

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

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)

Copyright: © 2017 Elsevier Ltd.

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

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>

1 Abstract

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

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

19 1. Introduction

20 Concerns for the future of ocean health and the sustainable exploitation of fish stocks are
21 rising and a large portion (estimate to be about 80% in Costello et al. 2012 [1]) of global fish catches still
22 come from unassessed fisheries. Fisheries policies around the globe (e.g., The Law of the Sea [2], the Code
23 of Conduct of the UN [3], the Magnuson-Stevens Act in the US [4], and the EU Common Fisheries Policy [5]
24 all mandate for responsible fisheries and sustainable harvesting via biomass based target and limit
25 reference points. The capacity to estimate biomass via population dynamics modelling has greatly

26 advanced [6,7] and is able to provide scientifically-derived information for management action. The
27 drawback of these models is that they typically require extensive and high-quality/quantity data. Data
28 collection requires time, investment, and appropriate scientific personnel, all in perpetuity, which are
29 typically lacking in developing countries [8,9]. Small-scale fisheries— often multi-species and multi-gear in
30 nature— epitomize the challenges of stock assessment in data-limited situations [10]. Low contribution to
31 Gross Domestic Product (GDP) is a major reason behind the limited quantitative information available for
32 small-scale fisheries [8]. This difference in data-collection investment, directly causes such fisheries to
33 require alternative methods to traditional approaches in quantitative stock assessment [11].

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

47 Some of the data-limited approaches are based on modified versions of classical stock
48 reduction analysis (SRA; [25,26]), which require prior information on the relative stock status, expressed in
49 the form of B_y/B_0 [27,28,29], where B is biomass (usually spawning biomass), y is year and 0 represents the
50 initial condition. In these methods, the stock status input is specified as a distribution, from which values

51 are resampled. The resampled status and the given catches, selectivity and biology of the stock solve for a
52 particular biomass time-series. Model performance evaluations [12,30] have shown that under- or over-
53 estimation of stock status leads to under- or over- estimation of potential yields respectively. Stock status
54 however, is typically a derived stock assessment output, so usually not available unless a stock assessment
55 has been carried out. Alternative ways of deriving this important stock metric would be a valuable
56 contribution to management. Cope et al. [31] developed a method that estimates a stock status prior using
57 Productivity-Susceptibility Analysis (PSA) vulnerability scores, being the first step towards solutions for
58 defining stock status in data-limited situations. Other possible alternatives for defining stock status, when
59 facing data limitations, include the utilization of information from species with similar exploitation/biology
60 [32].

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

76 stacking probability chips (e.g. 5% probability chip), or sliding bars of a given probability distribution in
77 order to depict their beliefs and construct the probability density [33,36,39].

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

92 2. Material and methods

93 Data collection

94 Stock assessments from the US. Pacific Fishery Management Council (PFMC) and Alaska
95 Fisheries Science Center (AFSC) were used as the basis for data of this analysis, where 18 stocks with
96 analytical assessments using Stock Synthesis (SS; [7]) or similar statistical catch-at-age models that cover a
97 variety of life history types were selected. For the PFMC stocks, SS files were obtained in order to extract
98 the necessary catch and length composition information. For the AFSC stocks, data for fishery length
99 compositions were obtained from the domestic observer program. Catch data were taken from the AFSC
100 assessment reports and if they were unavailable in the report, they were provided by the corresponding

101 assessment authors. Additional information on life history and management of all stocks was extracted
102 from the assessment reports or cited references. If information on maximum observed age was
103 unavailable, the last age group (i.e., the plus group) in the assessment was used instead. When a fishery
104 consisted of multiple fleets, length compositions were combined in order to create a single length
105 composition for the entire fishery in the following manner:

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

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

113 [FIGURE 1 ABOUT HERE]

114 [TABLE 1 ABOUT HERE]

115 Expert group

116 For the elicitation process, six experts with different levels of experience in stock assessment
117 were selected to quantify bias and imprecision in stock status determination. Experts were separated in
118 three groups of two: i) Level 1 (Experienced) with high experience in stock assessment, being formally
119 employed as stock assessors (Expert 1 and Expert 2), ii) Level 2 (Novice) being under training in stock
120 assessment with limited direct assessment experience (Expert 3 and Expert 4) and iii) Level 3
121 (Inexperienced) with a background in fisheries ecology (Expert 5 and Expert 6). All experts represent USA
122 west coast academic and federal research institutions and are familiar with the biology of west coast
123 groundfishes. Experts from different backgrounds, e.g. fishers or NGOs, were not considered eligible, as

124 the aim of this work was to imitate a real case of data-limited stock assessment conducted by fisheries
125 experts without any external input. This way we demonstrate how fisheries expert opinions could be
126 gathered, evaluated, and analyzed under such a scenario. However, obtaining a combined stock status prior
127 based on opinions of various backgrounds, is one way of conducting expert elicitation but not the one
128 considered here. Furthermore, as stock assessments are commonly carried out by a few experts (and most
129 likely only one)we were thus interested in individual expert performance and not cumulative performance.
130 Therefore, the number of experts participating in the elicitation process was small on purpose and should
131 not affect the results but be indicative of individual experts within their expertise level.

132 Simulated stocks

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

145 Data combinations

146 Case-study fisheries

147 Within a data-limited context, there are many potential data combinations that can be
148 tested, but for the purpose of this work, we selected a sub-set of catch time-series and fishery length

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

164 [FIGURE 2 ABOUT HERE]

165 Sophisticated stock assessment models account for sex-specific differences in life histories,
166 and sex-specific variation in growth is common [40,41,7]. On the other hand, given that data-limited
167 assessment tools are often more biologically simplified, they routinely use only single inputs for life-history
168 information (e.g natural mortality M , maximum length L_{max} , von Bertalanffy K , maximum age A_{max}) to
169 represent the entire stock across sexes [27,29]. To account for this methods' characteristic, and to make
170 sure that the information that experts can extract from length compositions and life history parameters
171 match each other, only females data were provided with the assumption that they are sufficient to make
172 inference for the entire stock. Given that spawning stock biomass is typically used to describe stock status,
173 this is not a poor assumption. However, there were a few cases where sex-specific length compositions

174 were not available for some fleets, so sex-combined compositions were used. These mix-sex data always
175 came from fleets with low contribution to the total catch and the effect on the final constructed length
176 compositions was negligible.

177 Simulated stocks

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

184 Expert groups

185 The experts were separated in two groups (A=Expert 1, Expert 3 and Expert 5 and B=Expert
186 2, Expert 4, Expert 6) and the first group received the simulated rockfish as first species and the simulated
187 sole as last species in the elicitation process, and vice versa for the second group. The 18 stocks were given
188 in the same order. Data sets allowed one to ask a) if and how additional data (i.e., longer time series)
189 would help experts to improve their belief about stock status and b) if during the elicitation process the
190 experts would gain experience and confidence that would lead in improved distribution estimates of stock
191 status.

192 Elicitation tool

193 The “Shiny” R package [42] was used to develop a web-based application for expert
194 elicitation. The two versions of the web application, corresponding to the two groups of experts mentioned
195 above, can be found in https://chrysafi1.shinyapps.io/Beta_shiny_1/ and
196 https://chrysafi1.shinyapps.io/Beta_shiny_2/ respectively. Step-by-step instructions on how to use the
197 application are provided within the tool. Experts were able to select the stock and data set, which produced

198 the respective graphs of catch and length compositions and background information. To construct a stock
199 status prior, the tool provided a beta distribution probability density function (PDF) and scroll bars for the
200 mean and standard deviation that users could freely move to capture the density that best fit their
201 perception of stock status. In addition, a beta cumulative distribution function (CDF) with the median and
202 interquartile range was provided in order to help experts visualize the distribution. The beta distribution is
203 commonly used when describing relative stock status priors [27] and is bounded from 0 to 1, similar to
204 value for expected stock status. The application also included a table, self-populated with the output of the
205 beta distribution values once the expert completed each data set. After all stocks and data sets were
206 completed, the final table of elicited stock status priors could be downloaded as a comma-separated-values
207 (csv) file and submitted. All experts have or were working towards a PhD and basic knowledge in
208 probability theory was assumed, though the visualization tool was developed to make the recommended
209 prior distribution clear. A detailed description of the project and the aims were provided, and experts were
210 given time to look thoroughly through the application to cover any arising questions before conducting the
211 elicitation experiment.

212 Performance measures

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

$$215 \quad RE = \frac{S_e - S_m}{S_m} \quad (1)$$

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

219 Linear and linear mixed models were used to identify which explanatory variables best
220 described expert bias and imprecision. RE data were log transformed ($\log(\log(\text{median RE} + 2))$ and
221 $\log(\log(\text{IQR} + 1))$ respectively) to fit the normal distribution. The variables of expert level, data set (being

222 each of the four data combinations), “true” stock status and life history (LH) were treated as categorical
223 explanatory variables (Table 2). LH categorization was derived using silhouette-based cluster validity
224 analysis [44,45] of the M/k ratio for each stock and LH was explored as either a fixed or random effect.
225 Backward stepwise model selection and the Akaike information criterion (AIC) and AIC with a correction for
226 finite sample sizes (AICc) were used for final model selection [46]. The final models were checked for
227 homogeneity and normality of residuals. All analyses were carried out using R [47] and the “lme4” package
228 [48]. Data from all 20 stocks (18 real and 2 simulated) were used for the analysis.

229 [TABLE 2 ABOUT HERE]

230 3. Results

231 Explanatory variables for expert performance

232 Life history traits were first treated as random effects, but this treatment did not explain any
233 variability in the data, so linear fixed-effects models were used instead for all subsequent exploration.
234 “Data set” and LH were not significant explanatory variables for either median RE nor IQR. For both cases,
235 the linear model with interactions between expert level and “true” stock status as explanatory variables
236 yielded the lowest AICc values (Table 3). A graphical inspection of the transformed median RE and IQR, for
237 all explored variables, verify the results of the linear models (Figure 3). For bias (median RE), “true” stock
238 status (F-test, DF=3, $p < 0.0001$) explained 60.5% of the variability and the interactions between true stock
239 status and expert level (F-test, DF=6, $p < 0.0001$) explained 2.5% of the variability (adj. $R^2 = 0.6257$). Expert
240 level (F-test, DF=2, $p < 0.0001$) alone explained only 0.3% of the variability in expert bias. For imprecision
241 (IQR), “true” stock status (F-test, DF=3, $p < 0.0001$) explained 57.3% of the variability, expert level (F-test,
242 DF=2, $p < 0.0001$) explained 8.5% of the variability, and the interactions between true stock status and
243 expert level (F-test, DF=6, $p = 0.01434$) explained 1.3% of the variability (adj. $R^2 = 0.6691$). Figure 4 illustrates
244 the effect of expert level on bias and imprecision from the interactions with “true” stock status. The
245 coefficients and p-values for all levels of main effects and individual interactions are provided in
246 supplementary material (Table S2). Final models for bias and imprecision were further tested for

247 redundancy by collapsing (e.g. merging of Stock Status 2 and 3 into one category) factor levels, but the
248 original treatment of factor levels was most supported by the data.

249 [TABLE 3 ABOUT HERE]

250 [FIGURE 3 ABOUT HERE]

251 [FIGURE 4 ABOUT HERE]

252 Individual expert performance

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

267 [FIGURE 5 ABOUT HERE]

268 The same pattern driven by “true” stock status is also observed for the simulated stocks, as
269 all experts overestimate stock status for the rockfish and underestimate stock status for the sole (Figure 6).
270 Moreover, experts that received the sole as the last species in the elicitation process, performed better in

271 comparison to the experts that received sole as the first species, supporting the improvement of expert
272 performance while completing it. However, the same finding was not observed with the simulated rockfish
273 (that had lower relative stock status), thus complicating the interpretation as to how and when experts may
274 improve their performance. More experts and further exploration of adding simulated species is needed to
275 gain a better understanding of whether experts improve as they work through more species.

276 [FIGURE 6 ABOUT HERE]

277 Expert calibration

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

291 [FIGURE 7 ABOUT HERE]

292 4. Discussion

293 Expert performance

294 The present study illustrates that the most important variables for expert bias is “true” stock
295 status and the interactions between “true” stock status and expert level. For stocks with low status, experts

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

315 Even though experience in stock assessment alone is not a significant factor for bias of stock
316 status, it does significantly determine the uncertainty in the constructed distribution, with the higher level
317 experts being most risk averse (Higher IQR values). One of the initial hypotheses was that experts with no
318 experience in stock assessment would be more uncertain regarding their response, but the opposite
319 appears to be the case (Figure 5). A possible explanation is that fisheries scientists that work with stock
320 assessments are trained to follow the precautionary approach incorporated in fisheries assessment [51,52].
321 People working with stock assessment are aware that data, regardless of their quality, are always

322 associated with uncertainty and even the most well fitted model outcomes contain a degree of uncertainty
323 [6]. This way of thinking starts with graduate education that trains people within a particular disciplinary
324 frame, here being fisheries stock assessment, where scientists must represent uncertainty in a way that
325 assures their audience that risks are tractable and manageable [53].

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

340 Another aspect that was explored is whether overestimation or underestimation of the stock
341 status is associated with a specific life history. Establishing a relationship between life-history type and
342 expert performance could assist in developing guidance for species with similar characteristics and under
343 severe data limitations to predict whether it is more likely to over- or underestimate stock status. However,
344 that link was not found and LH did not explain any variability in the data and was thus excluded from the
345 final models. That is not to say life history is not important, as there can be a correlation of stock status

346 with life history (e.g., long-lived, slow growing populations are more susceptible to being reduced to low
347 stock sizes [54]), but life history alone did not significantly influence stock status determination.

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

364 Fishery and life history data

365 In the present study, a sub-set of possible available data was selected in order to mimic a
366 data-limited situation and explore the information content of commonly available data sets. Hilborn and
367 Walters [57] warn that commercial fishery length compositions do not necessarily represent the length
368 composition of a whole stock and, thus, research surveys are usually more accurate at representing the
369 length composition of an exploited stock. However, for the majority of data-limited fisheries there is low
370 investment in research or data collection, such that non-commercial data are rarely available and

371 commercial ones are associated with higher uncertainty and bias, as sampling can be limited in comparison
372 to financially important species [8,58,59]. Regardless as to the source of the length compositions, key to its
373 interpreting, and thus the elicitation of population information, is an understanding of the underlying
374 selectivity curve [60].

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

381 Simplifying expert elicitation

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

393 Implications of expert performance

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

408 Conclusions and future directions

409 The present study is the first attempt to elicit expert knowledge to inform stock status, an
410 important and sensitive input to several data-limited methods used to inform fisheries management. The
411 aim was to keep the design simple and select a sub-set of possible data combinations that are typical of a
412 data-limited fishery: fishery-dependent length compositions and catch history. As the next step, one could
413 gradually add and remove different data types (e.g., fragmented catch time-series as it can be the case in
414 data-limited situations, longer series of commercial length compositions or abundance indexes that are
415 typically available for more data-rich stocks) provided to experts, in order to identify which type of data,
416 and the threshold where additional data significantly improve (or not) expert performance. Such an
417 exercise could help explore the value of information [6] and the amount of data required to make inference
418 with as little as possible associated bias. Furthermore, if the aim of the elicitation is to construct the best
419 possible stock status prior, industry-based experts such as local fishers, or other types of stakeholders could

420 be included. Moreover, even though the performance patterns were consistent, among and within the
421 expert groups and are indicative of the effects and potential implications of subjectivity and experience in
422 setting catch limits in data-limited situations, expert sample size could be further expanded. The elicitation
423 tool we presented here could also be used to compare expert opinions provided data from old vs new
424 assessments for data-rich stocks. It is not rare that even for data-rich stocks accumulating data can
425 drastically change the perceptions on stock status from one assessment to the next assessment. The web-
426 application is flexible enough to examine even more complicated elicitation experiments tailored on
427 different situations and can have both education and management potential uses.

428 Humans are not equipped with a competent mental statistical processor and as a
429 consequence expert judgments are often biased due to unconscious use of a variety of cognitive heuristics
430 [64]. Even though this is known, when the aim of an expert elicitation is to make inference on unknown
431 quantities, identifying the degree of bias is difficult. Here, we compared expert elicited judgments to “true”
432 stock status and we were able to identify the degree of bias in expert judgment. Even though expert bias is
433 still a limitation in constructing an accurate stock status prior, it is an important recognition from the
434 common assumption of a stock being at B_{msy} and thus shows the importance and usefulness of the
435 elicitation for assessing data-limited stocks. The next step is to address the problem of whether experts’
436 probability assessment can be improved by better understanding the assessment mechanism. Lindley [65]
437 showed that we can utilize all the knowledge gained from such an exercise, as the one showed here, and
438 correct experts’ assessments via Bayes theorem. Following [65] suggestions for improving probability
439 judgments we could create a bias-correction tool that can be applied in order to define corrected stock
440 status priors for data-limited cases. One could therefore imagine a correction applied to elicited low stock
441 status priors so the resultant prior is median-adjusted lower and/or with more uncertainty to avoid
442 overestimation of stock status at low population sizes. In addition to bias correction, elongating the tails at
443 the low end of the distribution to add additional precautionary capacity to an elicited stock status prior is a
444 further consideration.

445 Acknowledgements

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

449 References

- 450 [1] Costello C., Ovando D., Hilborn R., Gaines S. D., Deschenes O., Lester S.E. (2012). Status and solutions for
451 the world's unassessed fisheries. *Science* 338: 517-520.
- 452 [2] UNCLOS. 1982. United Nations Convention on the Law of the Sea. 1883 UNTS 3. Available from
453 http://www.un.org/depts/los/convention_agreements/texts/unclos/UNCLOS-TOC.htm.
- 454 [3] FAO (1995). Code of Conduct for responsible Fisheries. FAO, Rome, Italy. Available at:
455 <http://www.fao.org/docrep/005/v9878e/v9878e00.HTM>
- 456 [4] NMFS. Magnuson-Stevens Fishery Conservation and Management Act. Department Of Commerce, U. S.
457 O. A. (2007).
- 458 [5] EC (2011). Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the
459 Common Fisheries Policy. EC, Brussels, COM/2011/0425 final – 2011/0195 (COD). Available
460 from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0425:FIN:EN:PDF>.
- 461 [6] Magnusson, A., and Hilborn, R. (2007). What makes fisheries data informative? *Fish and Fisheries* 8:
462 337–358. doi:10.1111/j.1467-2979.2007.00258.x.
- 463 [7] Methot, R.D., Jr., and Wetzel, C.R. (2012). Stock synthesis: a biological and statistical framework for fish
464 stock assessment and fishery management. *Fisheries Research* 142: 86–89
465 doi:10.1016/j.fishres.2012.10.012.
- 466 [8] Salas, S., Chuenpagdee, R., Seij, J., and Charles, A. (2007). Challenges in the assessment and
467 management of small-scale fisheries in Latin America and Caribbean. *Fisheries Research* 87:
468 5–16. doi:10.1016/j.fishres.2007.06.015.

- 469 [9] Meissa, B., Gascuel, D., and Rivot, E. (2013). Assessing stocks in data-poor African fisheries: a case study
470 on the white grouper *Epinephelus aeneus* of Mauritania. *African Journal of Marine Science*,
471 35: 253–267. doi:10.2989/1814232X.2013.798244.
- 472 [10] Stergiou, K.I., Moutopoulos, D.K., Tsikliras, A.C., and Papaconstantinou, C. 2007. Hellenic marine
473 fisheries: A general perspective from the national statistical service data. Current state of the
474 fisheries sector.
- 475 [11] Wiedenmann, J., Wilberg, M.J., and Miller, T.J. (2013). An evaluation of harvest control rules for data-
476 poor fisheries. *North American Journal of Fisheries Management* 33: 845–860.
477 doi:10.1080/02755947.2013.811128.
- 478 [12] Carruthers, T., Punt, A., Walters, C., MacCall, A., McAllister, M., Dick, E., and Cope, J. (2014). Evaluating
479 methods for setting catch limits in data-limited fisheries. *Fisheries Research* 153: 48–68.
480 doi:10.1016/j.fishres.2013.12.014.
- 481 [13] Newman, D., J. Berkson, and L. Suatoni. 2015. Current methods for setting catch limits for data-limited
482 fish stocks in the United States. *Fisheries Research* 164: 86-93.
- 483 [14] Beddington, J.R., and Kirkwood, G.P. 2005. The estimation of potential yield and stock status using life-
484 history parameters. *Philosophical Transactions of the Royal Society B*, 360: 163– 170.
485 doi:10.1098/rstb.2004.1582
- 486 [15] Froese, R. 2004. Keep it simple: three indicators to deal with overfishing. *Fish and Fisheries* 5: 86–91.
487 doi:10.1111/j.1467-2979.2004.00144.x
- 488 [16] Cope M.J. and Punt, A.E. (2009). Drawing the lines: resolving fishery management units with simple
489 fisheries data. *Canadian Journal of Fisheries and Aquatic Sciences* 66(8): 1256-1273.
- 490 [17] Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2015b. A novel length-based empirical
491 estimation method of spawning potential ratio (SPR), and tests of its performance, for small-
492 scale, data-poor fisheries. *ICES Journal of Marine Science* 72(1): 217–231.
493 doi:10.1093/icesjms/fsu004.

- 494 [18] Kokkalis, A., Thygesen, U.H., Nielsen, A., and Andersen, K.H. 2014. Limits to the reliability of size-based
495 fishing status estimation for data-poor stocks. *Fisheries Research*. 171:4–11.
496 doi:10.1016/j.fishres.2014.10.007.
- 497 [19] Anderson, S.C., Branch, T.A., Ricard, D., and Lotze, H.K. 2012. Assessing global marine fishery status
498 with a revised dynamic catch-based method and stock assessment reference points. *ICES*
499 *Journal of Marine Science* 69: 1491–1500. doi:10.1093/icesjms/fss105.
- 500 [20] Branch T.A., Jensen O.P., Ricard D., Ye Y., Hilborn R. (2011). Contrasting Global Trends in Marine
501 Fishery Status Obtained from Catches and from Stock Assessments. *Conservation Biology* 25
502 (4): 777-786
- 503 [21] MacCall, A.D. 2009. Depletion-corrected average catch: a simple formula for estimating sustainable
504 yields in data-poor situations. *ICES Journal of Marine Science* 66: 2267–2271.
505 doi:10.1093/icesjms/fsp209.
- 506 [22] Thorson, J.T., Minto, C., Minto-Vera, C.V., Kleisner, K.M., and Longo, C. 2013. A new role for effort
507 dynamics in the theory of harvested populations and data-poor stock assessment. *Canadian*
508 *Journal of Fisheries and Aquatic Science* 70(12): 1829–1844. doi:10.1139/cjfas-2013-0280.
- 509 [23] Geromont, H.F., and Butterworth, D.S. (2015). A Review of assessment methods and the development of
510 management procedures for data-poor fisheries. FAO.
- 511 [24] Chrysafi A., Kuparinen A. (2015). Assessing abundance of populations with limited data: Lessons
512 learned from data poor fisheries stock assessment. *Environmental Reviews*
513 dx.doi.org/10.1139/er-2015-0044
- 514 [25] Kimura, D., Tagart, J. (1982). Stock reduction analysis, another solution to the catch equations.
515 *Canadian Journal of Fisheries & Aquatic Sciences* 39, 1467–1472.
- 516 [26] Kimura, D.K., Balsinger, J.W., and Ito, D.H. 1984. Generalized stock reduction analysis. *Canadian*
517 *Journal of Fisheries & Aquatic Sciences* 41(9): 1325–1333. doi:10.1139/f84-162.

- 518 [27] Dick, E.J., and MacCall, A.D. 2011. Depletion-Based Stock Reduction Analysis: A catch-based method
519 for determining sustainable yields for data-poor fish stocks. *Fisheries Research* 110:
520 331–341. doi:10.1016/j.fishres.2011.05.007
- 521 [28] Cope M.J. (2013). Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for
522 deriving overfishing limits in data-limited situations. *Fisheries Research* 142: 3-14.
- 523 [29] Martel, S., and Froese, R. (2013). A simple method for estimating MSY from catch and resilience. *Fish*
524 and *Fisheries* 14: 504–514. doi:10.1111/j.1467-2979.2012.00485.x.
- 525 [30] Wetzel C.R. and Punt A.E. (2011). Model performance for the determination of appropriate harvest
526 levels in the case of data-poor stocks. *Fisheries Research* 110: 342–355.
527 doi:10.1016/j.fishres.2011.04.024.
- 528 [31] Cope M.J., Thorson J.T., Wetzell C.R., DeVore J. (2014). Evaluating a prior on relative stock status using
529 simplified age-structured models. *Fisheries Research* 171: 101-109.
- 530 [32] Punt, A. E., Smith, D. C., and Smith, A. D. M. 2011. Among-stock comparisons for improving stock
531 assessments of data-poor stocks: the “Robin Hood” approach. – *ICES Journal of Marine*
532 *Science*, 68: 972–981.
- 533 [33] Uusitalo L., Kuikka S., Romakkaniemi A. (2005). Estimation of Atlantic salmon smolt carrying capacity
534 of rivers using expert knowledge. *ICES Journal of Marine Science* 62: 708-722.
- 535 [34] Roman HA, et al. (2008) Expert judgment assessment of the mortality impact of changes in ambient
536 fine particulate matter in the U.S. *Environmental Science & Technology* 42(7):2268–2274.
- 537 [35] Zickfeld K, Morgan MG, Frame DJ, Keith DW (2010) Expert judgments about transient climate
538 response to alternative future trajectories of radiative forcing. *PNAS* 107(28): 12451–
539 12456.
- 540 [36] Mäntyniemi S., Haapasaari P., Kuikka S., Parmanne R., Lehtiniemi M., Kaitaranta J. (2013).
541 Incorporating stakeholders’ knowledge to stock assessment: Central Baltic herring. *Canadian*
542 *Journal of Fisheries & Aquatic Sciences* 70: 591–599 dx.doi.org/10.1139/cjfas-2012-0316

- 543 [37] O'Hagan A., Oakley J.E. (2004). Probability is perfect, but we can't elicit it perfectly. *Reliability*
544 *Engineering and System Safety* 85: 239-248
- 545 [38] O'Hagan A., Buck C.A., Daneshkhah A., Richard Eiser J., Garthwaite P.H., Jenkinson D.J., Oakley J.E.,
546 Rakow T. (2006). *Uncertain Judgments: Eliciting experts' probabilities*. Wiley, Chichester, UK.
- 547 [39] Johnson S.R., Tomlison G.A., Hawker G.A., Granton J.T., Grosbein H.A., Feldman B.M. (2010). A valid
548 and reliable belief elicitation method for Bayesian priors. *Journal of Clinical Epidemiology* 63:
549 370-383.
- 550 [40] Kritzer J.P. (2004). Sex-specific growth and mortality, spawning season, and female maturation of
551 the striped bass (*Lutjanus carponotatus*) on the Great Barrier Reef.
552 *Fish Bulletin* 102: 94–107
- 553 [41] Williams A.J., Farley J.H., Hoyle S.D., Davies C.R., Nicol S.J. (2012). Spatial and Sex-Specific Variation in
554 Growth of Albacore Tuna (*Thunnus alalunga*) across the South Pacific Ocean. *PloS ONE* 7 (6):
555 e39318. doi:10.1371/journal.pone.0039318
- 556 [42] Chang W., Cheng J., Allaire J.J., Xie Y., McPherson J. (2015). Shiny: Web Application Framework for R.
557 R package version 0.11.1., <URL: <http://CRAN.R-project.org/package=shiny>
- 558 [43] Cope M.J. and Punt, A.E. (2011). Reconciling stock assessment and management scales under
559 conditions of spatially varying catch histories. *Fisheries Research* 107: 22-38.
- 560 [44] Cope, J.M., and Punt, A.E. 2009. Length-based reference points for data-limited situations:
561 Applications and restrictions. *Marine Coastal Fisheries* 1: 169–186. doi:10. 1577/C08-025.1.
- 562 [45] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2015). *cluster: Cluster Analysis Basics*
563 *and Extensions*. R package version 2.0.3.
- 564 [46] Montgomery D.C., Peck E.A., Vining G.G. (2012). *Introduction to Linear Regression Analysis*. Wiley,
565 Chichester, UK.
- 566 [47] R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for
567 Statistical Computing, Vienna, Austria. <URL: <http://www.R-project.org/>.

- 568 [48] Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models
569 Using lme4. *Journal of Statistical Software*, 67(1): 1-48 <[doi:10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)>
- 570 [49] Lichtenstein S., Fischhoff B., Phillips L. (1982). Judgment under uncertainty: heuristics and biases.
571 eds Kahneman D., Slovic P., Tversky A. (Cambridge Univ Press, New York), pp 306–334.
- 572 [50] Tversky, A. and Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and Biases." *Science*,
573 185(4157): 1124–1131
- 574 [51] Hilborn R., Maquire J., Parma A.M., Rosenberg A.A. (2001). The precautionary approach and risk
575 management: can they increase the probability of success in fisheries management?
576 *Canadian Journal of Fisheries & Aquatic Sciences* 58: 99–107
- 577 [52] González-Laxe F. (2005). The precautionary principle in fisheries management. *Marine Policy* 29: 495-
578 505.
- 579 [53] Degnbol P., Gislason H., Hanna S., Jentoft S., Nielsen J.R., Svendrup-Jensen S., Wilson D.C. (2006).
580 Painting the floor with a hammer: Technical fixes in fisheries management. *Marine policy* 30:
581 534-543
- 582 [54] Berkeley, S.A., M.A. Hixon, R.J. Larson, and M.S. Love. (2004). Fisheries Sustainability via Protection of
583 Age Structure and Spatial Distribution of Fish Populations. *Fisheries* 29(8): 23-32.
- 584 [55] Hansen G.J.A., Jones M.L. (2008). The value of information in fishery management. *Fisheries* 33: 340-
585 348
- 586 [56] Pauly D. (1995). Anecdotes and the shifting baseline syndrome of fisheries. *TREE* 10(10): 430
- 587 [57] Hilborn, R., and Walters, C.J. (1992). Quantitative fisheries stock assessment. Choice, Dynamics and
588 uncertainty. New York, NY: Chapman and Hall Inc.
- 589 [58] Leonart, J., and Maynou, F. (2003). Fish stock assessments in the Mediterranean: state of the art.
590 *Scientia Marina* 67: 37–49. doi:10.3989/scimar.2003.67s137.

- 591 [59] Espinoza-Tenorio, A., Espejel, I., Wolff, M., and Zepeda-Dominquez, J. (2011). Contextual factors
592 influencing sustainable fisheries in Mexico. *Mar. Policy*, 35: 343–
593 350.vdoi:10.1016/j.marpol.2010.10.014.
- 594 [60] Cope M.J. and Punt A. (2009). Length-based reference points for data-limited situations: Applications
595 and restrictions. *Marine and Coastal Fisheries: Dynamics, Management and Ecosystem*
596 *Science* 1: 169-186
- 597 [61] Froese R. and Pauly D. Editors. (2016). FishBase. Wolrd Wide Web electronic publication.
598 www.fishbase.org (01/2016)
- 599 [62] Patrick, W. S., J. M. Cope, and J. T. Thorson. 2014. Validating and improving life history data in
600 FishBase. *Fisheries* 39(4): 173-176.
- 601 [63] Voinov A., Bousquet F. (2010). Modelling with stakeholders. *Environmental Modelling & Software* 25:
602 1268-1281
- 603 [64] Morgan M.G. (2013). Use (and abuse) of expert elicitation in support of decision making for public
604 policy. *PNAS* 111 (20): 7176-7184
- 605 [65] Lindley D.V. (1982). The improvement of probability judgments. *Journal of Royal Statistical Society*
606 145: 117-126
- 607