Computational analysis of morphometric data of various fishes in order to predict success of invasion

Independent research project, COMP 401 (3 credits)

Sonia Ionescu, ID: 260665763

Supervisor: Dr. Anthony Ricciardi

January to April 2018

INTRODUCTION

An invasive species is defined by the Convention on Biological Diversity as a "species whose introduction and/or spread outside their natural past or present distribution threatens biological diversity." The "empty niche hypothesis" (Elton, 1958), posits that a particular invasive species' success in the location it is introduced to may be based on its exploitation of an empty niche, and therefore the reduction of niche overlap (Elton 1927, Azzurro *et al.* 2014). Conversely, high niche overlap would prevent establishment by the "limiting-similarity hypothesis" (Macarthur & Levins, 1967).

A species' niche is partially defined by its morphological characteristics, as its morphology is directly related to its interaction with resources, natural enemies, and the physical environment, the three main factors contributing to an invader's growth rate. These factors define "niche opportunities." Responses and effects with niche opportunities are the two defining aspects of an organism's niche (Shea & Chesson, 2002). Thus an invasive species with traits which allow it to exploit resources and physical environments in a way its competitors do not, and prosper despite natural enemies, will likely have high success (Olden *et al*., 2006).

One metric for success is establishment in the population, as measured by population abundance. To test the hypothesis that the population abundance of an introduced fish is related to morphological relationships, a morphospace (a spatial distribution of fishes based on morphometrics) (Farre *et al*., 2013)

is created and then analyzed. A morphospace can be defined using a Primary Component Analysis (PCA) in two dimensions, of landmark coordinates of a community of fish (Azzurro *et al*., 2014). This creates a convex hull, a measure which constitutes the minimum convex geometry including all observations, which is used to represent the space of traits occupied by a community (Azzurro *et al.*, 2014). The convex hull can then be decomposed into Voronoi polygons, which are used to estimate available ecological space around each species in a community.

In order to facilitate digitization of morphometric landmarks, PCA, application of Voronoi polygons, and analysis of the correlation between area of Voronoi polygons and abundance, various functions in R will be created. This will increase the speed and consistency of data analysis, and render it possible to look at data for a variety of different communities and at different scales. Upon creation of these convex hulls with Voronoi polygons overlayed, it is possible to perform ordered linear regression in order to see the correlation between the relative size and position of an invader's Voronoi polygon and its numeric abundance.

BACKGROUND

As humans are able to travel increasingly further, translocation of species beyond their natural ranges, or within their geographical range to a system not previously inhabited, is increasingly possible. Invasive species constitute a considerable anthropogenic contribution to the current biodiversity crisis, (Azzurro *et al*., 2014), and are a leading cause of animal extinctions (Clavero and García-Berthou, 2005), only second to habitat destruction (Canonico, 2005). Invasive species may reduce the abundance of native inland water species through predation, hybridization, parasitism, or competition, and may otherwise "alter community structure and ecosystem processes" (Canonico, 2005).

The success of different species in invasion may depend on the morphological disparity between invaders and native species. Morphological disparity is a measure of the amount of morphological variation in a group of samples (Farre *et al*., 2013), and is often used for quantifying the species

distribution within morphospace. Morphological disparity has generally been of interest in determining the influence of phylogeny and taxonomy on shape (Clabaut *et al*., 2006, Foote 1993), the relationship between morphology and habitat complexity (Willis *et al*., 2004), and, more recently, the relationship between morphology and success of invasion (Farre *et al.*, 2013, Azzurro *et al.*, 2014).

In order to quantify morphological disparity, the field of geometric morphometrics has been developed. Geometric morphometrics evolved from traditional morphometrics, an approach characterized by the application of multivariate statistical methods to sets of variables; these variables usually correspond to various measured distances on an organism, such as lengths and widths of structures, and distances between certain landmarks. In traditional morphometrics, the shape of the original form is not possible to recover, and it is not used in analyses. In contrast, in geometric morphometrics, data are recorded to capture the geometry of the structure being studied in the form of two-dimensional or three-dimensional coordinates of morphological landmark points (Rohlf, 1993).

The results obtained by performing geometric morphometric techniques on data can then be compared to other data regarding the organism in order to analyze the relationship between morphology and other characteristics of interest. For example, it is believed that area and location of Voronoi polygons applied over a convex hull should correspond with niche opportunity, and thus abundance. The niche opportunity (and abundance) is expected to be high when the introduced species occupies a morphological domain not occupied by the resident assemblage, which is likely if they are outside the convex hull (category III), medium when occupying a larger Voronoi polygon than its neighbor, but still residing within the convex hull (category II), and low when occupying a smaller Voronoi polygon than its neighbors (category I) (Azzurro *et al*., 2014). In fact, it has been found that external morphology can be used to predict which introduced species will develop abundant populations and which will not (Azzurro *et al.*, 2014). This may be a useful tool in deciding which invasive species must be considered as a serious threat to biodiversity, and which will not be successful.

METHODS

Digitization of landmarks

Images of fish caught at the Saint Lawrence River Trail (SLRT) site were taken with the fish facing the left, and a ruler situated in the photograph. Photographs without rulers cannot be used, as the photographs cannot be scaled from pixels to inches, and thus there is no meaningful way of comparing photographs which may have been taken with different cameras, which might have different pixel densities, or at different distances from the fish (which is often done in order to maintain detail in photographs of fish of different sizes). In this case, 23 species were looked at; two species (the common carp, *Cyprinus carpio*, and the bluegill, *Lepomis macrochirus*) were excluded due to there being a lack of photos with rulers in them (Table 2).

Landmarks on these fish are then marked (Figure 1). There are 28 landmarks which have been marked; 27 of these have been used as landmarks on marine fish (Farre *et al*., 2016), and one landmark has been added in order to account for the adipose fin, which is present in some teleost fishes (Stewart *et al.*, 2014) (Table 1). These landmarks are marked in R. The input of the function, *digitize_landmark*, is a .csv file which has the file names of each .jpg. This is used to plot the .jpg's, which must be in the working directory. The first .jpg is plotted using a function which reads the file and displays it as a two dimensional array of pixels. Then, the user is prompted to enter the species of the fish, and then to click the points they perceive to correspond to the landmarks that they are prompted to. The prompt tells the user which points to click (as described in table 1), five at a time. Then, the user is prompted to click two points one inch apart on the ruler. Once the user has done this, the points are scaled and added to a data frame, and this process is repeated for the next photo. This goes on until all of the files in the .csv has been plotted. The output of this function is a .csv file which contains the species, filename, and 56 labeled data points: 28 x coordinates, and 28 corresponding y coordinates.

Once the landmarks of each fish have been digitized, they need to be scaled and rotated. General Procrustes Analysis (GPA), a method of generalized least-squares (GLS) superimposition, is performed on all of the samples. This ensures that all specimens are translated to a common centroid position in the coordinate system, that they are scaled to unit centroid size, and that they are rotated to minimise the distances between corresponding landmarks (Azzurro *et al.,* 2014). GLS methods are based on the idea of overlaying images – or in this case, landmarks – so that their homologous landmarks match as closely as possible according to an optimality criterion. Procrustean methods, such as GPA, seek to reduce biases in methods of inferring scale, translation, and rotation from the whole dataset of landmarks. GPA aims to minimize the sum of squared distances between similar landmarks of configurations by allowing size, rotation, and translation to be adjusted (Claude, 2008). It does so by translating all specimens to the origin, scaling them to unit-centroid size, and optimally rotating them until the coordinates of corresponding points align as closely as possible (Adams, 2017). In this project, GPA was performed using the function *gpagen* from the package geomorph (Adams, 2017). The results of this scaling and centering on select species is shown below.

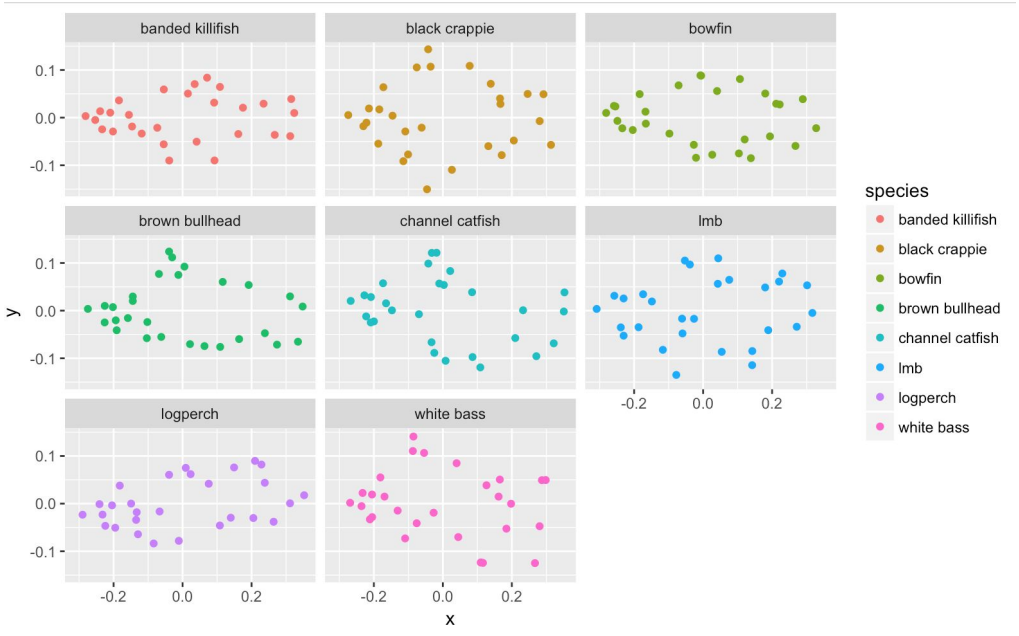


Figure A. Result of running *gpagen* on fish of different species.

In order to rotate all of the landmarks of each fish in the same way, GPA is performed before creating a consensus morphology within each species; if this were done after creating a consensus morphology, each species would be rotated differently, as the Procrustes distance would be optimized within a species, not across an average. After conducting GPA, a function, *species_average* is created, in which each species is looked at individually and each coordinate is averaged in order to create a consensus morphology. The output of this function is a three-dimensional array, with the x and y coordinates of each landmark of each consensus morphology of each species. Once this array has been created, it is possible to run PCA and create a morphospace.

Creation of a morphospace

The simplest way to describe variation and covariation within the community is to perform PCA of the superimposed coordinates of landmarks within the community. PCA transforms the data to a new coordinate system such that the greatest variance of the data is explained by the first transformed component, the second greatest variance on the second transformed component, and so on (Claude, 2008). The score of a given observation on a given axis corresponds to the projection of the data on that axis. A plot of the PCA then provides a visual sense of the degree to which landmark configurations are similar to, or differ from, the reference configuration (in this case, the community average) (MacLeod, 2010).

This PCA is done on the array which is the output of *species_average*; therefore, there is one score for each species, which corresponds to the consensus morphology of that species. A function called *do_PCA* was created which performs PCA and plots the data along the first and second principal components, outputting a list containing a PCA object and a list of species. The PCA is calculated using the *prcomp* function available in base R. *prcomp* uses the singular value decomposition (which examines the covariances and correlations between individuals), rather than the spectral decomposition approach (which examines the covariances and correlations between variables) (Anderson, 2013). Singular value decomposition has better numerical accuracy, and is therefore the preferred function. In this function, the

results of *prcomp* are then plotted using *autoplot.pca_common*, a function of the package ggfortify (Tang, 2018) .

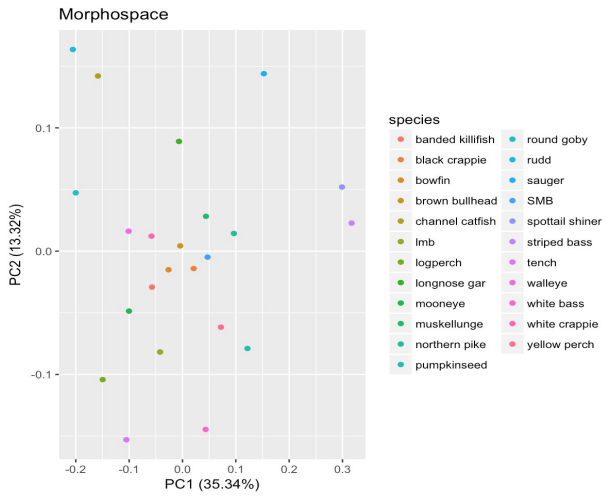


Figure B. The result of plotting the output of *prcomp* using *autoplot.pca_common* on the SLRT data.

Alternately, a function called *create_RelWarp* was created, which uses principal warps derived from each species' thin plate spline, and performs and plots a PCA of the weight matrices of the principal warps. This was tested using a function *relWarps* from the package Morpho (Schlager, 2017). A function *create_relWarp* was created which used this function and plotted it in order to create an alternate morphospace. This gave very different results, both in terms of numerical scores – the axes were about twice as long – and in terms of which fish were considered to occupy similar niches.

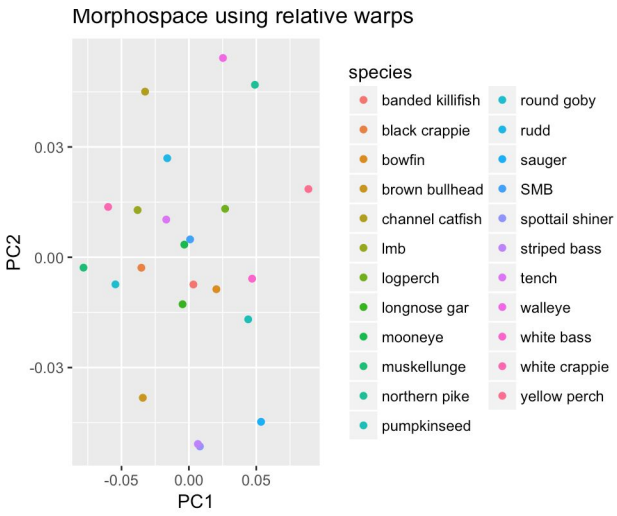
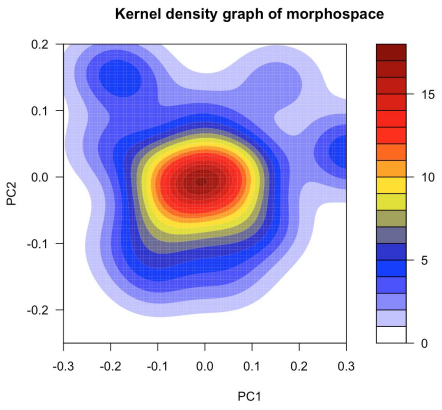


Figure C. Plotting of the first and second component axes of the *relWarps* function

Analysis of the morphospace

Once a morphospace has been created, it can be analysed to see the areas of greatest density, and Voronoi polygons may be analysed. This is useful because it can be used to anticipate the effect of an invader; an invader which adds a plot point in a low density area is predicted to be more successful than one which adds a plot point in a high density area because it will have a greater niche opportunity. An invader which increases this size of the convex hull by adding a plot point outside the original convex hull is predicted to have the highest niche opportunity.

Several functions were created in order to facilitate the analysis of the morphospace. The function *continuous_kernelDensity* takes as input a PCA object as output by *do_PCA* and puts it in the format required by the *kde2d* function in the MASS package (Ripley, 2018). The function *contour_kernelDensity* creates an equivalent graph with *ggplot2* (Wickham, 2013). A mountain plot can also be created for a three-dimensional visualization of the density of the morphospace, using the function *mountain_graph*, which relies on the *kde2d* and *persp* functions in the MASS package (Ripley, 2018). The function *Voronoi_polygon* overlays Voronoi polygons on the output of *do_PCA*, using the function *deldir*, in the package deldir (Turner, 2018). Voronoi polygons are of particular interest because it is hypothesized that there is a correlation between occupied Voronoi area and abundance (Du *et al.*, 2012). Using these different functions, it is possible to deduce the predicted effect of an invader on a community.



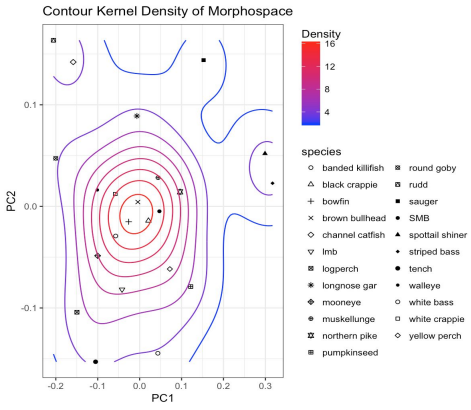


Figure D. The result of running *continuous_kernelDensity* on the output of *do_PCA*

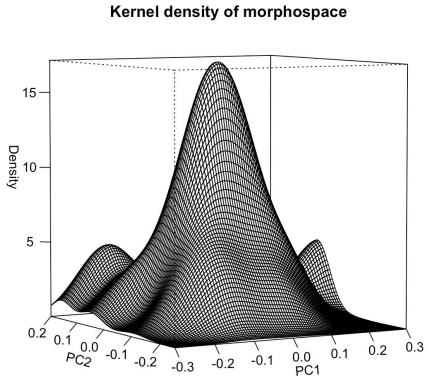


Figure E. The result of running *contour_kernelDensity* on the output of *do_PCA*

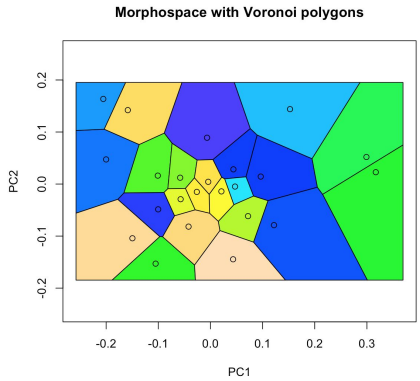


Figure F. The result of running *mountain_graph* on the output of *do_PCA*

Figure G. The result of running *Voronoi_polygon* on the output of *do_PCA*

Linear regression

In order to test the result of the hypothesis, it is possible to test the relationship between the area of Voronoi polygon and abundance. A function *linear_Voronoi* was created which uses the areas of the Voronoi polygons found in *Voronoi_polygon* and plots them against their abundances. The abundances are to be listed in a .csv file. *linear_Voronoi* then allows the user to input what site, and what output of *Voronoi_polygon* to plot against; it then goes through the .csv and sums up the fish across nets for each site, and then creates an object with only the fish abundances from the site of interest. It can then perform linear regression for the site of interest, plotting the area of the Voronoi polygons against the abundances of the fish. In order for the abundances to be correctly matched up with the area of the Voronoi polygons it is important for the common name of the fish to be used, and for these to match up correctly (except for capitalization); for example, the use of an acronym such as SMB for smallmouth bass is only acceptable if it says SMB in both the objects created before (which will be how it was entered when it was digitized, in

the function *digitize_landmark*), as well as the .csv which is the input of this function. If it says SMB in one location and smallmouth bass in another, they will not be correctly matched up.

RESULT

Principal component analysis

The first axis of the PCA – which accounts for 35.34% of variation – is composed primarily of the x coordinate of posterior insertion of the dorsal fin (with a relative weight of -.39), the x coordinate of the anterior insertion of the anal fin (with a weight of .22), and the x coordinate of distal tip of the pelvic fin (with a weight of .21). The second axis of the PCA – which ccounds for 13.32% of variation – is composed primarily of the x coordinate of the posterior insertion of the pectoral fin (with a weight of -.28), the y coordinate of the distal tip of the pelvic fin (with a weight of .26), and the x coordinate of the posterior insertion of the dorsal fin (with a weight of .23). These weights are multiplied by their corresponding measurements in each fish, and then these are summed across each coordinate in order to obtain the value of that component for each fish. This tells us that these measurements account for most of the independent variation seen between species. The other 21 principal component axes correspond to the remaining 52.34% of variation.

Rudimentary linear regression

The linear regression between area of the Voronoi polygons and the abundances do not show a significant relationship, as there is a p-value of .6577.

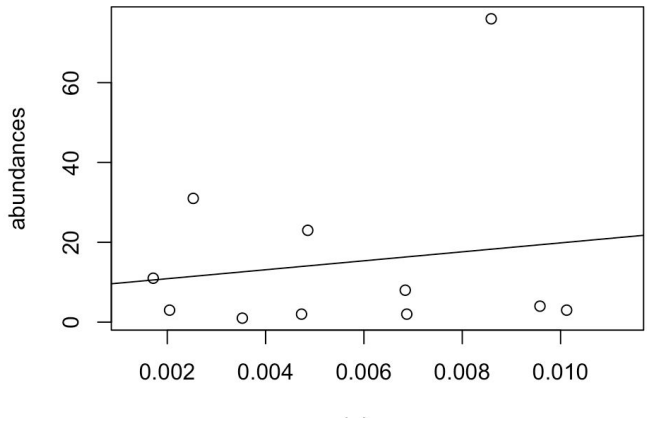


Figure H. Linear regression of areas and abundances of fish at the St. Ours site.

Potential impact of functions/package

The input and output of these functions is very easily available for the user to visualize and interpret. Where in tpsDig and tpsRelw, the software commonly used for morphometric analyses such as these, much of the analysis is done by the software and not overseen by the user, here it is possible to see exactly what is done by each function, merely by looking at the code, or viewing the output of each step. It is also easier to customize and adapt an analysis, as the code can be altered easily due its to straightforward layout. The process is easy to individualize for a specific scenario, and by looking at the code it is clear to see the mathematical principles behind each step, which allows for greater understanding of the results.

Another benefit of these functions is the speed of analysis. As the output of one is then meant to be the input of another, it is an extremely streamlined process. Digitization of the photographs is by far the most time consuming element of the process, after which analysis of these measurements is extremely quick. In order to estimate the effect of an invader using the hypothesis, it is only necessary for a photograph of the potential invader to be digitized, added to the .csv with the landmark coordinates of the other fish in the community, and then to be analyzed using *species_average*, *do_PCA*, and *Voronoi_polygon*. The output of *Voronoi_polygon* on a new data set can then be compared to the output of *Voronoi_polygon* prior to introduction of an invasive species in order to classify the invader as category I, II, or III. This can be a useful tool in order to predict the success of an invader within a community.

DISCUSSION

Use of different superimposition methods, and relative warps

Superimposition methods are important, for without them a shape space cannot be defined, and a consensus configuration cannot meaningfully be created (Rohlf, 1996). Procrustes methods of

superimposition, such as GPA, are based on the idea of overlaying the images of multiple specimens so that their homologous landmarks match as closely as possible (Rohlf, 1993). In order for this to be done, scaling to centroid size is applied; centroid size is the square root of the sum of squared distances of all the landmarks of an object from their centroid. GPA scales to centroid size 1.0 and performs transposition and rotation so that the sum of squared distances between homologous landmarks is minimal (Klingenberg, 2016). By using GPA, the "size variable" is effectively eliminated; for the sake of this report, this means that a fish with the same relative position of features, but significantly larger, will be treated as occupying the same niche as its smaller counterpart. This can be useful within a species; by doing this and creating species average after GPA is applied, static allometry, or variation within a single ontogenetic stage, is controlled for (Klingenberg, 2016). However, because this requires the creation of a consensus specimen of centroid size, this also treats different species as though they were of the same size, effectively only looking at the effect of proportion on niche. It could be possible to effectively "undo" this effect by storing a "scale" variable which consists of its length from head to distal tip of dorsal fin, and then scaling the coordinates by this length after GPA has performed scaling, rotation, and transformation, while maintaining the ratio of distances between coordinates, and maintaining the location of the "centroid" as determined by GPA. It is unclear how meaningful it would be to compare landmark coordinates when they are not aligned.

In this project, both PCA based on the landmarks, and relative warps based on the thin plate splines of specimens, were done. When using the function *do_PCA*, each coordinate is transformed onto the first and second principal component axes, so that these axes reflect the maximum of variation and covariation (Claude, 2008). In these cases, each coordinate is an individual landmark, and the landmarks are multiplied by each weight and then summed in order to find the score of a particular fish; this gives a score which primarily accounts for the landmarks which have been determined to account for the most variation. There are, however, other ways to find a score for a particular fish. One of these ways is the

method of thin plate splines (Bookstein, 1989). Thin plate splines were developed in order to describe deformations in a grid, fitting the differences in the positions of landmarks in one organism relative to their positions in another (Rohlf, 1993). These deformations can be described in the form of a matrix mapping $\mathbb{R}^2 \rightarrow \mathbb{R}^2$. This matrix is composed of "bending energies," namely the energy which would be required had the landmark displacements been normal to the plane of the figure (Bookstein, 1989). In this way, a matrix can be constructed which explains the differences between two shapes, such as the difference between a consensus configuration and one particular species. Thin plate splines are useful because they ordinate entire shapes on the basis of their similarity (Rohlf, 1993). Rather than describing shapes as a compilation of landmarks, it describes an organism as the interaction between landmarks; in thin plate splines, deformation of a landmark has an effect on the space around it, and one can interpolate change in other regions of the shape being considered (Claude, 2008). They are a more holistic and realistic description of a morphology than what is done by just PCA.

Relative warps analysis finds the thin plate spline transformations that map a reference configuration of landmarks onto each specimen (Rohlf, 1993). Then, the parameters of these transformations can be used as variables in traditional multivariate statistical techniques, such as PCA. Thin plate splines are transformations which transform a reference organism, described as a grid of parallel lines, into an organism of interest, described as a deformation of this grid. This transformation is described as a function which minimizes the bending energy, which describes how much energy would be needed to account for the deformation at the landmark points. The eigenvectors of the bending energy matrix, which then correspond to the non-affine transformations, describe the principal warps, which expresses the proximity of landparks to each other in the shape coordinate space (Macleod 2010). From these principal warps, the matrix of deviations is multiplied in, yielding the weight matrix. A classic relative warps analysis then takes the complete set of scores for both the x and y coordinates of the weight matrix, and uses these as input into a PCA (Macleod, 2010). In order to test whether the use of relative

warps gave significantly different results, an alternate morphospace was created using a function which uses thin plate splines in order to find relative warps scores. Using such a function, which relies on thin plate splines, allows for each fish to be given an individual score based off of bending energies, rather than merely the coordinate location of each landmark. Another use of this function could be when there is an expectation that important deformations will occur at particular scales (Rohlf, 1993), as relative warps allows for a weighting factor, which if greater than 0 will weigh large-scale variations more highly, and if smaller than 0 will weigh small-scale variations more highly (Macleod, 2010). The use of relative warps will result in a summary of shape variation which ecompasses variation at all spatial scales. Relative warps is ideal as it ordinates landmark-based shape configurations on the basis of their mutual similarity, treats configurations as unified geometric entities (as opposed to *do_PCA,* which treats it as a combination of individual landmarks), and respects the Kendall shape space (Macleod, 2010).

Overall, the use of relative warps and thin plate splines is the most comprehensive evaluation of a morphology and is probably the best way to estimate similarity between species. However, reliance on scaling may impact which fish are seen as most similar; for example, in the morphospace calculated using the function *create_relWarps*, the species spottail shiner and striped bass were considered to be extremely similar, as were mooneye and smallmouth bass. While these do appear to have similar characteristics, they have extremely different sizes. A new function (or series of functions) could be created which rotates the fish, translates them, but scales them to an accurate measure of relative length. This potentially would result in the most accurate creation of a morphospace, although it is unclear whether taking into account size would result in a more accurate description of niche opportunity.

Linear regression

The linear regression performed by the function *linear_Voronoi* does not take into account location of the Voronoi polygon, or size in relation to its neighbor; it solely analyzes sizes relative to other members of the community. Human input would be required in order to categorize the Voronoi polygons into category

I, II, or III, (Azzurro et al., 2014), and to see whether that correlates to a certain level of abundance, which may also need to be categorical rather than numerical. Additionally, there was only abundance data for 11 of the 23 fish which were included in the Voronoi polygons; in order for an accurate linear regression to be found, there should be abundance data for all of the fish which are included in the morphospace, as their relation to each other is what results in the relative areas of the Voronoi polygons; this would not be the same if there were fewer fish included in the morphospace.

CONCLUSION

Under the hypothesis that analysis of a morphospace can be used to predict the success of an invader, the creation and subsequent analysis of the morphospace must be as accurate as possible. If a species' morphology describes its niche opportunities, the morphospace should also describe the relative similarities of species' niches. Thus, if the generation of the morphospace does not accurately place species relative to each other, the similarities of niche opportunities of these species will be incorrectly described. In order to most accurately describe these niche opportunities, it is prudent to use thin plate splines and relative warps, which describe each species as a deformation from the species average. The landmarks of each species, as well as the community average, are transformed, scaled, and rotated, using General Procrustes Analysis, which aligns homologous landmarks points as closely as possible using a generalized least-squares superimposition. Then these landmarks are used as input into thin plate splines, and a relative warps analysis is generated; as relative warps can be an input for PCA, a PCA can then be done. This PCA is then an accurate morphospace, a spatial distribution of species based on morphology. This morphospace can then be analyzed, such as by the overlaying of Voronoi polygons. Although a linear regression of area of Voronoi polygon and abundance found the relationship to be insignificant, it is possible that use of categorical variables to describe relative size of Voronoi polygons, and location of polygon, could be implemented in order to perform a linear regression which may more accurately align with the theory of prediction of success of establishment by the creation of a unique niche. Overall, the

creation of tools to create and analyze a morphospace, has enabled the investigation of the methods which are used to describe and compare morphologies, as well as to test hypotheses surrounding them.

WORKS REFERENCED

Adams, Dean. "Geomorph v3.0.5." *Geomorph Package | R Documentation*, 8 Apr. 2017, www.rdocumentation.org/packages/geomorph/versions/3.0.5.

Anderson, Gregory B. "Principal Component Analysis in R." *Instituto De Matematica e Estatistica*, Universidad De Sao Paulo, www.ime.usp.br/~pavan/pdf/MAE0330-PCA-R-2013.

Azzurro, Ernesto, et al. "External Morphology Explains the Success of Biological Invasions." *Ecology Letters*, vol. 17, no. 11, 2014, pp. 1455–1463., doi:10.1111/ele.12351.

Bookstein, F.l. "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, June 1989, pp. 567–585., doi:10.1109/34.24792.

Bower, L. M., and K. R. Piller. "Shaping up: a Geometric Morphometric Approach to Assemblage Ecomorphology." *Journal of Fish Biology*, vol. 87, no. 3, 12 June 2015, pp. 691–714., doi:10.1111/jfb.12752.

Canonico, Gabrielle C., et al. "The Effects of Introduced Tilapias on Native Biodiversity." *Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 15, no. 5, 2005, pp. 463–483., doi:10.1002/aqc.699.

Ciampaglio, Charles N., et al. "Detecting Changes in Morphospace Occupation Patterns in the Fossil Record: Characterization and Analysis of Measures of Disparity." *Paleobiology*,

vol. 27, no. 4, 2001, pp. 695–715., doi:10.1666/0094-8373(2001)027<0695:dcimop>2.0.co;2.

Clabaut, Céline, et al. “Geometric Morphometric Analyses Provide Evidence For The Adaptive Character Of The Tanganyikan Cichlid Fish Radiations.” *Evolution*, vol. 61, no. 3, 2007, pp. 560–578., doi:10.1111/j.1558-5646.2007.00045.x.

Claude, Julien. *Morphometrics with R*. Springer, 2008.

Clavero, M, and E Garciaberthou. “Invasive Species Are a Leading Cause of Animal Extinctions.” *Trends in Ecology & Evolution*, vol. 20, no. 3, 2005, pp. 110–110., doi:10.1016/j.tree.2005.01.003.

Du, Feng, et al. “The Relationship between Aboveground Biomass and Voronoi Area of Coexisting Species in an Old-Field Community.” *Polish Journal of Ecology*, vol. 60, no. 3, Sept. 2012, pp. 479–489.

Elton, Charles S. *Animal Ecology*. University of Chicago Press, 1927.

Farré, Marc, et al. “Geometric Morphology as an Alternative for Measuring the Diversity of Fish Assemblages.” *Ecological Indicators*, vol. 29, 2013, pp. 159–166., doi:10.1016/j.ecolind.2012.12.005.

Farré, Marc, et al. “Selection of Landmarks and Semilandmarks in Fishes for Geometric Morphometric Analyses: a Comparative Study Based on Analytical Methods.” *Scientia Marina*, vol. 80, no. 2, Jan. 2016, pp. 175–186., doi:10.3989/scimar.04280.15a.

Foote, Mike. “Discordance and Concordance between Morphological and Taxonomic Diversity.” *Paleobiology*, vol. 19, no. 02, 1993, pp. 185–204., doi:10.1017/s0094837300015864.

Klingenberg, Christian Peter. “Size, Shape, and Form: Concepts of Allometry in Geometric Morphometrics.” *Development Genes and Evolution*, vol. 226, no. 3, Jan. 2016, pp. 113–137., doi:10.1007/s00427-016-0539-2.

Macarthur, Robert, and Richard Levins. “The Limiting Similarity, Convergence, and Divergence of Coexisting Species.” *The American Naturalist*, vol. 101, no. 921, 1967, pp. 377–385., doi:10.1086/282505.

MacLeod, Norman. “PalaeoMath: Part 21 - Principal Warps, Relative Warps, and Procrustes PCA.” *PalaeoMath: Part 21 - Principal Warps, Relative Warps, and Procrustes PCA | The Palaeontological Association*, Palaeontology Newletter, 2010, www.palass.org/publications/newsletter/palaeomath-101/palaeomath-part-21-principal-warps-relative-warps-and-procrustes-pca.

Mosimann, James E. “Size Allometry: Size and Shape Variables with Characterizations of the Lognormal and Generalized Gamma Distributions.” *Journal of the American Statistical Association*, vol. 65, no. 330, 1970, p. 930., doi:10.2307/2284599.

Neha. “Sizing the Shape: Understanding Morphometrics.” *Journal of Clinical and Diagnostic Research : JCDR* 9.1 (2015): ZC21–ZC26. *PMC*. Web. 10 Apr. 2018.

Olden, Julian D., et al. “Life-History Strategies Predict Fish Invasions And Extirpations In The Colorado River Basin.” *Ecological Monographs*, vol. 76, no. 1, 2006, pp. 25–40., doi:10.1890/05-0330.

Ripley, Brian. “Support Functions and Datasets for Venables and Ripley's MASS.” *The Comprehensive R Archive Network*, 23 Feb. 2018, cran.r-project.org/web/packages/MASS/MASS.pdf.

Rohlf, F. James, and Leslie F. Marcus. “A Revolution in Morphometrics.” *Trends in Ecology & Evolution*, vol. 8, no. 4, Apr. 1993, pp. 129–132., doi:10.1016/0169-5347(93)90024-j.

Rohlf, F. James, et al. “Morphometric Analysis of Old World Talpidae (Mammalia, Insectivora) Using Partial-Warp Scores.” *Systematic Biology*, vol. 45, no. 3, 1996, p. 344., doi:10.2307/2413569.

Schlager, Stefan. “Morpho v2.5.1.” *Morpho Package | R Documentation*, 19 Apr. 2017, www.rdocumentation.org/packages/Morpho/versions/2.5.1.

Shea, K. “Community Ecology Theory as a Framework for Biological Invasions.” *Trends in Ecology & Evolution*, vol. 17, no. 4, 1 Apr. 2002, pp. 170–176., doi:10.1016/s0169-5347(02)02495-3.

Stewart, T. A., et al. “The Origins of Adipose Fins: an Analysis of Homoplasy and the Serial Homology of Vertebrate Appendages.” *Proceedings of the Royal Society B: Biological Sciences*, vol. 281, no. 1781, May 2014, pp. 20133120–20133120., doi:10.1098/rspb.2013.3120.

Tang, Yuan. “Data Visualization Tools for Statistical Analysis Results.” *Package 'Ggfortify'*, cran.r-project.org/web/packages/ggfortify/ggfortify.pdf.

Willis, S. C., et al. “Habitat Structural Complexity and Morphological Diversity of Fish Assemblages in a Neotropical Floodplain River.” *Oecologia*, vol. 142, no. 2, Feb. 2004, pp. 284–295., doi:10.1007/s00442-004-1723-z.

APPENDIX

Figure 1. Illustrations of different landmark placements

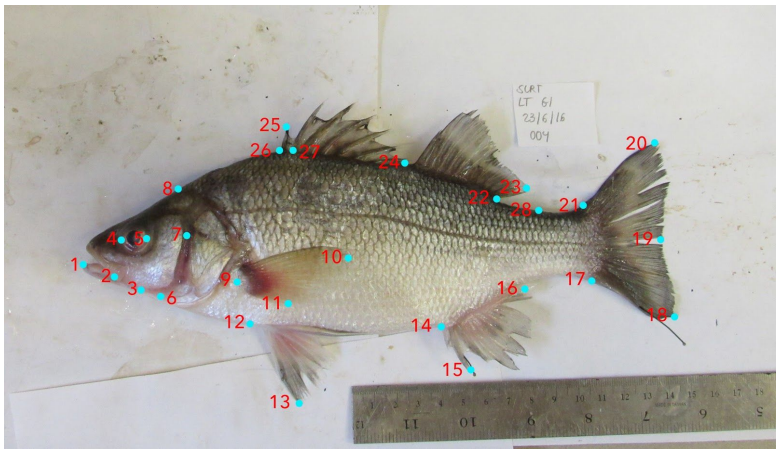


Figure 2. Results of PCA with images of fish placed over each point, for easier visualization

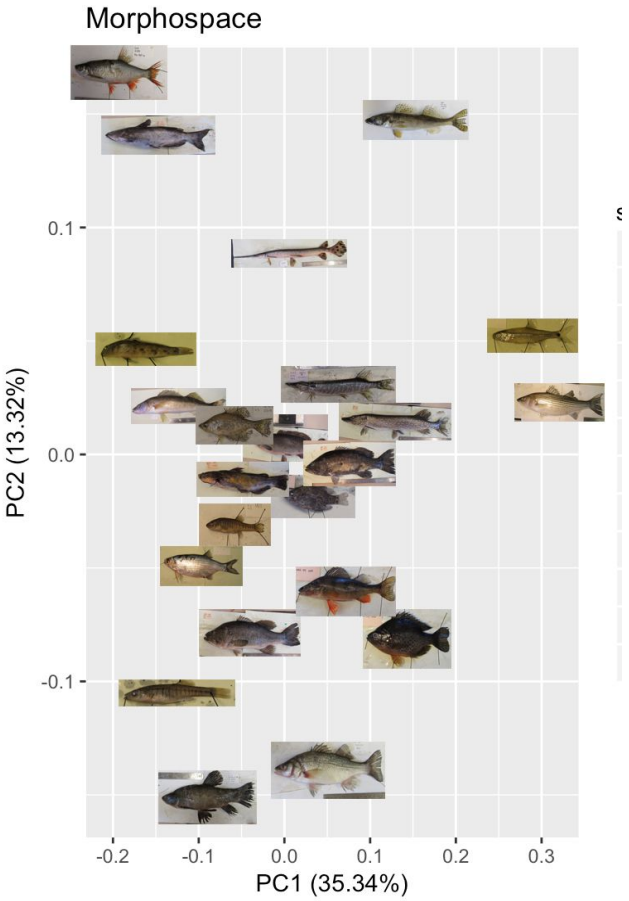


Table 1. **Description of morphological meaning of landmarks**

Landmark	Description
1	Anterior tip of the mouth
2	Posterior tip of the mouth
3	Distal tip of barbel projected at 80 degrees of the body (or projection in the lower jaw inferior margin of the position of the hyomandibular insertion, when no barbels)

4	Anterior margin in the maximum eye width
5	Posterior margin in the maximum eye width
6	Ventral margin in the end of head
7	Posterior margin in the end of head
8	Dorsal margin in the end of the head
9	Central point in the baseline of the pectoral fin
10	Posterior tip of the pectoral fin when the fin is in position of maximum extension
11	Ventral margin of the pectoral fin when the fin is in position of maximum extension
12	Insertion of the pelvic fin
13	Distal tip of the pelvic fin when the fin is in position of maximum extension
14	Anterior insertion of the anal fin
15	Distal tip from the anterior insertion of the anal fin when the fin is in position of maximum extension
16	Posterior insertion of the anal fin
17	Ventral insertion of the caudal fin
18	Distal tip of the ventral lobe of the caudal fin when the fin is in position of maximum extension

19	Posterior margin of the caudal fin between dorsal and ventral lobes
20	Distal tip of the dorsal lobe of the caudal fin when the fin is in position of maximum extension
21	Dorsal insertion of the caudal fin
22	Posterior insertion of the dorsal fin (second dorsal fin if it exists)
23	Posterior tip of the dorsal fin (second dorsal fin if it exists) when the fin is in position of maximum extension
24	Transition point between spines and soft rays in the dorsal fin (if only one dorsal fin exists) or central point in the gap between the two dorsal fins on the dorsal margin of the body (if second dorsal fin exists)
25	Distal tip of the first spine of the dorsal fin (first dorsal fin if second fin exists) when the fin is in position of maximum extension
26	Anterior insertion of the dorsal fin (first dorsal fin if second fin exists)
27	Point of maximum body height in the body margin
28	Dorsal distal tip of adipose fin, or point projected on the body between landmarks 21 and 22, when no adipose fin

Table 2. **List of species of the SLRT analysed in this project**

	Species	**Common name**
1	*Fundulus diaphanus*	Banded killifish
2	*Pomoxis nigromaculatus*	Black crappie

3	*Lepomis macrochirus*^	Bluegill
4	*Amia calva*	Bowfin
5	*Ameiurus nebulosus*	Brown bullhead
6	*Ictalurus punctatus*	Channel catfish
7	*Cyprinus carpio**^	Common carp
8	*Micropterus salmoides*	Largemouth bass
9	*Percina caprodes*	Logperch
10	*Lepisosteus osseus*	Longnose gar
11	*Hiodon tergisus*	Mooneye
12	*Percina caprodes*	Muskellunge
13	*Percina caprodes*	Northern pike
14	*Lepomis gibbosus*	Pumpkinseed
15	*Neogobius melanostomus**	Round goby
16	*Leuciscus scardafa**	Rudd
17	*Sander canadensis*	Sauger
18	*Micropterus dolomieu*	Smallmouth bass
19	*Notropis hudsonius*	Spottail shiner
20	*Micropterus punctulatus*	Striped bass
21	*Tinca tinca**	Tench

22	*Sander vitreus*	Walleye
23	*Morone chrysops*	White bass
24	*Pomoxis annularis*	White crappie
25	*Perca flavescens*	Yellow perch

*Non-indigenous species.

^Not included because of inadequate photographs.