

## **Response to Reviewer Comments – Khan et al.**

Thank you for the opportunity to respond to the review. This response letter is formatted to give each editor/reviewer comment (numbered) followed by a numbered response to each comment in blue and quotes from the main text in italics.

### **Comment #1 (Editor):**

We have received an extremely thorough review wherein the reviewer outlines a number of issues with the manuscript which mean it can not be accepted in its current form. However, the reviewer gives clear guidelines for improvement which we hope you are willing to address. These comments relate to both the discussion of the statistical methods AND the biological significance of the results.

Due to requests from the editorial office we have also changed the reference style and removed the map figure (Figure 1).

### **Response #1:**

Thank you for the opportunity to respond to the reviewer comments. The comments were very useful and helped us to improve the paper by refining the focus of the paper. The major changes which we made were:

- Bringing a trained statistician (Ben Maslen) into the authorship group which ensured we are now comparing equivalent statistical tests. This involved removing the random forest analysis and replacing it with a distance-based which we then use to show how otolith data sometimes violates the assumptions of distance-based analyses and then directly compare this to the MGLM method.
- Revising the conclusions made regarding stock structure and being more careful in our interpretation. The conclusions are now more cautious, suggesting investigation of potential demographic differences rather than immediately calling for management changes.

### **Reviewer comments:**

#### **Comment #2**

The manuscript aimed at assessment of the chemical composition of otoliths and otolith shape variability as markers for the stock discrimination of the striped snakehead in India. The Authors applied a recently established statistical method (multivariate generalized linear model with Tweedy error distributions) to the otolith chemical composition and otolith shape data (wavelet coefficients). They identified differences in the chemical composition and shape of otolith between fish in three rivers. These results in the Authors' opinion suggest that there are at least three stocks of striped snakehead in India and future management considerations should be made at a regional scale.

**Response #2:** This is an accurate summary of our manuscript.

### Comment #3

I found the manuscript interesting, but unfortunately, the Authors failed to demonstrate how the proposed method is advantageous to the currently used methods of the analysis. Authors claim that they “see a number of advantages in applying MGLMs to otolith data”, while only two are listed in the manuscript discussion:

- “Firstly, the assumptions of MGLMs can be easily checked and the appropriateness of the models assessed before any inference is made from the results”

- “The latent variable model-based ordinations successfully visualized the multivariate differences identified in the MGLMs”

In my opinion, both aspects (check of model assumptions and visualization of the multidimensional differences) can be easily obtained with the available methods. Therefore, I do not see any real advantage of using this method with the cost of so high computational time (e.g. line 279: “138 hours for the shape and element data”). In my opinion, if one of the main aims in the study is to show the applicability of the MGLM to the otolith shape and chemical data, there is a necessity of profound arguments that would be convincing for the Reader. In this case, the Authors applied both MGLM and already established in the field method of random forests, giving the performance metrics (accuracy) for random forests, but not for MGLM. Some kind of comparison of the performance of the proposed MGLM framework and current methods should be presented, together with a clear description of the MGLM properties which are advantageous for the analysis of the otolith data.

### Response #3:

The comparison of MGLMs with a directly comparable widely used alternative was a great suggestion. We have now brought a trained statistician into the authorship group (Ben Maslen) and with their help we now directly compare a distance based PERMANOVA and NMDS approach with the MGLM and latent variable ordinations.

This has resulted in the removal of the random forest analysis as it is not a directly comparable approach to the hypothesis testing of MGLMs.

The advantages of taking an MGLM approach over a conventional PERMANOVA approach has been discussed in far greater detail throughout the paper and in particular in the discussion section. The main advantage of taking an MGLM approach is highlighted as the ability to take into account the mean-variance relationship of the data. By failing to take into account the strong mean-variance relationship of the otolith data the PERMANOVA approach can provide misleading results (which we demonstrate with our own data); confounding differences in variance with differences in the mean within ordinations, higher type 1 error rates and lower statistical power, as well as a failure to detect multivariate effects unless it is expressed in high-variance responses (Warton et al. 2012).

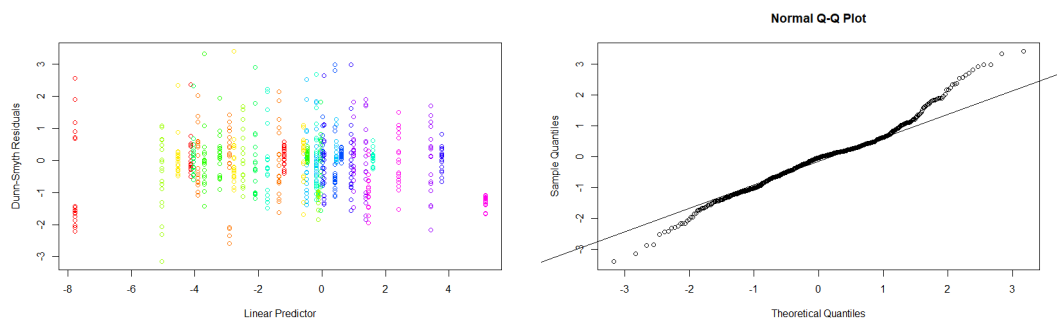
### Comment #4

Also, the main references cited by the Authors regarding MGLM (Warton et al. 2012, Warton 2008 and Wang et al. 2012) present MGLM applications for multivariate analysis of ecological count data (abundances). The specific adaptation of the method for the data which are normally distributed or can be transformed to achieve normal distribution should be more deeply described in the manuscript. The Tweedie distribution offers a flexible alternative to the quasi-Poisson and negative binomial distributions as a response distribution when modeling count data (see e.g. Miller et al., 2013), but here it is applied for positive continuous data (line 166). When running provided R code (files “Combined\_Tweedie.R” or “elements\_tweedie.R”, in lines 27 and 28 of the script), qqplot shows considerable deviations of the model residuals. It probably suggests that otolith shape and

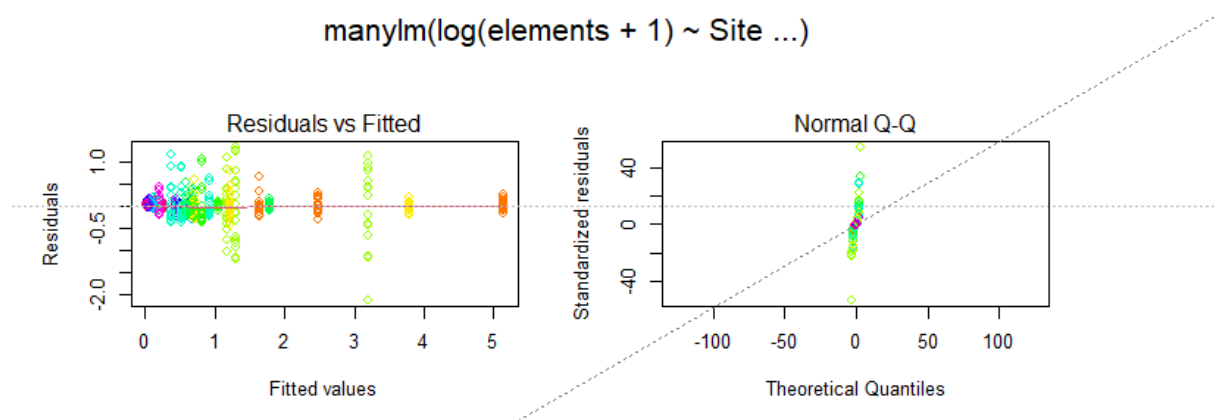
chemical data do not follow the assumed Tweedie distribution, while the Authors claim that “residual plots were inspected for each model. No strong pattern were visible and the models were deemed to be accurately representing our data (Figure 3), allowing the use of these models to address our hypothesis.” (lines 176-179). In fact, Figure 3 presents a different type of the residual diagnostics (based on the randomized quantile residuals of Dunn and Smyth), but contrasting both pictures is confusing.

#### Response #4:

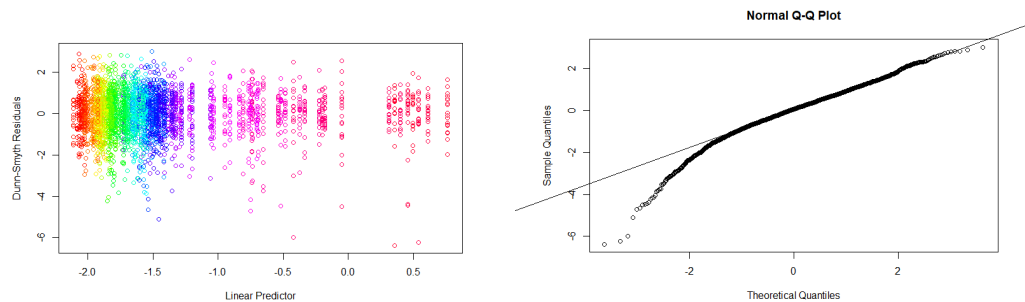
The error distributions for the data have been re-assessed by the Statistician that has been brought into the authorship group. He has tuned the var.power parameter for the otolith chemistry data to satisfy the mean-variance relationship of the data resulting in a much better fit with the Tweedie distribution which now provides a much better fit (although still with some deviation in the tails of the qqplot plot):



We note in the paper that taking a  $\log(y+1)$  transform and modelling this data in a multivariate gaussian manner is not appropriate as this transform does not take into account the null observations, nor is it properly dealing with the mean-variance trend. This can be demonstrated by taking a  $\log(y+1)$  tranform and fitting a ‘many1m’ model resulting in the following diagnostics which is clearly not appropriate:

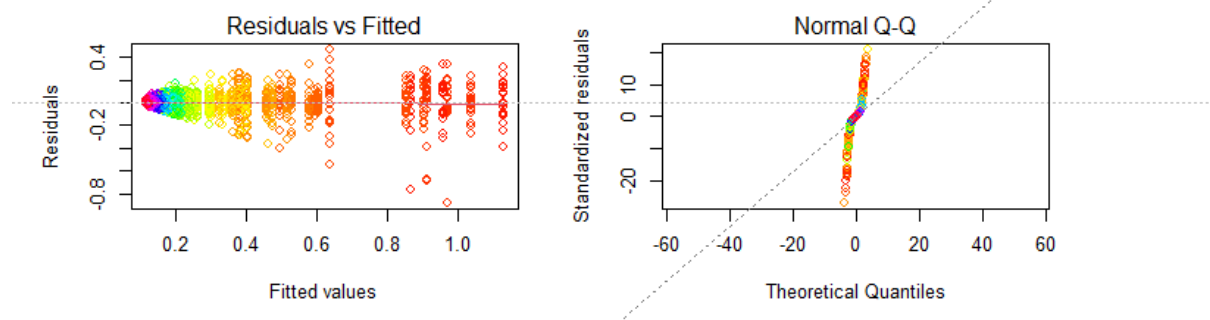


The shape only data has then been changed to fit a gamma MGLM (due to the absence of zeros in the data set) which vastly improved the model fit, with the following model diagnostic plots:



Again taking a log +1 transform and running a multivariate normal model is also not sufficient here:

`manylm(log(elements + 1) ~ pop ...)`



#### Comment #5

There is a lack of a clear description of the concept of stock and population in the manuscript and both terms are used interchangeably, while they have different meanings (Cadrin et al., 2014). Giving the high differences in the river environments, it is not fully unexpected to see high differences, especially in the otolith chemistry, but also in otolith shape. Such differences can be visible even in the closely located tributaries (e.g. Zeigler and Whitledge, 2011). In my opinion, these differences are, however, not enough to claim the existence of separate stocks and the necessity of the separate management actions for each river. These are the consecutive differences in the life-history parameters and the dynamics of the subgroups that should be evaluated in the next step. A stock can be defined variously by its ecological, technical, recreational, economic, or fishery attributes (Cadrin et al., 2014). Therefore, in my opinion, the main conclusions of the manuscript (suggesting serious consequences for the management practices) are not supported by the presented data.

#### Response #5:

Thank you for the references. This poor use of terminology (stocks) has now been revised throughout the manuscript. We have also removed any recommendations for management changes and instead suggest further research into potential differences in demographic dynamics as suggested by the reviewer.

## Specific comments:

### Comment #6

-line 23: Tweedie should be capitalized, here and through the manuscript

**Response #6:** This change has been made.

### Comment #7

-lines 24-25: only the accuracy of the random forest was reported, not for MGLM, so the methods were not directly compared

**Response #7:** As the reviewer points out, the random forest model and the MGLM were not directly compared in terms of accuracy. As these are fundamentally different analysis techniques (hypothesis testing against classification analysis), the random forest was entirely removed from this paper and replaced with the most common distance-based hypothesis testing method, PERMANOVA. See comment # 3 for more information on this change.

### Comment #8

-lines 29-41: please be more specific, give some examples

**Response #8:** We think this refers to the original lines 39-41 rather than 29-41 and have added some examples here. The line now reads:

*"Otoliths are a common tool used for stock discrimination and numerous studies have shown the potential of otoliths in addressing research problems related to successful fishery resource management (Tracey et al. 2006, Ferguson et al. 2011). Both otolith shape and elemental composition have become popular and successful tools in discriminating fish stocks (Campana and Casselman 1993; Begg et al. 2001; Miyan et al. 2016a; Nazir and Khan 2019)."*

### Comment #9

-lines 58-59: please be more precise and give the explanation of "more subtle differences"

**Response #9:** This sentence has been revised. It now reads:

*"By combining both otolith shape and chemistry data in the same analysis the ability to differentiate groups of fish can sometimes be improved (Fowler et al., 2015)."*

### Comment #10

-line 67: are they really "becoming the preferred analysis framework"?

**Response #10:** We believe this to be true but we have toned down the language here to be more cautious. This line now reads:

*"For ecological studies using multivariate abundance data such as species abundances, multivariate generalised linear models (MGLMs) are becoming more popular as they allow increased certainty and interpretability of the results, flexibility, and efficiency compared to distance-based methods (Warton et al 2015)"*

### Comment #11

-lines 67-68: "allow increased certainty and interpretability of the results, flexibility, and efficiency". Please say increased in comparison to what?

**Response #11:** In comparison to distance-based methods. This has been added to the text and now reads:

*"For ecological studies using multivariate abundance data such as species abundances, multivariate generalised linear models (MGLMs) are becoming more popular as they allow increased certainty and interpretability of the results, flexibility, and efficiency compared to distance-based methods (Warton et al 2015)."*

**Comment #12**

-line 71: consider a change from “fitting” to “choosing”, or “selecting”

**Response #12:** This sentence has been removed.

**Comment #13**

-lines 72-73: “and may increase certainty in stock discrimination scenarios”. It is not clear how it may increase certainty.

**Response #13:** This sentence has been removed.

**Comment #14**

-line 104-105: Basic characteristic of the samples (length distribution) is presented in Supplementary materials. It would be suitable to investigate the differences in the length of an individual between rivers. Although the effect of individual length was considered in further analysis and a correction was made for the length of the fish (lines 127-128), and this issue was addressed in the discussion (lines 317-324), the differences in length between rivers may influence the results and the observed differences in the shape of the otolith (see e.g. Vignon, 2012).

**Response #14:** As the reviewer states, there is a natural possibility of length biasing the analysis but this was controlled for and also addressed in the discussion. We do not believe there is any advantage to the reader to further show that fish at one site were longer than the other sites as this is not the main point of the paper

**Comment #15**

-Lines 114-116: It should be clearly described how outlines were extracted first. Was e.g. any smoothing applied?

**Response #15:** We have added the additional detail to the paragraph. The full analysis and method can also be reproduced with the provided code.

The paragraph now reads;

*“... All otoliths were photographed using a light microscope and reflected light with the otolith placed with distal surface up on a black background. The procedure followed is fully detailed in Libungan and Pálsson (2015) although some photos of otoliths needed manual editing to accurately capture the otolith outlines. Once the photos were captured the outlines of the otoliths were smoothed to remove high frequency pixel noise around the otolith outlines using the smoothout() function with 100 iterations. The wavelet method then fitted a series of approximating functions within restricted domains to quantify the outline shapes (Graps 1995)....”*

**Comment #16**

-Lines 124-125: The wavelet method was found to more accurately reproduce the shape of the otoliths and there was used for the remaining analysis. How this was tested? Please indicate the results of the preliminary test.

**Response #16:** We have now added in the details of this comparison. The test reads:

*“Both methods result in coefficients which can be used to quantify the shape. Using 10 wavelets (63 wavelet coefficients), >99 % of otolith shape was explained as opposed to the elliptical Fourier transformed coefficients which were only able to reproduce 95% of the shape (Fourier transformed results not shown) and we therefore proceeded only with the wavelet analysis.”*

**Comment #17**

-Lines 141-142: Consider scientific notation.

**Response #17:** We have not made this suggestion as we would prefer to not use scientific notation as we believe the current decimal layout is easier for readers without strong mathematical background.

**Comment #18**

-Lines 158-166: This description seems to be a bit awkward but maybe it is a matter of taste.

**Response #18:** This section has been reorganised and we now believe it to be clearer for the reader.

**Comment #19**

-Lines 182-183: How multicollinearity of the predictors was assessed and addressed in the study?

**Response #19:** There was no collinearity as we only 1 predictor (Site) in this analysis.

**Comment #20**

-Lines 196-200: The concept of the visualization based on the latent variables is a fresh and interesting element of this study. Please explain the concept and statistical background for the latent variable ordinations. I assume that many Readers would be not familiar with this specific approach.

**Response #20:** To give the readers more context for this method the following has been added to the methods section:

*“The boral package takes generalised linear models for each response variable, using Bayesian Markov chain Monte Carlo methods to estimate latent variables that account for between response correlation, which can then be used to visualise multivariate differences on a low-dimension plot (Hui 2016). By using generalised linear models, this approach can align the visualisation model with the testing model, check assumptions and specify mean-variance relationships.”*

**Comment #21**

-lines 213-215: I have the impression cited paper by Dushoff et al., 2019 refers to the language used in the scientific articles, rather than the way of reporting p-values. Reporting of the exact p-values is the approach currently preferred in many scientific journals.

**Response #21:** this is a correct summary and we agree exact p-values are preferred. For clarity we have removed the reference to Dushoff et al 2019 and provide exact p-values to 3 decimal places throughout the paper.

**Comment #22**

-Line 220: This figure doesn't allow for the assessment of how much of the variance is explained by the coefficients.

**Response #22:** We have added this information to the figure caption. The caption now reads:

*“Figure 4 Mean otolith shape from the three sites. The solid black line represents Agra, dashed red line represents Lucknow and the dotted blue line represents Narora. The wavelet coefficients recreated over 99% of the variance in otolith shape.”*

**Comment #23**

-Lines 240-248: So, there is a natural question that many Readers would ask: what is the classification accuracy for the proposed method?

**Response #23:** there is no accuracy as such for an hypothesis test. This misunderstanding has been fixed by removing the random forest classification analysis.

**Comment #24**

-Line 253: Something is missing in the sentence.

**Response #24:** This sentence has been removed.



#### **Comment #25**

-Lines 255-256: Please see my general comment. Differences in the otolith chemistry and shape between sites not necessarily directly imply that subgroups need to be managed separately. These are the consecutive differences in the life-history traits and the dynamics of the subgroups that should be evaluated in the next step. In my opinion, this study doesn't provide "strong evidence that *C. striata* should be managed on a regional rather than at a national scale".

**Response #25:** We have significantly changed the wording of our findings and no longer suggest that management should be changed. The new section reads:

*"The distinct otolith chemistry and shapes suggests that *C. striata* in these three rivers are not regularly mixing. This confirms recent research which used truss morphometry based upon body shape of *C. striata* to suggest that the same three groups analysed in the current study may be distinct sub-populations (Khan et al. 2019). Further research should next examine key demographic dynamics at each of these three sites including growth rates and age of maturity. If the demographics at each site also differ then management changes may be required."*

#### **Comment #26**

-Lines 256-259: Some examples needed. And what is the advantage of the proposed method over the previously used methods (e.g. classical LDA)?

**Response #26:** We have added some more detail here. It now reads:

*"We showed that distance-based analyses including PERMANOVA are not appropriate for some otolith chemistry data due to violations of the assumption of homogeneity of variance stemming from a non-linear mean variance relationship in the data. This mean-variance relationship can be directly modelled with MGLMs which we then use to show that *Channa striata* from three sites have clear differences in both otolith chemistry and otolith shape based."*

#### **Comment #27**

Table 1 shows the results of the univariate GLMs applied to the chemical data. There is a lack of the actual discussion of the results presented in Table 1. Why the differences between the rivers for some of the elements are higher than the others? Different contributions of the studied elements should be discussed in the appropriate section of the manuscript.

**Response #27:** This is discussed in the discussion section 4.2. We are unable to comment on some elements specifically due to uncertainty in the processes of elemental uptake in the otoliths. As this paper was primarily concerned with multivariate differences we do not consider addressing each element and their drivers to be an important addition and it is likely to be addressed in future research.

*"The unusually high concentrations of some elements in the otoliths likely reflects a heavily polluted environment as it is known that in India there continues to be concerns around pollution of waterways (Sengupta 2006). The Yamuna river is very polluted due to many cities lying on its bank and pouring sewage and other industrial effluents directly into the river. For this reason, the Yamuna river is recognized as one of the most polluted in the world (Bhardwaj et al., 2017). Our fish from the Agra site were located on the Yamuna river and their otoliths are reflective of the heavily polluted state with high concentrations of many elements, particularly heavy metals. It should be noted that fish at the Agra site were also bigger than the other sites (Table S1) but as we used whole otolith elemental composition and controlled for length in the shape analysis, the comparison of differences remains valid as there were very large differences between all three sites, particularly in the elemental composition of the otoliths. There were variations many elements which contributed to the multivariate differences discussed in the current paper and the drivers behind the specific elemental differences, whether natural or potential pollution present the opportunity for future study."*



**Comment #28**

Consider the description of the Dunn-Smyth residual plots presented in Figure 3 in the method section.

**Response #28:** To give more context to the use of Dunn-Smyth residuals, the following has been added to the introduction section:

*“Assumptions for these models can also be readily checked by plotting Dunn-Smyth residuals (Dunn and Smyth 1996) which are randomised quantile residuals that have been shown to be effective at detecting many forms of model misspecification for generalised linear models (Dunn and Smyth 2018)”*