Comments on: MS *"Impact of ocean acidification and elevated temperatures on early juveniles of the polar shelled pteropod Limacina helicina: mortality, shell degradation, and shell growth"* Lischka et al. BGD doi:10.5194/bgd-7-1-2010

General comments:

This manuscript reports the results of an important and timely investigation of the combined effects of temperature and pCO_2 on the shell growth of a common Arctic pteropod. The work is not novel in design, however there is a critical lack of information on the responses of many common pteropod species to climate change, and the results presented here make a substantial contribution to our understanding of a key species in polar and boreal marine ecosystems.

The first three sections of the manuscript are very well presented, providing the reader with a clear, logical, informed, and concise understanding of the problem and the experimental approach. The Title accurately reflects the content of the manuscript, and the Abstract is especially well-written. The scientific methods are clearly described and reproduceable form the information provided, and the results are generally well presented. The Discussion, unfortunately, lacks the clarity of logic and language in the remainder of the manuscript and considerable re-working is required here (see below). In this particular context, the distinction between the authors' own work and conclusions /assumptions based on the work of others needs to be clarified through reference to the authors' own Tables / Figures and to the literature (respectively). The references are relevant and comprehensive without being extensive.

The primary conclusions reached are generally supported by the data, however there are some issues of interpretation (see Specific Comments, below). Notably, the authors' choice of a compound non-metric score for "shell degradation" has confounded the interpretation of the results and the analysis presented is of unknown value. I recommend strongly that this particular aspect of the work be re-analysed using a non-metric ordination technique (e.g. NMDS) and MANOVA (rather than ANOVA) as these tools have the capacity to reveal patterns of response to the temperature and pCO_2 treatments that have been masked by summing the scores for the different shell degradation categories. I've made specific recommendations below.

In summary the manuscript is clearly within the scope of *Biogeosciences* and of sufficient quality to warrant publication. However the issues raised above must be addressed and the many avoidable errors in language and interpretation outlined below must be corrected before the manuscript is acceptable for publication.

Jon. Havenhand

Specific Comments:

p 3 li 2 Use of "rose" is odd here. I'd suggest "have risen", and (later on the same line) "have decreased".

p 3 li 6 The subject of this sentence is surface waters of the Arctic, not undersaturated surface waters of the Arctic. Therefore the statement in parentheses "(and even more widespread . . .)" is inappropriate and out of context. I suggest deleting this.

p 5 li 1 I cannot see the justification for the argument here that fitness reductions will be especially likely if metabolism is low. Although possible, it doesn't logically follow that adaptively advantageous low winter metabolism will necessarily increase stress-susceptibility, and the authors present no reference to justify their assertion. As this argument isn't important for the overall point of this paragraph – juvenile pteropods, which are a vulnerable life-stage, will be

exposed to winter saturation levels whether stressed or not – I suggest deletion. If the authors wish instead to keep this argument they should justify it clearly.

p 6 li 1 The sentence beginning here is gramatically challenging. I suggest rewriting to something like "The combination of three temperatures and four pCO_2 levels permitted conclusions regarding the separate, as well as combined, effects of temperature and CO_2."

p 7 li 9 ". . . for each of the twelve." should read ". . . for each of the twelve treatment combinations."

p 7 li 14 What is meant here by "glasses"? Do the authors mean "jars"? If so they should use "jar" as this has been defined earlier. If not, then the authors should explain what a "glass" is.

p 7 li 18 Replace "(total alkalinity (A_T) and nutrients)" with "for total alkalinity (A_T), nutrients . . .".

p 7 li 23 Again there is a reference to "(A_T and nutrients)" which lacks adequate context. The reader is left to assume that the authors took the water samples in order to test for A_T and nutrients. This should be clarified.

p 8 li 14-15 Reference to measured *versus* target "dicksons" is terminology with which I (and I assume most readers) am not familiar. I assume the authors mean the difference between measured and target A_T in the CRMs from Andrew Dickson. This should be clarified.

p 8 li 15 What is meant by "(s.a.)" ? There is also no full-stop at the end of this sentence.

p 9 - 10 The "measure" of shell degradation stage is a compound classification and non-metric. i.e. it does not have traditional arithmetic properties - a value of 8 is not half the shell degradation of a value of 16, for example. Moreover, any given value of this "measure" can be obtained in multiple ways: a value of 8 could be obtained by a score of 2 on each of four categories, a score of 4 on only two categories, or any of a number of other combinations. Consequently, the use and interpretation of traditional, metric, statistical analyses such as ANOVA and multiple regression will be extremely problematic and possibly wholly misleading.

I suggest that the authors analyse the raw scores for each category using an ordination technique such as non-metric multidimensional scaling (NMDS), classifying the data according to treatment. Such an analysis will permit assessment of how temperature and pCO_2 influence the different categories of degradation, as well as degradation as a whole.

For the reasons above I recommend strongly against attempts to analyse the compound "shell degradation" score. Instead, MANOVA and multiple regression using all five degradation categories as response variables would preclude the obvious problems of the compound score.

Note that any analysis of the compound score must include consideration in the Discussion of the potentially misleading properties of this score (ie how the same score can reflect very different degrees of decomposition in the different categories). At present this is missing.

p 12 li 1 The transformation for Box-Cox is represented strangely (and has one too few parentheses). This should be corrected. "box-cox" should also be "Box-Cox" as these are proper names.

p 12 li 7 Multiple regression is a robust technique, however the metrics of shell-degradation used here are not arithmetically related (see above). Any quantitative analysis (such as multiple regression or ANOVA) will give spurious results at best, and present considerable difficulties in interpretation. Consequently this analytical approach should be avoided.

If the authors wish to understand the relationship between mortality and shell-degradation & growth, I recommend re-analysing these data using the separate scores for the different degradation categories rather than the combined total score. This approach would also permit

the use of Hierarchical Partitioning to assess which of the independent variables had the greatest partial correlation with mortality.

p 12 li 23 Reference to statistics should be accompanied by reference to the relevant table (Table 2). (Note this error appears throughout the manuscript and should be corrected throughout.)

p 12 li 24 The assertion that "the temperature effect was stronger . . .". is only supported by reference to Figs 5a and b. ANOVA results only allow interpretation of statistical difference, not magnitude of effect. The text needs to be amended to include reference either to Figs 5a,b or to indicate that the issue is dealt with below.

p 13 li 1 The fact that significant values in Table 2 are in italics should be stated in the legend to this table, not here in the text.

p 13 li 18 F-values cited in the text without reference to an ANOVA table should include degrees of freedom for the F-value. (This error should be corrected throughout the manuscript)

p 13 li 23 the use of "neither - nor" is inappropriate here. Suggest rewriting to ". . . but there was no significant effect of temperature (*F-value, df, p-value*) and no significant interaction (*F-value, df, p-value*) . . ."

Note that the text as originally written here is incorrect: a non-significant result from a statistical test does not mean there was "no effect" (as written) but rather that the result was inconclusive (ie the experiment/test could not detect an effect). Interpreting this result as "no effect" is a common, but fatal, error. The absence of evidence (ie a non-significant result) is not evidence of absence (ie no effect). See R.A. Fisher's 1939 book for more details.

p14 li 7 If the authors wish to claim the lack of a correlation they should provide the relevant statistics (presumably *r*), degrees of freedom and probability value.

p 15 li 8 Here again the authors refer to statistics without reference to the relevant Table or Figure. This is frustrating for the reader, and should be avoided. Cite the relevant Table or Figure here.

p 15 li 10 As noted earlier, this multiple regression analysis is unreliable for shell degradation. This discussion should be removed.

p 16 li 7 Calcein is widely known to show active calcification, whether this be at a growing margin or areas of repair. The statement here implies a "discovery" which is inappropriate. This should be rewritten accordingly.

p 16 li 20 Throughout this paragraph the language is unusual and here the use of "In principle" is out of context. Either these data show growth of juvenile *Limacina* or they don't. There is no "in principle" issue here. This should be rewritten.

p 16 li 27 "that means according to" should be replaced with "*sensu*"

p 17 li 4 The phrase "true shell increment" used here is required because the authors elsewhere use the term "increment" to indicate a lengthening of the shell spiral. This should be defined earlier, and the term "thickening" used here. This will avoid the obvious confusion here regarding "increment" (which actually means, "growth" or "increase" and is therefore equally applicable to length and thickness).

p 17 li 11 Again, the authors should refer the reader here to the relevant Table or Figure that shows this result.

p 17 li 12 "In that" is a strange use of English. "Consequently" or "Therefore" would be better here.

p 17 li 17 The “two questions” are actually three. This is perhaps a pedantic point, but the authors should be accurate in their statements.

Later in this same paragraph the use of “afford” is again unusual. This needs to be rewritten.

p 17 li 25 There was no *statistically* significant effect of temperature on shell growth in these experiments, however for the reasons outlined above this does not mean that “temperature had no significant effect”. The authors should clarify this issue here.

p 18 li 3 The language here is again unecessarily long and complex. The long clause “and furthermore . . .750 µatm” should be replaced with the much shorter “especially at 1100 µatm.”

p 18 li 4 I don’t see the logical justification for the sentence “Hence, shell degradation . . .” The authors’ own results show no statistically significant relationship between shell degradation state and growth or mortality (albeit through an admittedly problematic analysis - see above). While I accept that it’s likely that shell degradation will affect fitness, the test of this was non-significant and therefore there is no support for this assertion. The statement here implies that the authors have *shown* this to be the case. This statement should be toned down to reflect the fact that it’s a prediction rather than a result.

p 18 li 12 The opening statement of this paragraph needs to be rewritten. Contrary to the authors’ claim, the results as presented do <u>not</u> “suggest temperature to also have some influence” (on shell degradation). Table 3 shows clearly that the effect of temperature was not statistically significant, yet this fact is omitted from this discussion. The authors could claim that temperature *may* have an influence on shell degradation however in doing so they should highlight that they found no statistically significant support for this assertion. This section should be rewritten to include reference to Table 3 and to highlight the non-significant nature of this trend.

p 18 li 13 “have” should be “has”

This sentence is gramatically incorrect. The initial “If” requires a conditional “would be” later in the sentence rather than the “is” that is present here. I suggest changing the sentence to read: “Temperature did not significantly influence shell degradation (Table 3) however plots of degradation against temperature revealed a negative trend across all pCO_2 treatments (data not shown)”.

“data” is a plural and therefore “degradation data was” should be “degradation data were”.

p 18 li 15 The sentence beginning here is also badly constructed and confusing. I presume the authors are attempting to state that the highest incidence of perforated shells was at ambient temperatures and high pCO_2, and lower at higher temperatures.

This paragraph also raises the problem of how different values for the different components of the compound “measure” of shell degradation can yield the same overall value (many perforations and no corrosion is not equivalent to much corrosion and no perforations, yet they are numerically identical in the compound score). This problem with the shell degradation metric should be discussed here.

p 18 li 20 It’s not at all clear to me that calcification rate will *necessarily* be driven by Q_{10} rather than other temperature dependent processes. While I accept that this is possible, and perhaps even plausible, the wording here is too strong. “presumably” should be replaced with “perhaps”.

p 18 li 21 Again the language is confusing and gramatically incorrect. What do the authors mean here ? Can “by that being able to” be replaced by “therefore”? This makes sense to me, but it’s not clear that this is what the authors mean.

p 18 li 22 "Similar to this" can be replaced by "Similarly".

p 19 li 3 I don't see how this study has shown "plasticity" of juvenile *Limacina*. None of the results here demonstrate "plasticity" to the treatment variables, and nowhere else is "plasticity" discussed. Consequently I recommend removing "and plasticity" from this sentence.

p 19 li 5 "and" should be replaced with "but" – otherwise this sentence is syntactically incorrect.

p 19 li 7 Again the language here is awkward. I suggest this sentence be rewritten to begin: "pCO_2 change scenarios projected for the near future . . ."

p 19 li 13 The issue of metabolism and temperature arises here without citation and therefore this must be intended as a conclusion from this work. The study has, however, made no investigation of metabolic rate and temperature, and therefore this can only be conjecture. These sentences should be rewritten to reflect that, and appropriate references included to justify the assertions made.

p 19 li 17 The final sentence needs to be rewritten so that it is more easily understood by an English-speaking audience. I suggest: "To address these questions, detailed physiological studies of the effects of rising CO_2 and temperature on different life-history stages of *Limacina helicina* are needed."

Tables / Figures

Tables Significant results are represented in italics. This should be stated in all Table legends.

Only three significant figures are required in ANOVA tables.

Table 3 Legend states this is a test of pCO_2 on shell degradation, but table shows results for pCO_2 and temperature. Legend should be modified accordingly.

Table 4 The legend doesn't mention temperature. It should.

Table 5 Why is the ANOVA table presented here? A typical regression analysis table presenting the partial regression coefficients for each variable with respective *F* and *p*-values would be far more informative.

Figure 5 Legend defines vertical and horizontal error bars for both 5a and 5b yet the x-axis is different in these figures. Presumably the horizontal error bars in 5a are temperature? This needs to be corrected.

Figure 9 Are the "right" and "left" images reversed? The legend would imply so.