Revised Data Analysis
Texas Precipitation Gradient (2017)

For PeerJ revision

By Sean Kinard

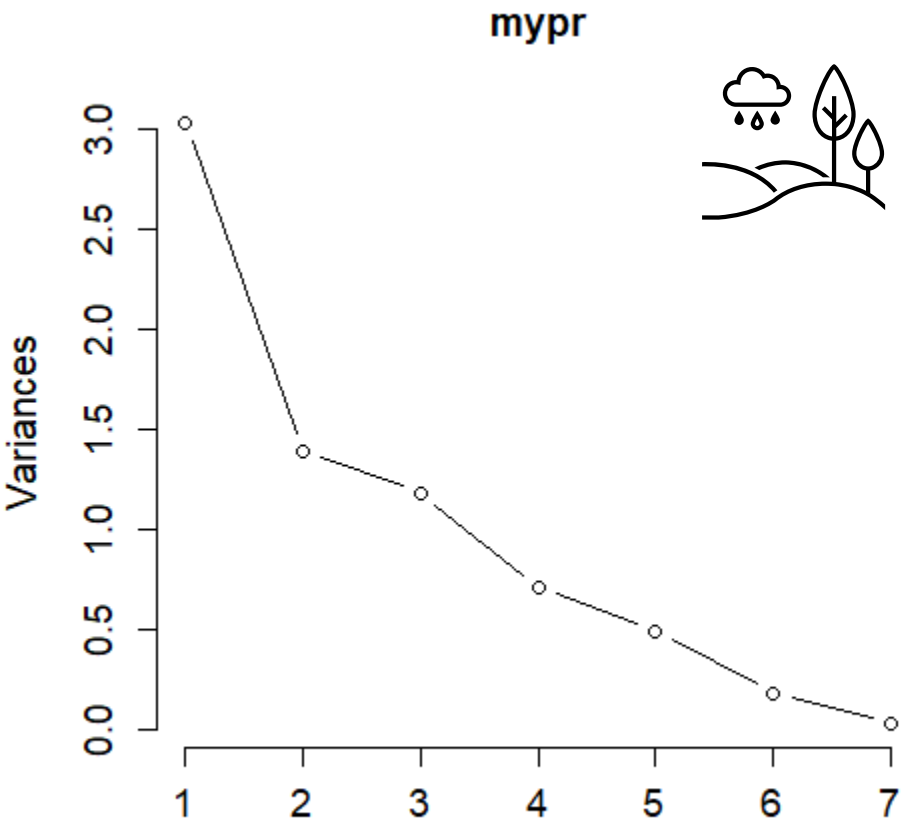
While it is useful to include many environmental variables for the sake of orientation to the survey region and transparency, the limited number of sample sites (only ten in this study) will require few variables to be evaluated using regression and multivariate methods. So, I will report all variables in a site x environment matrix as an appendix and include a more thorough introduction to subtropical basin and stream characteristics in the methods. Then I will proceed with a limited number of interpretable variables selected to evaluate climate, hydrologic, water quality, and local habitat features as drivers of community assembly.

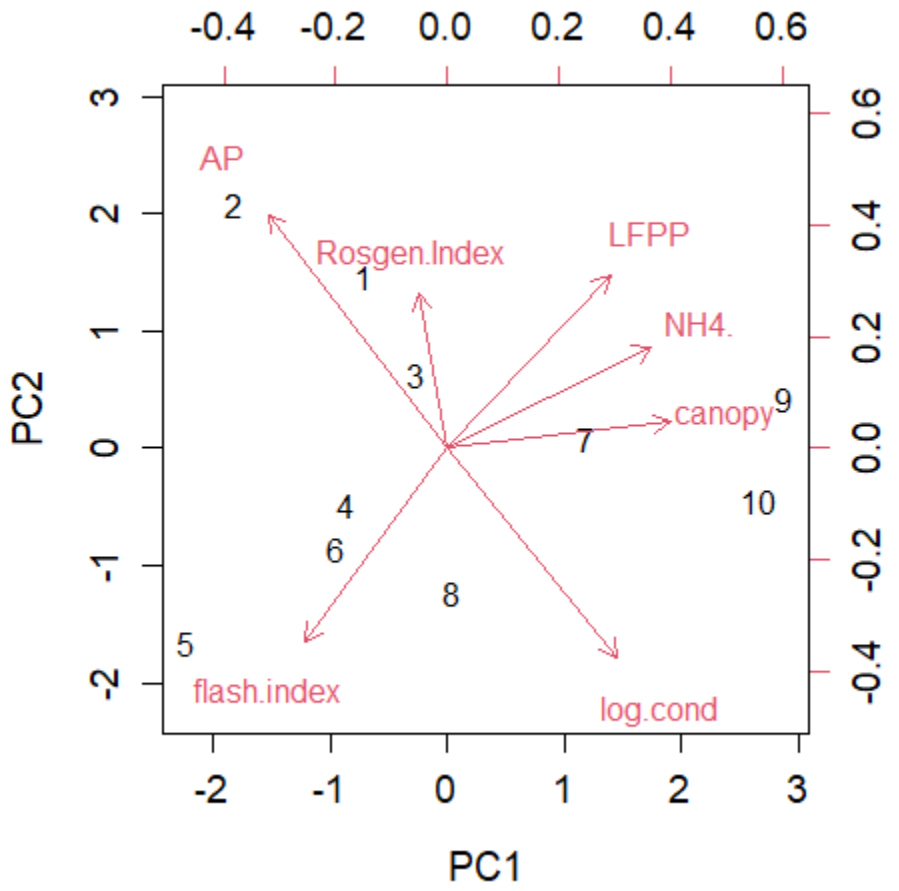
A-piori variable selection for regression and ordinations

7 variables to evaluate with community diversity metrics and community composition ordinations
Precipitation: Annual precipitation (30-year average)
Flood disturbance regime: flash index
Drought disturbance regime: Low Flow Pulse Percent
Biogenic Pollutants: NH_4^+
Osmotic stressors: Conductivity
Canopy effects: Canopy Coverage
Stream morphology: Rosgen index

1. *Use PCA to discern patterns in environmental variation among sites.*
2. *Use linear regressions to identify significant correlations with fish and invertebrate community diversity.*
3. *Use multivariate regressions to create predictive models for fish and invertebrate community diversity.*
4. *Use PCoA ordinations to visualize patterns in community composition and identify coincidental environmental predictors*

1. Use PCA to discern patterns in environmental variation among sites.





variable	PC1
canopy	0.867128
NH4.	0.792109
AP	-0.69812
log.cond	0.662801
LFPP	0.633236
flash.index	-0.55337
Rosgen.Index	-0.11131

variable	PC2
AP	0.615274
log.cond	-0.54994
flash.index	-0.50805
LFPP	0.455821
Rosgen.Index	0.410933
NH4.	0.265428
canopy	0.068905

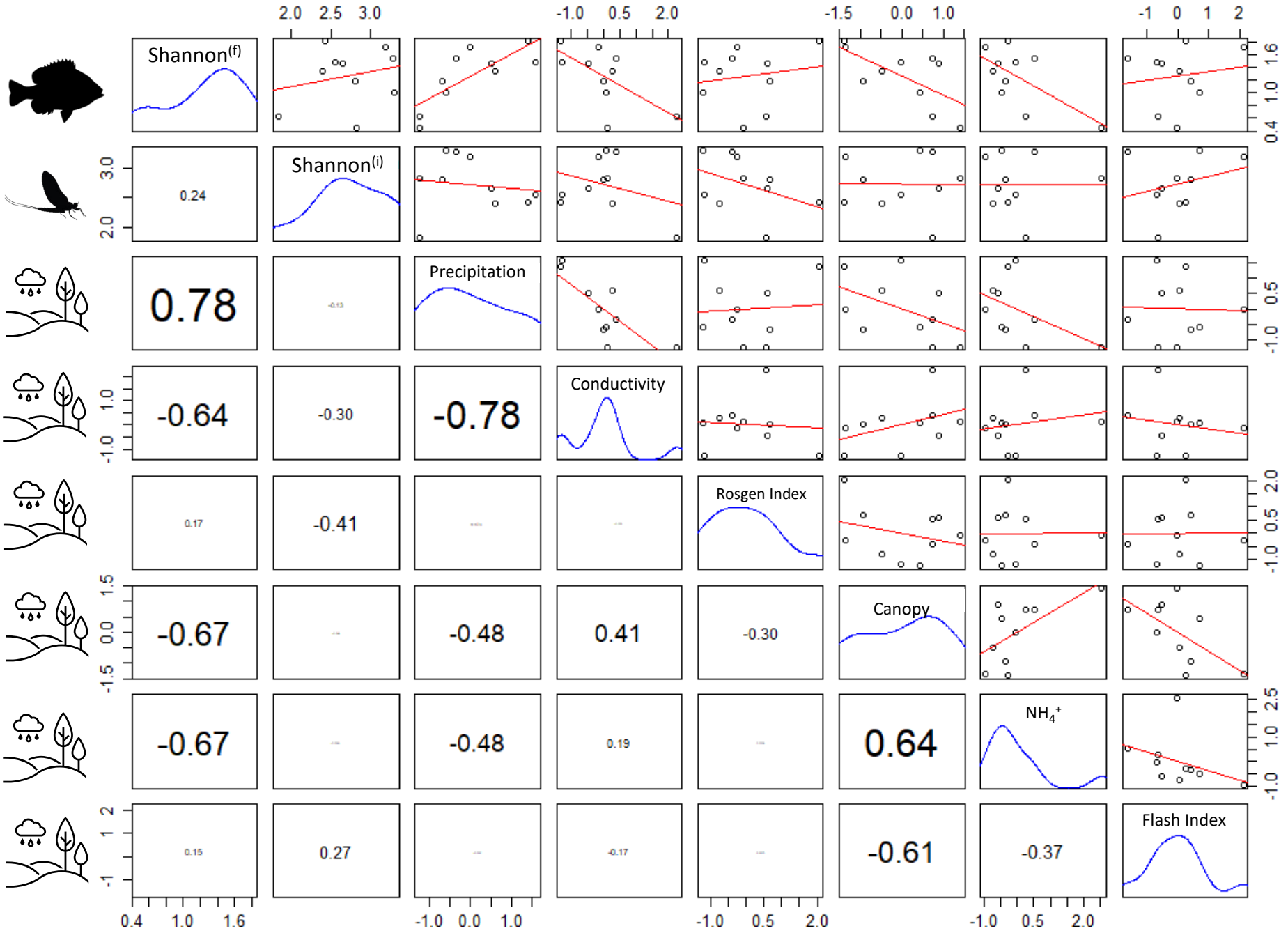
metric	Standard deviation	Proportion of Variance	Cumulative Proportion
PC1	1.739434199	0.43223	0.43223
PC2	1.179388223	0.19871	0.63094
PC3	1.085778668	0.16842	0.79936
PC4	0.839789846	0.10075	0.90011
PC5	0.699978246	0.07	0.9701
PC6	0.424007876	0.02568	0.99579
PC7	0.171748545	0.00421	1

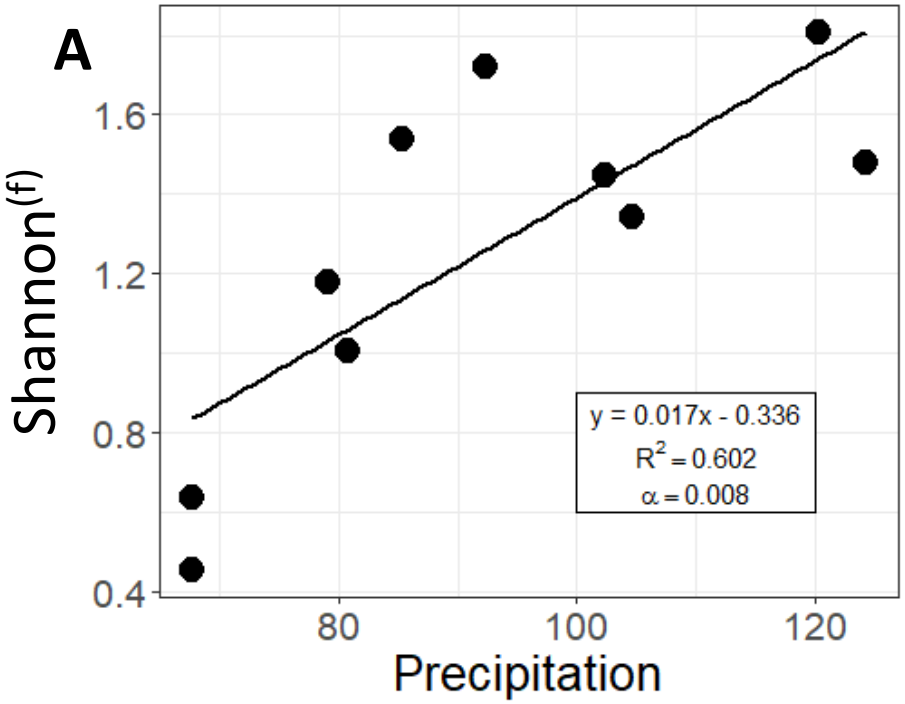
Scree plot (left) and biplot (right) of Principle Component Analysis for 7 environmental predictors to be evaluated as predictors of fish and invertebrate assembly. Left, table contains cumulative variance explained by principle components. Above, Tables display correlations of environmental predictors with principle component axis 1 (top) and principle component 2 (middle) in descending order.

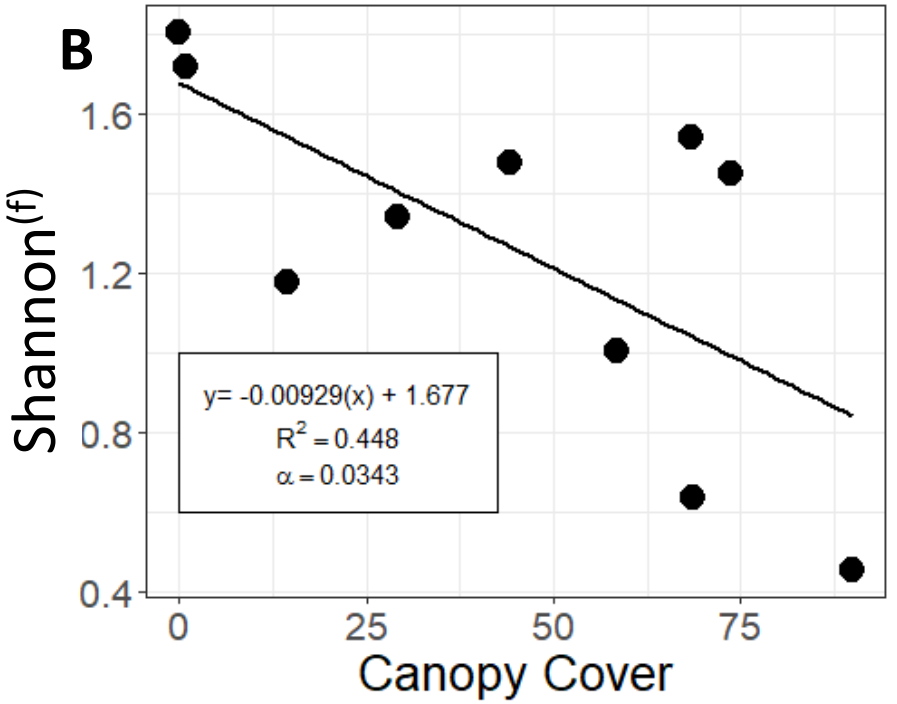
2. Use linear regressions to identify significant correlations with fish and invertebrate community diversity.

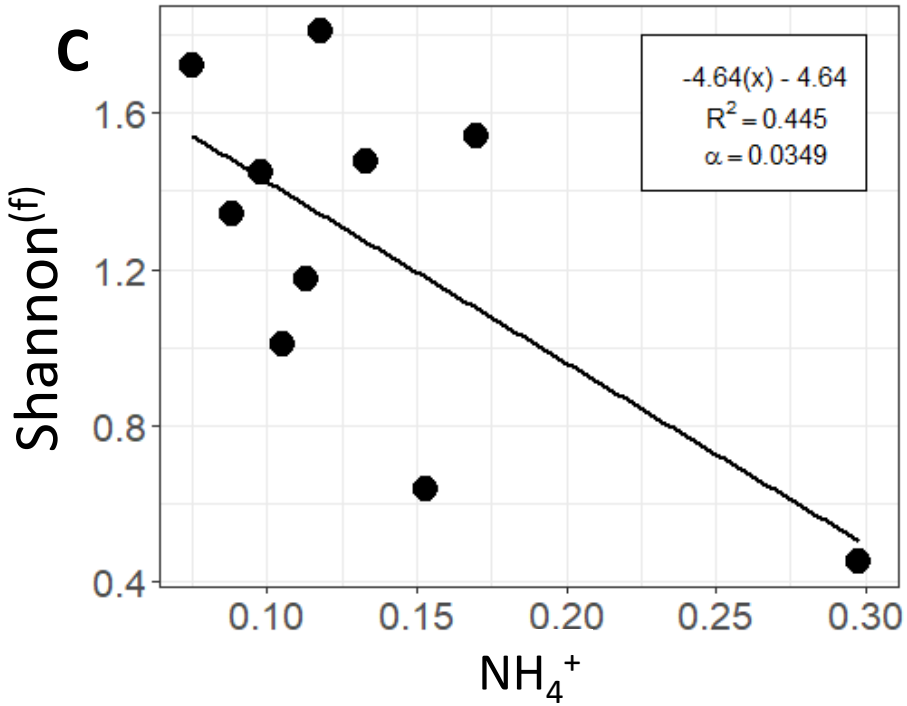
Pairs plots for fish and invertebrate Shannon index versus 7 scaled environmental predictors. Scatterplots and linear regressions are plotted in the top. Correlation coefficients are displayed in the bottom plots.

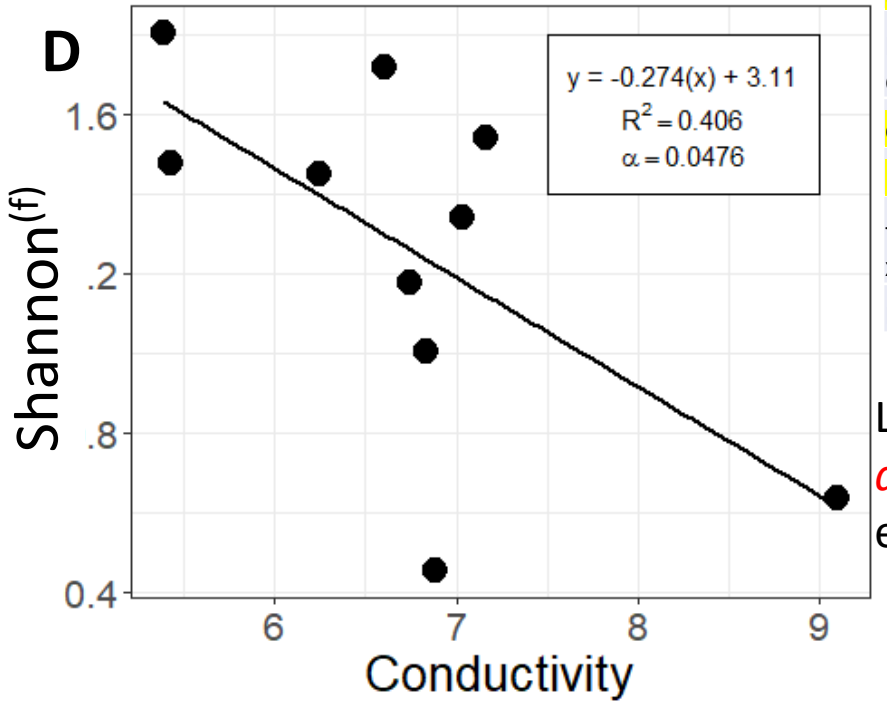
This is a useful tool to identify strong correlations with our dependent variables (first 2 rows) and to see collinearities between independent variables (bottom 6 rows). AP covaries with conductivity, canopy, and NH4+ concentrations.











Fish diversity plotted against (A) annual precipitation (cm/yr), (B) Canopy cover (%), (C) NH4+ (mg/L), (D) log-transformed conductivity (µS/cm).

predictor	estimate	df	r2	f.stat	p.valu
AP	0.017271	2 8	0.602017	12.10136	0.008
log.cond	-0.27437	2 8	0.405813	5.463779	0.047
Rosgen.Index	0.021205	2 8	0.027242	0.224035	0.648
canopy	-0.0093	2 8	0.447887	6.489777	0.034
NH4.	-4.64708	2 8	0.445651	6.431349	0.034
flash.index	0.338843	2 8	0.023591	0.193286	0.67
LFPP	-0.03083	2 8	0.338929	4.101581	0.077

Linear regression statistics for *a-priori* selected environmental variables

3. Use multivariate regressions to create predictive models for fish and invertebrate community diversity.

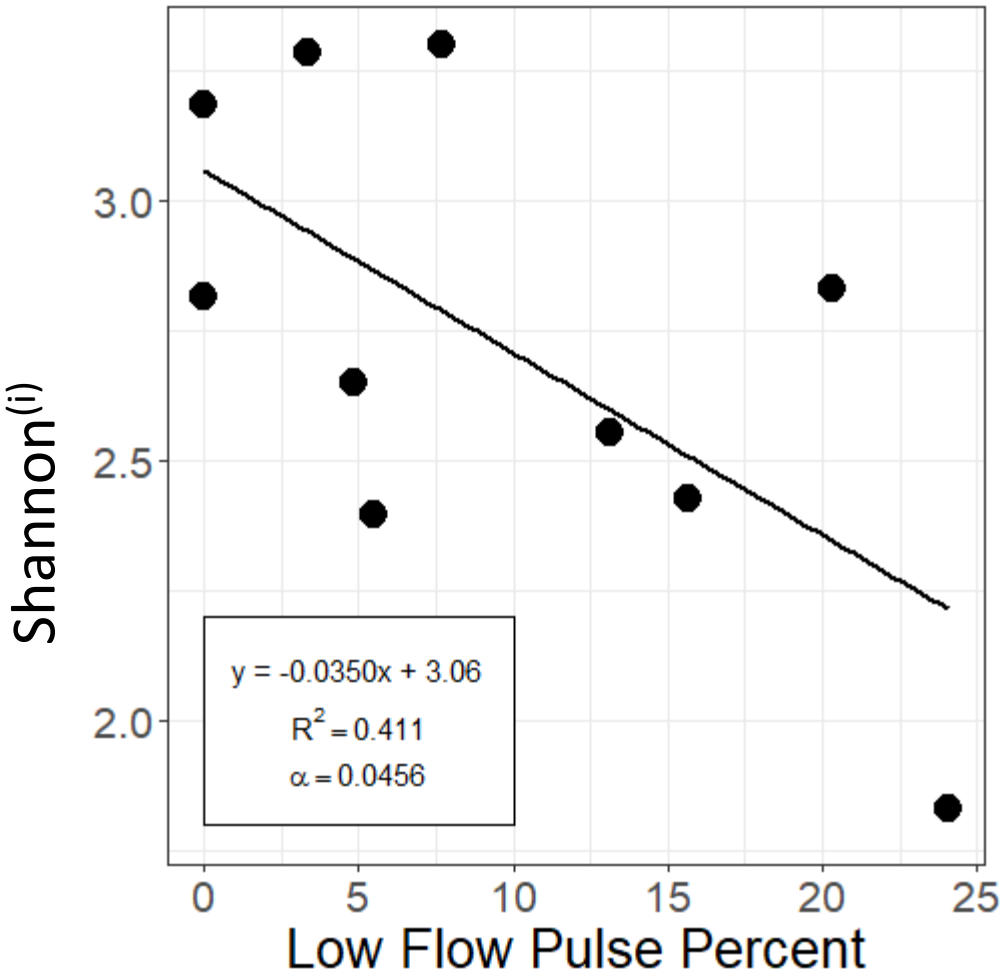
model	(Int)	AP	cnp	fls.ind	LFPP	log.cnd	NH4.	Rsg.Ind	df	logLik	AICc	delta	weight
10	1.261	0.3127			-0.2098				4	2.912	10.2	0	0.429
2	1.261	0.3462							3	-0.985	12	1.79	0.175
4	1.261	0.2634	-0.1715						4	0.69	14.6	4.44	0.046
34	1.261	0.2641					-0.1711		4	0.688	14.6	4.45	0.046
49	1.261					-0.2352	-0.2521		4	0.649	14.7	4.52	0.045
3	1.261		-0.2986						3	-2.622	15.2	5.07	0.034
33	1.261						-0.2979		3	-2.642	15.3	5.11	0.033
17	1.261					-0.2842			3	-2.989	16	5.8	0.024
74	1.261	0.3014			-0.2338			0.1025	5	4.501	16	5.82	0.023
1	1.261								2	-5.592	16.9	6.72	0.015

Top ten Multivariate regression models predicting fish Shannon diversity.

model_name	var1	estimate	var2	estimate2	var3	estimate3	adjusted_r2	f.stat	df	p.value
Fm10	AP	0.312661569	LFPP	-0.209790458	na	na	0.765305514	15.67386332	2 7	0.00259869
Fm4	AP	0.263436925	canopy	-0.171468311	na	na	0.633976268	8.794284779	2 7	0.012310529
Fm34	AP	0.264055079	NH4.	-0.171113979	na	na	0.633850528	8.790062783	2 7	0.012325337
Fm49	log.cond	-0.235193285	NH4.	-0.252107645	na	na	0.631013908	8.695581603	2 7	0.012662788
Fm74	AP	0.30138	LFPP	-0.23384	Rosgen.Index	0.10252	0.8007	13.05	3 6	0.004866

Summary statistics for top five multivariate regressions predicant fish Shannon diversity

Precipitation is a predictor in the top multivariate models and has the largest estimated slope compared to all other scaled environmental variables in the exhaustive multivariate regression.
One multivariate regression utilizes 2 water quality parameters (conductivity and NH4).
Positive predictors of fish diversity include AP and Rosgen Index
Negative predictors of fish diversity include LFPP, canopy, NH4., log.cond



Invertebrate Shannon index plotted against Low Flow Pulse Percent.

predictor	estimate	df	r2	f.stat	p.value
AP	-0.00302	2 8	0.01736	0.141333	0.716733
log.cond	-0.13388	2 8	0.091107	0.801919	0.396664
Rosgen.Index	-0.05419	2 8	0.167748	1.612478	0.239833
canopy	-0.00022	2 8	0.000243	0.001946	0.965892
NH4.	-0.06524	2 8	8.28E-05	0.000662	0.980096
flash.index	0.622914	2 8	0.075173	0.650264	0.443328
LFPP	-0.03498	2 8	0.411386	5.591259	0.045631

Linear regression statistics for invertebrate Shannon index versus *a-priori* selected environmental variables

model	(Int)	AP	cnp	fls.ind	LFP	log.cnd	NH4.	Rsg.Ind	df	logLik	AICc	delta	weight
9	3.057				-0.03498				3	-3.236	16.5	0	0.299
1	2.727								2	-5.886	17.5	1.01	0.18
41	2.676				-0.05369		4.133		4	-0.967	17.9	1.46	0.144
65	3.596							-0.05419	3	-4.968	19.9	3.46	0.053
17	3.629					-0.1339			3	-5.409	20.8	4.34	0.034
5	2.159			0.6229					3	-5.495	21	4.52	0.031
11	2.926		4.41E-03		-0.04203				4	-2.524	21	4.58	0.03
73	3.628				-0.03162			-0.03754	4	-2.538	21.1	4.6	0.03
10	3.587	-0.00551			-0.03708				4	-2.733	21.5	4.99	0.025
2	3.006	-0.00302							3	-5.799	21.6	5.12	0.023

Top ten Multivariate regression models predicting invertebrate Shannon diversity.

model_name	var1	estimate	var2	estimate2	adjusted_r2	f.stat	df	p.value
im1	LFPP	-0.05369	NH4.	4.133124	0.519315	5.86164	2 7	0.031953
im2	canopy	0.004406	LFPP	-0.04203	0.343686	3.35647	2 7	0.095036
im3	AP	-0.00551	LFPP	-0.03708	0.315609	3.075187	2 7	0.110042

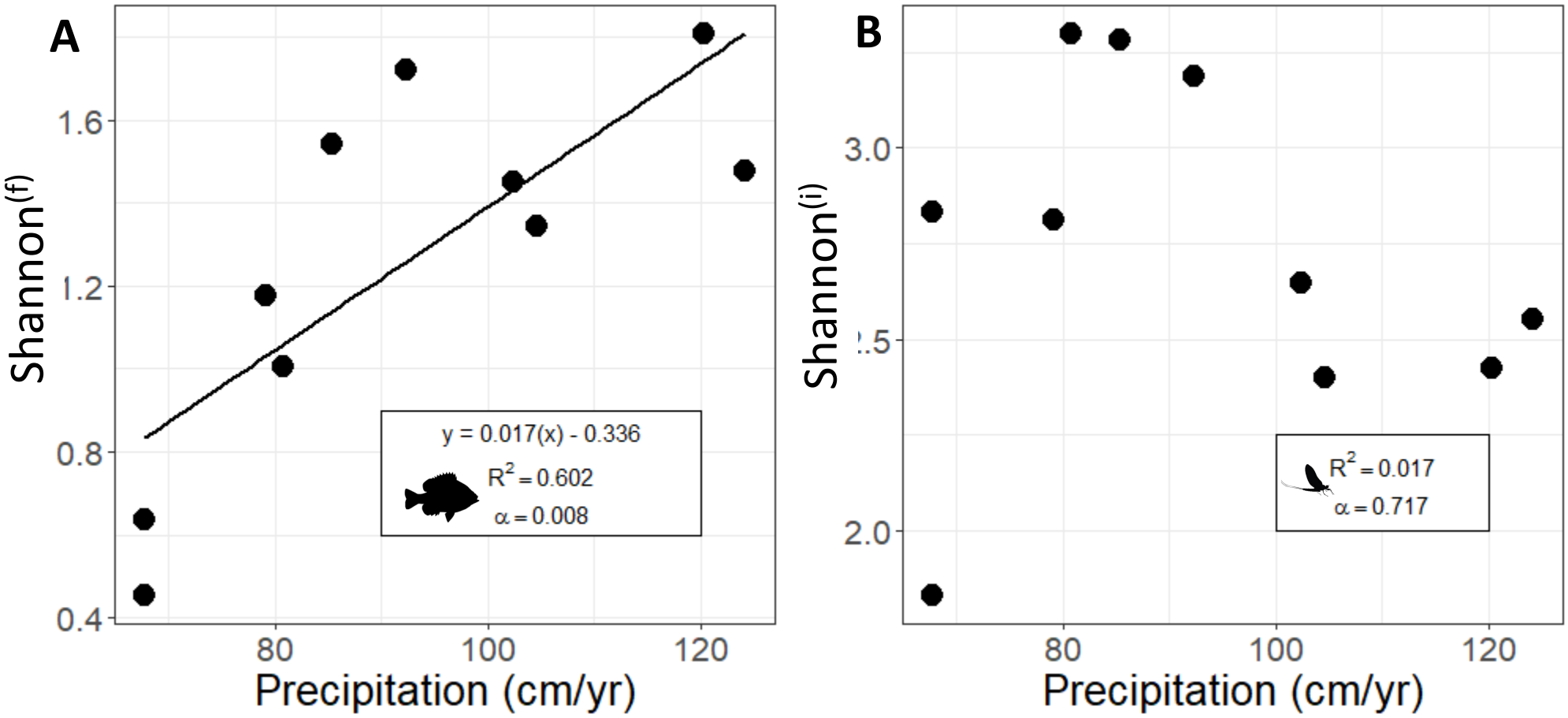
Summary statistics for top five multivariate regressions predicant invertebrate Shannon diversity

7/10 top AICc models contained one or fewer environmental predictors. Previous univariate regressions indicate no statistically significant correlations

1/3 of the top AICc ranked multivariation regression models has a p value less than 0.05 and has an adjusted r^2 of 0.519 (-LFPP + NH4)

Positive predictors of invertebrate diversity include NH4 and canopy
Negative predictors of invertebrate diversity include LFPP and AP

Hypothesis test: precipitation correlates with diversity?



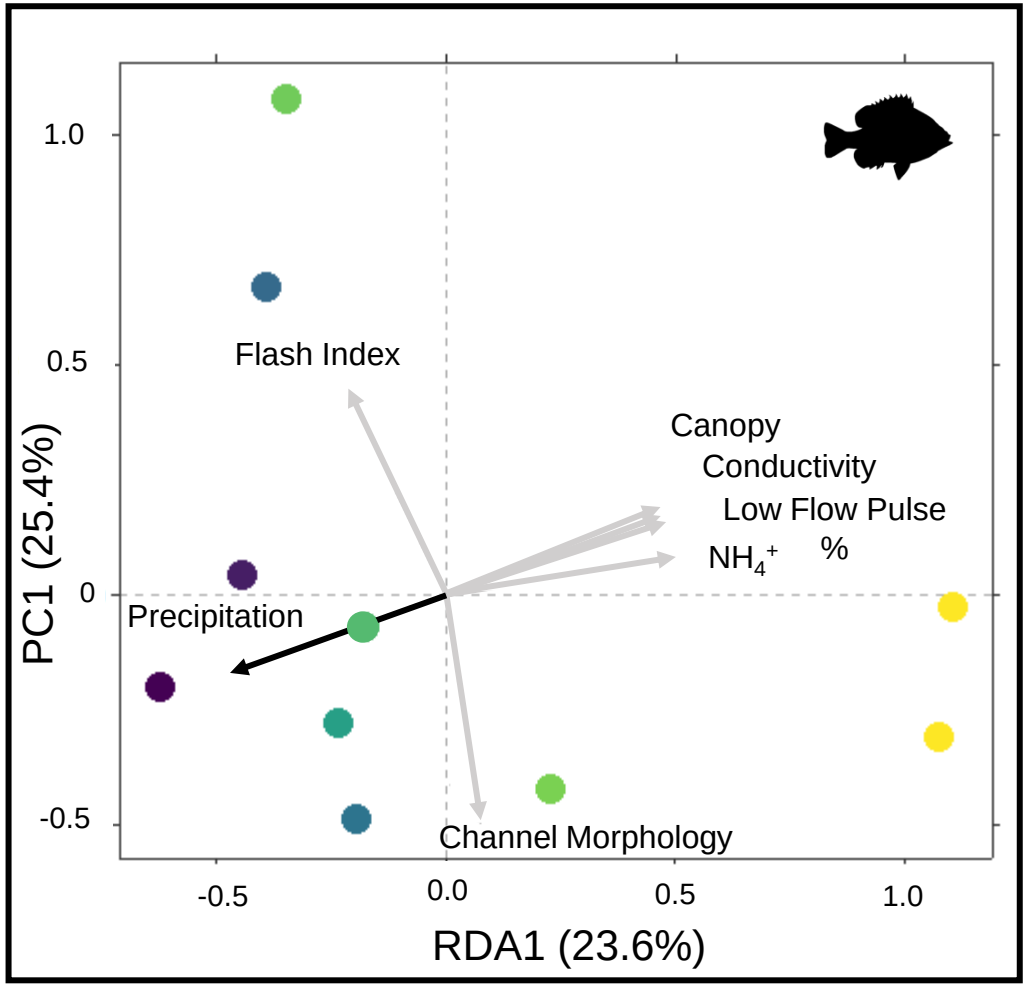
Shannon diversity of fish (A) and invertebrate (B) communities for each site plotted against annual precipitation.

Fish Shannon index correlates positively with annual precipitation but invertebrate Shannon index does not significantly correlate.

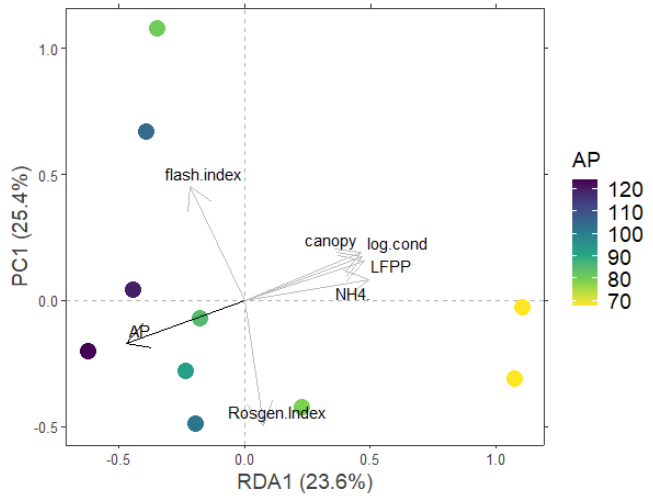
4. Use PCoA ordinations to visualize patterns in community composition and identify coincidental environmental predictors

Replace NMDS with PCoA
Replace envfit with correspondence anlaysis

climate drives marine fish communities in slack paper
one plot of sites
one plot of vectors ccoa
table of results cca1, cca2, r2, df, chix2, fval



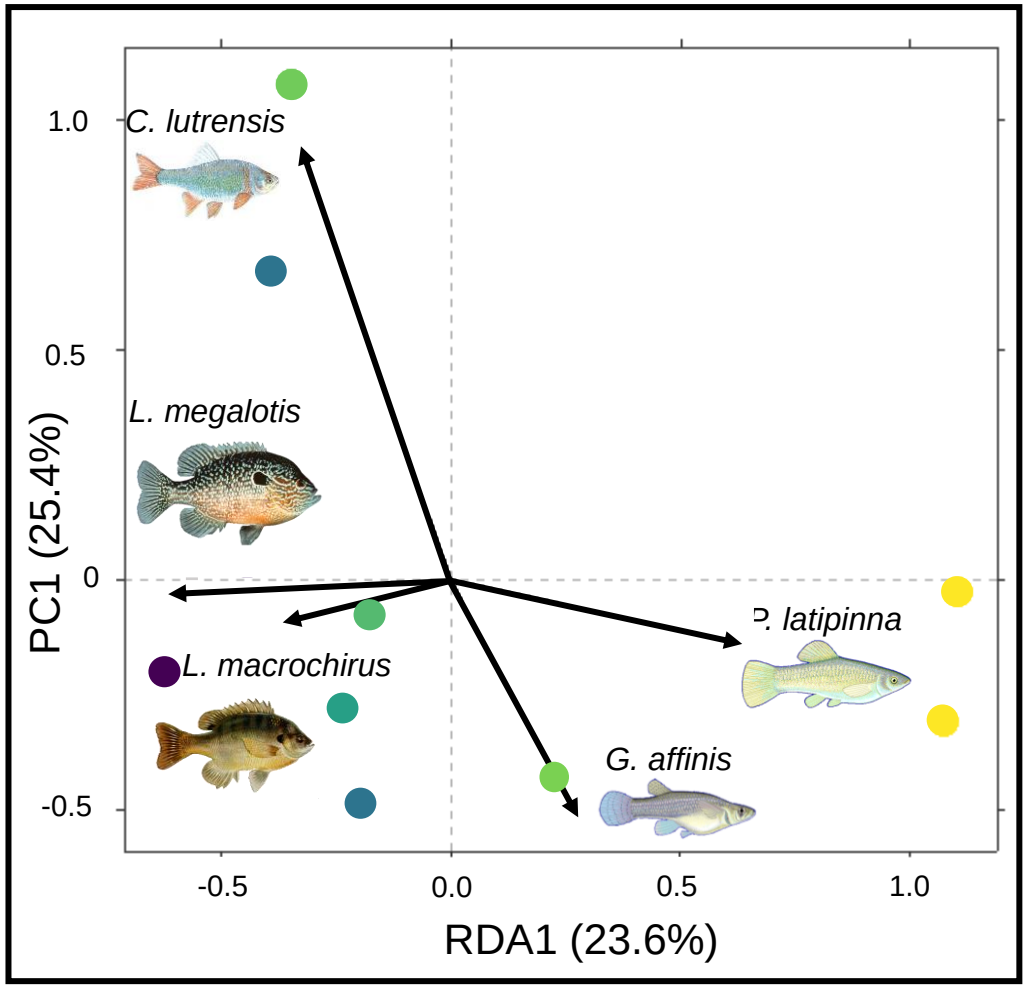
species	R^2	p-value	axis1	axis2
A.melas	0.23023817	0.423	-0.022387738	0.049458134
A.rostrata	0.04895361	0.875	-0.004901997	0.018920955
H.cyanoguttatus	0.11674133	0.703	0.044499826	0.06724042
C.lutrensis	0.8797549	0.007	-0.217359354	0.626652466
C.venusta	0.19174898	0.496	-0.074923046	-0.018880656
E.gracile	0.01702391	1	0.023746088	-0.009510118
F.notatus	0.26996839	0.244	-0.096394548	-0.061708444
G.affinis	0.61787634	0.041	0.187553123	-0.339810377
L.auritis	0.0954481	0.747	0.036414193	0.080138119
L.cyanellus	0.27043676	0.32	-0.072507811	-0.166251083
L.gulosus	0.17486578	0.476	-0.099324212	-0.010412806
L.humilis	0.16582945	0.594	-0.017865623	-0.034986287
L.macrochirus	0.75601603	0.018	-0.240268678	-0.061350821
L.marginatus	0.16582945	0.594	-0.025265807	-0.049478082
L.megalotis	0.72038963	0.013	-0.407334824	-0.020938185
P.latipinna	0.8374159	0.016	0.414714616	-0.089880974
P.vigilax	0.02933178	0.914	-0.136897484	0.033999261
T.maculatus	0.23023817	0.423	-0.012925567	0.028554667



Redundancy Analysis : Hellinger-transformed fish community matrix
Constrained by Annual Precipitation

Sites are colored by annual precipitation (cm/year)

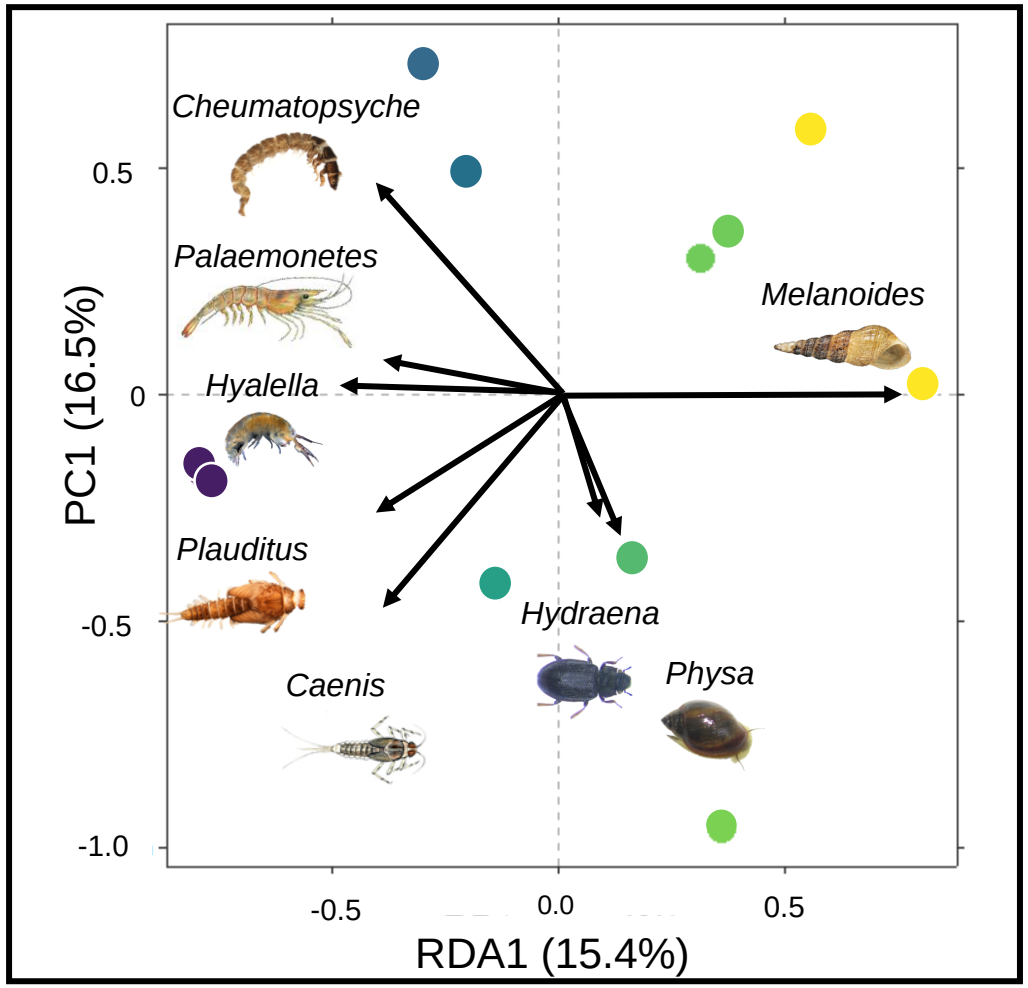
Arrows: The positions of species are interpreted as arrows generated by maximum correlation in ordination space (see table for correlation outputs). Only vectors with *p*-values <0.05 are plotted to improve figure clarity.



predictor	R^2	p-value	axis1	axis2
AP	0.725622	0.013	-0.940598	-0.3395
log.cond	0.538635	0.051	0.9384577	0.34539
Rosgen.Index	0.240689	0.363	0.1488126	-0.9889
canopy	0.292377	0.302	0.9238926	0.38265
NH4.	0.493665	0.104	0.985893	0.16738
flash.index	0.03314	0.877	-0.433603	0.9011
LFPP	0.356723	0.214	0.948997	0.31529

The projections of points onto vectors have maximum correlation with corresponding environmental variables, and the factors show the averages of factor levels.

- Redundancy Analysis : Hellinger-transformed fish community matrix
  - Constrained by Annual Precipitation
- Sites are colored by annual precipitation (cm/year)
- Arrows: Environmental predictors are interpreted as arrows generated by maximum correlation in ordination space (see table for correlation outputs).

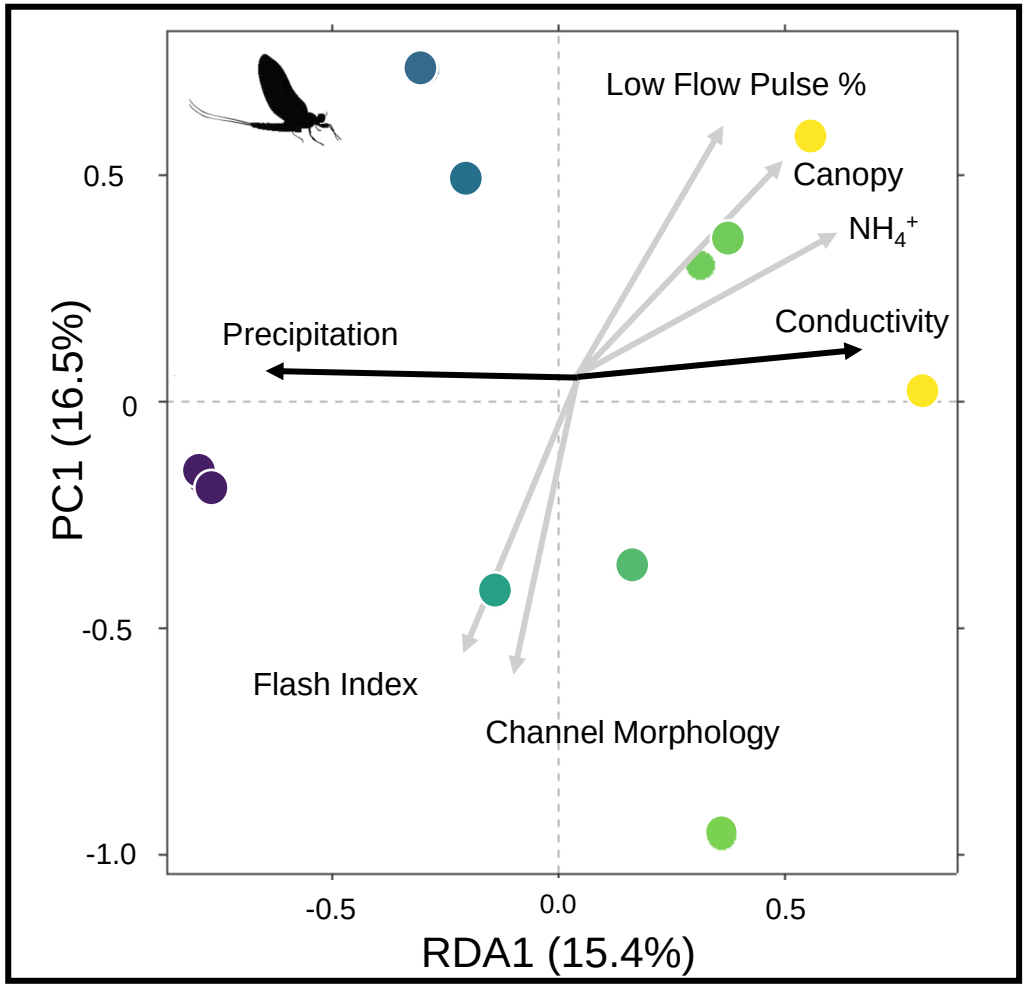


Species	R^2	p-value	axis1	axis2
Alisotrichia	0.278124	0.497	-0.0787	-0.01637
Ameletus	0.09205	0.697	-0.05197	-0.01859
Ametropus	0	1	0	3.23E-17
Amnicola	0.557658	0.072	0.067816	0.19706
Amphiagrion	0.496395	0.102	0.033336	-0.12992
Ancyronyx	0.009591	0.968	-0.00869	0.014862
Argia	0.37329	0.205	0.147555	-0.05412
Baetis	0.013734	0.975	0.026729	-0.00221
Baetodes	0.278124	0.497	-0.0787	-0.01637
Belostoma	0.270257	0.602	-0.07326	-0.01926
Berosus	0	1	0	0
Biomphalaria	0.517251	0.076	0.119822	-0.08838
Bithynia	0.174315	0.501	-0.01336	-0.1058
Brachycerus	0.29809	0.312	-0.07302	-0.05883
Brechmorhoga	0.392815	0.167	0.070816	0.064984
Caecidotea	0.278631	0.391	0.073592	0.066135
Caenis	0.642217	0.034	-0.13507	-0.16165
Calopteryx	0.071467	1	0.014141	-0.0277
Camelobaetidius	0.081087	0.9	0.026557	0.025767
Campeloma	0.197974	0.544	0.048071	0.090109
Centroptilum	0.470995	0.109	0.033093	-0.09035
Cercobrachys	0.09205	0.697	-0.05197	-0.01859
Cheumatopsyche	0.740952	0.009	-0.13976	0.158003
Chromagrion	0.071467	1	0.014141	-0.0277
Cloeon	0	1	0	0
Corbicula	0.164334	0.551	-0.09754	0.057242
Cyphon	0.251092	0.34	-0.04735	0.103009
Dubiraphia	0.30962	0.265	-0.02857	0.121229
Elodes	0.125092	0.694	-0.02283	0.042559
Erpetogomphus	0.392815	0.167	0.100149	0.091902
Erythemis	0.470995	0.109	0.033093	-0.09035
Fallceon	0.624436	0.063	-0.17455	-0.03264
Farrodes	0.078979	0.737	0.033091	-0.03746
Ferrissia	0.041635	0.869	0.010086	0.033082
Fossaria	0.305494	0.351	0.10087	-0.02573
Gammarus	0.392815	0.167	0.070816	0.064984
Glaenocorisa	0	1	0	0
Gyraulus	0	1	0	0
Gyretes	0.071467	1	0.014141	-0.0277
Hebrus	0.278631	0.391	0.052037	0.046764
Helicina	0.081087	0.9	0.018779	0.01822
Helisoma	0.102486	0.7	-0.03842	0.048578
Hetaerina	0	1	0	0
Heterelmis	0.071467	1	0.014141	-0.0277
Hyalella	0.780031	0.005	-0.16756	0.005313
Hydraena	0.552657	0.065	0.033265	-0.11833
Hydrobius	0.470995	0.109	0.033093	-0.09035
Hydropsyche	0.081087	0.9	0.018779	0.01822
Isonychiidae	0.081087	0.9	0.026557	0.025767
Leptoceridae	0.055541	0.793	0.046556	-0.0134
Leptohyphes	0.302371	0.265	-0.05895	-0.07494
Leptonema	0.278631	0.391	0.052037	0.046764
Limnocoris	0.071467	1	0.014141	-0.0277
Limnoporus	0.087723	0.8	0.000297	-0.04846
Linisa	0.071467	1	0.014141	-0.0277
Lipogomphus	0.081087	0.9	0.018779	0.01822
Macrelmis	0.081087	0.9	0.018779	0.01822
Macronychus	0.125092	0.694	-0.02283	0.042559
Melanoides	0.691493	0.017	0.265552	-0.00212
Menetus	0	1	0	0
Mesovelia	0.470995	0.109	0.033093	-0.09035
Metrichia	0.087723	0.8	0.000171	-0.02798
Microvelia	0.087723	0.8	0.000171	-0.02798
Morphocorixa	0	1	0	0
Neoelmis	0	1	0	0
Neoneura	0.278631	0.391	0.052037	0.046764
Neoplea	0.012323	0.925	0.01895	-0.00976
Orconectes	0.278631	0.391	0.052037	0.046764
Palaemonetes	0.574876	0.042	-0.13503	0.023684
Pelocoris	0.071467	1	0.014141	-0.0277
Peltodytes	0.510337	0.089	-0.01776	-0.17184
Philopotamidae	0.081087	0.9	0.026557	0.025767
Physa	0.638867	0.03	0.047477	-0.15761
Physella	0.158681	0.555	-0.05431	-0.02902
Pisidium	0.279282	0.303	-0.03866	0.088146
Planorbula	0	1	0	0
Plauditus	0.661025	0.045	-0.13765	-0.09131
Potamyia	0	1	0	0
Procloeon	0.071467	1	0.014141	-0.0277
Pseudocloeon	0.087723	0.8	0.000171	-0.02798
Pseudosuccinea	0.016499	0.957	-0.0148	0.028147
Ranatra	0.470995	0.109	0.033093	-0.09035
Rhagovelia	0.392815	0.167	0.100149	0.091902
Rhematobates	0.081087	0.9	0.018779	0.01822
Scirtes	0.071467	1	0.014141	-0.0277
Siphlonurus	0.278631	0.391	0.052037	0.046764
Smicridea	0	1	0	0
Sphaeriidinae	0	1	0	0
Stactobiella	0.279282	0.303	-0.03866	0.088146
Stenelmis	0.22286	0.413	-0.02798	0.125115
Stenonema	0.125092	0.694	-0.03229	0.060187
Strobilops	0.087723	0.8	0.000171	-0.02798
Synaptonecta	0	1	0	0
Trepobates	0.15686	0.595	0.014383	-0.06727
Valvata	0.470995	0.109	0.033093	-0.09035

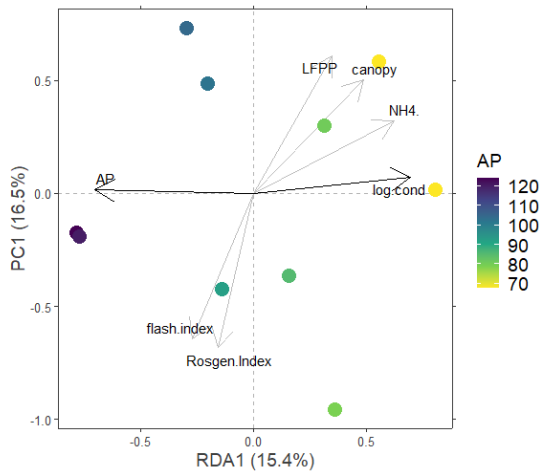
Redundancy Analysis : Hellinger-transformed invertebrate community matrix
Constrained by Annual Precipitation

Sites are colored by annual precipitation (cm/year)

Arrows: The positions of species are interpreted as arrows generated by maximum correlation

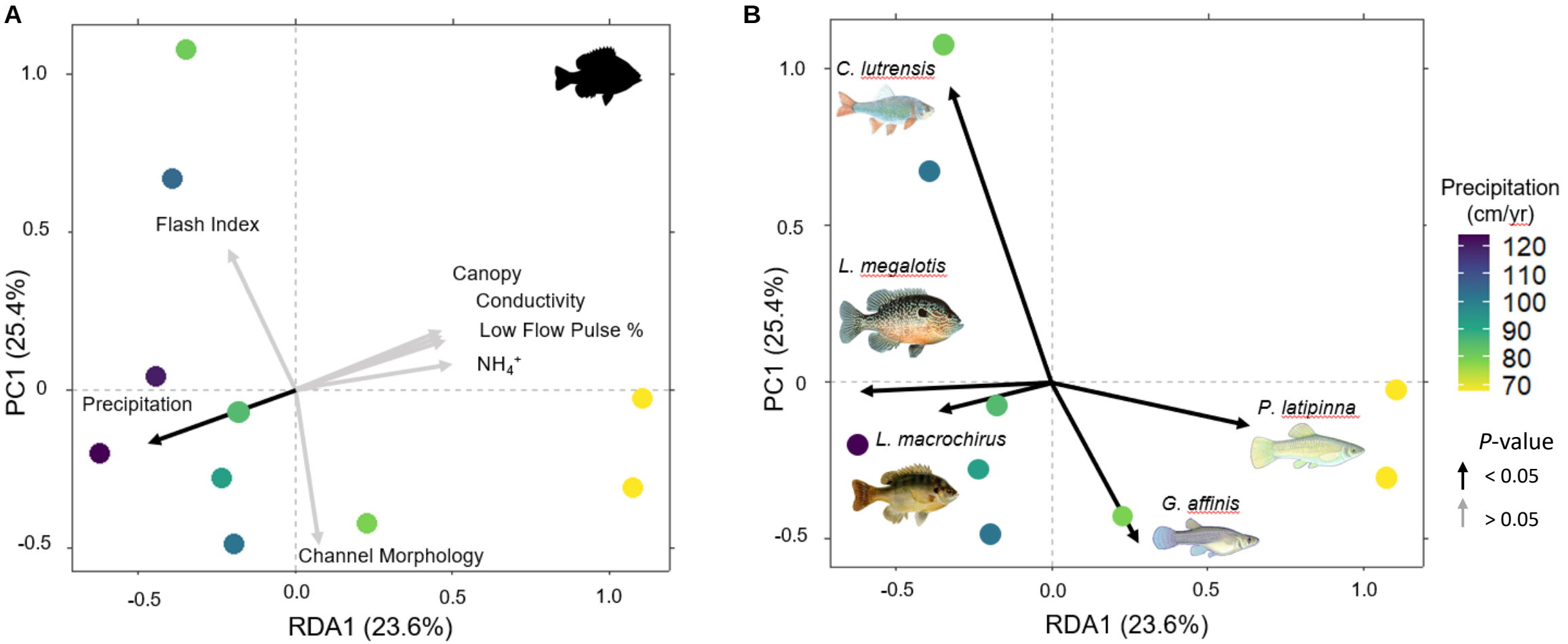


Predictor	axis1	axis2	R^2	p-value
AP	-0.9997573	0.02203034	0.97630889	0.001
log.cond	0.9947189	0.10263696	0.71112442	0.01
Rosgen.Index	-0.2220796	-0.97502853	0.07441911	0.766
canopy	0.6974804	0.7166039	0.54993554	0.069
NH4.	0.8878623	0.46010922	0.24973926	0.362
flash.index	-0.3886091	-0.92140273	0.02211306	0.932
LFPP	0.4940106	0.86945586	0.17072192	0.512

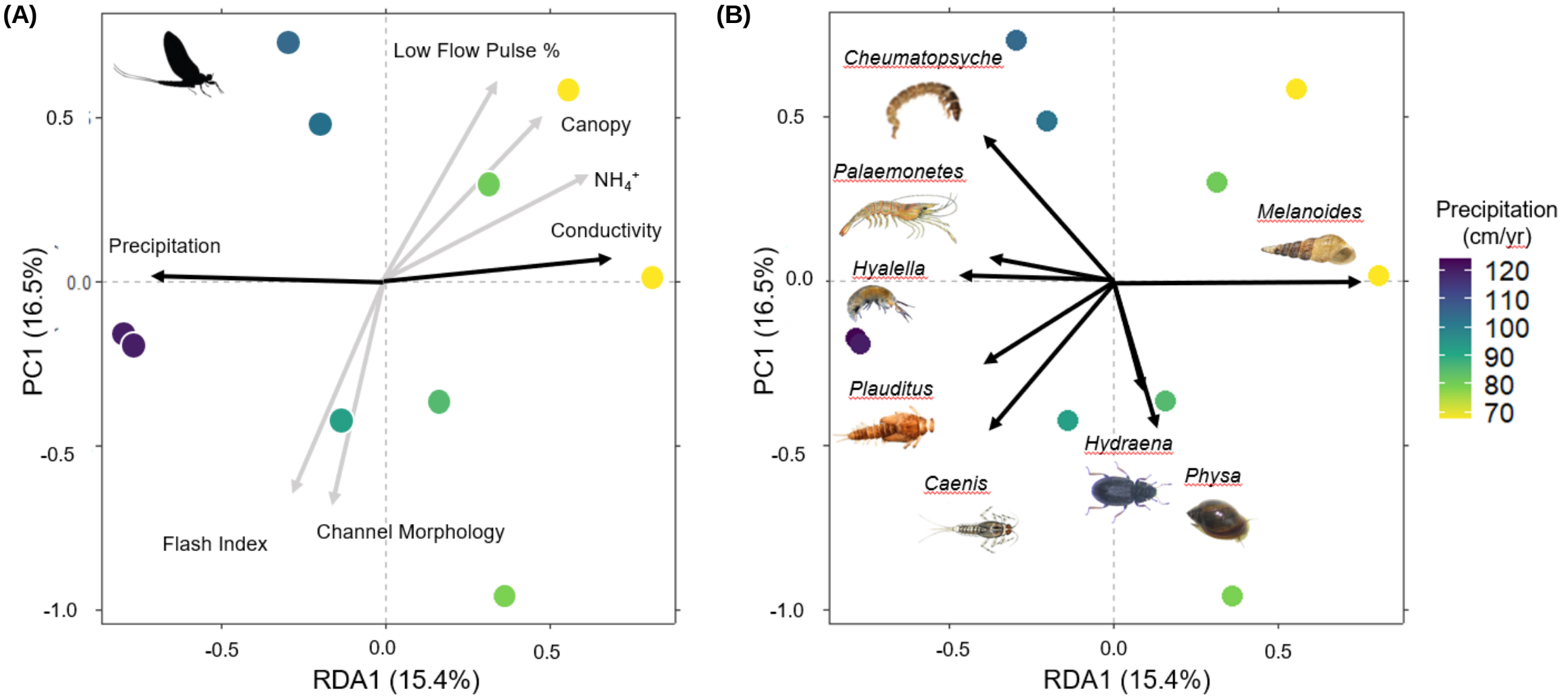


Redundancy Analysis : Hellinger-transformed invertebrate community matrix
Constrained by Annual Precipitation

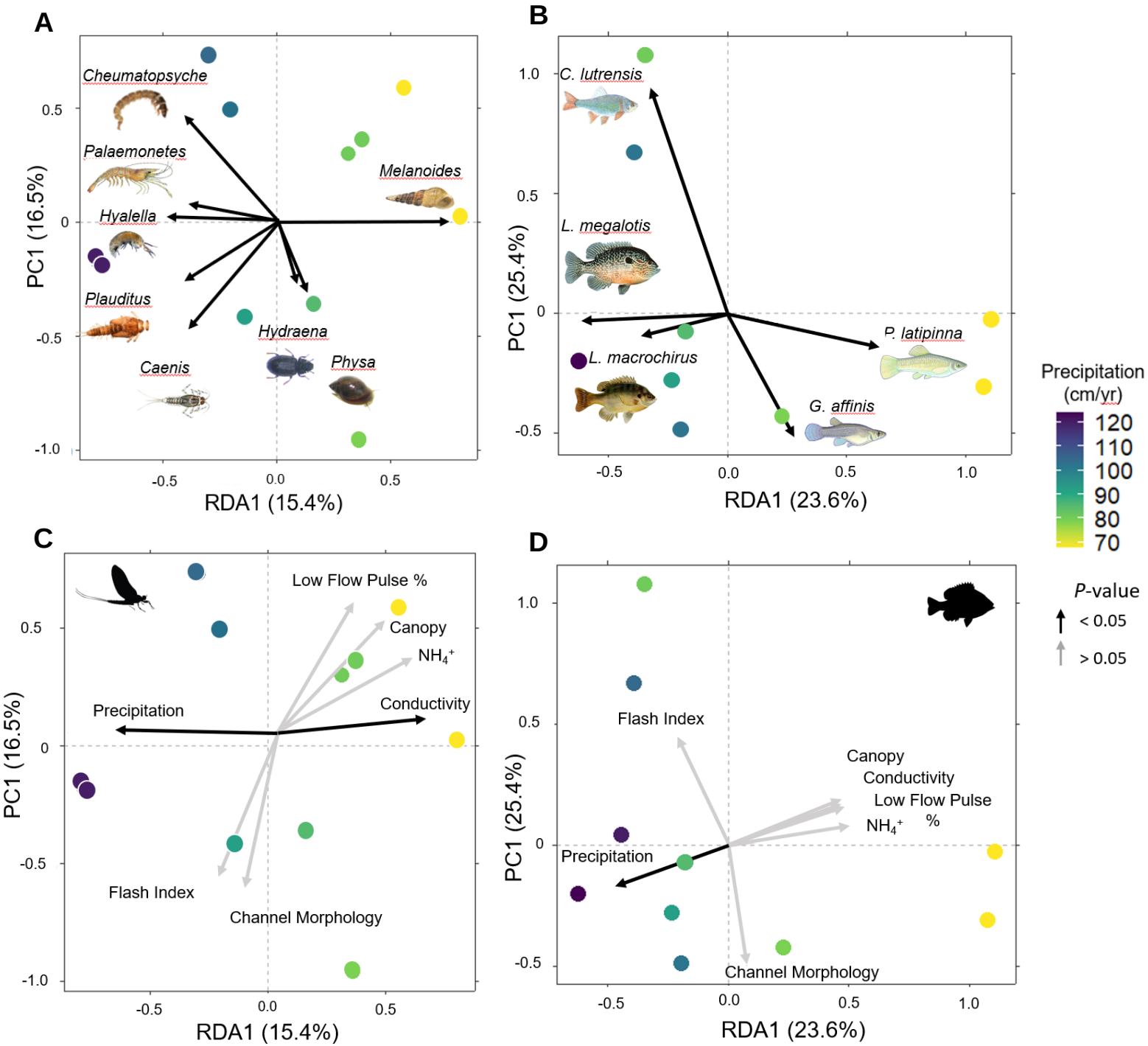
Sites are colored by annual precipitation (cm/year)



Fish community ordinations using Hellinger transformation and redundancy analysis. Dots represent sites with color determined by annual precipitation. Arrows represent fitted vectors for (A) environmental predictors and (C) species which can be visually interpreted based on their direction. Black arrows indicate statistically significant (*p*-value < 0.05) correlations. Only significant species vectors were plotted to improve figure clarity.



Macroinvertebrate community ordinations using Hellinger transformation and redundancy analysis. Dots represent sites with color determined by annual precipitation. Arrows represent fitted vectors for (A) environmental predictors and (B) species which can be visually interpreted based on their direction. Black arrows indicate statistically significant (p-value < 0.05) correlations. Only significant species vectors were plotted to improve figure clarity.



Ordinations of invertebrate (left: A & C) and fish (right: B & D) communities using Hellinger transformation and redundancy analysis. Dots represent sites with color determined by annual precipitation. Arrows represent fitted vectors for species (Top: A & B) and environmental predictors (bottom: C & D) which can be visually interpreted based on their direction. Black arrows indicate statistically significant (p-value < 0.05) correlations. Only significant species vectors were plotted to improve figure clarity.

Final Figures

~~Table1: Significant Environmental linear relationships (v1, v2, R2, p, +/-, fstat, df)~~

~~Figure1: Environmental PCA~~

Figure2: Fish Diversity and Ordinations

Table2: Fish multivariate regression outputs (v1, v2, v3, +/-, coef, R2, pm fstat, df)

Figure3: Invertebrate Diversity and ordination

Table3: Invertebrate multivariate regression tables

Appendix tables:

~~"T1_E_LM" ~ All env regression outputs~~

~~"F1_E_PCA" ~ All Environmental PCA outputs~~

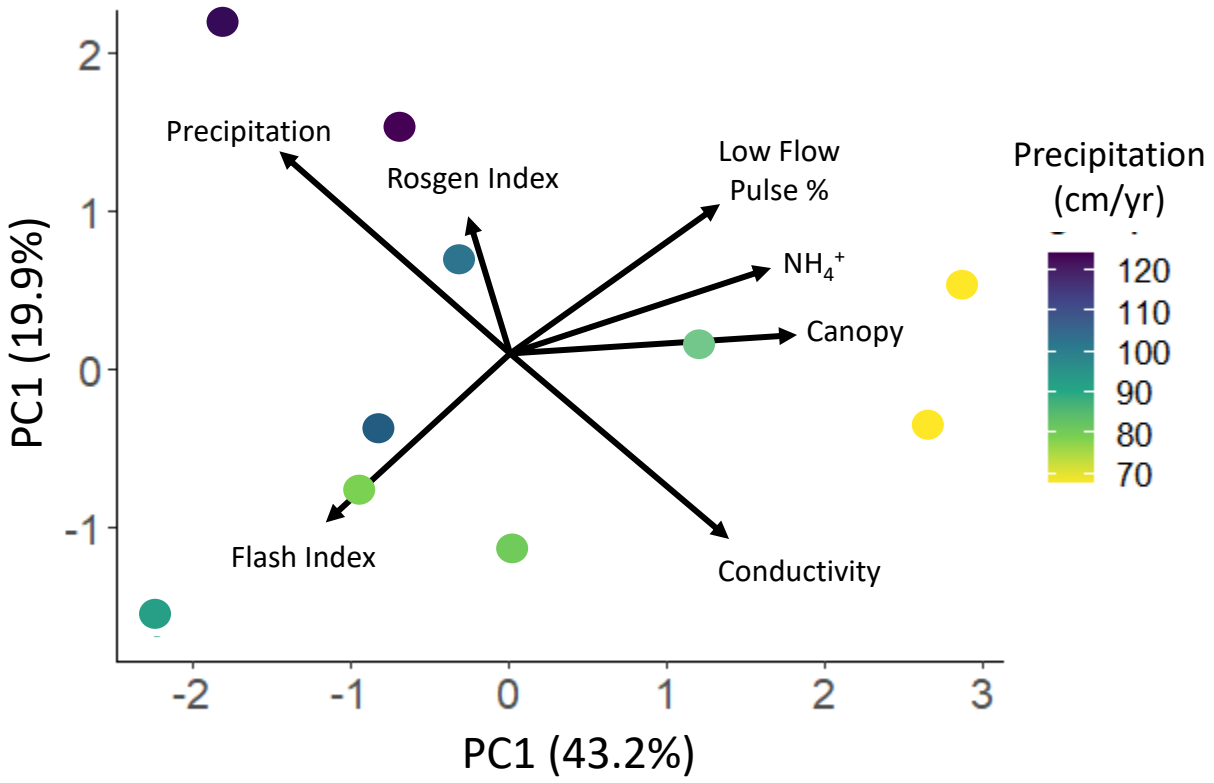
"F2_F_LM" ~ fish regression outputs

"F2_F_RDA" ~ ordination outputs (envfits)

"F3_I_LM" ~ Inv regression outputs

"F3_I_LM" ~ ordination outputs

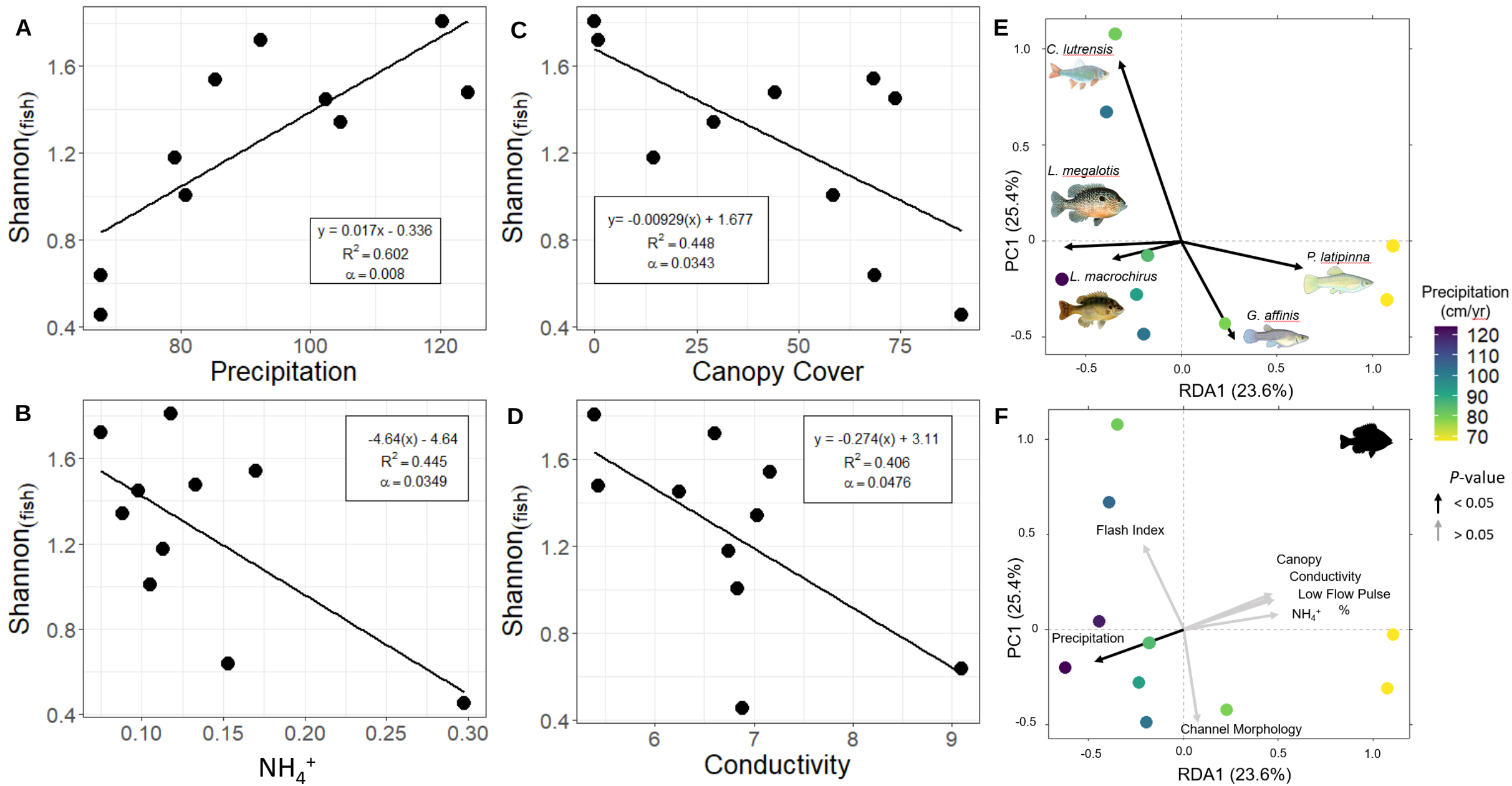
"S1_E_data" ~ supplemental environmental data



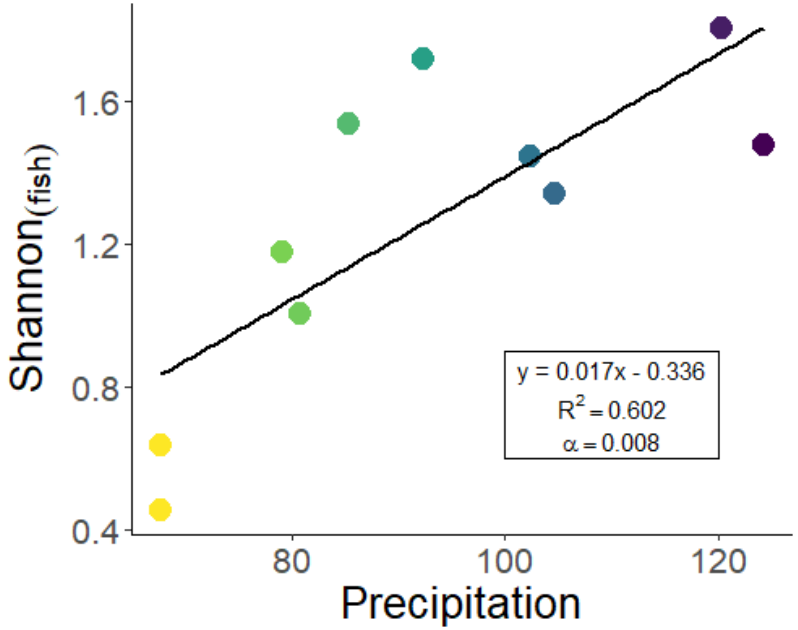
Principal Component Analysis of environmental predictors at 10 sites spanning a precipitation gradient along the Texas Coastal Prairie. Circles representing sample sites are colored based on their annual precipitation. Axes labels include the percentage of variance explained by PC1 (horizontal axis) and PC2 (vertical axis).

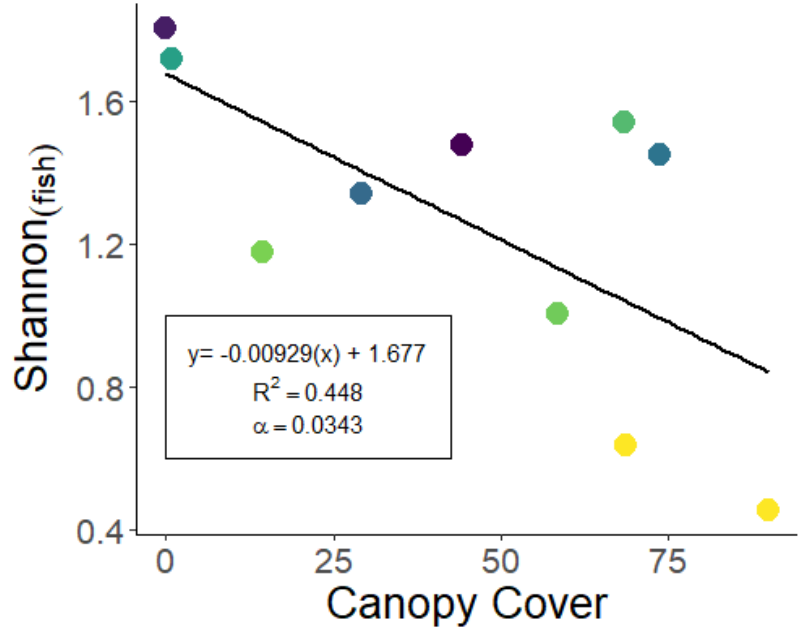
Input	Response	Sign	R^2	Slope	df	F-stat	p-value
Precipitation	Canopy	-	0.233	-0.483	2, 8	2.429	0.158
Precipitation	Flash Index	-	0.001	-0.025	2, 8	0.005	0.945
Precipitation	Low Flow Pulse %	-	0.026	-0.160	2, 8	0.210	0.659
Precipitation	Conductivity	-	0.616	-0.785	2, 8	12.817	0.007 *
Precipitation	NH4	-	0.230	-0.480	2, 8	2.395	0.16
Precipitation	Rosgen Index	+	0.005	0.073	2, 8	0.042	0.842
Canopy	Flash Index	-	0.375	-0.612	2, 8	4.796	0.06
Canopy	Low Flow Pulse %	+	0.176	0.420	2, 8	1.712	0.227
Canopy	Conductivity	+	0.168	0.410	2, 8	1.617	0.239
Canopy	NH_4^+	+	0.414	0.643	2, 8	5.649	0.045 *
Flash Index	Low Flow Pulse %	-	0.096	-0.309	2, 8	0.847	0.384
Flash Index	Conductivity	-	0.030	-0.174	2, 8	0.251	0.63
Flash Index	NH4	-	0.139	-0.373	2, 8	1.291	0.289
Flash Index	Rosgen Index	+	0.000	0.007	2, 8	0.000	0.984
Low Flow Pulse %	Conductivity	+	0.081	0.285	2, 8	0.706	0.425
Low Flow Pulse %	NH_4^+	+	0.354	0.595	2, 8	4.383	0.07
Low Flow Pulse %	Rosgen Index	+	0.047	0.217	2, 8	0.395	0.547
Conductivity	Rosgen Index	-	0.006	-0.074	2, 8	0.044	0.838
NH_4^+	Conductivity	+	0.038	0.194	2, 8	0.315	0.59
NH_4^+	Rosgen Index	+	0.000	0.003	2, 8	0.000	0.994
Rosgen.Index	Canopy	-	0.091	-0.301	2, 8	0.798	0.398

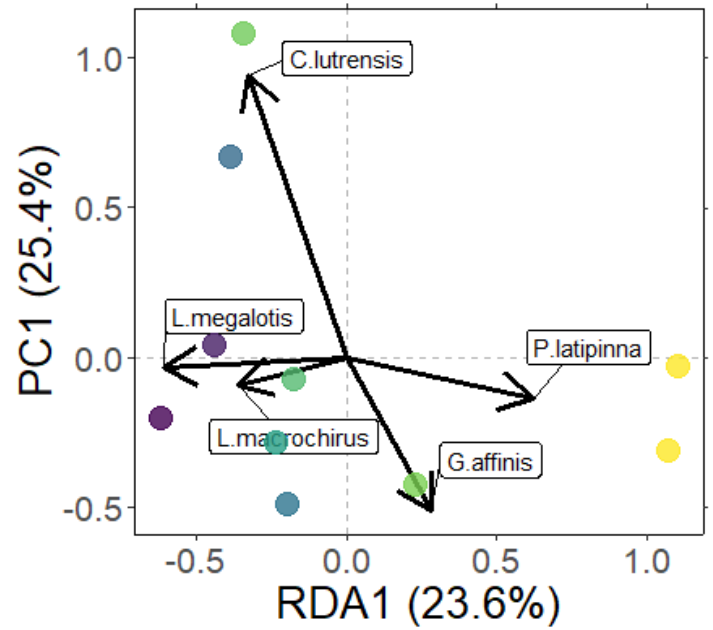
Linear regression statistics to evaluate correlations among environmental predictors. * denotes a statistically significant correlation (p-value < 0.05).

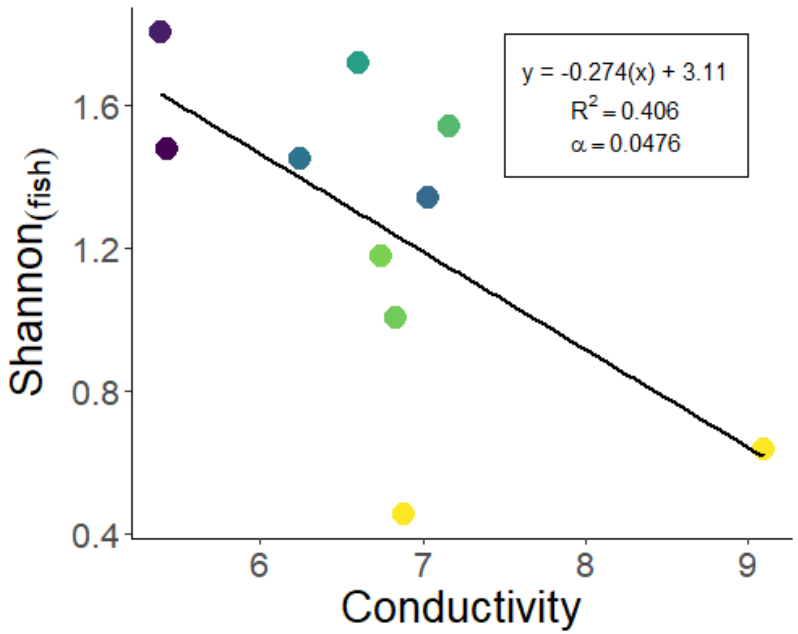


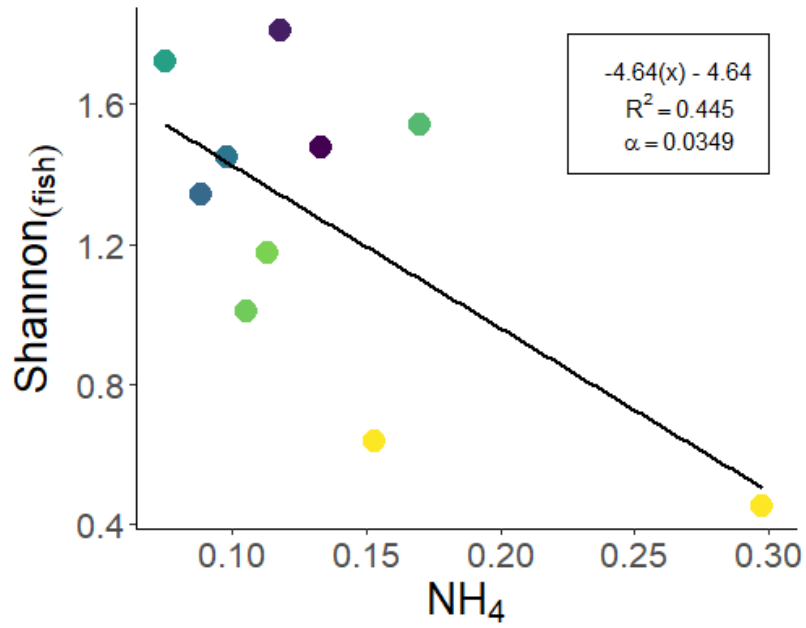
Fish diversity plotted against (A) annual precipitation (cm/yr), (B) NH_4^+ (mg/L), (C) canopy cover (%), (D) log-transformed conductivity (µS/cm). (E & F) Fish community ordinations using Hellinger transformation and redundancy analysis. Axes labels contain proportion of the variance explained as a percentage. Colored circles represent sites with color determined by annual precipitation. Arrows depict fitted vectors for (E) species and (F) environmental predictors which can be visually interpreted based on their direction. Black arrows indicate statistically significant (*p*-value < 0.05) correlations. Only significant species vectors were plotted to improve figure clarity.

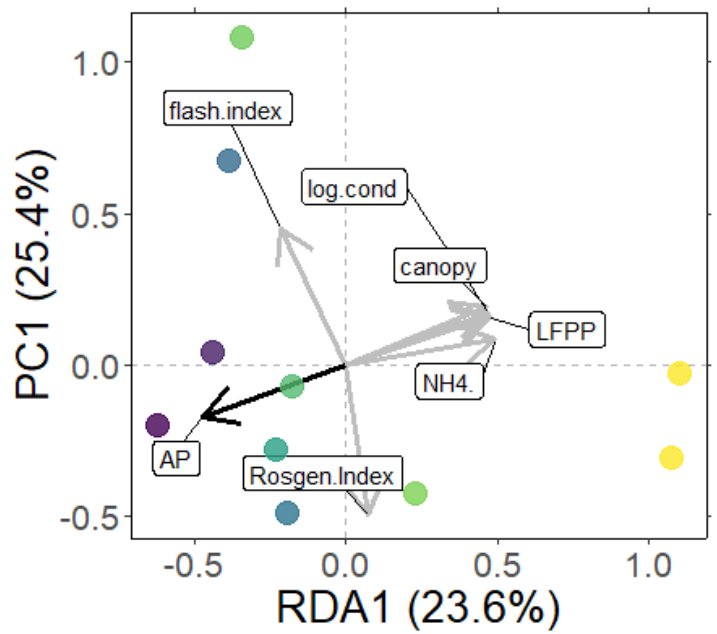

Shannon(fish)
Precipitation
y = 0.017x - 0.336
R2 = 0.602
α = 0.008

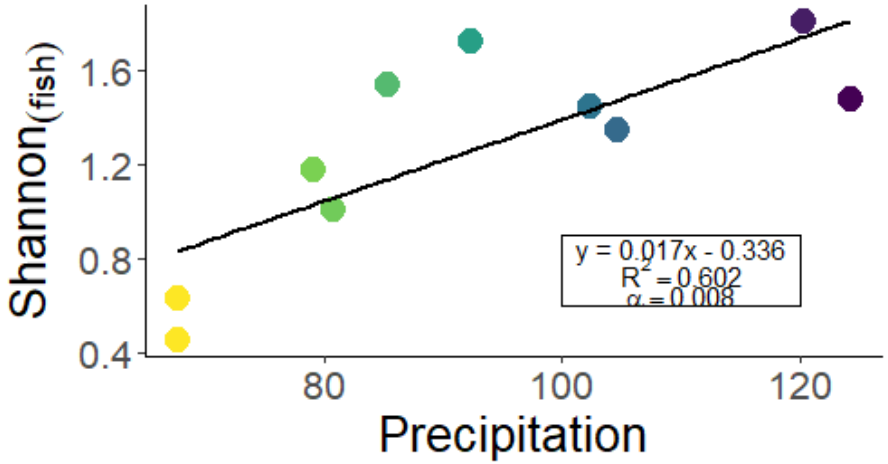

Shannon(fish)
Canopy Cover
y= -0.00929(x) + 1.677
R2 = 0.448
α = 0.0343

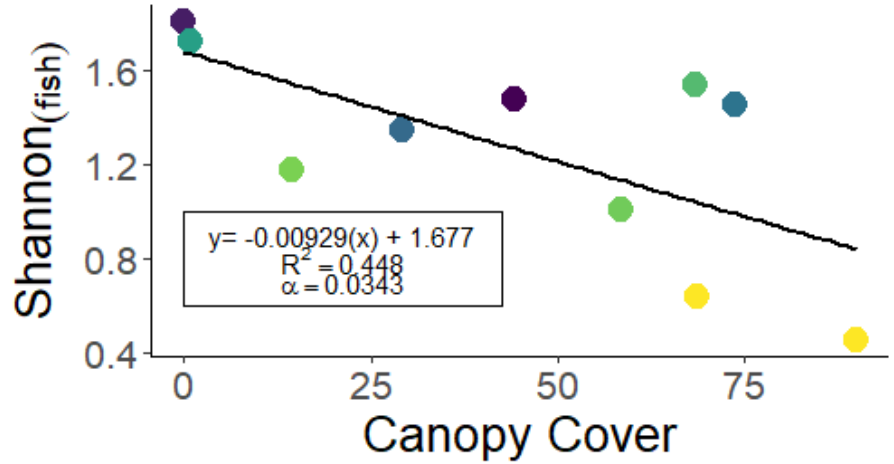

PC1 (25.4%)
RDA1 (23.6%)
C.lutrensis
L.megalotis
P.latipinna
L.macrochirus
G.affinis

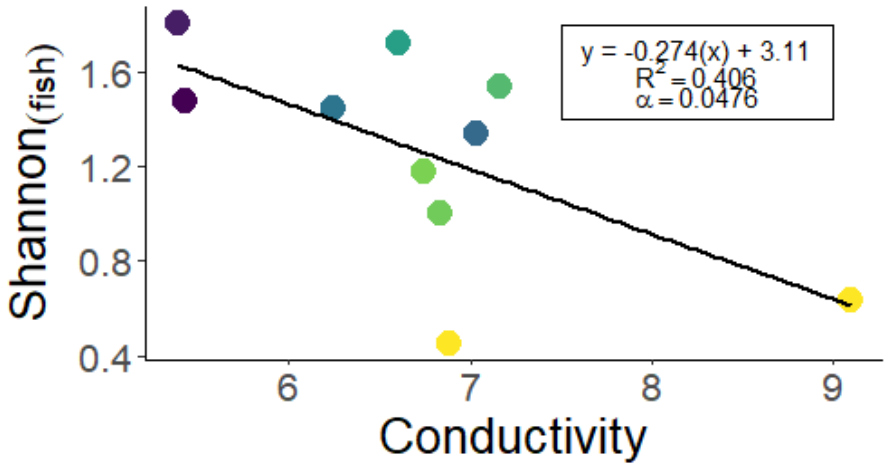

Shannon(fish)
Conductivity
y = -0.274(x) + 3.11
R2 = 0.406
α = 0.0476

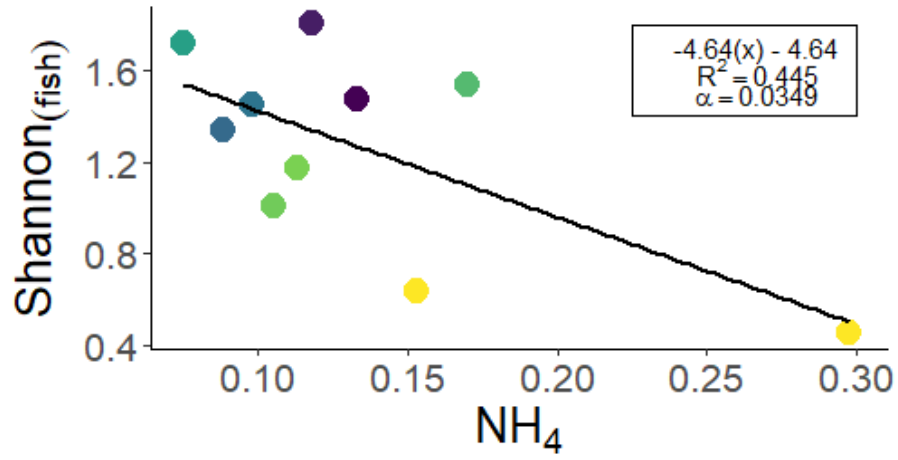

Shannon(fish)
NH4
-4.64(x) - 4.64
R2 = 0.445
α = 0.0349

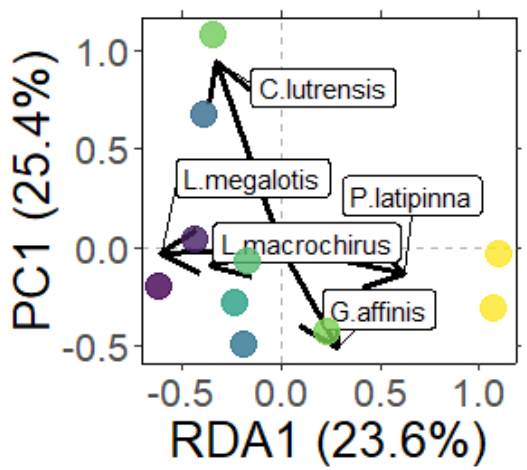

PC1 (25.4%)
RDA1 (23.6%)
flash.index
log.cond
canopy
LFPP
NH4.
AP
Rosgen.Index

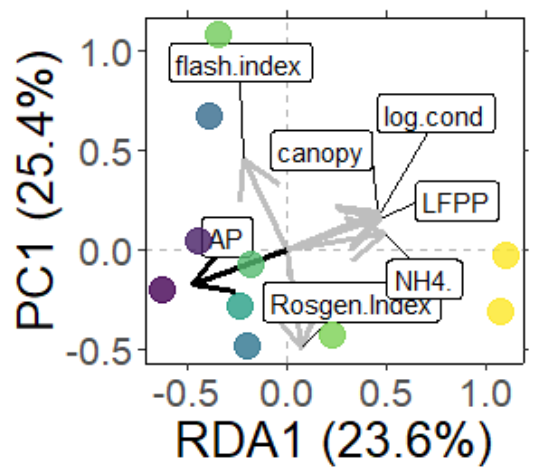

Shannon(fish)
1.6
1.2
0.8
0.4
80
100
120
Precipitation
y = 0.017x - 0.336
R2 = 0.602
α = 0.008


Shannon(fish)
1.6
1.2
0.8
0.4
0
25
50
75
Canopy Cover
y= -0.00929(x) + 1.677
R2 = 0.448
α = 0.0343


Shannon(fish)
1.6
1.2
0.8
0.4
6
7
8
9
Conductivity
y = -0.274(x) + 3.11
R2 = 0.406
α = 0.0476


Shannon(fish)
1.6
1.2
0.8
0.4
0.10
0.15
0.20
0.25
0.30
NH4
-4.64(x) - 4.64
R2 = 0.445
α = 0.0349


PC1 (25.4%)
1.0
0.5
0.0
-0.5
-0.5
0.0
0.5
1.0
RDA1 (23.6%)
C.lutrensis
L.megalotis
P.latipinna
L.macrochirus
G.affinis


PC1 (25.4%)
1.0
0.5
0.0
-0.5
-0.5
0.0
0.5
1.0
RDA1 (23.6%)
flash.index
log.cond
canopy
LFPP
AP
NH4.
Rosgen.Index

Input	Response	Sign	R2	Slope	df	F-stat	p-value
AP	$Shannon_{Fish}$	-	0.602	0.017	2, 8	12.101	0.008 *
log.cond	$Shannon_{Fish}$	-	0.406	-0.274	2, 8	5.464	0.048 *
Rosgen.Index	$Shannon_{Fish}$	-	0.027	0.021	2, 8	0.224	0.649
canopy	$Shannon_{Fish}$	-	0.448	-0.009	2, 8	6.490	0.034 *
NH4.	$Shannon_{Fish}$	-	0.446	-4.647	2, 8	6.431	0.035 *
flash.index	$Shannon_{Fish}$	-	0.024	0.339	2, 8	0.193	0.672
LFPP	$Shannon_{Fish}$	-	0.339	-0.031	2, 8	4.102	0.077

Summary statistics for linear regressions of fish Shannon Index values versus environmental predictors. * denotes a statistically significant correlation (*p*-value < 0.05).

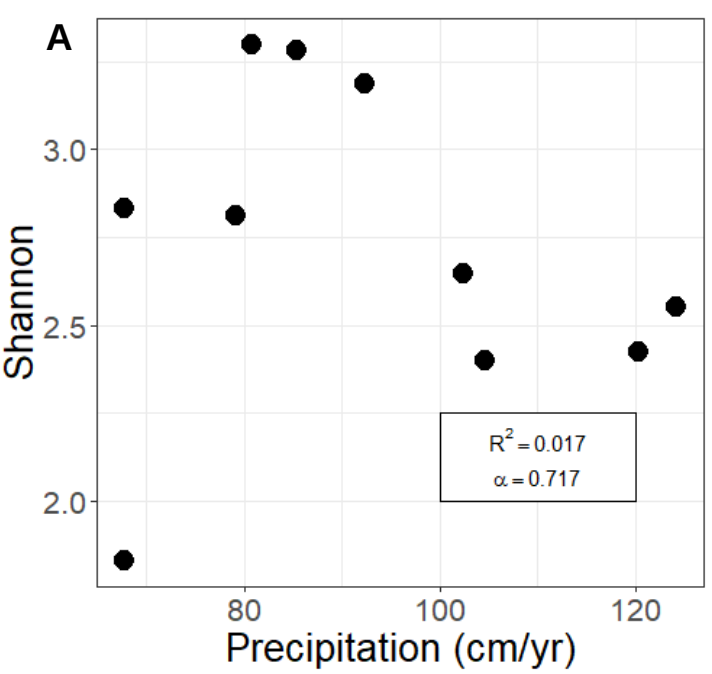
model	(Int)	AP	cnp	fls.ind	LFPP	log.cnd	NH4.	Rsg.Ind	df	logLik	AICc	delta	weight
10	1.261	0.3127			-0.2098				4	2.912	10.2	0	0.429
2	1.261	0.3462							3	-0.985	12	1.79	0.175
4	1.261	0.2634	-0.1715						4	0.69	14.6	4.44	0.046
34	1.261	0.2641					-0.1711		4	0.688	14.6	4.45	0.046
49	1.261					-0.2352	-0.2521		4	0.649	14.7	4.52	0.045
3	1.261		-0.2986						3	-2.622	15.2	5.07	0.034
33	1.261						-0.2979		3	-2.642	15.3	5.11	0.033
17	1.261					-0.2842			3	-2.989	16	5.8	0.024
74	1.261	0.3014			-0.2338			0.1025	5	4.501	16	5.82	0.023
1	1.261								2	-5.592	16.9	6.72	0.015

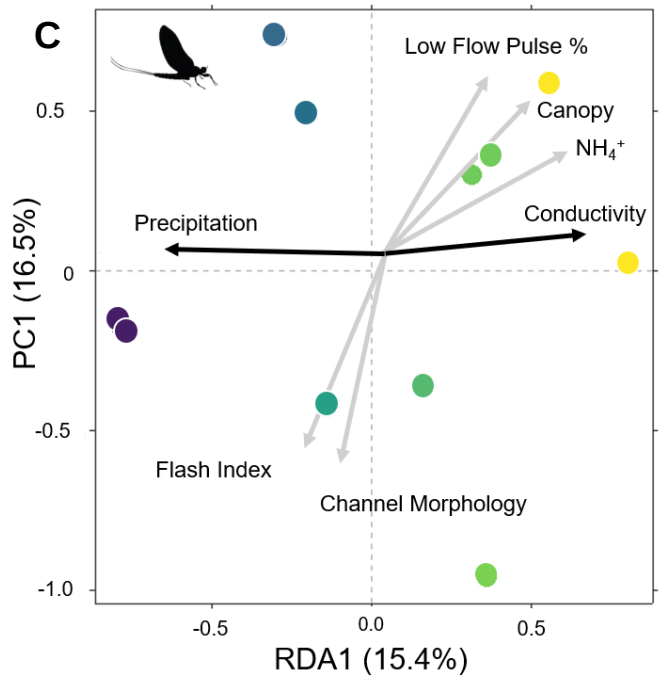
Top ten Multivariate regression models predicting fish Shannon diversity.

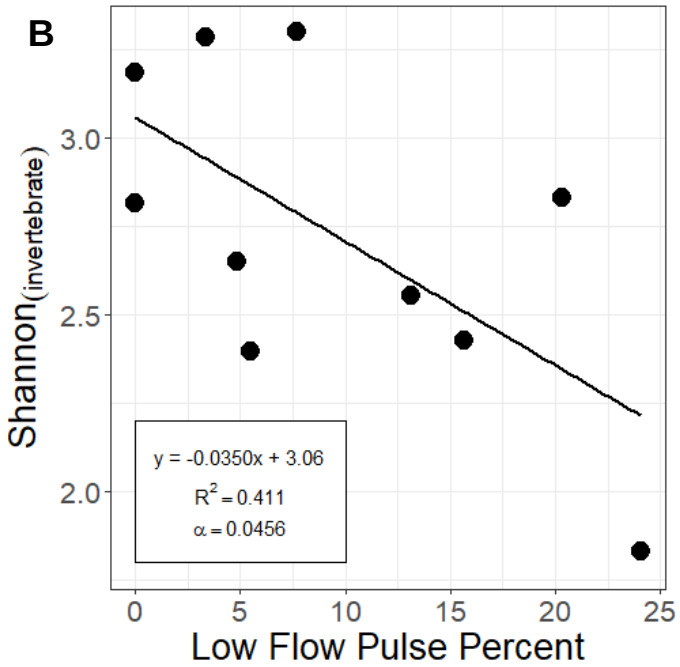
model_name	var1	estimate	var2	estimate2	var3	estimate3	adjusted_r2	f.stat	df	p.value
Fm10	AP	0.312661569	LFPP	-0.209790458	na	na	0.765305514	15.67386332	2 7	0.00259869
Fm4	AP	0.263436925	canopy	-0.171468311	na	na	0.633976268	8.794284779	2 7	0.012310529
Fm34	AP	0.264055079	NH4.	-0.171113979	na	na	0.633850528	8.790062783	2 7	0.012325337
Fm49	log.cond	-0.235193285	NH4.	-0.252107645	na	na	0.631013908	8.695581603	2 7	0.012662788
Fm74	AP	0.30138	LFPP	-0.23384	Rosgen.Index	0.10252	0.8007	13.05	3 6	0.004866

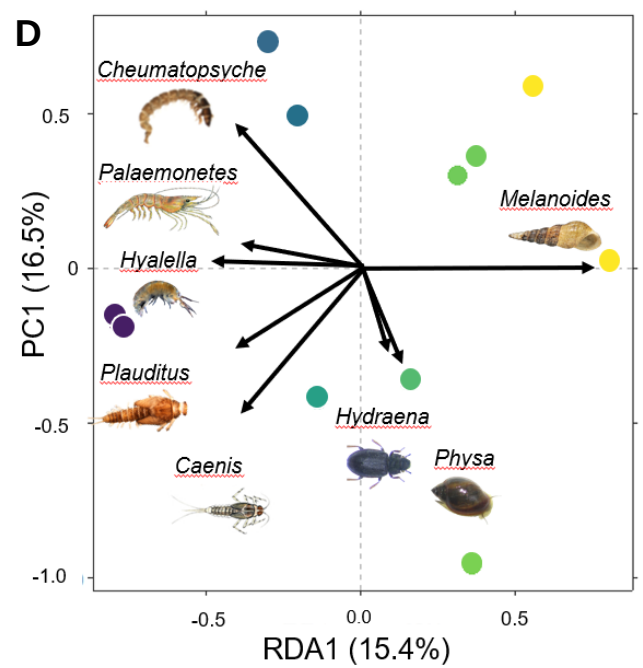
Summary statistics for top five multivariate regressions predicant fish Shannon diversity

Precipitation is a predictor in the top multivariate models and has the largest estimated slope compared to all other scaled environmental variables in the exhaustive multivariate regression.
One multivariate regression utilizes 2 water quality parameters (conductivity and NH4).
Positive predictors of fish diversity include AP and Rosgen Index
Negative predictors of fish diversity include LFPP, canopy, NH4., log.cond









Macroinvertebrate diversity plotted against (A) annual precipitation (cm/yr) and (B) Low Flow Pulse Percent. (C & D) Fish community ordinations using Hellinger transformation and redundancy analysis. Axes labels contain proportion of the variance explained as a percentage. Dots represent sites with color determined by annual precipitation. Arrows represent fitted vectors for (E) environmental predictors and (F) species which can be visually interpreted based on their direction. Black arrows indicate statistically significant (p-value < 0.05) correlations. Only significant species vectors were plotted to improve figure clarity.

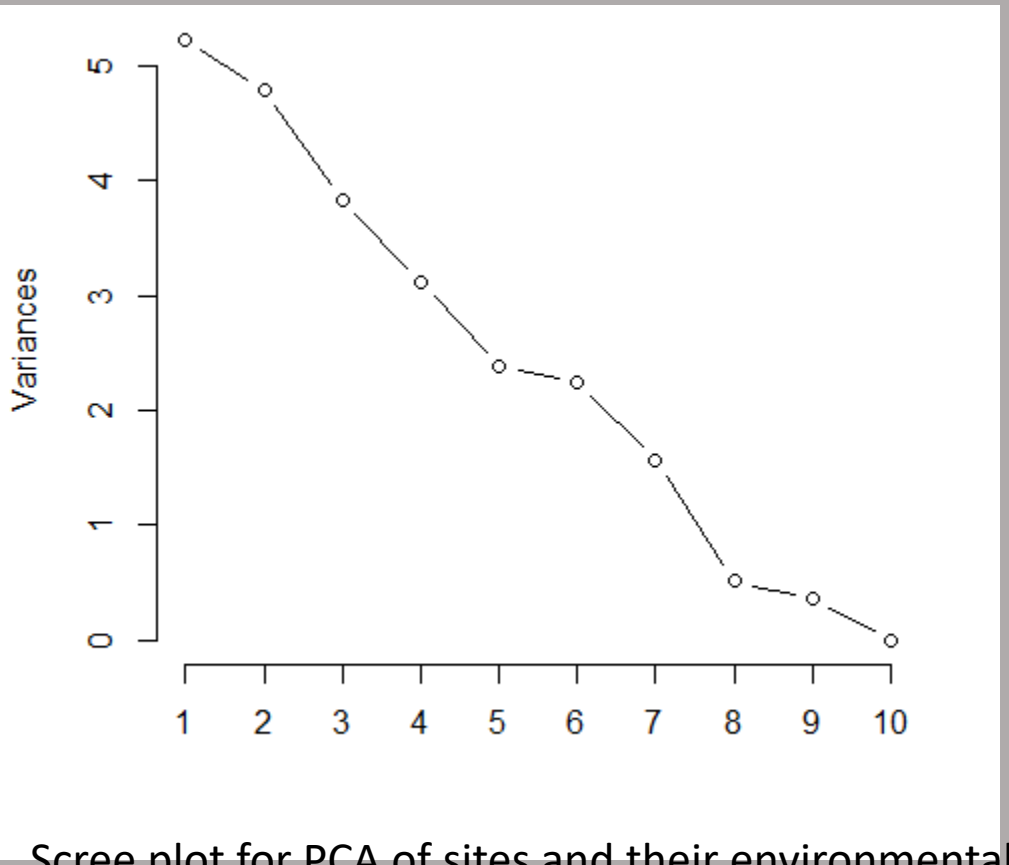
Site.Name	STAID	AP	RH	Bas.dev	Bas.forest	Bas.plant	Soil.Perm	Soil.Org	Rip.forest	log.cond	DO	turbidity	Twater	pH	bank.height	Rosgen.Index	canopy	d50	NH4.	PO4.	log.no3	flash.index	HFPP	LFPP	Av.Flow
Bear Branch	8068390	124.19	71.4	76.71	9.98	0.82	3.75	0.48	14.33	5.425501	8.173	67.455	17	6.763	0.893	12.011	44.144	2	0.133	0.168	-3.91202	0.777	0.25	13.146	25.875
Big Creek	8115000	120.31	71.6	10.88	2.57	77.35	0.65	2.55	9.06	5.390213	7.881	99.084	15.875	6.963	0.361	23.15	0	2	0.118	1.668	1.064711	0.961	0.24	15.631	43.538
Garcitas	8164600	102.41	74.2	4.28	17.53	50.34	1.75	0.75	19.86	6.249493	5.587	66.909	22.575	7.135	0.339	18.161	73.649	2	0.098	0.408	-3.03655	0.806	0.285	4.867	43.985
Placedo	8164800	104.65	75	11.81	2.57	78.16	0.5	0.9	5.28	7.034828	8.596	70.66	24.625	7.313	0.438	13.38	29.279	25	0.088	0.298	-0.82098	0.921	0.256	5.482	44.717
Perdido	8177300	92.37	72.1	2.45	34.85	27.41	1.17	0.94	32.9	6.602249	7.203	66.199	18.875	7.183	0.288	15.152	0.901	2	0.075	0.298	-3.77226	1.339	0.286	0	5.413
Medio	8189300	79.13	72	5.24	1.44	40.31	1.16	0.95	2.27	6.74788	4.206	68.798	19.025	6.895	0.691	18.405	14.414	2	0.113	0.405	-4.82831	0.993	0.43	0	3.119
Mission	8189500	85.36	73.4	3.72	8.43	35.06	1.1	0.82	11.05	7.163172	5.277	68.613	23.6	7.215	0.46	14.701	68.468	2	0.17	0.315	-3.21888	0.58	0.208	3.357	115.13
Aransas	8189700	80.77	74.1	8.28	3.58	52.39	0.96	0.8	7.41	6.834647	7.649	66.351	19.05	7.078	0.54	11.781	58.559	32	0.105	4.128	-0.44942	1.053	0.112	7.664	35.038
San Fernando	8211900	67.75	69.6	4.74	1.48	28.01	1.36	0.91	3.92	6.886532	5.829	67.172	23.275	6.428	0.591	15.762	89.865	2	0.298	4.275	0.587787	0.907	0.173	20.306	13.834
Tranquitas	8212300	67.75	69.6	4.74	1.48	28.01	1.36	0.91	3.92	9.096387	9.971	66.132	22.525	7.988	0.75	17.988	68.769	2	0.153	0.125	-2.65926	0.781	0.313	24.06	1.763

PET and runoff.factor were removed because they were derived from precipitation data
Similarly, I think I should reduce the variables included in my analysis:
Relative humidity is redundant with precipitation
Soil.Org is redundant with Soil permeability
Rip.forest is redundant with Bas.forest
D50 should be reported for orientation, but is not a useful predictor due to no variation
Bas.dev isn't an informative variable. It has one outlier and little variation otherwise

In my revised analysis, the number of environmental predictors is reduced from 25 to 17.

1. I will use PCA to ascertain which variables characterize variation among the sample locations
2. I will run univariate regressions using precipitation and variables identified in PCA
3. I will run multivariate nonlinear regressions using variables identified in PCA
4. I will run PCoA ordinations on community abundance data

Goal: Conduct PCA on environmental variables to discern patterns of variation among the sites

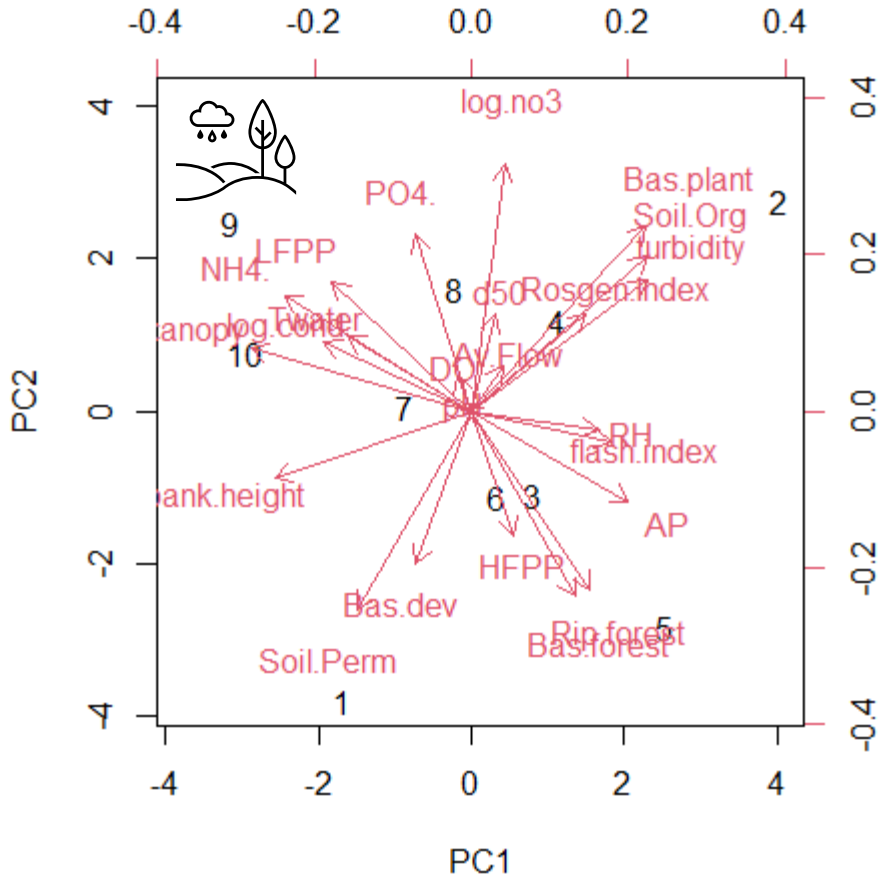


Scree plot for PCA of sites and their environmental variables

Principle Component Axis	Standard deviation	Proportion of Variance	Cumulative Proportion
PC1	2.4957	0.2595	0.2595
PC2	2.2056	0.2027	0.4622
PC3	1.9068	0.1515	0.6137
PC4	1.7469	0.1272	0.7409
PC5	1.5634	0.1018	0.8427
PC6	1.32689	0.07336	0.91607
PC7	1.15941	0.05601	0.97208
PC8	0.65057	0.01764	0.98971
PC9	0.49691	0.01029	1
PC10	2.80E-15	0.00E+00	1.00E+00

Summary of PCA

- Unlike most PCAs, this scree plot does not have an exponential decay. PC1 + PC2 only accounts for 41.69% of the variation among sites.



PCA of 10 sites with selected environmental variables

variable	PC1	PC2
canopy	-0.80414	0.220828
bank.height	-0.71586	-0.2356
NH4.	-0.68092	0.404682
log.cond	-0.53953	0.239143
LFPP	-0.50937	0.454223
Twater	-0.45219	0.261157
Soil.Perm	-0.41537	-0.69273
PO4.	-0.20534	0.615143
Bas.dev	-0.2015	-0.53301
DO	-0.05358	0.122313
pH	-0.02303	0.01263
d50	0.084812	0.336731
Av.Flow	0.113341	0.160583
log.no3	0.122872	0.862227
HFPP	0.152055	-0.43224
Bas.forest	0.376102	-0.64802
Rosgen.Index	0.418434	0.340295
Rip.forest	0.43299	-0.62211
RH	0.463994	-0.06327
flash.index	0.508531	-0.10762
AP	0.572453	-0.31467
Bas.plant	0.635744	0.643684
turbidity	0.639434	0.456038
Soil.Org	0.640082	0.543809

PCA correlations ordered by PCA1

variable	PC1	PC2
Soil.Perm	-0.41537	-0.69273
Bas.forest	0.376102	-0.64802
Rip.forest	0.43299	-0.62211
Bas.dev	-0.2015	-0.53301
HFPP	0.152055	-0.43224
AP	0.572453	-0.31467
bank.height	-0.71586	-0.2356
flash.index	0.508531	-0.10762
RH	0.463994	-0.06327
pH	-0.02303	0.01263
DO	-0.05358	0.122313
Av.Flow	0.113341	0.160583
canopy	-0.80414	0.220828
log.cond	-0.53953	0.239143
Twater	-0.45219	0.261157
d50	0.084812	0.336731
Rosgen.Index	0.418434	0.340295
NH4.	-0.68092	0.404682
LFPP	-0.50937	0.454223
turbidity	0.639434	0.456038
Soil.Org	0.640082	0.543809
PO4.	-0.20534	0.615143
Bas.plant	0.635744	0.643684
log.no3	0.122872	0.862227

PCA correlations ordered by PCA2

- PCA1 has correlations greater than 0.6 with canopy, bank.height, NH4
- PCA2 has correlations greater than 0.6 with Soil.Perm, Bas.forest, & Rip.forest
- With only ten sample sites, 6 environmental variables to be assessed using linear regression include the top 3 covariates with PC1 and PC2: “canopy", “bank,height", “NH4", “Soil.Org", “Bas.forest", “Rip.forest”

model	(Int)	AP	bnk.hgh	Bas.frs	cnp	NH4.	Rip.frs	Sol.Org df	logLik	AICc	delta	weight
2	-0.336	0.01727						3	-0.985	12	0	0.241
4	0.2589	0.01605	-0.9008					4	1.461	13.1	1.11	0.138
6	-0.3085	0.01554		0.0158				4	1.108	13.8	1.81	0.097
34	-0.2682	0.01438					0.01817	4	0.984	14	2.06	0.086
10	0.2849	0.01314			-0.00534			4	0.69	14.6	2.65	0.064
18	0.4036	0.01317				-2.67		4	0.688	14.6	2.65	0.064
9	1.678				-0.0093			3	-2.622	15.2	3.27	0.047
17	1.889					-4.647		3	-2.642	15.3	3.31	0.046
33	0.9466						0.02858	3	-3.34	16.7	4.71	0.023
1	1.261							2	-5.592	16.9	4.93	0.02

Top ten dredge outputs for fish Shannon versus 7 PCA-selected environmental variables

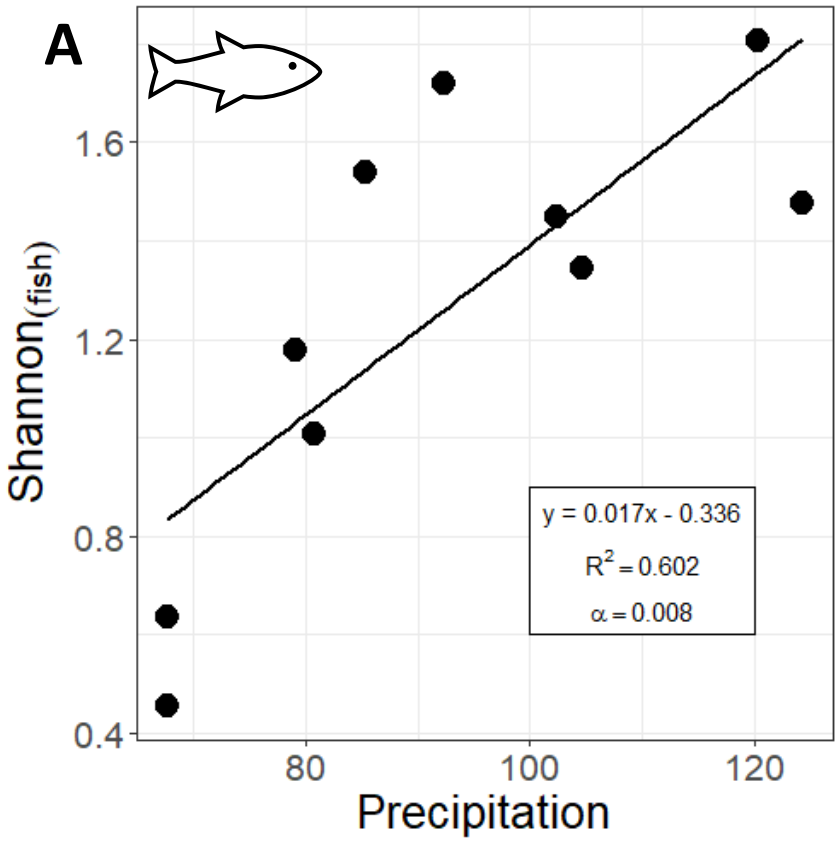
model_name	var1	estimate	var2	estimate2	adjusted_r2	f.stat	df	p.value
fm3	AP	0.321722	bank.height	-0.17679	0.6863	10.84493	3	0.007175
fm6	AP	0.311504	Bas.forest	0.168235	0.663337	9.866483	3	0.009187
fm34	AP	0.288174	Rip.forest	0.170771	0.654924	9.540603	3	0.010016
fm10	AP	0.263437	canopy	-0.17147	0.633976	8.794285	3	0.012311
fm18	AP	0.264055	NH4.	-0.17111	0.633851	8.790063	3	0.012325

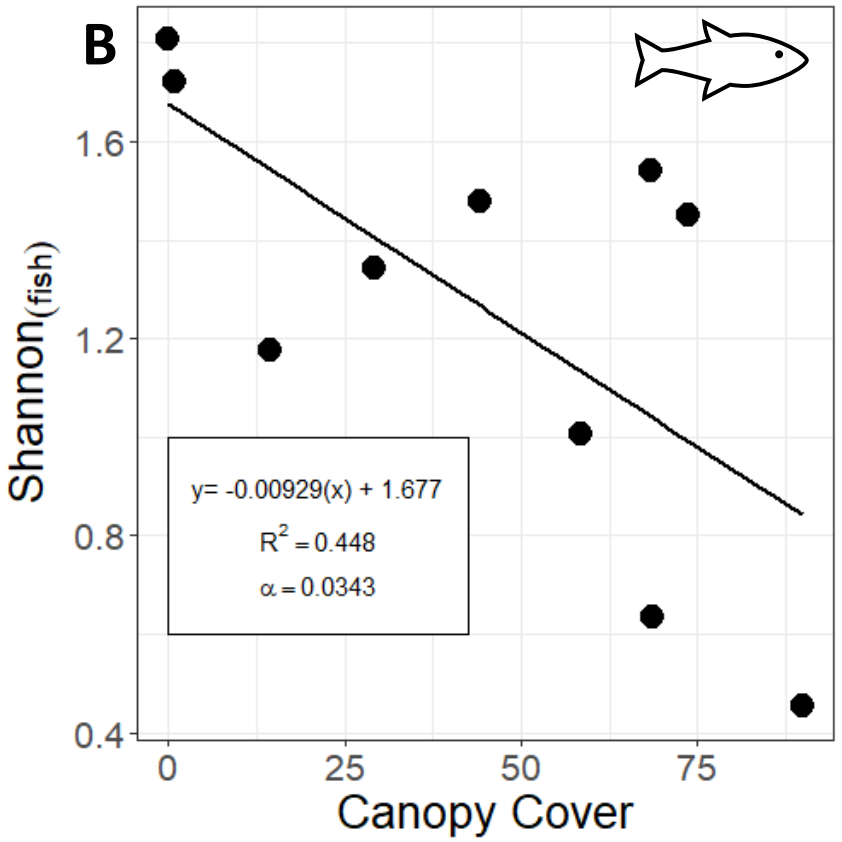
Summary statistics for top five AICc ranked multivariate regression models for predicting fish Shannon index using 7 PCA-selected environmental variables

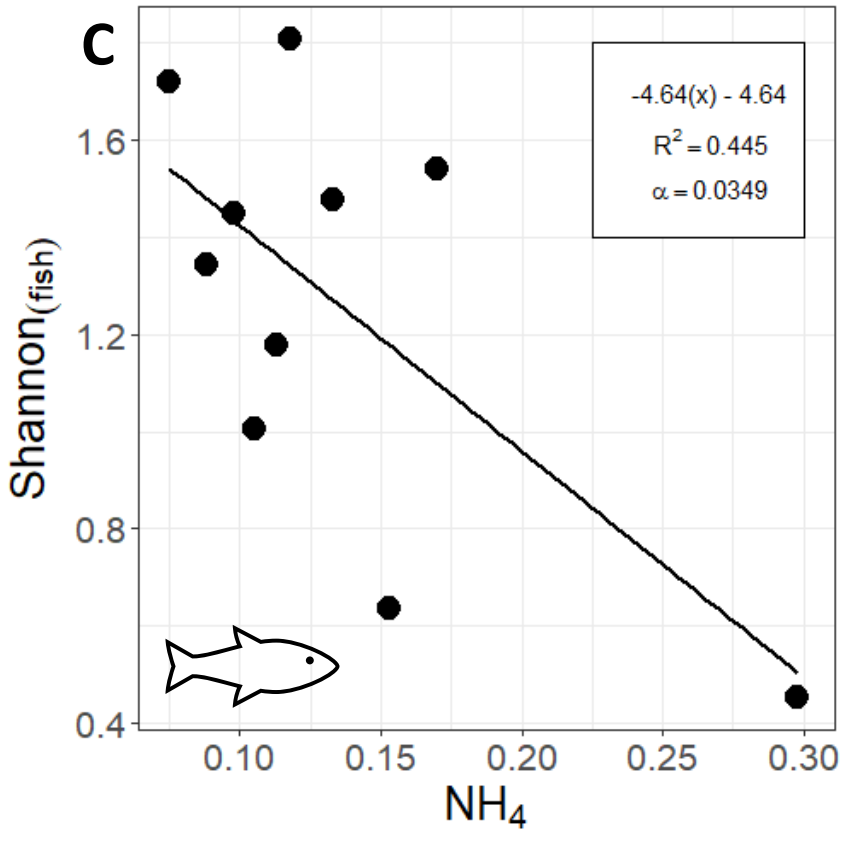
Precipitation is a predictor in the top 5 multivariate models and has the largest estimated slope compared to all other scaled environmental variables in the exhaustive multivariate regression.

Positive predictors of fish diversity include Bas.forest and Rip.forest

Negative predictors of fish diversity include bank height, canopy, and NH4+





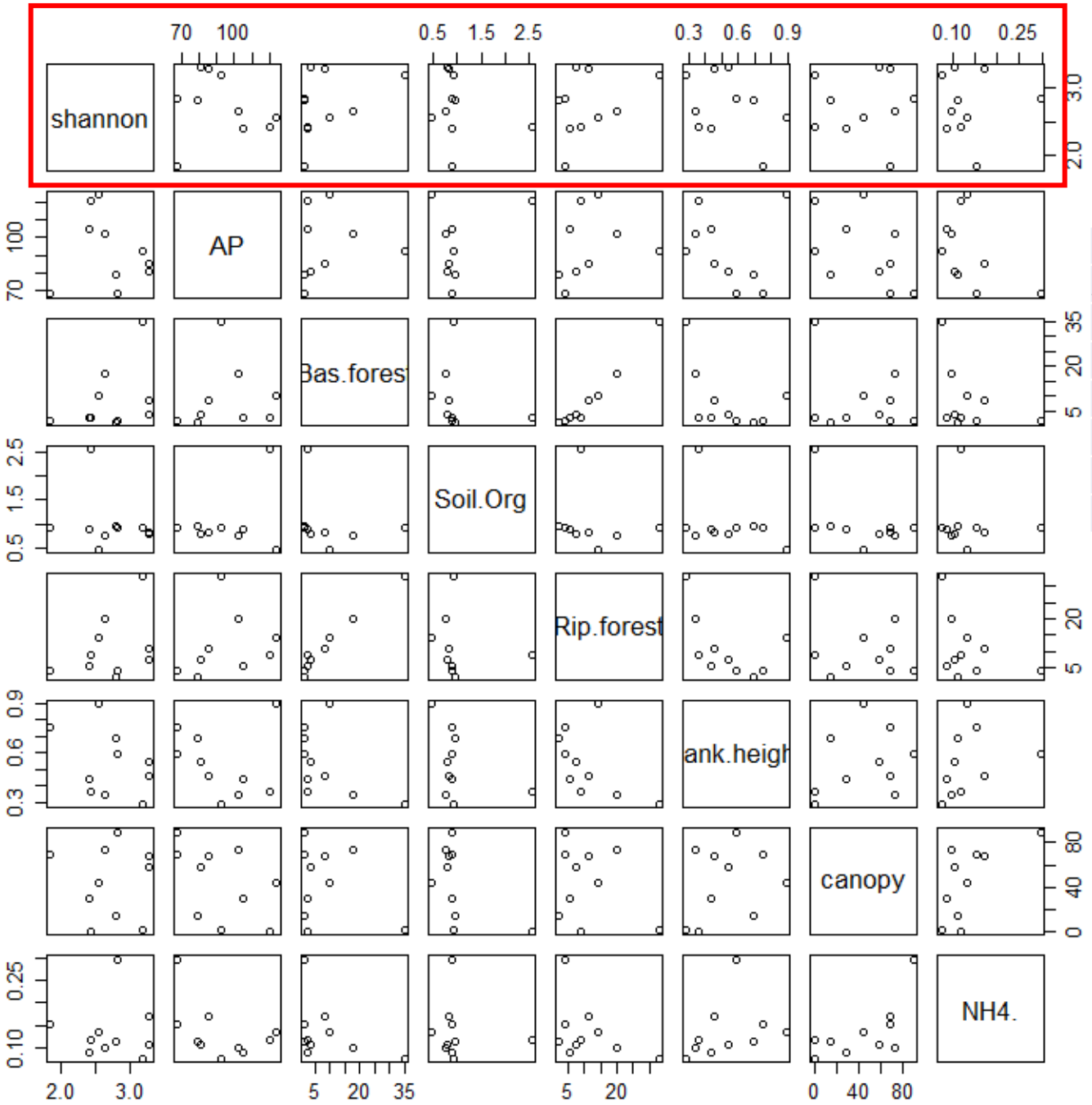


Above: Shannon diversity of fish communities plotted against (A) annual precipitation (cm/yr) (B) canopy coverage (%) and (C) NH4+ (mg/L).

3 statistically significant regressions :
+ precipitation
- canopy
- NH4

predictor	estimate	df	r2	f.stat	p.value
AP	0.017271	2 8	0.602017	12.10136	0.008336
Bas.forest	0.021825	2 8	0.271423	2.980307	0.122558
Soil.Org	0.276431	2 8	0.121132	1.102616	0.324376
Rip.forest	0.028585	2 8	0.362586	4.550714	0.065462
bank.height	-1.12768	2 8	0.246026	2.610439	0.144825
canopy	-0.0093	2 8	0.447887	6.489777	0.034308
NH4.	-4.64708	2 8	0.445651	6.431349	0.034927

Linear regression statistics for fish Shannon diversity against PCA-selected variables



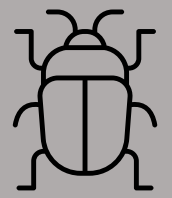
predictor	estimate	df	r2	f.stat	p.value
AP	-0.00302	2 8	0.01736	0.141333	0.716733
Bas.forest	0.016587	2 8	0.147817	1.38766	0.272652
Soil.Org	-0.17543	2 8	0.046001	0.385754	0.551821
Rip.forest	0.01767	2 8	0.130638	1.202154	0.304796
bank.height	-0.8301	2 8	0.125697	1.150145	0.314801
canopy	-0.00022	2 8	0.000243	0.001946	0.965892
NH4.	-0.06524	2 8	8.28E-05	0.000662	0.980096

Above: Regression statistics for invertebrate Shannon index versus PCA-selected environmental predictors

Left: Pairs plots for invertebrate diversity and environmental variables

No significant correlations with invertebrate diversity

model	(Int)	AP	bnk.hgh	Bas.frs	cnp	NH4.	Rip.frs	Sol.Org	df	logLik	AICc	delta	weight
1	2.727								2	-5.886	17.5	0	0.397
5	2.727			0.1767					3	-5.086	20.2	2.69	0.104
33	2.727						0.1661		3	-5.186	20.4	2.89	0.094
3	2.727		-0.1629						3	-5.215	20.4	2.94	0.091
65	2.727							-0.09855	3	-5.651	21.3	3.81	0.059
2	2.727	-0.06054							3	-5.799	21.6	4.11	0.051
9	2.727				-0.00717				3	-5.885	21.8	4.28	0.047
17	2.727					-0.00418			3	-5.886	21.8	4.28	0.047
67	2.727		-0.2427					-0.1969	4	-4.25	24.5	7.01	0.012



Top ten AICc ranked multivariate regressions for invertebrate Shannon index using 7 PCA-selected variables

model_name	var1	estimate	var2	estimate2	adjusted_r2	f.stat	df	p.value
im67	bank.height	-0.24268	Soil.Org	-0.19688	0.073198	1.355403	3	0.318016

10th AICc-ranked multivariate regression predicting invertebrate Shannon index

9/10 top AICc models contained one or fewer environmental predictors. Previous univariate regressions indicate no statistically significant correlations

Invertebrate model 67 indicates bank height and soil organic content as negative predictors of invertebrate shannon index. However, the p value for the model is 0.318 and the correlation coefficient is 0.073. So this multivarite regression is neither accurate nor statistically significant.

Regression analysis fails to provide useful environmental predictors for stream macroinvertebrate diversity in this survey.