# Last update: August 2019

# Subject: Spring 2017 Fish Community analysis along Precipitation Gradient in South Texas

# Analyst: Sean Kinard

# Create Species x Site matrix: After evaluating with bio.infer, I decided to use genus.species matrix directly from the data. OTU assignment reduces species pool by half resulting in compression of diversity.

# Simple linear regressions (diversity vs ppt): 4 sites have n <10. By removing the 4 sites, I lose resolution on the potential threshold between 70-90 cm annual rainfall in the diversity metrics. However, most correlations are greater in intensity and significance.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ~ AvgPPT | slope | R^2 | R^2.adjust | pr(>|t|) | F-stat | DF |
| shannon | 0.007002 | 0.4413 | 0.3482 | 0.07236 | 4.739 | 6 |
| simpson | 0.006892 | 0.4151 | 0.3176 | 0.0846 | 4.259 | 6 |
| PIE | -0.006953 | 0.4102 | 0.3119 | 0.0871 | 4.173 | 6 |
| rich | 0.05252 | 0.2081 | 0.07614 | 0.256 | 1.577 | 6 |
| rar.rich | 0.007265 | 0.4208 | 0.3243 | 0.0818 | 4.36 | 6 |

-----------------------------------------------------------------------------------------------------------------------------------------

# PCA indicates the following variables stratify observed differences among the sites:

sort.PC1 # Note winter precipitation is large contributor to PC1 (dec, nov, feb, jan, mar) LNG\_GAGE, PPTAVG\_BASIN, PPTAVG\_SITE, WD\_BASIN, RIP100\_43, WD\_SITE

sort.PC2 # WDMAX\_BASIN, RIP100\_82, NITR\_APP\_KG\_SQKM, PHOS\_APP\_KG\_SQKM, RIP100\_PLANT, NO3.NO2, turbidity.adjusted, conductivity, salinity, oxygen.saturation, dissolved.oxygen, pH

sort.PC3 # HYDRO\_DISTURB\_INDX, PIE, NH3.N, RIP100\_95, PO43, NO3.NO2

sort.PC4 # Dissolved.oxygen, rich, shannon, RH\_Basin

# It doesn't appear that skipping OUT assignments affects the PCA. The same variables compose each of the PCA in similar weights as before.

----------------------------------------------------------------------------------------------------------------------------------------

# linear regressions indidcate the following variables relate to diversity metrics: "HYDRO\_DISTURB\_INDX", "RH\_BASIN", "NO3.NO2", "pH", "NH3.N","PO43.", "dissolved.oxygen", "PPTAVG\_SITE"

# Identified need to log transform: RIP100\_43, NITR\_APP\_KG\_SQKM, NO3.NO2, conductivity, PO43.

# PIE = Probability of interspecies encounter

full.model.fish <- lm(PIE~ PPTAVG\_SITE + dissolved.oxygen + pH + lg.cond + lg.PO43. + lg.NH3.N, data = trm3.fish) backward.model <- traditional.backward(full.model.fish)

summary(backward.model)

# Coefficients:

# Estimate Std. Error t value Pr(>|t|)

# (Intercept) -7.459056 1.527194 -4.884 0.00814 \*\*

# dissolved.oxygen -0.005570 0.002524 -2.207 0.09194 .

# pH 1.635106 0.282295 5.792 0.00442 \*\*

# lg.NH3.N 3.330467 0.474818 7.014 0.00218 \*\*

# Multiple R-squared: 0.934, Adjusted R-squared: 0.8846

# F-statistic: 18.88 on 3 and 4 DF, p-value: 0.007977

# Summary of PIE glms: Water chemistry paramaters (pH, and ammonia concentrations) are significant exlanatory variables of variation in PIE across the sites. Removing outlier sites with n < 10 significantly reduces the scope of multivariate linear analysis, but the results may be more robust and easier to interpret.

# Summary Shannon glms: Significant predictor variables include DO, pH, lg.NH3N.

# Summary Richness: variation in species richness among the sites is largely explained by ammonia concentrations and dissolved oxygen.

# ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

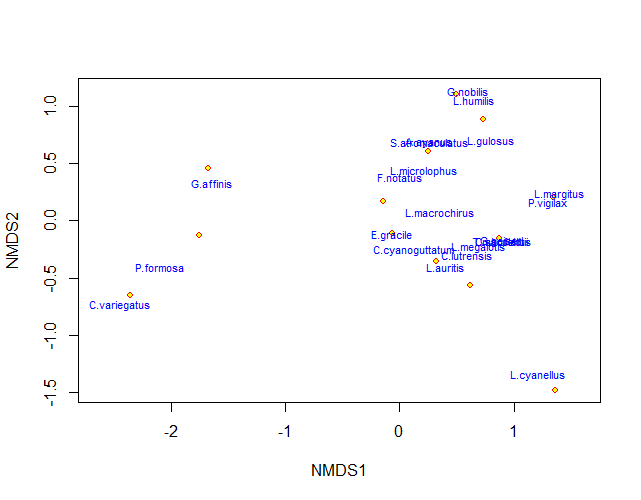
# Fish glm conclusions:

# variation in PIE, SI, richness among the sites in the gradient is largely explained by water quality paramters such as ammonia(colinear with pH) anddissolved oxygen. This conclusion is much simpler and the results are more robust when using 8/13 sites due to n < 10 at 5 sites.

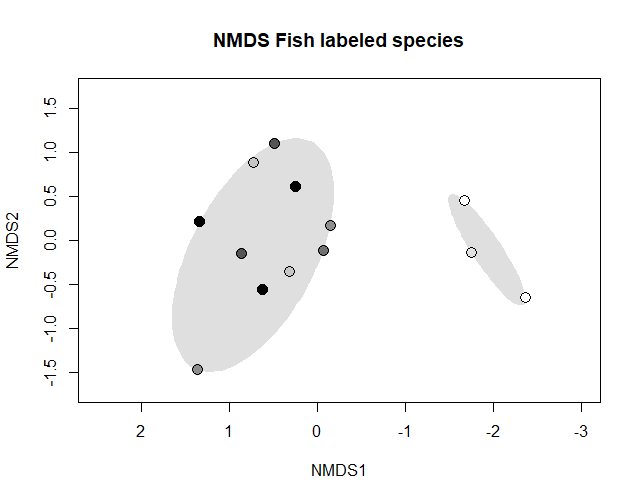
# These conclustions are different from 13 site genus.species glms.

# These conclusions are different from the analysis that used simplified OTUs (genus-level).

# ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::



NMDS ordination of fish communities in 13 streams surveyed in the Texas Coastal Prairie in 2017. Species are labeled in blue text and sites are plotted as red-yellow dots.



NMDS ordination of fish communities in 13 streams surveyed in the Texas Coastal Prairie in 2017. Sites are shaded darker with increasing average annual precipitation at site.

# ::::::::::::::::::::::::::::::::::::::::::::::::

# Final figure entails: remove env variable labels and replace in paint or ppt.

# axis recommendations?

# Title recommendation?

# Any species labels? too much clutter?

# which to remove: NH3N or pH in plot to simplify (they are colinear)?

# This concludes the fish community analysis of 13 streams along the South Texas Coastal Prairie. In it, 103 environmental variables were evaluated as predictors of variation in 5 different fish community diversity metrics. I used PCA to identify influential variables and trim the data set down to <20 variables to conduct multivariate linear regressions. Finally, I used NMDS to reveal clusters of fish community composition in conjunction with the climate variable (annual precipiation). Environmental variables were fit to the NMDS and species labels added for interprettations and conclusions.

# Closing Coding Comments:

# This analysis involved several reboots. First, I ran PCA and glm on 13 sites using bio.infer at the start. Then, I realized the correlations and nmds were not as clear as they had been during my preliminary data analysis. The second go involved bypassing the bio.infer OTU assignment code. This produced double the species and produced stronger R and p-vals, but again seemed weak for simple linear regressions. So, the third and final script involved removing sites with fewer than 10 sample size for the PCA, linear regression, and mulivariate glm analysis. Ulitmately, each script improved R and p-vals, simplified results, leading to simpler conclusions. For the NMDS plots at the end for visualization of these conclusions, I reset the community data to include 13 of 13 sites. This results in 2 distinct clusters of fish community composition. Arid streams contain G.affinis, P.formosa, and C. variegatus. These communities are evidently constrained by environmental variables: PPT, Hydro-Dist, and ammonia concentrations (colinear with pH).