57.3% of the variability, expert level (F-test, DF=2, $p<0.0001$) explained 8.5% of the variability, and the interactions between true stock status and expert level (F-test, DF=6, $p=0.01434$) explained 1.3% of the variability (adj. $R^2=0.6691$). Figure 4 illustrates the effect of expert level on bias and imprecision from the interactions with "true" stock status. The coefficients and p-values for all levels of main effects and individual interactions are provided in supplementary material (Table S2). Final models for bias and imprecision were further tested for redundancy by collapsing (e.g. merging of Stock Status 2 and 3 into one category) factor levels, but the original treatment of factor levels was most supported by the data.

[TABLE 3 ABOUT HERE]

[FIGURE 3 ABOUT HERE]

[FIGURE 4 ABOUT HERE]

Individual expert performance

For the 18 real stocks, Experienced and Novice experts tended to include "true" stock status more often in their prior distribution, in comparison to Inexperienced experts, as evidenced by the higher IQR range (Figure 5). Expert 2 (Experienced) shows wider RE distributions within the experience level and in comparison to the other two expert experience levels. This difference could be attributed to the fact that Expert 2 overtly recognizes the uncertainty inherent in predicting stock status given the minimal amount of data provided and the lack of information about the fishery over time, including such factors as selectivity. Novice experts are less uncertain in their opinions, but appear more conservative relative to the Experienced experts as they tend to score more negative RE values. Inexperienced experts appear to be more biased, in both directions, and more overconfident in comparison to experts from the other two levels for some of the cases (Figure 5). All experts, regardless of their level of experience, appear to follow a similar pattern in over- or under-estimating stock status depending on the "true" stock status (e.g., overestimate status for low "true" stock status; underestimate for high "true" stock status). Furthermore, expert performance seems to improve over the course of completing the task, as experts appear to gain experience/confidence during the elicitation process.

[FIGURE 5 ABOUT HERE]

The same pattern driven by "true" stock status is also observed for the simulated stocks, as all experts overestimate stock status for the rockfish and underestimate stock status for the sole (Figure 6). Moreover, experts that received the sole as the last species in the elicitation process, performed better in comparison to the experts that received sole as the first species, supporting the improvement of expert performance while completing it. However, the same finding was not observed with the simulated rockfish (that had lower relative stock status), thus complicating the interpretation as to how and when experts may improve their performance. More experts and further exploration of adding simulated species is needed to gain a better understanding of whether experts improve as they work through more species.

[FIGURE 6 ABOUT HERE]

Expert calibration

Lichtenstein et al. [49] explain that an assessor (here being our fisheries expert) is well calibrated "if, over the long run, for all propositions assigned a given probability, the proportion that is true equals the probability that is assigned" meaning that expert probabilities should lie on the diagonal when plotted against the true probabilities for an event. Here, as experts had to give a probability distribution for stock status the median assessed value for stock status was compared with stock "true" status (Figure 7). Moving from more experienced to less experienced in stock assessment experts, the median response becomes more distant from the diagonal line (being perfect expert calibration) and evenly distributed in the space of "true" stock status against assessed stock status. Inspection of the Locally Weighted Scatterplot Smoothing (LOWESS) for each experience level shows clearly that expert calibration shifts to the opposite direction of the perfect calibration diagonal (Figure 7). This effect is especially prominent for stocks with low stock status values (Stock Status 0 and 1) where inexperienced experts appear to assign significantly higher stock status values in comparison to experts from the other levels, as is shown also from the effect of expert level-stock status interaction (Figure 4).

[FIGURE 7 ABOUT HERE]

# 4. Discussion

Expert performance

The present study illustrates that the most important variables for expert bias is "true" stock status and the interactions between "true" stock status and expert level. For stocks with low status, experts exhibit consistent overestimation, which could prove to be a dangerous practice. If the data appear to support a stock at very low stock size, that could be because it is indeed a very low stock size or because interpreting the data is causing a bias. On the other hand, stock status for stocks that are moderately or under-exploited is more likely to be underestimated (Figure 5), as experts seem to include the probability that the information is actually uninformative and thus, are mixing the apparent information with an uninformed prior on stock status. Experts appear to be least biased when estimating the status of stocks with intermediate “true” status (Figure 5). That indicates that experts perform better for stocks around $B_{msy}$, being the biomass that produces maximum sustainable yield (MSY), which is close to a common practice of assuming a biomass at 40% ($B_{40\%}$) of the virgin biomass, when no information on stock status is available [27]. If that is indeed the case, it seems possible that experts make use of the adjustment and anchoring heuristic, where an expert starts from an anchor (here being the $B_{40\%}$) and then adjusts around this value. The adjustments are typically inefficient [50], a conclusion that bears out in this study. However, this pattern could be independent of common assumptions on stock status, and it is just the effect that on average, experts' estimates tend to be average. These findings indicate that experts are biased towards middle values of stock status, that is consistent for all expert levels and need to be considered when interpreting the stock status priors for actual assessment applications. However, even though the inherit bias in expert opinions is still a limitation in obtaining accurate priors on stock status, it is an important recognition from the common assumption of a stock being at $B_{msy}$ and can be used to construct priors (even if at a course scale) for stock assessments.

Even though experience in stock assessment alone is not a significant factor for bias of stock status, it does significantly determine the uncertainty in the constructed distribution, with the higher level experts being most risk averse (Higher IQR values). One of the initial hypotheses was that experts with no experience in stock assessment would be more uncertain regarding their response, but the opposite appears to be the case (Figure 5). A possible explanation is that fisheries scientists that work with stock assessments are trained to follow the precautionary approach incorporated in fisheries assessment [51,52]. People working with stock assessment are aware that data, regardless of their quality, are always associated with uncertainty and even the most well fitted model outcomes contain a degree of uncertainty [6]. This way of thinking starts with graduate education that trains people within a particular disciplinary frame, here being fisheries stock assessment, where scientists must represent uncertainty in a way that assures their audience that risks are tractable and manageable [53].

Another initial hypothesis was that additional data would help experts improve their performance. However, this was not supported given the data levels provided in this present study, thus the amount of information provided did not significantly affect expert performance. A potential explanation for this could be the fact that in all four data sets, the amount of information was rather limited (catch time-series and one or two years of length compositions for an aggregated fleet). If more information was available to experts, e.g. effort or survey abundance index data, we could potentially observe a significant improvement in expert performance and could be a consideration for future work to improve expert elicitation of stock status. If we liken experts to a stock assessment model, where we feed it with data in order to estimate certain quantities, the uncertainty and bias about the estimated quantities reflects the information contained in the available data, following the general rule of statistical inference, where more data lead to less uncertainty [6]. This work begins the attempt to discover a lower limit of useful data inference in order to balance the need for higher level derived quantities in data-limited situations. Future work could look to identify additional data types and time series that may have an affect on stock status estimates.

Another aspect that was explored is whether overestimation or underestimation of the stock status is associated with a specific life history. Establishing a relationship between life-history type and expert performance could assist in developing guidance for species with similar characteristics and under severe data limitations to predict whether it is more likely to over- or underestimate stock status. However, that link was not found and LH did not explain any variability in the data and was thus excluded from the final models. That is not to say life history in not important, as there can be a correlation of stock status with life history (e.g., long-lived, slow growing populations are more susceptible to being reduced to low stock sizes [54]), but life history alone did not significantly influence stock status determination.

A particularly noteworthy finding was the severe overestimation of stock status when "true" stock status was very low. A possible reason for the inability of experts to identify the severe stock reduction lies in the short time series of the provided data relative to the generation time of the stock. An inspection of the length compositions of the entire 50 years of exploitation history for simulated rockfish showed a rapid decrease of large individuals in the end of the second decade, but these data were not provided to the experts. As the given length compositions consisted of the later exploitation period, they did not contain enough contrast observed in length compositions during the earlier years. Yet, the lack of contrast that caused poor expert performance for the simulated rockfish could be due to limitations of the OM. Nevertheless, this brings us to the value of information and the trade-offs between investing in data-collection or other activities. In this case, it is likely that increased investment in data collection to reduce uncertainty may be justified by the benefits in reduced risk of over-exploitation [55]. However, in cases of complete lack of baseline information the increased amount of contemporary data might faulty improve our perception about the stock status, due to the "shifting baseline syndrome" [56]. This is especially relevant to data-limited cases, where financial resources are low [8], thus the challenge remains to prioritize information needs, and to correctly identify and collect the optimal type of information (both historical and contemporary) required to support good management decisions.

Fishery and life history data

In the present study, a sub-set of possible available data was selected in order to mimic a data-limited situation and explore the information content of commonly available data sets. Hilborn and Walters [57] warn that commercial fishery length compositions do not necessarily represent the length composition of a whole stock and, thus, research surveys are usually more accurate at representing the length composition of an exploited stock. However, for the majority of data-limited fisheries there is low investment in research or data collection, such that non-commercial data are rarely available and commercial ones are associated with higher uncertainty and bias, as sampling can be limited in comparison to financially important species [8,58,59]. Regardless as to the source of the length compositions, key to its interpreting, and thus the elicitation of population information, is an understanding of the underlying selectivity curve [60].

Furthermore, the auxiliary information on life history was taken from assessment reports, cited literature and assessment models to aid the process and imitate a real data-limited assessment. Even if no local studies have provided value for life history traits, this information can be obtained from databases such as FishBase [61], though caution is needed when interpreting those values [62]. For the above reasons, the results presented here could be sensitive to the available data and provided background information, though we attempted to provide the most up to date information possible to each expert.

## Simplifying expert elicitation

Many traditional approaches in expert elicitation [33,34,38,39], can be time consuming as it can require experts meet in person with the analyst or receive extensive background materials, making the task seem daunting and complicated. Here, we presented an approach that uses the "Shiny" R package [42] to develop an on-line elicitation tool that experts can access with no logistical limitations. The application provides all the background information, fishery data, and tools for probability distribution construction to assist the experts in the elicitation process. Once the expert has completed the task, all answers are saved in a file and easily sent to the analyst. The benefits of this simple method of eliciting expert knowledge increases efficiency and logistical convenience, which could potentially decrease time investment requirements, user fatigue and hopefully encourages expert compliance. Quite often low levels of motivation for participating in such activities is caused by the degree of investment required from experts for the elicitation process [63].

## Implications of expert performance

Model performance evaluations have shown that data-limited methods that require stock status as an input are highly sensitive to its misspecification, leading to poor performance that cause over or under-estimation of yields, depending on the direction of bias [12,30]. Providing a way to specify informed stock status priors could thus be very valuable, and expert elicitation could be one way of specifying that prior. This work demonstrated that experts may be able to provide such information, but treatment of the expert elicitation will be important. This works highlights the magnitude of the true underlying stock status as the biggest factor causing elicitation bias. Experts are more likely to overestimate the status of severely reduced stocks and to underestimate the status of stocks with high relative biomass. It is though, more desirable to be precautionary and underestimate stock status even if it leads to yield loss than to overestimate it, which could have other negative effects on the fishery [30], especially when the stock is already reduced. At the same time, the constructed stock status prior is affected significantly by the experience level of each expert. These findings are not documented in the available literature, and understanding them are a necessary advance if one wants to consider eliciting stock status from experts and to explore the effects of expert subjectivity and experience.

## Conclusions and future directions

The present study is the first attempt to elicit expert knowledge to inform stock status, an important and sensitive input to several data-limited methods used to inform fisheries management. The aim was to keep the design simple and select a sub-set of possible data combinations that are typical of a data-limited fishery: fishery-dependent length compositions and catch history. As the next step, one could gradually add and remove different data types (e.g., fragmented catch time-series as it can be the case in data-limited situations, longer series of commercial length compositions or abundance indexes that are typically available for more data-rich stocks) provided to experts, in order to identify which type of data, and the threshold where additional data significantly improve ( or not) expert performance. Such an exercise could help explore the value of information [6] and the amount of data required to make inference with as little as possible associated bias. Furthermore, if the aim of the elicitation is to construct the best possible stock status prior, industry-based experts such as local fishers, or other types of stakeholders could be included. Moreover, even though the performance patterns were consistent, among and within the expert groups and are indicative of the effects and potential implications of subjectivity and experience in setting catch limits in data-limited situations, expert sample size could be further expanded. The elicitation tool we presented here could also be used to compare expert opinions provided data from old vs new assessments for data-rich stocks. It is not rare that even for data-rich stocks accumulating data can drastically change the perceptions on stock status from one assessment to the next assessment. The web-application is flexible enough to examine even more complicated elicitation experiments tailored on different situations and can have both education and management potential uses.

Humans are not equipped with a competent mental statistical processor and as a consequence expert judgments are often biased due to unconscious use of a variety of cognitive heuristics [64]. Even though this is known, when the aim of an expert elicitation is to make inference on unknown quantities, identifying the degree of bias is difficult. Here, we compared expert elicited judgments to “true” stock status and we were able to identify the degree of bias in expert judgment. Even though expert bias is still a limitation in constructing an accurate stock status prior, it is an important recognition from the common assumption of a stock being at $B_{msy}$ and thus shows the importance and usefulness of the elicitation for assessing data-limited stocks.The next step is to address the problem of whether experts’ probability assessment can be improved by better understanding the assessment mechanism. Lindley [65] showed that we can utilize all the knowledge gained from such an exercise, as the one showed here, and correct experts’ assessments via Bayes theorem. Following [65] suggestions for improving probability judgments we could create a bias-correction tool that can be applied in order to define corrected stock status priors for data-limited cases. One could therefore imagine a correction applied to elicited low stock status priors so the resultant prior is median-adjusted lower and/or with more uncertainty to avoid overestimation of stock status at low population sizes. In addition to bias correction, elongating the tails at the low end of the distribution to add additional precautionary capacity to an elicited stock status prior is a further consideration.

Acknowledgements

This work has received funding from the University of Helsinki and Academy of Finland. We would like to thank O.S. Hamel, J.T. Thorson, I.J. Stewart, R. Hilborn, A. Punt and other University of Washington affiliates for helpful comments.

## References

[1] Costello C., Ovando D., Hilborn R., Gaines S. D., Deschenes O., Lester S.E. (2012). Status and solutions for the world's unassessed fisheries. Science 338: 517-520.

[2] UNCLOS. 1982. United Nations Convention on the Law of the Sea. 1883 UNTS 3. Available from http://www.un.org/depts/los/convention_agreements/texts/ unclos/UNCLOS-TOC.htm.

[3] FAO (1995). Code of Conduct for responsible Fisheries. FAO, Rome, Italy. Available at: http://www.fao.org/docrep/005/v9878e/v9878e00.HTM

[4] NMFS. Magnuson-Stevens Fishery Conservation and Management Act. Department Of Commerce, U. S. O. A. (2007).

[5] EC (2011). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the Common Fisheries Policy. EC, Brussels, COM/2011/0425 final – 2011/0195 (COD). Available from http://eur-lex.europa.eu/LexUriServ/ LexUriServ.do?uri=COM:2011:0425:FIN:EN:PDF.

[6] Magnusson, A., and Hilborn, R. (2007). What makes fisheries data informative? Fish and Fisheries 8: 337–358. doi:10.1111/j.1467-2979.2007.00258.x.

[7] Methot, R.D., Jr., and Wetzel, C.R. (2012). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142: 86–89 doi:10.1016/j.fishres.2012.10.012.

[8] Salas, S., Chuenpagdee, R., Seij, J., and Charles, A. (2007). Challenges in the assessment and management of small-scale fisheries in Latin America and Caribbean. Fisheries Research 87: 5–16. doi:10.1016/j.fishres.2007.06.015.

[9] Meissa, B., Gascuel, D., and Rivot, E. (2013). Assessing stocks in data-poor African fisheries: a case study on the white grouper Epinephelus aeneus of Mauritania. African Journal of Marine Science, 35: 253–267. doi:10.2989/1814232X.2013.798244.

[10] Stergiou, K.I., Moutopoulos, D.K., Tsikliras, A.C., and Papaconstantinou, C. 2007. Hellenic marine fisheries: A general perspective from the national statistical service data. Current state of the fisheries sector.

[11] Wiedenmann, J., Wilberg, M.J., and Miller, T.J. (2013). An evaluation of harvest control rules for data-poor fisheries. North American Journal of Fisheries Management 33: 845–860. doi:10.1080/02755947.2013.811128.

[12] Carruthers, T., Punt, A., Walters, C., MacCall, A., McAllister, M., Dick, E., and Cope, J. (2014). Evaluating methods for setting catch limits in data-limited fisheries. Fisheries Research 153: 48–68. doi:10.1016/j.fishres.2013.12.014.

[13] Newman, D., J. Berkson, and L. Suatoni. 2015. Current methods for setting catch limits for data-limited fish stocks in the United States. Fisheries Research 164: 86-93.

[14] Beddington, J.R., and Kirkwood, G.P. 2005. The estimation of potential yield and stock status using life-history parameters. Philosophical Transactions of the Royal Society B, 360: 163– 170. doi:10.1098/rstb.2004.1582

[15] Froese, R. 2004. Keep it simple: three indicators to deal with overfishing. Fish and Fisheries 5: 86–91. doi:10.1111/j.1467-2979.2004.00144.x

[16] Cope M.J. and Punt, A.E. (2009). Drawing the lines: resolving fishery management units with simple fisheries data. Canadian Journal of Fisheries and Aquatic Sciences 66(8): 1256-1273.

[17] Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2015b. A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. ICES Journal of Marine Science 72(1): 217–231. doi:10.1093/icesjms/fsu004.

[18] Kokkalis, A., Thygesen, U.H., Nielsen, A., and Andersen, K.H. 2014. Limits to the reliability of size-based fishing status estimation for data-poor stocks. Fisheries Research. 171:4–11. doi:10.1016/j.fishres.2014.10.007.

[19] Anderson, S.C., Branch, T.A., Ricard, D., and Lotze, H.K. 2012. Assessing global marine fishery status with a revised dynamic catch-based method and stock assessment reference points. ICES Journal of Marine Science 69: 1491–1500. doi:10.1093/icesjms/fss105.

[20] Branch T.A., Jensen O.P., Ricard D., Ye Y., Hilborn R. (2011). Contrasting Global Trends in Marine Fishery Status Obtained from Catches and from Stock Assessments. Conservation Biology 25 (4): 777-786

[21] MacCall, A.D. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable yields in data-poor situations. ICES Journal of Marine Science 66: 2267–2271. doi:10.1093/icesjms/fsp209.

[22] Thorson, J.T., Minto, C., Minte-Vera, C.V., Kleisner, K.M., and Longo, C. 2013. A new role for effort dynamics in the theory of harvested populations and data-poor stock assessment. Canadian Journal of Fisheries and Aquatic Science 70(12): 1829–1844. doi:10. 1139/cjfas-2013-0280.

[23] Geromont, H.F., and Butterworth, D.S. (2015). A Revie of assessment methods and the development of management procedures for data-poor fisheries. FAO.

[24] Chrysafi A., Kuparinen A. (2015). Assessing abundance of populations with limited data: Lessons learned from data poor fisheries stock assessment. Environmental Reviews dx.doi.org/10.1139/er-2015-0044

[25] Kimura, D., Tagart, J. (1982). Stock reduction analysis, another solution to the catch equations. Canadian Journal of Fisheries & Aquatic Sciences 39, 1467–1472.

[26] Kimura, D.K., Balsinger, J.W., and Ito, D.H. 1984. Generalized stock reduction analysis. Canadian Journal of Fisheries & Aquatic Sciences 41(9): 1325–1333. doi:10.1139/f84-162.

[27] Dick, E.J., and MacCall, A.D. 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fisheries Research 110: 331–341. doi:10.1016/j.fishres.2011.05.007

[28] Cope M.J. (2013). Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fisheries Research 142: 3-14.

[29] Martel, S., and Froese, R. (2013). A simple method for estimating MSY from catch and resilience. Fish and Fisheries 14: 504–514. doi:10.1111/j.1467-2979.2012.00485.x.

[30] Wetzel C.R. and Punt A.E. (2011). Model performance for the determination of appropriate harvest levels in the case of data-poor stocks. Fisheries Research 110: 342–355. doi:10.1016/j.fishres.2011.04.024.

[31] Cope M.J., Thorson J.T., Wetzell C.R., DeVore J. (2014). Evaluating a prior on relative stock status using simplified age-structured models. Fisheries Research 171: 101-109.

[32] Punt, A. E., Smith, D. C., and Smith, A. D. M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the “Robin Hood” approach. – ICES Journal of Marine Science, 68: 972–981.

[33] Uusitalo L., Kuikka S., Romakkaniemi A. (2005). Estimation of Atlantic salmon smolt carrying capacity of rivers using expert knowledge. ICES Journal of Marine Science 62: 708-722.

[34] Roman HA, et al. (2008) Expert judgment assessment of the mortality impact of changes in ambient fine particulate matter in the U.S. Environmental Science &Technology 42(7):2268–2274.

[35] Zickfeld K, Morgan MG, Frame DJ, Keith DW (2010) Expert judgments about transient climate response to alternative future trajectories of radiative forcing. PNAS 107(28): 12451–12456.

[36] Mäntyniemi S., Haapasaari P., Kuikka S., Parmanne R., Lehtiniemi M., Kaitaranta J. (2013). Incorporating stakeholders’ knowledge to stock assessment: Central Baltic herring. Canadian Journal of Fisheries & Aquatic Sciences 70: 591–599 dx.doi.org/10.1139/cjfas-2012-0316

[37] O'Hagan A., Oakley J.E. (2004). Probability is perfect, but we can't elicit it perfectly. Reliability Engineering and System Safety 85: 239-248

[38] O'Hagan A., Buck C.A., Daneshkhah A., Richard Eiser J., Garthwaite P.H., Jenkinson D.J., Oakley J.E., Rakow T. (2006). Uncertain Judgments: Eliciting experts' probabilities. Wiley, Chichester, UK.

[39] Johnson S.R., Tomlison G.A., Hawker G.A., Granton J.T., Grosbein H.A., Feldman B.M. (2010). A valid and reliable belief elicitation method for Bayesian priors. Journal of Clinical Epidemiology 63: 370-383.

[40] Kritzer J.P. (2004). Sex-specific growth and mortality, spawning season, and female maturation of the striped bass (Lutjanus carponotatus) on the Great Barrier Reef. Fish Bulletin 102: 94–107

[41 Williams A.J., Farley J.H., Hoyle S.D., Davies C.R., Nicol S.J. (2012). Spatial and Sex-Specific Variation in Growth of Albacore Tuna (Thunnus alalunga) across the South Pacific Ocean. PloS ONE 7 (6): e39318. doi:10.1371/journal.pone.0039318

[42] Chang W., Cheng J., Allaire J.J., Xie Y., McPherson J. (2015). Shiny: Web Application Framework for R. R package version 0.11.1., <URL: http://CRAN.R-project.org/package=shiny

[43] Cope M.J. and Punt, A.E. (2011). Reconciling stock assessment and management scales under conditions of spatially varying catch histories. Fisheries Research 107: 22-38.

[44] Cope, J.M., and Punt, A.E. 2009. Length-based reference points for data-limited situations: Applications and restrictions. Marine Coastal Fisheries 1: 169–186. doi:10. 1577/C08-025.1.

[45] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2015). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.3.

[46] Montgomery D.C., Peck E.A., Vining G.G. (2012). Introduction to Linear Regression Analysis. Wiley, Chichester, UK.

[47] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <URL: http://www.R-project.org/.

[48] Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1): 1-48 <doi:10.18637/jss.v067.i01>

[49] Lichtenstein S., Fuschhoff B., Phillips L. (1982). Judgment under uncertainty: heuristics and biases. eds Kahneman D., Slovic P., Tversky A. (Cambridge Univ Press, New York), pp 306–334.

[50] Tversky, A. and Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and Biases." Science, 185(4157): 1124–1131

[51] Hilborn R., Maquire J., Parma A.M., Rosenberg A.A. (2001). The precautionary approach and risk management: can they increase the probability of success in fisheries management? Canadian Journal of Fisheries & Aquatic Sciences 58: 99–107

[52] González-Laxe F. (2005). The precautionary principle in fisheries management. Marine Policy 29: 495-505.

[53] Degnbol P., Gislason H., Hanna S., Jentoft S., Nielsen J.R., Svendrup-Jensen S., Wilson D.C.(2006). Painting the floor with a hammer: Technical fixes in fisheries management. Marine policy 30: 534-543

[54] Berkeley, S.A., M.A. Hixon, R.J. Larson, and M.S. Love. (2004). Fisheries Sustainability via Protection of Age Structure and Spatial Distribution of Fish Populations. Fisheries 29(8): 23-32.

[55] Hansen G.J.A., Jones M.L. (2008). The value of information in fishery management. Fisheries 33: 340-348

[56] Pauly D. (1995). Anecdotes and the shifting baseline syndrome of fisheries. TREE 10(10): 430

[57] Hilborn, R., and Walters, C.J. (1992). Quantitative fisheries stock assessment. Choice, Dynamics and uncertainty. New York, NY: Chapman and Hall Inc.

[58] Lleonart, J., and Maynou, F. (2003). Fish stock assessments in the Mediterranean: state of the art. Scientia Marina 67: 37–49. doi:10.3989/scimar.2003.67s137.

[59] Espinoza-Tenorio, A., Espejel, I., Wolff, M., and Zepeda-Dominquez, J. (2011). Contextual factors influencing sustainable fisheries in Mexico. Mar. Policy, 35: 343–350.vdoi:10.1016/j.marpol.2010.10.014.

[60] Cope M.J. and Punt A. (2009). Length-based reference points for data-limited situations: Applications and restrictions. Marine and Coastal Fisheries: Dynamics, Management and Ecosystem Science 1: 169-186

[61] Froese R. and Pauly D. Editors. (2016). FishBase. Wolrd Wide Web electronic publication. www.fishbase.org (01/2016)

[62] Patrick, W. S., J. M. Cope, and J. T. Thorson. 2014. Validating and improving life history data in FishBase. Fisheries 39(4): 173-176.

[63] Voinov A., Bousquet F. (2010). Modelling with stakeholders. Environmental Modelling & Software 25: 1268-1281

[64] Morgan M.G. (2013). Use (and abuse) of expert elicitation in support of decision making for public policy. PNAS 111 (20): 7176-7184

[65] Lindley D.V. (1982). The improvement of probability judgments. Journal of Royal Statistical Society 145: 117-126