Response to reviewers

Dear Recommender, Dear Jean-Olivier Irisson

We are grateful for your dedication and support in reviewing our manuscript entitled 'Assessing precocious maturation in salmon using ultrasound imaging'. We thank the two reviewers for their thorough reviews and their helpful comments. We have made every effort to address the reviewers' comments. In particular, i) we revised the analysis of the data and rearrange the result section accordingly, ii) we clarified the link between the probability for a mature parr to produce milt and the probability to detect precocious maturation with the phenotypic method, iii) we put more emphasis on the false-positive and false-negative errors with each method and iv) in the conclusion we recommend using the two methods together to assess precocious maturation in salmon parrs.

We look forward to your assessment and hope that the paper is now suitable for recommendation in PCI.

Yours sincerely,

Marie Nevoux

Response to reviewer 1

I think that your paper is original and globally well written. This study is very interesting because it proposes a new approach for gathering data on fish populations. You showed here that the proportion of precocious young salmon is highly variable each year which is an important demographic feature of the populations. I really support such approach that makes field measurements more reliable.

In the comparison of phenotypic and ultrasound methods (Li158 – 178) it is difficult to understand clearly how environmental factors, which play an important (significant) role on precocious sexual maturation in young salmons, may influence the accuracy of each method. You presented results on the "probability to produce milt". But your objective was to compare two different approaches for detecting the precocious state, not the drivers for this precocious state. You have to clarify the link between the "probability to produce milt" and the accuracy of each approach in rightly estimating if a salmon is a precocious individual. You detected a strong year effect. That means that among the three sampling years the proportion of precocious salmon was highly variable. How this variability was assumed to influence the accuracy of both methods? One option could be to split your Table 1 in three parts, i.e. the same three lines repeated for 2015, 2016 and 2017.
> In the methods and the results sections, we clarified the link between the probability to produce milt in mature males and the probability to detect precocious maturation as follow: "Because the probability to detect precocious maturation with the phenotypic method depends on the probability for a mature male to produce milt at the time of capture, an analysis is conducted to investigate which biological factor drive the this probability "
As recommended, we split the data presented in table 1 by year. We thus pointed out at this interannual variability in the text by adding a sentence: "Hence, the ultrasound method unveil that only 16.8% and 13.2% of the mature males are identified as precocious parrs at the time of capture in 2015 and 2017, and this proportion rise to 87.6% in 2016 (Table 1)."

some minor comments
Li4: Your title is short. You can add some words. Here is a proposition:
Field assessement **of** precocious maturation in salmon **parr** using ultrasound imaging
> This proposition improves the description of the study, we changed our title accordingly.

Li26: Atlantic salmon **parr**
> done

Li29: in the wild **such as individual sexual maturation**.
> done

Li40: in the future **under different climatic scenarios**.
> done

Li40: , predictive **population dynamics** models
> done

Li44: monitoring programs **without invasive or lethal methods**.

> done

Li57: empirical studies **and in salmon population dynamics models**.
> done

Li69: milt **(sperm)**
> done

Li71: However, Hence,
> done

Li86: young salmon **(i.e. < 2 years)**
> we rather described « young » salmon as freshwater resident fish, because, parrs are of age 0, 1 or 2 in this study

Li118: oriented moved (?)
> "aligned with the lateral line"

Li124: use the same than Li158
> done

Li126-127: how did you calculated the probability?
> The significance of the above-mentioned effects was assessed using the z-value.

Table 1: see before (split in three parts ; one for each year of sampling)
> done

Figure 1: The figure 1 could be divided in two columns: left = phenotypic view and right = ultrasound image of the same sexual maturation state (three lines for three different stages). This could help us to better appreciate the differences between approaches
> We added three phenotypic views of the parrs to the figure.

Response to reviewer 2
The paper compares assessment of sexual maturation in Salmon parrs through visual and ultrasound examination. It concludes that ultrasound imaging of gonads provides a reliable assessment of sexual maturity, but requires skill and still presents a small percentage of error
(<3%).

General remarks
The study objectives are clear. The text is short and to the point. The results are unambiguous
regarding the quality of ultrasound imaging. So the study is already a success overall. Here are a few remarks, which, hopefully, should help the authors further improve it.

36–45 The first paragraph is very broad and somewhat vague. It would benefit from a few

concrete examples (possibly taken from papers already cited) of the impact of global change on variables measured in long term monitoring programs, of the difficulty to parameterise models due to lack of data, etc. This would make the motivation for this particular study more
apparent.
> As recommended, we added some example to motivate our study: "In the North Pacific, analyses confirm climate-related shifts in the abundance of most salmonid species since the 1940's, associated with reported ecosystem regime shifts (Irvine and Fukuwaka 2011). In the North Atlantic, the widespread decline in salmon abundance go with a decrease in survival and age at maturity (Chaput 2012) (…) tock assessments form the basis for catch advice for salmon fisheries, but they mask regional differences and annual river-specific stock assessments are only available for some 25% of the rivers (Chaput 2012)."

50 "frequency-dependent selection": do you consider this equivalent to "density-dependent selection" or is that something else? If non-trivial please explain.
> we explained this point as follow: "(…) frequency dependent selection, i.e. the reproductive success of precocious parrs increased with decreasing frequency of precocious maturation in the population (Berejikian et al. 2010)."

"mate competition": is that
the competition *for mates* (e.g. among males for access to females) or the competition *between mates* (i.e. between male and female). Make it explicit.
> We made it more explicit by referring to "intra-sexual competition for mates"

54 Define "smoltify", for readers out of the field.
> We changed it to "migrate at sea"

61–62 Is the 60% loss in stock estimated by Myers a general, average figure or an estimation based on only a few data points. If it is the former, this is major and should be emphasised more; if it is the latter, it should be specified.
> We explained in more details the data and spatial extent of this result as follow: "For instance, using more than 280,000 records from Little Codroy River parrs, Myers (1984) estimated that in Newfoundland 60% of the stock of adult male salmon was lost because of increased mortality due to precocious maturation".

125 How were the numbers of mature parrs detected by each method compared? Did you not compare the proportions rather than the numbers? Also, this sentence calls for a statistical test. Was one used (to compare the numbers/proportions)? If yes, specify which one.
> We compared the proportions of mature parrs detected by one or the other method each year with a Ch^2 test. To improve clarity, we define the number of mature parrs as the sum of individuals whose gonads were detected with the ultrasound and individuals whose gonads were not detected with the ultrasound but producing milt. We also added a sentence in the result section to compare the proportion of mature parrs detected by each of the two methods.

130 How was the probability to miss a mature male tested. I suppose with a GLM as

previously. Please specify it.
> We specified that the probability to miss a mature male was tested with a GLM

132 It is unlikely a R version from 2004 was used. Please update the citation (the R Core Team updates annually, see citation() for the up to date citation for R).
> We updated this citation

160–161 The interaction of age * length is not significant, which is actually good news for the interpretation of the effect of length by itself. And this pure factor effect is barely significant. So the discussion about the effect of length should be reviewed.
> Indeed the age effect appears more structuring than the length effect (but see new results below). Thus, we put more emphasis on this result and discuss the potential implication of the age-specific milting probability in mature males at the time of capture on our understanding of precocious maturation in parrs: "Empirical evidence have shown an increase in the proportion of mature parr as the parr are getting older (Baglinière and Maisse 1985; Myers et al. 1986). But this study suggests that mature parrs develop and produce milt earlier in the season than young ones, which may accentuate even further the above mentioned age-specific pattern reported with the phenotypic method."

161 The main difference in the probability to produce milt seems to be the among-years variability. While this is an important finding, it also suggests that this strong variability is likely
to hide the effect of other variables. It should be easier to partition that effect out, before investigating the effects of other variables, by using a mixed-effects model (and setting year and a random effect on either slope or, more likely and interestingly, intercept only).
*> as suggested, we separated out the test of the interannual variability and the test of the other variables. We explained this in the method section: "Empirical records show a large interannual variability in the proportion of maturing parrs (Baglinière and Maisse 1985; Myers et al. 1986). Similarly, the year of capture is expected to affect milting within mature parrs. Thus, a We tested the effect of year, age, body length in a generalized linear model (GLM) with a binomial distribution of errors is used to test for a year effect. Then, a mixed model with year as a random intercept and a binomial distribution of errors is used to test for the effect of age, body length, and their interaction." Actually, the age*length effect became more clear under this model framework. The result section was updated accordingly: "A strong year effect is detected ($p < 0.001$) in Tthe probability to produce milt at the time of capture in mature males. When accounting for this year effect as a random factor, a strong increase with age (estimate = 5.582, SE = 1.031, $p < 0.001$) and a negative interaction with body length within each age class (estimate = -0.031, SE = 0.008, $p < 0.001$) are detected. This means that a higher proportion of mature males produce milt in autumn in old parrs than in young ones, and within a given age class in small parrs than in large ones."*

165 Explain "uncertain": were those the parrs that looked mature but did not produce milt. Make this more explicit (and probably merge with the next sentence).
> we made this sentence more explicit:" In our study, the maturation state of 106 individuals was is uncertain according to the phenotypic method, i.e. individuals that look like mature parrs. However, following ultrasound examination, we it is found that 89.6 % of those parrs that looked like mature parrs were are actually indeed mature."

The operators, visually, missed 5.6% of the mature parrs; the ultrasound examination missed 2.7%. The difference between the two methods is not that large. So, in the conclusions, I would emphasise further the fact that the ultrasound approach is more objective while the visual inspection is more subjective and (possibly) requires more training.

> in the text, me put more emphasis on false positive and false negative data to clarify the amount and the type of errors encountered within each method. By accounting for both false positive and false negative data, we bring out a larger difference between the two methods (16% of errors with the phenotypic method, and 2.7% of errors with the ultrasound method). In the conclusion, we put more emphasis on the level of subjectivity/objectivity of the two methods: "Assessing precocious maturation in parr from phenotypic observation can reach a high level of accuracy in trained operators. However, this method relies on some subjective criteria and gut feeling, which render the level of accuracy in the data difficult to quantify. (…) Ultrasound imaging is an objective, easily transferable and replicable approach to investigate precocious maturity in parr as it relies on key diagnostic features".

However, this small difference makes it questionable to determine which method should be taken as the reference to determine the "mistakes" of the other. You seem to have taken the ultrasound as reference to compute the mistakes of the operators but could that be imperfect? Can there be false
positives?

> We assume that the ultrasound method cannot produce false positive data, because we only find gonads large enough to be detected in mature parrs. As a consequence, we defined our reference in the method section follow: "The number of mature parrs is defined as the sum of individuals whose gonads are detected with the ultrasound and individuals whose gonads are not detected with the ultrasound but producing milt."

Related to this, would you recommend to assess maturity through the production of milt (which is a objective and practical technique) and resort to ultrasound only for the ambiguous
cases, or do you consider that ultrasound-only is a better approach in terms of accuracy and fish well being? I would like to see a bit more opinionated discussion about this in the conclusion.

> We added to the conclusion our recommendation for assessing maturity in parrs: "In the end, this study calls for the assessment of precocious maturation in Atlantic salmon parrs using the two methods together: 1) testing whether parrs produce milt is a quick examination in the field and provides an objective and undeniable evidence of maturity, and 2) ultrasound imaging should remove any ambiguity about the state of maturation in parrs that do not produce milt. Therefore, reducing uncertainty in empirical data should offer new opportunities for further research on this alternative breeding strategy in salmon and improve our understanding of this key biological process."

Finally, in the conclusion, you write that the ultrasound imaging approach is "easily transferable". What do you mean exactly by that? Earlier, it was mentioned that the operation
of the ultrasound scanner and the reading of the ultrasounds required a skilled and trained

operator. Is this training easy to transfer (especially compared to the training to assess the fish visually, which seems to work pretty well too in the end)?
> To improve the clarity of the ms, the advices regarding ultrasound training were moved to the result and discussion paragraph dedicated to ultrasound. Instead, we described in more details the degree of "transferability" of the two methods: "A naïve operator can gain a good expertise and assess precocious maturation with the ultrasound method within a day, whereas more training would be required to achieve a similar level of expertise with the phenotypic method."

Language/formatting remarks
Many of these are *suggestions* for improved readability, written as imperative for shortness. Some are actual errors, but they should be easy to spot. The changes are marked as:
deletion
and **addition** or **change**.
Throughout the paper:
please ensure a correct sequence and homogeneity of tenses. Methods are first described in the present tense, then switch to past; the same is true for results. Pick one tense and stick with it. Past tense is usually easier. I personally find present tense more engaging if you can make it work

prefer passive forms to active ones, as usual in scientific papers. e.g. l. 96: "We assessed age through scale reading" -> "Age was assessed through scale reading".
> Done.

"gonads" should always be plural; there are two in each animal.
> This has been changed.

p-values are usually noted with a lowercase, italic, "p" rather than a capital "P".
> Done.

28 Reformulate as "By **allowing to see** the unseen"
> Done, here and in the conclusion.

37 "global changes": usually singular "global change". Change everywhere.
> Done.

38 now, this is plural "monitoring program**s**" "major change**s**"
> Done.

39 "Indeed" : no real causation here. "trigger" reads strange. Reformulate as "This global change makes it particularly important to be able to predict…"
> Done.

43 "sex" -> "gender"
> Done.

45 "individual-**level** data"

> Done.

49 “would” -> “can”
> Done.

50 “selection, as affected by **the** intensity”
> Done.

51 “threshold **that triggers precocious maturation**”
> Done.

52 “access to mating with large anadromous females **for mating**”
> Done.

53 “but**,** as”
> Done.

55 “frequency” -> “prevalence”?
> This has been changed by “occurrence”.

55 “space**,** and”: a comma before and is acceptable in enumerations of 3 or more elements and I find it makes them clearer. Also you used it on occasions so it needs to be consistent.
> this has been changed here and in the rest of the text.

56 “1986) but precise” break the sentence “1986). Yet, precise”
> Done.

57 “remain**s**” verbs at the third person of the present tense take an “s”. Beware, this is a frequent mistake of yours.
> Done.

58 “and **in the** management **of** Atlantic salmon **fisheries**”
> Done.

58 Delete the double parenthesis after the citation
> Done.

60 “survival **as well as** salmon” to avoid repeating a “and”
> Done.

62 “parr**s**”
> Done.

63 “of adult reproduction **by adults**”
> Done.

64 “secretive” what do you mean by that? Is that the fact that they are mature “in secret” (i.e.
not visible from their aspect), or that they “secrete” sperm hence are mature? I am not sure that “secretive” is necessary in either case but if you want to specify something, please make it more explicit.
> This has been deleted.

64–66 “parrs **is due** to the complexity of salmon**the** life cycle **of Atlantic salmon**. **But it** could also be partly due to the difficulty of quantifying the proportion of precocious**ly mature**
parrs in wild salmon populations” alternative ending “the frequency of occurrence of precious
maturation in wild salmon populations.”
> Done.

69 “extract milt **(sperm)**”
> Done.

70 “look” -> “aspect”
> Done.

72 “expert opinion**s** data in”
> Done.

72–73 “due to **the potential for incorrect diagnosis and strong operator bias**.”
> Done.

75 “offer**s**”
> Done.

76 “as **it** provides”
> Done.

77–78 remove “, e.g.” and add “in fish **for example**”
> Done.

81 “non-mature” + “salmon**s**”
> Done.

83 “offer**s**” + “out into the wild” -> “in the field”
> Done.

84 “scanner**s**” + “enhance” -> “increase”
> Done.

85 “salmonids population**s**”
> Done.

86 “field**,** using”
> Done.

87 “by testing” -> “and test”
> Done.

87–88 “operator-related”
> Done.

88 “factors would affect” + “assessment of **parr maturity**”
> Done.

89 “we provided” + “and **on** the interpretation”
> Done.

93 “the **r**iver Oir”
> Done.

94 “(parr**s**)”
> Done.

95 “standardized” + “protocol**. T**hey are **then** placed in a light an**a**esthetic **solution**,”
> Done.

96 “scanned**,**” + “0.5 hour” -> “30 min” + delete “maximum” which is redundant with “within”
> Done.

97 “2017**,**” + “examined **a total of** 850”
> Done.

98 “using **a** traditional**,** phenotypic (**i.e.** external)**,** approach”
> Done.

104 “of milt (sperm)” was moved earlier, the first time the word “milt” was used.
> Done.

105 “gonad**s** does not”
> Done.

106 “looks like”
> Done.

108 “but they do not” + “rely” -> “relies”
> Done.

109 “general**,** and subjective**,** appreciation”
> Done.

114–116 “**T**he default setting**,** for muscular examination**, was selected**.”
> Done.

115 “transmit**s** ultrasounds, we do not need to use ultrasound transmission gel **was not needed**”.
> Done.

116 “freshwater, and” a comma before “and” is not acceptable here, this is not an enumeration.
> Done.

120 “when the gonad**s were** detected”
> Done.

120–121 “when the gonad**s were too small to be** detected.” and delete the rest of the sentence.
> Done.

121 “Images” -> “Snapshots” or “Ultrasound snapshots”; change everywhere.
> Done.

122 “exported **to** a USB” + “for post-**processing**.”
> Done.

125 “non-mature” -> “immature”; you use both forms, please homogenise (and “immature” seems most natural) + “one or the other” -> “each”
> Done.

126–128 Merge into one sentence to make it immediately obvious what the “multivariate analysis” is: “A Generalized Linear Model (GLM) with a binomial distribution of errors was used to test which biological factors (among year, age, and body length) influenced the probability that a mature male was producing milt at the time of capture.”
> This paragraph has been reworded as the analysis was split into 2 (GLM then GLMM)

129 “mature but parr”
> Done.

130 “identity**,** and time of the day”
> Done.

131 “variability” -> “variance”
> Done.

137–138 Reformulate as “immature individuals, the liver, stomach, caecas, and intestine

could be identified." (also note that caeca is already plural, no s).
> Done.

138 "on the live images"
> Done.

139 "than on **the snapshots displayed here.**"
> Done.

140 "diagnostic **of immature individuals**, as well"
> Done.

141 "appear**s**" + "parr**s**"
> Done.

141–143 "males), as gonads virtually fill the whole **abdominal** cavity **and** ultrasound images" + "(Figure 1b,c)**; d**igestive organs are hardly visible and the shape of the cavity is convex."
> Done.

143 "does"
> Done.

146 "provides **the same** objective **diagnosis** of maturation status in"
> Done.

147 "and the ones that **did** (Figure 1c)."
> Done.

148–149 "discriminate between truly mature parr**s**" + "can mistakenly confuse **for** mature male**s upon visual inspection**, because of"
> Done.

152 "one-**year**-old"
> Done.

153 "young-**of**-**the**-year"
> Done.

154 "**r**iver Oir"
> Done.

155–156 "not shown; **therefore the left flank is towards the bottom of the image and the right towards the top**.
> Done.

156 "identified **by the letter "G"**."
> Done.

159 “of thematur**e** male**s did** not produc**e** milt.”
> Done.

166 “looked like mature parrs were **indeed** mature.”
> Done.

168 “operator**s**”
> Done.

170 “probability” -> “proportion” + Rephrase as “**This highlights how difficult it is to assess** parr maturation”
> Done.

172 “None of the other variable**s** had a significant effect”; move this right after (p=0.008) on line 170
> Done.

174–175 “**T**he sample size is too small to **assess the potential effect of external factors through multivariate analysis. Still, this shows that** ultrasound imaging”
> Done.

176 “The **correct** interpretation”
> Done.

177 “requires a good training to improve the accuracy of the information. **In hindsight, w**e also”
> Done.

178 “all **ultrasounds sh**ould”
> Done.

Table 1: “Parr” -> “Parrs” + “looks” -> “looked”
> Done.

185 “By **allowing to see** the unseen” + “offer**s**”
> Done.

187 “characterize **the** true biological state **of individuals**”
> Done.

188 “parr**s,** as”
> Done.

194 “issue**s**.”
> Done.