**Title:** Stock delineation of Striped Snakehead, *Channa striata* using multivariate generalised linear models with otolith shape and chemistry data

**Authors:** Salman Khan1#, Hayden T. Schilling2,3#, Mohammad Afzal Khan1, Devendra Kumar Patel4, Kaish Miyan1

1. Section of Fishery Science and Aquaculture, Department of Zoology, Aligarh Muslim University, Aligarh -202 002, India
2. Centre for Marine Science & Innovation, UNSW Australia, Sydney 2052, Australia
3. Sydney Institute of Marine Science, Chowder Bay Road, Mosman 2088, Australia
4. Indian Institute of Toxicology Research, Council of Scientific and Industrial Research, Lucknow-226 001, India

# SK and HTS are joint first authors.

Corresponding author: Hayden T. Schilling, [h.schilling@unsw.edu.au](mailto:h.schilling@unsw.edu.au) , +614 02 9435 4600 (no fax)

**Abstract:**

Otoliths are commonly used to discriminate between fish stocks, through both elemental composition and otolith shape. Both otolith elemental composition and shape are multivariate datasets which are suitable for use within a multivariate generalised linear model (MGLM) framework, yet MGLMs have never been applied to otolith data. Here we apply MGLMs to a case study of *Channa striata* (Striped Snakehead) in India and show that when using a tweedie error distribution, MGLMs meet all assumptions and are appropriate for both otolith shape and elemental composition data, giving similar results to a random forest analysis method. Consistent differences between 3 groups of *C. striata* were identified using both otolith shape, otolith chemistry and a combined otolith shape and chemistry dataset. This suggests that there are at least 3 stocks of *C. striata* in India and future management considerations should be made at a regional scale. The MGLM method is widely applicable and could be applied to any multivariate otolith shape or elemental composition dataset.

**Keywords:**

Otolith shape, otolith chemistry, multivariate generalised linear models, India, stock discrimination

1. **Introduction:**

Natural markers such as genetic, elemental or morphological markers can be used to delineate populations or stocks, providing important information for fisheries management (Turner et al. 2015). 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. 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. 2016; Nazir and Khan 2019).

Differences in the shape of otoliths can help to discriminate between groups of fish that are at least partly separated, inhabiting different environments (Bird et al., 1986; Campana and Casselman, 1993; Begg et al., 2001; Smith et al., 2002) Variations in otolith shape increase with the extent of genetic discreteness or geographic separation (Castonguay et al. 1991; Friedland and Reddin 1994), although disentangling the physiological and environmental influences is often complicated (Vignon and Morat 2010). Similarly, the elemental composition of otoliths can also be used to distinguish between fish populations (Campana et al. 2000; Longmore et al, 2011). Minor and trace elements laid down within the protein matrix become a permanent record of the chemical characteristics of the environment experienced by the fish (Elsdon and Gillanders 2003; Stransky et al. 2005). While both physiological and environmental factors influence the elemental composition of otoliths (Grammar et al 2017, Izzo et al 2018), if fish inhabit different water masses or environments for a certain period of time they can be differentiated via the elemental composition of their otoliths (Elsdon and Gillanders 2004; Khan et al. 2012; Miyan et al., 2014; 2016a; 2016b). Even more subtle differences between groups of fish can be detected by combining both otolith shape and chemistry data (Fowler et al., 2015).

Both otolith shape and otolith chemistry data are usually multivariate with traditional analyses either using distance-based methods (eg. PERMANOVA; Schilling et al. 2018) or model-based methods which assume a gaussian error distribution (eg. MANOVA or LDA; Maguffee et al 2019). More recently however, machine learning classification techniques have become prominent as they are robust to many assumptions that are often hard for traditional methods to satisfy (Jones et al., 2017). For ecological studies using multivariate abundance data such as species abundances, multivariate generalised linear models (MGLMs) are becoming the preferred analysis framework as they allow increased certainty and interpretability of the results, flexibility, and efficiency (Warton et al 2015). While MGLMs are now common for abundance data (Wang et al 2012), they are rarely used for other datasets despite the flexibility of the method which allows users to specify model parameters to fit a dataset. By fitting appropriate error distributions to MGLMs, this method is suitable for use in the analysis of otolith shape and elemental data and may increase certainty in stock discrimination scenarios.

*Channa striata*, locally known in India as “Dharidar-Sol” or “striped snakehead”, is commercially important in food, ornamental and sport fisheries along with other species of the family Channidae. *C. striata* is one of the main food fishes in Asian countries including India. In the last few years, due to increasing anthropogenic activities, unstrained harvesting and habitat alterations, the natural stocks of the fish have decreased severely (Rahman and Awal, 2016). Consequently, feeding and natural breeding grounds of this economically important fish species have been reduced, which has caused a shrinkage in wild populations (Rahman and Awal, 2016). The present study was carried out with the dual aim of firstly, demonstrating the use of MGLMs with otolith chemistry and otolith shape data, and secondly, using both otolith chemistry and shape data to assess if there are multiple stocks of *C. striata* in India.

1. **Methods and Materials**

*2.1 Study species, region, and sample collection*

The striped snakehead, *Channa striata* is native to east and southeast Asia. It is found in India, Pakistan, southern Nepal, Sri Lanka, Bhutan, southern China, Bangladesh, and all the countries of southeast Asia. It is also native to the major western islands of the Malay Archipelago, including Sumatra, Borneo and Java. The species has been introduced to the Philippines, eastern islands of Indonesia, New Caledonia, New Guinea, Fiji, south-eastern Russia and South Korea (Phen et al., 2004; War et al., 2011). *C. striata* can be found in many types of slow-moving freshwater habitat, including rivers, ponds, lakes, creeks, canals, flooded rice paddies, swamps, and irrigation reservoirs (Cagauan 2007).

Eighteen *C. striata* were collected from each of three locations. Each site was located on a different major river in northern India with fish collected regularly from each site between October 2017 and November 2018 using cast nets (25mm mesh) and drag nets (28mm mesh). The three locations were Narora (27° 30' N; 78° 25' E) on the river Ganga, Agra (27.1767° N; 78.0081°E) on the river Yamuna and Lucknow (26° 55' N; 80° 59' E) on the river Gomti (Figure 1). Identification of the fish was based on the descriptions of Jayaram (1999); Talwar and Jhingaran (1991). Total length was measured to the nearest mm. Otoliths were extracted using forceps, cleaned in fresh water and stored dry before subsequent shape and chemical analysis. Full details of fish used in this study can be seen in Table S1.

All methods were carried out in accordance with the relevant guidelines and regulations. The target fish species is a commercially exploited common food fish in India; therefore the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA) 2018, Ministry of Environment, Forests and Climate Change, Government of India, does not require ethical approval to be given for this study.

*2.2 Otolith shape*

The shape of the otoliths was quantified using wavelet coefficients using R v3.6.0 (R Core Team, 2019). The R package *‘shapeR*’ was used to calculate both Normalized Elliptic Fourier and the discrete wavelet coefficients using photographs of each otolith which create mathematical representations of the otolith outlines. 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. Briefly, the wavelet method fits a series of approximating functions within restricted domains to quantify the outline shapes (Graps 1995). The elliptical Fourier method by contrast fits a number of harmonic functions to capture crenulations and lobes on the edges of the otoliths (Tracey et al. 2006). Both methods result in coefficients which can be used to quantify the shape. The wavelet method was found to more accurately reproduce the shape of the otoliths and there was used for the remaining analysis.

To visualise the difference in mean shape between the three sites, the mean shape was reconstructed using the mean wavelets for each site. Wavelet coefficients were standardised for fish length as per Libungan and Pálsson (2015) before analysis to test for differences between the three sites.

*2.3 Otolith chemistry*

To remove any surface contamination, otoliths were soaked in 3 % hydrogen peroxide for 5 min and immersed for 5 min in 1 % HNO3. Otoliths were then flooded with ultra-pure water for 5 min to remove the acid. After decontamination, the otoliths were dried under a laminar flow hood and weighed to the nearest 0.1 mg (Turan 2006; Khan et al. 2012). For analysis, the decontaminated otoliths were dissolved in 10 ml of 37 % HNO3 and the volume was brought up to 25 ml with Milli Q water. Elemental composition of whole otoliths were analysed using inductively coupled plasma atomic emission spectrometry (ICP–AES; Thermo Electron IRIS Intrepid II XSP DUO). Blank samples were used to correct for background noise in readings. The elements (and detection limits in ppm) measured from the otoliths included: Ca (0.005), Na (0.05), Mg (0.0005), Sr (0.0005), Ba (0.0005), Mn (0.001), Fe (0.005), Pb (0.05), Ni (0.005), Zn (0.005), Cd (0.005), Cr (0.005) and K (0.1). All elements were consistently above minimum detection levels. Internal standards Indium (In) and Gallium (Ga) were added in samples and blanks, which were used to correct for the remaining matrix effect and to compensate for instrument drift. Multi elemental standards were prepared with high purity ICP multi-element standard solution IV certiPUR (NIST SRM) obtained from Merck (Germany) using Milli-Q water and analytical grade 2%v/v HNO3 for external calibration. Standards were run every 10 samples. A calibration blank was also prepared in the same procedure. The calibration curve was obtained for five points. The concentration of elements in the sample and blank were calculated and expressed as µg g-1 (ppm) on dry weight basis (Turan, 2006; Miyan et al., 2016b). All elemental concentrations were converted from ppm to ratios of element:Calcium (mmol:mol) to control for the size of each analysed otolith.

*2.4 Statistics*

All analysis was performed in R v3.6.0 (R Core Team, 2019). For the model-based MGLM analysis we followed the analysis guidelines provided in Warton et al. (2015), following a defined modelling process. We first identified our question: Are there differences in otolith chemistry or otolith shape between the three groups of *C. striata*? Secondly, we considered our data. We had only one predictor variable, Site (a categorical variable) and many response variables (all the elemental concentrations and shape coefficients). Thirdly, we conducted exploratory data analysis but as we only had a single categorical predictor variable this was limited. Next, we selected an appropriate model for the question. Our goal was to compare means between three groups using multivariate data and our *a priori* hypothesis was that there will be multivariate differences between the three sites. Both the otolith chemistry and shape data are positive continuous data, therefore, tweedie error distributions were considered as the most appropriate fit for our data. We therefore used multivariate generalised linear models (MGLMs) with a tweedie error distributions (variance power 1.01) to test for our hypothesis. When using multivariate models it is important to understand the relationship between the mean of each response variable and the observed variance (Warton, 2008, Warton et al., 2012). To investigate this relationship in our data, we created mean variance plots which show how the variance changes with the mean of each variable. The mean variance plots identified that for both chemistry and shape data, as the mean increased, the variance also increased (Figure 2). As a final step prior to inspecting the results, we assessed our models. To assess if the MGLMs with tweedie error distributions accurately captured the properties of our data, 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. All MGLM models were run using the *‘manyany’* function in the *‘mvabund’* R package (Wang et al., 2012).

To compare the effectiveness of otolith chemistry and otolith shape in discriminating the three sites, three MGLMs were run. One only used otolith chemistry data, one only used otolith shape data and one combined both chemistry and shape data. For the two MGLMs involving the elemental data, univariate generalised linear models (GLMs; also with a tweedie error distribution) were also ran for each variable to identify which variables were driving the differences. This was conducted within the ‘*manyany*’ function. The influence of each variable in driving the differences (similar objective to a distance-based SIMPER analysis) was quantified using the individual contribution to the Sum-of-LR (Warton et al. 2012), whereby variables with a larger likelihood ratio value are more influential. For the GLMs which included shape data there is no meaningful interpretation of the univariate GLMs as the wavelet shape coefficients cannot be interpreted individually but it does allow the relative contributions of otolith chemistry and shape to be assessed in the combined model. Posthoc tests to identify which sites had showed evidence of differences in specific otolith elemental concentrations were run manually using two sites at a time using the same *‘manyany’* GLM process and adjusting the *P*-values using the Bonferroni method with the ‘*padjust*’ funtion. To visualise the multivariate differences between the 3 fish groups (as an alternative to the commonly applied distance-based ordinations), two factor model-based latent variable ordinations were produced using the *boral* R package (Hui 2016), again using tweedie error distributions with the assumptions being visually assessed (Hui et al. 2015).

To confirm the validity of the MGLM method we also conducted a random forest classification analysis as an alternative method of stock delineation. Using the ‘randomForest’ R package (Liaw & Weiner 2002), we again conducted three separate analyses, for the otolith chemistry data, otolith shape data and combined otolith chemistry and shape data. For each analysis we generated 5000 trees using all the variables in each dataset. At each split in the tree the number of variables was determined by the square root of total number of the variables in the analysis (number of variables at each split = 3 for the otolith chemistry data, 7 for the otolith shape data and 8 for the combined otolith chemistry and shape data; Breiman, 2001). To determine the classification accuracy of assigning fish to the correct collection site and therefore if the sites represented distinct populations, we calculated the average Out Of Bag (OOB, a cross validation method, Breiman, 2001) classification error.

In order to move away from a hard *P*-value threshold, all statistical results in this paper are reported through the lens of ‘statistical clarity’ with exact *P*-values (Dushoff et al., 2019). The code and data used in these analysis is available at: <https://github.com/HaydenSchilling/MGLMs-Otoliths>

1. **Results**

*3.1 MGLM Analyses*

Using 10 wavelets (63 wavelet coefficients), >99 % of otolith shape was explained (Figure 4) as opposed to the elliptical Fourier transformed coefficients which were only able to reproduce 95% of the shape. Using the wavelet coefficients, the MGLM analysis showed clear difference in otolith shape between all three sites (*LR*: 12.385, *P* < 0.001; Figure 4).

Otolith chemistry was also clearly different between the three sites (*LR* = 7161, *P* < 0.001; Figure 5; Table 1). Large differences in mean concentration were observed for many elements with the Agra site having the highest concentrations of 10 of the 12 tested elements (Figure 5). The Narora site had the highest concentrations of the other two elements (Zinc & Magnesium; Figure 5). The Lucknow site showed the lowest concentrations of all elements (Figure 5).

The combined analysis of otolith chemistry and shape also revealed clear differences between all three sites (*LR* = 7173.4, *P* < 0.001). Within this combined analysis most of the differences were driven by the chemistry data (99.8% of the *LR* ratio was made up by the element data).

The differences identified by the MGLMs between sites were visible in the latent variable ordinations (Figure 6). Similar patterns were visible to those identified in the multivariate generalised linear models with larger differences evident in the otolith chemistry data (Figure 6a) than the otolith shape data (Figure 6b). When both datasets were combined, the sites were the most tightly grouped and distinct (Figure 6c).

*3.2 Random Forest Analysis*

Both the random forest analysis using otolith chemistry data and the random forest analysis using combined otolith chemistry and shape data had 100% classification accuracy (0% OOB classification error) clearly suggesting the three sites are distinct populations. This strong classification was clearly driven by the differences in otolith chemistry as the random forest analysis using only otolith shape had a lower but still high overall accuracy of 74.07% (25.93% OOB classification error). This is more than double the accuracy that would be expected if assignment was driven by chance. The OOB classification error rate in the analysis based on shape data ranged from 16.67 % for fish from the Agra site to 38.89% for fish from the Naorora site (Table 2).

1. **Discussion**

This study demonstrated how multivariate generalised linear models (MGLMs) can be applied to otolith chemistry and otolith shape data to test for differences between groups of samples. Using MGLMs we clearly showed that three groups of *Channa striata* have clear differences in both otolith chemistry and otolith shape based. This was confirmed using a random forest classification approach which also identified differences between all three groups. Both the MGLM and random forest classification results provide strong evidence that *C. striata* should be managed on a regional rather than at a national scale. The MGLM method applied here to a simple test between three groups could be easily adapted and expanded to answer other ecological questions requiring more complex model frameworks as is currently done in the broader field of ecology.

*4.1 The MGLM method for otolith data*

This study has demonstrated the potential for MGLMs to be used as an analysis tool for otolith chemistry and/or otolith shape data, for example in fisheries stock discrimination. We successfully applied a model-based multivariate analysis method to a case study in India and identified differences between three populations of *C. striata* using both otolith chemistry data and otolith shape data. The MGLM framework which we have used can be considered a robust alternate to the more widely used distance-based analyses including permutational ANOVAs (PERMANOVAs) which are commonly used when the assumptions of MANOVA can not be met. The advantages for using GLMs over distance based methods are well documented in Warton et al. (2012), but briefly we see a number of advantages in applying MGLMs to otolith data as well as a potential disadvantage.

Firstly, the assumptions of MGLMs can be easily checked and the appropriateness of the models assessed before any inference is made from the results. We demonstrated this using mean-variance plots and residual plots in our case study where we demonstrated that the MGLM with a tweedie error distribution was an appropriate fit to both the otolith chemistry and otolith shape data (Figures 2 & 3).

The main downside of using this approach is that computational time can be long when there are large numbers of variables. Our examples with a small sample size of only 48 otoliths took 2 hours (chemistry data only), 137 hours (shape data only) and 138 hours (shape and element data) to run on the UNSW High Performance Cluster using a single core (8 gb RAM). It may take longer if the model needs adjusting, for example, the variance parameter in the tweedie error distribution may need to be varied so that it appropriately fits the data and it is useful to test the appropriateness of the models on a small subset of data before running the full analysis. While these calculations can be run on regular computers the time factor is a trade-off which individual researchers will need to consider, particularly if they do not have access to large computing resources, although with advances in computing software and technology this is likely to become faster and more accessible.

The latent variable model-based ordinations successfully visualised the multivariate differences identified in the MGLMs (and random forest analyses). The ordinations visually matched the model results with the elemental data clearly driving the separation but the overall separation improving when shape data was combined with the elemental data. While the current study used a Bayesian model based latent variable method (Hui et al. 2015), an alternative ordination method directly based upon the MGLM model could be produced using Gaussian copula graphical models which can be run using the ‘ecoCopula’ R package (Popovic et al., 2019). Both these ordination methods provide an alternative to traditional based distance-based ordination methods and by following the code provided with this paper, the MGLMs and model-based ordination methods can easily be applied in future studies.

*4.2 Implications for C. striata in India*

Both otolith elemental composition and shape data showed clear differences between the three sampling locations. Otolith chemistry showed the largest differences but the differences in shape were also clear in both the MGLMs and random forest classification analyses. This suggests that these three groups of fish do not belong to a single stock and that the *C. striata* in these three rivers are different populations. This confirms recent research which used truss morphometry based upon body shape of *C. striata* to show strong evidence that the three groups analysed in the current study are distinct (Khan et al. 2019). We believe that any management of this species should therefore occur at a regional scale, rather than at the national level.

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. The random forest classification analysis based upon otolith shape also supports the lack of size driven bias with the Agra site containing the largest fish (and otoliths) having more classification errors than Lucknow which had similar-sized fish to Narora.

*4.3 Conclusion*

This study has successfully demonstrated the use of multivariate generalised linear models with otolith data by discriminating populations of *C. striata* in India based upon otolith chemistry and otolith shape data. These results suggest that management of this species should occur at a regional scale. This method (and code provided with this paper) is highly flexible and has the potential to be applied to many ecological questions using multivariate otolith data.

**Author Contributions**

SK, HS & KM conceived the idea, SK, MK, DP & KM collected the data, HS analysed the data, SK & HS wrote the manuscript, MK, DP & KM critically reviewed the manuscript. All authors approved publication.

**Funding**

The stipend for HS was partially funded by a UNSW Faculty of Science writing scholarship, a NSW Research Attraction and Acceleration Program grant to the Sydney Institute of Marine Science, the UNSW Research Infrastructure Scheme Network Lab for Ocean Collaboration and an Australian Research Council Linkage Project (LP150100923). SK is grateful to the Council of Scientific and Industrial Research, New Delhi, for funding the study and providing financial assistance to the first author as Senior Research Fellowship (ACK No. 113724/2K18/1).

**Acknowledgments**

This research includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney. Thanks to Ben Maslen (UNSW Stats Central) for many useful conversations. This is contribution no. XXX of the Sydney Institute of Marine Science.

**References**

Begg, G.A., Friedland, K.D., Pearce, J.B., 1999. Stock identification and its role in stock assessment and fisheries management: an overview, Fish. Res. 43, 1–8.

Begg, G.A., Overholtz, W.J., Munroe, N. J., 2001. The use of internal otolith morphometrics for identification of haddock (*Melanogrammus aeglefinus*) stocks on Georges Bank. Fish. Bull. 99, 1–14.

Bhardwaj, R., Gupta, A., & Garg, J. K. 2017. Evaluation of heavy metal contamination using environmetrics and indexing approach for River Yamuna, Delhi stretch, India. *Water science*, *31*(1), 52-66.

Bird, J. L.; Eppler, D. T.; and Checkley, D. M., Jr., 1986: Comparisons of herring otoliths using Fourier series shape analysis. Can. J. Fish. Aquat. Sci. 43, 1228– 1234.

Cagauan, A. 2007. Exotic aquatic species introduction in the Philippines for aquaculture- A threat to biodiversity or a boon to the economy?. Journal of Environmental Science and Management, 10/1: 48-62.

Campana, S.E., and Casselman, J.M., 1993. Stock discrimination using otolith shape analysis. Can. J. Fish. Aquat. Sci. 50, 1062-1083.

Castonguay, M., Simard, P., Gagnon, P., 1991. Usefulness of Fourier analysis of otolith shape for Atlantic mackerel (Scomber scombrus) stock discrimination. Can. J. Fish. Aquat. Sci. 48, 296–302.

Dushoff, J., Kain, M. P. & Bolker, B. M. 2019. I can see clearly now: Reinterpreting statistical significance. *Methods in Ecology and Evolution,* 10**,** 756-759.

Elsdon TS, Gillanders BM 2004. Fish otolith chemistry influenced by exposure to multiple environmental variables. J Exp Mar Biol Ecol 313:269–284

Elsdon, T. S., and B. M. Gillanders. 2003. Reconstructing migratory patterns of fish based on environmental influences on otolith chemistry. Rev. Fish Biol. Fish. 13: 219– 235. doi:10.1023/B:RFBF.0000033071.73952.40

Fowler, A. M., P. I. Macreadie, D. P. Bishop, and D. J. Booth. 2015. Using otolith microchemistry and shape to assess the habitat value of oil structures for reef fish. Marine Environmental Research **106**:103-113.

Friedland, K. D., and D. G. Reddin. 1994. Use of otolith morphology in stock discriminations of Atlantic salmon (Salmo salar). Can. J. Fish. Aquat. Sci. 51:91–98.

Grammer, G. L., J. R. Morrongiello, C. Izzo, P. J. Hawthorne, J. F. Middleton, and B. M. Gillanders. 2017. Coupling biogeochemical tracers with fish growth reveals physiological and environmental controls on otolith chemistry. Ecological Monographs 87: 487– 507.

Graps, A. 1995. An introduction to wavelets. IEEE computational science and engineering, 2(2), 50-61.

Hui, F. K. C. 2016. boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. Methods in Ecology and Evolution 7:744-750.

Hui, F. K. C., S. Taskinen, S. Pledger, S. D. Foster, and D. I. Warton. 2015. Model-based approaches to unconstrained ordination. Methods in Ecology and Evolution 6:399-411.

Izzo, C., Reis-Santos, P., and Gillanders, B. M. 2018. Otolith chemistry does not just reflect environmental conditions: a meta-analytic evaluation. Fish and Fisheries 19, 441–454. doi:10.1111/FAF.12264

Jayaram, K.C., 1999. The Freshwater Fishes of the Indian Region. Narendra Publishing House, Delhi, India.

Jones, C. M., Palmer, M., & Schaffler, J. J. 2017. Beyond Z ar: the use and abuse of classification statistics for otolith chemistry. *Journal of Fish Biology*, *90*(2), 492-504.

Khan, M.A., Miyan, K., Khan, S., Patel, D.K., Ansari, N.G., 2012. Studies on elemental profile of otoliths and truss network analysis for stock discrimination of the threatened stinging catfish, *Heteropneustes fossilis* (Bloch 1794), from the river Ganga and its tributaries. Zool. Stud. 51, 1195‒1206.

Khan, M.A., Khan, S., Miyan, K., 2019. Stock identification of the *Channa striata* inhabiting the Gangetic river system using truss morphometry. Russ. J. Ecol. 50, 391–396.

Liaw, A and Wiener, M., 2002. Classification and Regression by randomForest. R News 2(3), 18--22.

Libungan, L. A., and S. Pálsson. 2015. ShapeR: an R package to study otolith shape variation among fish populations. PLOS ONE **10**:e0121102.

Maguffee, A. C., Reilly, R., Clark, R., & Jones, M. L. 2019. Examining the potential of otolith chemistry to determine natal origins of wild Lake Michigan Chinook salmon. Canadian Journal of Fisheries and Aquatic Sciences, *76*(11), 2035-2044.

Miyan, K., Khan, M.A., Khan, S., 2014. Stock structure delineation using variation in otolith chemistry of snakehead, *Channa punctata* (Bloch, 1793), from three Indian rivers. J. Appl. Ichthyol. 30, 881‒886.

Miyan, K., Khan, M.A., Patel, D.K., Khan, S., Ansari, N.G., 2016a. Truss morphometry and otolith microchemistry reveal stock discrimination in *Clarias batrachus* (Linnaeus, 1758) inhabiting the Gangetic river system. Fish. Res. 173, 294‒302.

Miyan, K., Khan, M.A., Patel, D.K., Khan, S., Prasad, S., 2016b. Otolith fingerprints reveal stock discrimination of *Sperata seenghala* inhabiting the Gangetic river system. Ichthyol. Res. 63, 294‒301.

Nazir, A., Khan, M.A., 2019. Spatial and temporal variation in otolith chemistry and its relationship with water chemistry: Stock discrimination of Sperata aor. Ecol. Freshw. Fish. 28, 499–511.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. & Wagner, H. 2019. vegan: Community Ecology Package. R package version 2.5-5.

Phen, C., T. Thang, E. Baran, L. Vann. 2004. Biological Reviews of Important Cambodian Fish Species, Based on Fishbase. Penang, Malaysia: WorldFish Center.

Popovic, G. C., D. I. Warton, F. J. Thomson, F. K. C. Hui, and A. T. Moles. 2019. Untangling direct species associations from indirect mediator species effects with graphical models. Methods in Ecology and Evolution **10**:1571-1583.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahman, M.A., Awal, S., 2016. Development of Captive Breeding, Seed Production and Culture Techniques of Snakehead Fish for Species Conservation and Sustainable Aquaculture. Int. J. Adv. Agric. Environ. Eng. 3, 117-120.

Sarkar, U. K., Gupta, B. K., Lakra, W. S., 2010. Biodiversity, ecohydrology, threat status and conservation priority of the freshwater fishes of river Gomti, a tributary of river Ganga (India). The Environmentalist. 30, 3-17.

Smoliński, S., Schade, F. M. & Berg, F. 2019. Assessing the performance of statistical classifiers to discriminate fish stocks using Fourier analysis of otolith shape. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/cjfas-2019-0251

Schilling, H. T., P. Reis-Santos, J. M. Hughes, J. A. Smith, J. D. Everett, J. Stewart, B. M. Gillanders, and I. M. Suthers. 2018. Evaluating estuarine nursery use and life history patterns of *Pomatomus saltatrix* in eastern Australia. Marine Ecology Progress Series **598**:187-199.

Sengupta, B., 2006. Water quality status of Yamuna River (1999–2005). Assessment and Development of River Basin Series: ADSORBS/41/2006-07. CPCB Ministry of Environment and Forests, Delhi. 136 pp.

Stransky, C., 2005. Geographic variation of golden redfish (Sebastes marinus) and deep-sea redfish (S. mentella) in the North Atlantic based on otolith shape analysis. ICES J. Mar. Sci. 62, 1691–1698.

Talwar, P. K., Jhingran, A.G., 1991. Insland Fishes of India and Adjacent Countries. Vol. 1-2. Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi, Calcutta, India.

Tracey, S. R., Lyle, J. M., & Duhamel, G. 2006. Application of elliptical Fourier analysis of otolith form as a tool for stock identification. *Fisheries Research*, *77*(2), 138-147.

Turan, C., 2006. The use of otolith shape and chemistry to determine stock structure of Mediterranean horse mackerel *Trachurus mediterraneus* (Steindachner). J. Fish. Biol. 69(Suppl. C), 165-180.

Turner, S. M., Manderson, J. P., Richardson, D. E., Hoey, J. J., and Hare, J. A. 2015. Using habitat association models to predict Alewife and Blueback Herring marine distributions and overlap with Atlantic Herring and Atlantic Mackerel: can incidental catches be avoided? ICES J. Mar. Sci. 73, 1912–1924. doi: 10.1093/icesjms/fsv166

Vignon M, Morat F (2010) Environmental and genetic determinant of otolith shape revealed by a non-indigenous tropical fish. Mar Ecol Prog Ser 411:231-241. <https://doi.org/10.3354/meps08651>

Wang, Y., U. Naumann, S. T. Wright, and D. I. Warton. 2012. mvabund– an R package for model-based analysis of multivariate abundance data. Methods in Ecology and Evolution **3**:471-474.

War, M., K. Altaff, M. Haniffa. 2011. Growth and survival of larval snakehead Channa striates (Bloch, 1793) fed different live feed organisms. Turkish Journal of Fisheries and Aquatic Sciences, 11: 523-528.

Warton, D. I. 2008. Raw data graphing: an informative but under-utilized tool for the analysis of multivariate abundances. Austral Ecology **33**: 290-300.

Warton, D. I., S. D. Foster, G. De’ath, J. Stoklosa, and P. K. Dunstan. 2015. Model-based thinking for community ecology. Plant Ecology **216**:669-682.

Warton, D. I., S. T. Wright, and Y. Wang. 2012. Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution **3**:89-101.

**Tables**

**Table 1** Univariate GLM results for the otolith chemistry analysis. For each element, the Likelihood Ratio test statistic (LR; 2 dp) and *P*-value (3 dp) are shown as well as the % contribution to the multivariate differences (2 dp).

|  |  |  |  |
| --- | --- | --- | --- |
| **Element** | **LR** | ***P*-value** | **% Contribution** |
| Na | 4952.27 | <0.001 | 69.16 |
| Sr | 1295.93 | <0.001 | 18.10 |
| Mg | 675.02 | <0.001 | 9.43 |
| Ba | 120.32 | <0.001 | 1.68 |
| Ni | 27.64 | <0.001 | 0.39 |
| Mn | 26.80 | <0.001 | 0.37 |
| Cr | 18.55 | <0.001 | 0.26 |
| Pb | 16.08 | <0.001 | 0.22 |
| Cd | 10.42 | <0.001 | 0.15 |
| Fe | 8.36 | 0.013 | 0.12 |
| Zn | 5.27 | <0.001 | 0.07 |
| K | 4.39 | 0.022 | 0.06 |
| Total | 7161.02 | NA | 100 |

**Table 2** Confusion matrix for the random forest classification analysis using otolith shape data. The numbers show how the fish were reclassified from each site. The OOB classification error is the overall error rate for fish from the each site (overall OOB was 25.93%). The OOB of the random forest analysis using otolith chemistry data and the random forest analysis using combined otolith shape and chemistry data was 0%.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Reclassified Site | | |  |
|  |  | Agra | Lucknow | Narora | OOB Classification Error (%) |
| Original Site | Agra | 14 | 2 | 2 | 22.22 |
| Lucknow | 0 | 15 | 3 | 16.67 |
| Narora | 5 | 2 | 11 | 38.89 |

**Figure Captions:**

**Figure 1** – Map showing the locations where samples of *C. striata* were collected. The 3 coloured dots represent the sample locations on the 3 different rivers.

**Figure 2** Mean variance plots showing non-linear relationships for a) the otolith chemistry dataset, b) the otolith shape dataset, and c) the combined otolith chemistry and shape dataset. Note the log scale on both axes.

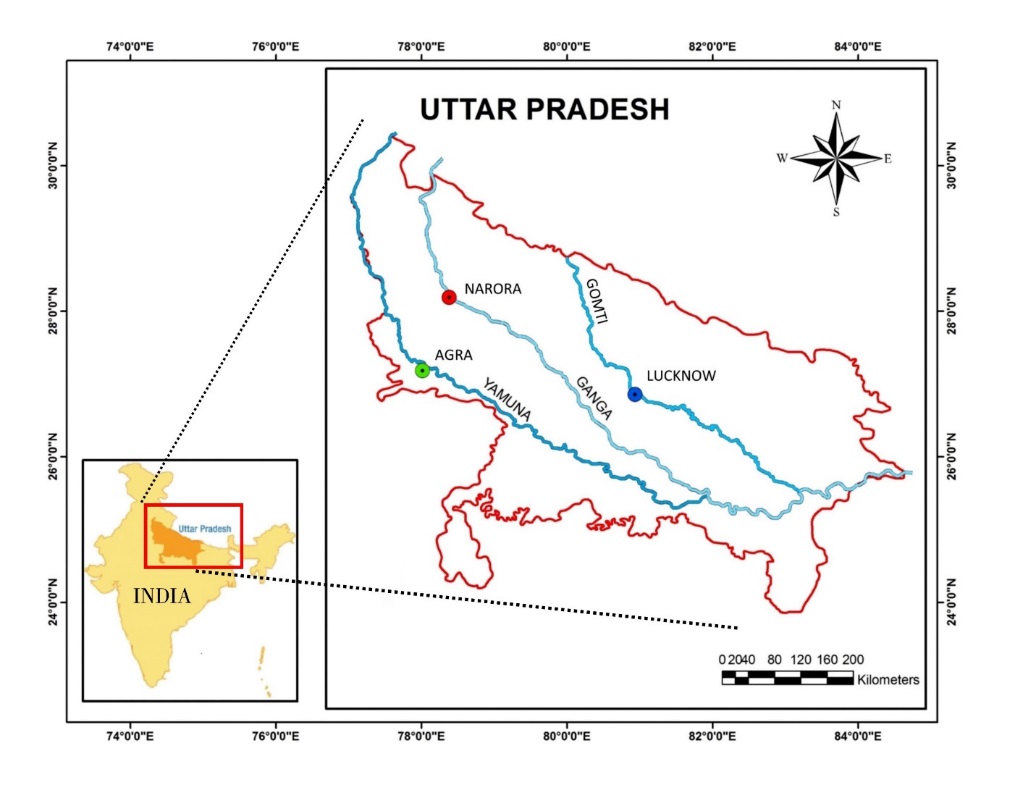
**Figure 3** Dunn-Smyth Residual plots for a) the otolith chemistry dataset, b) the otolith shape dataset, and c) the combined otolith chemistry and shape dataset. No strong patterns are visible in any of the subplots, suggesting that our GLM models were appropriate. Colours show the variables in the analysis.

**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.

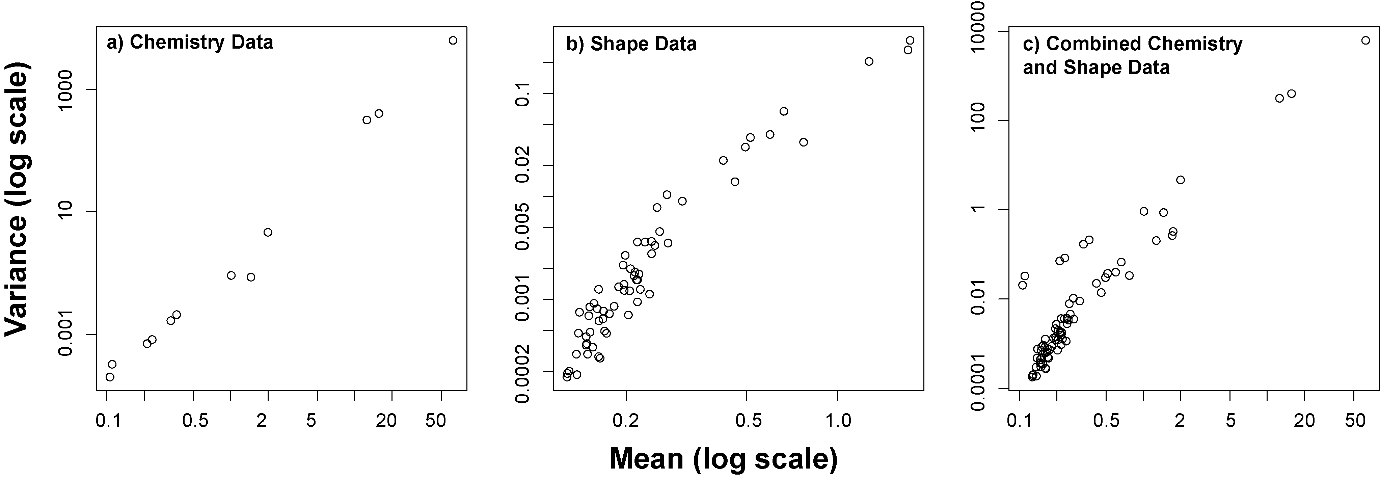
**Figure 5** Mean otolith element concentrations (mmol:mol Ca) for each of the three populations. Error bars show 1 standard error. Within a subplot, bars which do not share a common letter are clearly different (*P* < 0.05).

**Figure 6** Model-based latent variable ordinations of a) the otolith chemistry dataset, b) the otolith shape dataset, and c) the combined otolith chemistry and shape dataset. Colours and shapes represent the three groups of *C. striata*.

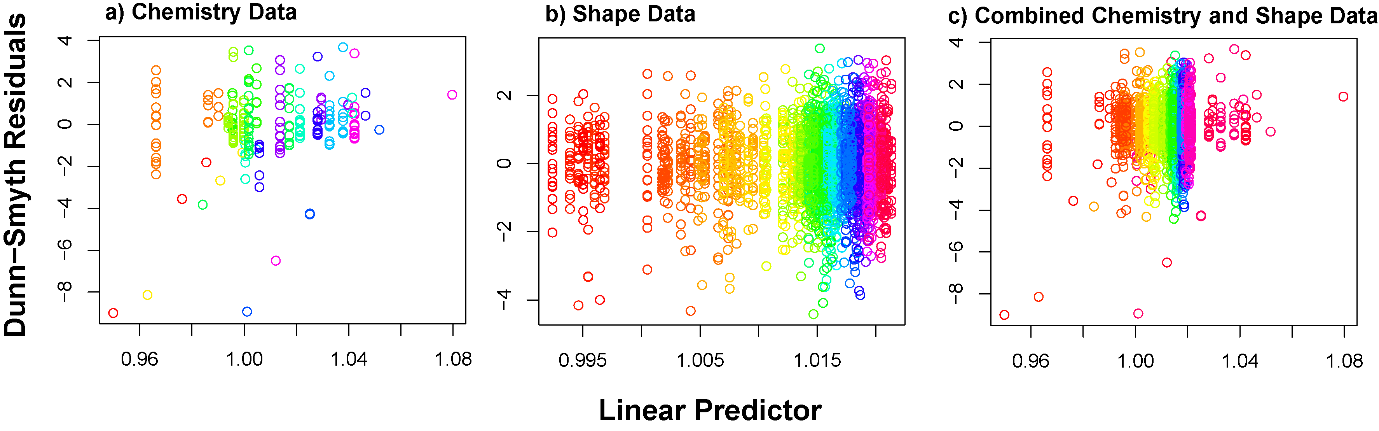
**Figures:**



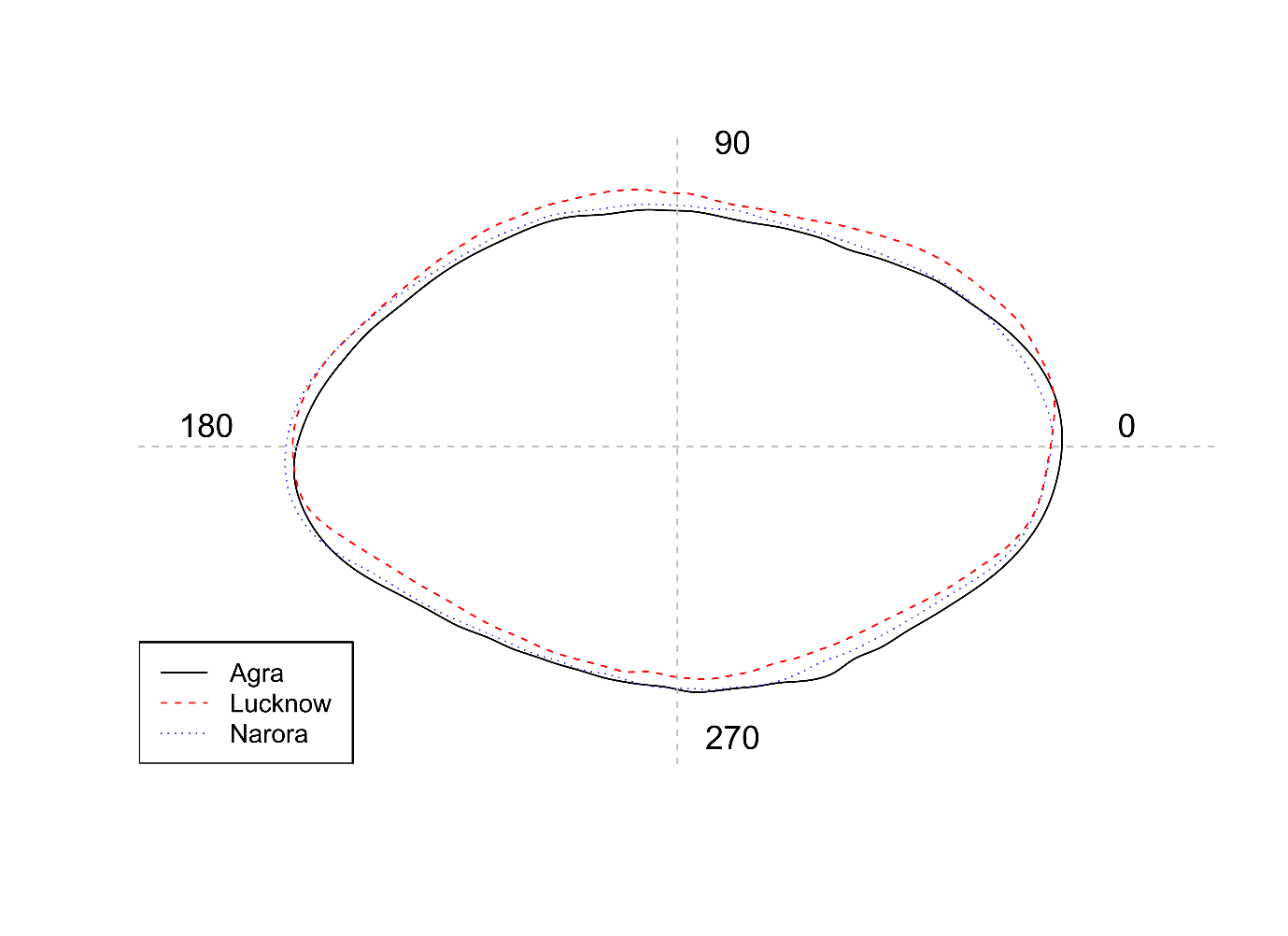
**Figure 1** – Map showing the locations where samples of *C. striata* were collected. The 3 coloured dots represent the sample locations on the 3 different rivers.



**Figure 2** Mean variance plots showing non-linear relationships for a) the otolith chemistry dataset, b) the otolith shape dataset, and c) the combined otolith chemistry and shape dataset. Note the log scale on both axes.



**Figure 3** Dunn-Smyth Residual plots for a) the otolith chemistry dataset, b) the otolith shape dataset, and c) the combined otolith chemistry and shape dataset. No strong patterns are visible in any of the subplots, suggesting that our GLM models were appropriate. Colours show the variables in the analysis.

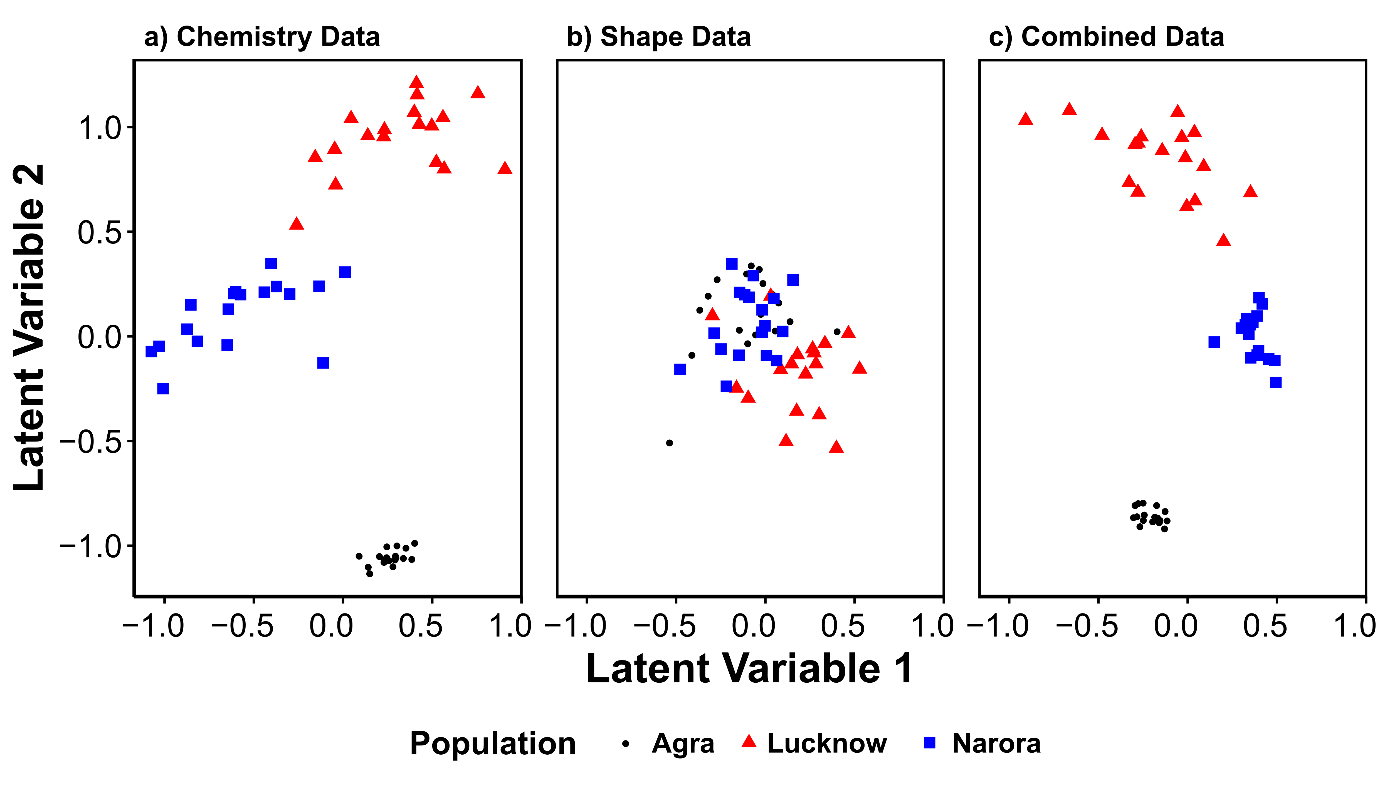


**Figure 4** Mean otolith shape from the three sites. The solid black line represents Agra, dashed red line represents Lucknow and the dotted blue line represents Narora.

A screenshot of a cell phone

Description automatically generated

**Figure 5** Mean otolith element concentrations (mmol:mol Ca) for each of the three populations. Error bars show 1 standard error. Within a subplot, bars which do not share a common letter are clearly different (*P* < 0.05). For univariate GLM results see Table 1.



**Figure 6** Model-based latent variable ordinations of a) the otolith chemistry dataset, b) the otolith shape dataset, and c) the combined otolith chemistry and shape dataset. Colours and shapes represent the three groups of *C. striata*.