Acoustic indices as proxies for biodiversity: a meta-analysis

Supplementary material

A neat presentation of the supplementary material can be found in https://irene-alcocer.github.io/Acoustic-Indices/

Data descriptions

Supplementary Table 1: Dataset used in the study. Due to the size of this table it is only available at https://irene-alcocer.github.io/Acoustic-Indices/

Supplementary Table 2: Variable descriptions for Supplementary Table 2. Available at https://irene-alcocer.github.io/Acoustic-Indices/

Supplementary Table 3: List of the 35 features used to characterize studies that tested the relationship between acoustic indices and diversity metrics.

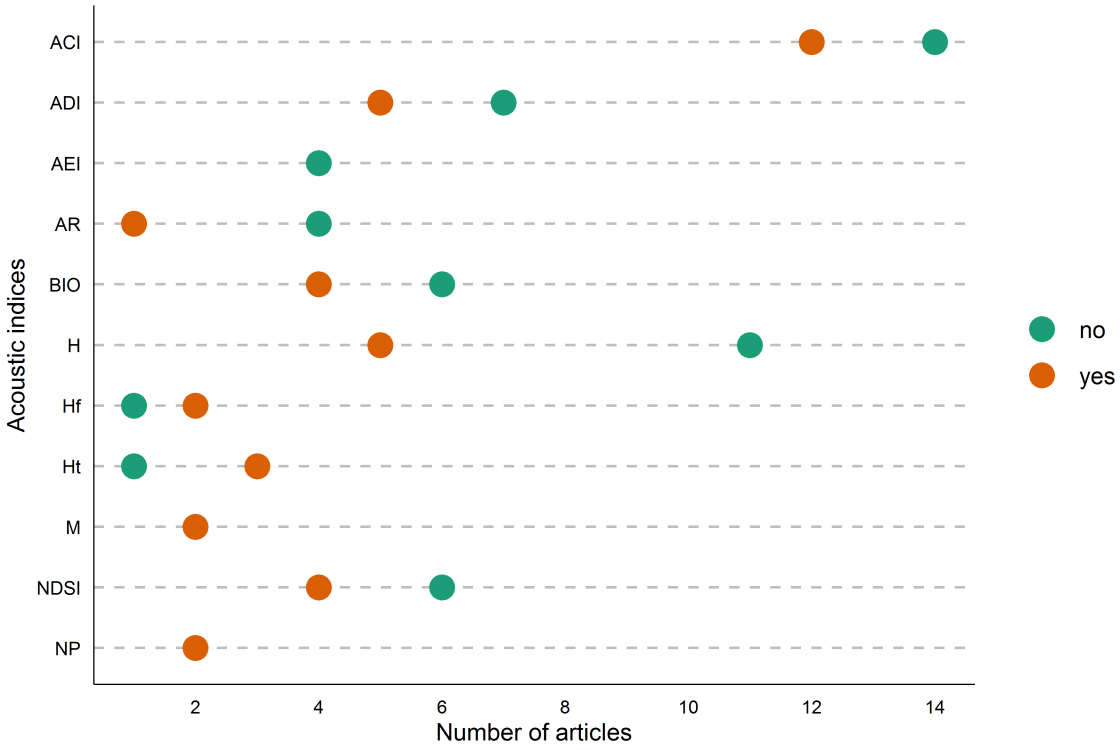
Category	Features	Description
Publication	Authors	
	Title	
	Journal	
	Year of publication	
	Peer reviewed	Whether the study was subjected to peer review (Yes or No)
Biological data	Environment	Ecosystem type where recordings were collected (aquatic or terrestrial)
	Taxonomic group	Primary studied group (invertebrates, fish, anurans, mammals, birds, or several)
	Diversity metric	Species abundance, species richness, species diversity, abundance of sounds, or diversity of sounds
	Diversity source	Method applied to obtain the diversity metric (acoustic or non-acoustic)
Acoustic data	Acoustic index	ACI, AEI, ADI, AR, BIO, H, Ht, Hf, M, NP, or NSDI
	Frequency range	Range (in Hz) used for index calculation
	FFT size	Window size of the Fast Fourier Transformation (FFT)
	Noise treatment	Audio pre-processing related to noise (noise filtering, noise addition, or exclusion of noisy recordings)
Recording	Sampling rate	Number of audio samples per second (in kHz) used for index calculation
	Audio format	Format of audio files (.wav, .mp3, etc.)
	Recording length	Length of each recording (in seconds) used for index calculation
	Recording method	Non-programmed (continuous), programmed (periodic) or manual (by an operator)
Sampling design	Study sites	Number of study sites (= spatial replicates)
	Distance between sites	Minimum distance between study sites (in meters)
	Recorders per site	Number of recording units per study site
	Recording days	Number of recording days per study site (= temporal replicates)
	Daily period	Period recorded within the day (dawn, morning, midday, evening, dusk, night, or all day)
	Daily sample	Number of recordings collected within a day per study site
Statistics	Statistical test	Statistical analysis used to test the relationship between acoustic indices and diversity metrics
	Independence	Whether the statistical test was considered independent from other tests of the same study
	R^2	Coefficient of determination (for regression analysis)
	r	Correlation coefficient (for Pearson or Spearman correlation)
	b	Regression coefficient (for linear regression analysis)
	t-statistic	Statistic value for Student's t-test
	Standard error	Standard error of the test coefficient
	Sample size	Number of observations included in the statistical test
	Pseudoreplication	Inadequate specification of the number of true replicates in the statistical test (Yes or No)
	Pseudoreplication type	Spatial, temporal, or spatial-temporal pseudoreplication
	Adjusted sample size	Suitable specification of the number of true replicates (for pseudoreplicated studies)

Supplementary Table 4: Number of effect sizes collected from each of the 34 studies included in the meta-analysis. ID corresponds to the study identification number in our dataset.

ID	Study	Effect_sizes
2740	Mammides et al. 2017	84
53	Moreno-Gomez 2019	42
87	Eldridge et al. 2018	28
80	Staaterman et al. 2017	24
90	Ferreira et al. 2018	24
96	Izaguirre et al. 2018	24
10	Buscaino et al. 2016	22
2	Desjonquères et al. 2015	12
11	Bertucci et al. 2016	12
89	Gage et al. 2017	12
2977	Jorge et al. 2018	12
70	Bolgan et al. 2018	11
9	Harris et al. 2016	6
86	Indraswari et al. 2018	6
427	Fuller et al. 2015	6
14	Wa Maina et al. 2016	4
60	Patrick Lyon et al. 2019	4
77	Fairbrass et al. 2017	4
92	Torti et al. 2018	4
15	Roca & Proulx 2016	3
45	McLaren 2012	3
13	McWilliam & Hawkin 2013	2
41	Paisley-Jones 2011	2
44	Machado et al. 2017	2
251	Buxton et al. 2016	2
4	Parks et al. 2014	1
6	Boelman et al. 2007	1
17	Zhang et al. 2015	1
37	Picciulin et al. 2016	1
1132	Depraetere et al. 2012	1
1177	Joo et al. 2011	1
1262	Pieretti et al. 2011	1
2745	Sueur et al. 2008	1
2986	Raynor et al. 2017	1

Supplementary Table 5: Number of effect sizes and studies per moderator levels.

Moderator levels	Effect sizes	Studies
Acoustic indices		
ACI	113	25
ADI	38	12
AEI	34	8
AR	18	5
BIO	36	10
H	55	16
Hf	12	3
Ht	15	4
M	5	2
NDSI	33	10
NP	5	2
Diversity metrics		
Species abundance	27	6
Species diversity	49	9
Species richness	187	21
Abundance of sounds	66	11
Diversity of sounds	35	3
Diversity source		
Acoustic	200	26
Non-acoustic	164	11
Environment		
Aquatic	95	10
Terrestrial	269	24



Supplementary Figure 1: Pseudoreplication summary. The data is representing the total number of articles for each index. Color orange represents number of pseudo-replicated studies and green non pseudo-replicated studies. The article (Papin et al., 2019b) was withdrawal from the pseudoreplication analysis due to impossibility of obtaining the pseudoreplication data causing some variation on the total number of each acoustic index.

Overall Effect Size

Model output

```
res_main <- rma.mv(z, var, random = ~1 | id/entry, data = df_tidy)
res_main
```

```
##
## Multivariate Meta-Analysis Model (k = 364; method: REML)
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed    factor
## sigma^2.1  0.0458  0.2139     34     no        id
## sigma^2.2  0.1755  0.4190    364     no  id/entry
##
## Test for Heterogeneity:
## Q(df = 363) = 2220.9097, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.3461  0.0577  6.0014  <.0001  0.2331  0.4591  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

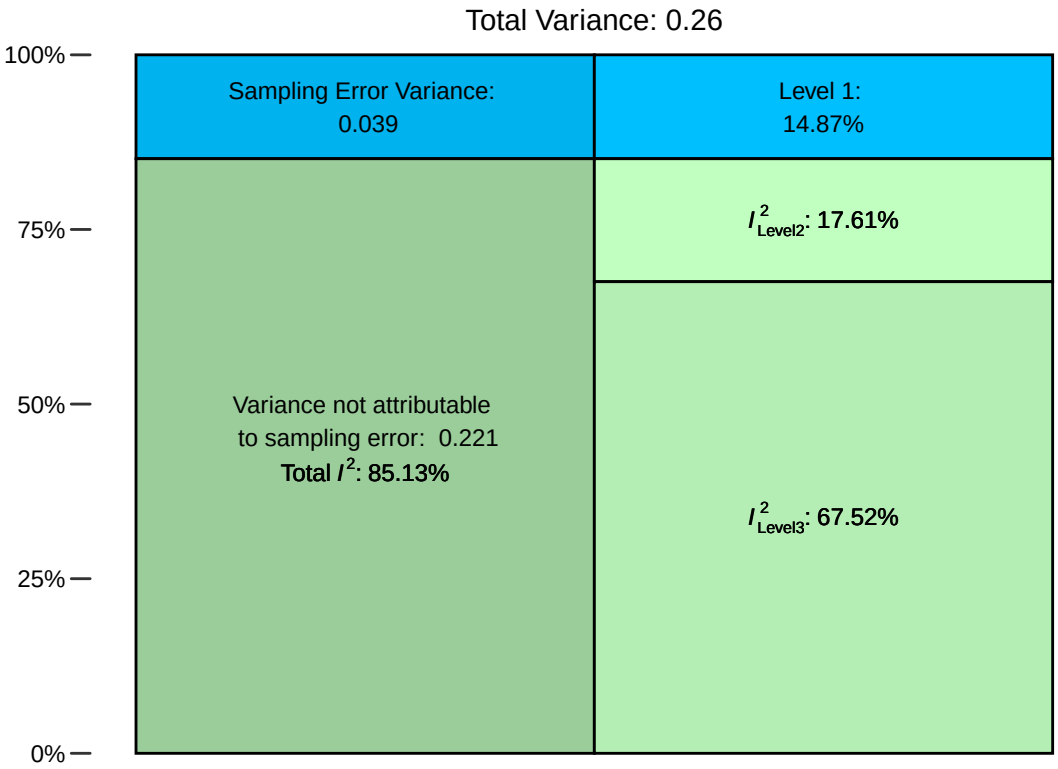
Supplementary Table 6: Resulting estimates from intercept-only model converted to Pearson's correlation. 'Estimate' is the Pearson's r summary effect size. 'CI.lb' and 'CI.ub' are the confidence intervals lower and upper bounds, respectively.

Parameter	Value
Estimate	0.33
CI.lb	0.23
CI.ub	0.43

Heterogeneity

Supplementary Table 7: Uncounted heterogeneity of the intercept-only model as measured by I^2 statistic. Within study heterogeneity (level 2) corresponds to the uncounted variation that is found on effect sizes within studies, and between study heterogeneity corresponds to the uncounted variation between studies (level 3).

	Within study	Between study
% Unexplained variation	17.61	67.52



Supplementary Figure 2: Visual representation of how variance was distributed over the multilevel structure of the intercept-only model. Within study heterogeneity (level 2) corresponds to the unnacounted variation that is found on effect sizes within studies, and between study heterogeneity corresponds to the unnacounted variation between studies (level 3).

Subgroup analysis

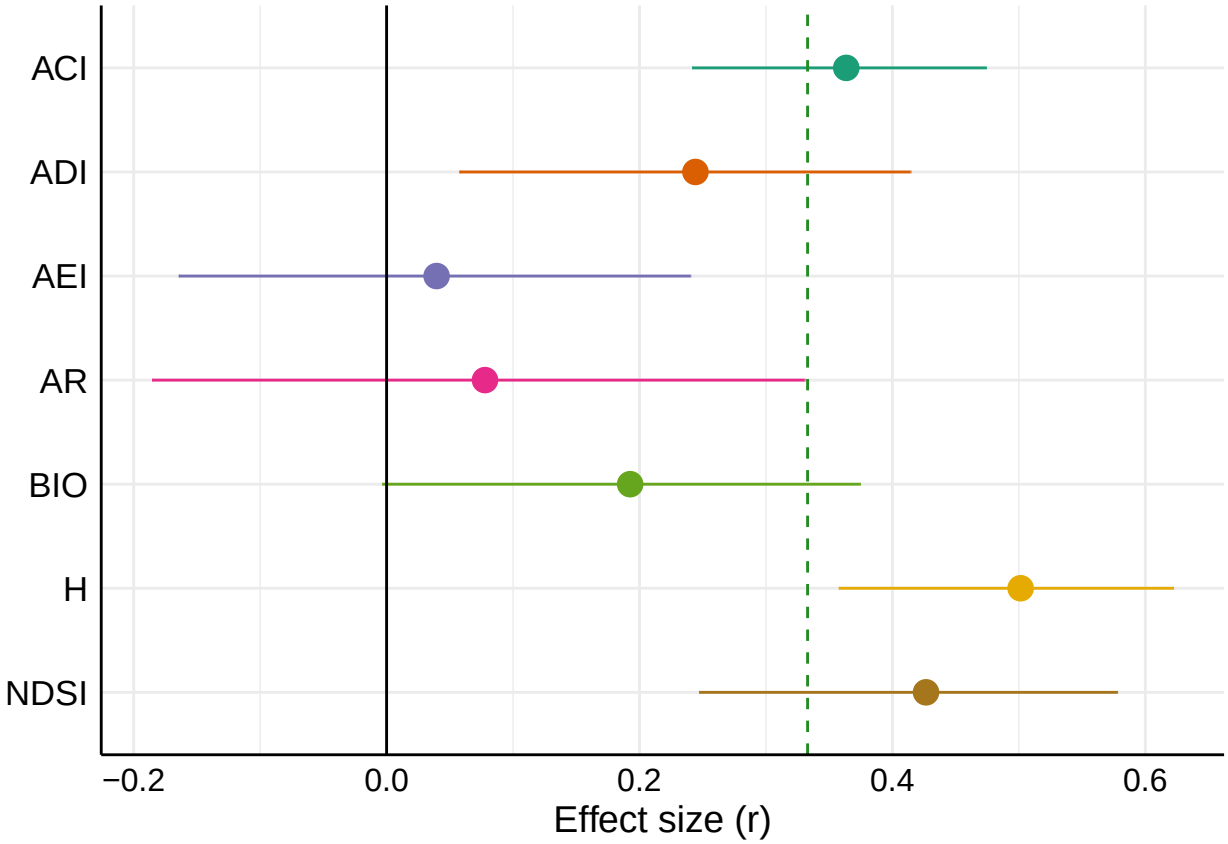
Model output

```
res_indices <- rma.mv(z, var, random = ~ 1 | id/entry, mods = ~ index - 1, data = df_indices)
res_indices
```

```
##
## Multivariate Meta-Analysis Model (k = 327; method: REML)
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed    factor
## sigma^2.1  0.0295  0.1716     34     no        id
## sigma^2.2  0.1710  0.4135    327     no  id/entry
##
## Test for Residual Heterogeneity:
## QE(df = 320) = 1876.0325, p-val < .0001
##
## Test of Moderators (coefficients 1:7):
## QM(df = 7) = 70.5454, p-val < .0001
##
## Model Results:
##
##            estimate      se    zval    pval    ci.lb   ci.ub
## indexACI     0.3809  0.0685  5.5596  <.0001   0.2466  0.5152  ***
## indexADI     0.2493  0.0977  2.5506  0.0108   0.0577  0.4408    *
## indexAEI     0.0396  0.1048  0.3774  0.7059  -0.1658  0.2449
## indexAR      0.0780  0.1354  0.5756  0.5649  -0.1875  0.3434
## indexBIO     0.1950  0.1012  1.9266  0.0540  -0.0034  0.3934    .
## indexH       0.5511  0.0903  6.1036  <.0001   0.3742  0.7281  ***
## indexNDSI    0.4557  0.1037  4.3944  <.0001   0.2524  0.6589  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Supplementary Table 8: Resulting estimates from sub-group analysis. The 'Estimate' column is the Pearson correlation effect size; SE is the standard error and CI.lb and CI.up; the lower and upper bounds of the confidence intervals, respectively.

Index	Estimate	SE	CI.lb	CI.ub
ACI	0.363	0.068	0.242	0.474
ADI	0.244	0.097	0.058	0.414
AEI	0.040	0.104	-0.164	0.240
AR	0.078	0.135	-0.185	0.331
BIO	0.193	0.101	-0.003	0.374
H	0.501	0.090	0.358	0.622
NDSI	0.427	0.103	0.247	0.578



Supplementary Figure 3: Effect size mean estimates (circles) and corresponding 95% confidence intervals (horizontal lines) obtained from sub-group meta-analysis with acoustic indices as the grouping factor. Estimated effect sizes whose 95% confidence intervals do not overlap zero (black vertical line) indicate a positive correlation between acoustic indices and diversity if they are to the right of zero, or a negative correlation if they are to the left of zero. The dashed green vertical line represents the summary effect size obtained from the intercept only meta-analysis.

Meta-regression

Model output

```
res_full <- rma.mv(z, var, random = ~1 | id/entry,
                   mods = ~ index + bio + environ + diversity_source,
                   data = df_full)

res_full
```

```
##
## Multivariate Meta-Analysis Model (k = 296; method: REML)
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed    factor
## sigma^2.1  0.0355  0.1884     33     no        id
## sigma^2.2  0.1738  0.4168    296     no  id/entry
##
## Test for Residual Heterogeneity:
## QE(df = 284) = 1730.2577, p-val < .0001
##
```

```
## Test of Moderators (coefficients 2:12):
## QM(df = 11) = 27.4277, p-val = 0.0040
##
## Model Results:
##
##                           estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt                     0.3590  0.1416   2.5359  0.0112   0.0815   0.6365
## indexADI                   -0.1294  0.1171  -1.1049  0.2692  -0.3590   0.1001
## indexAEI                   -0.2916  0.1231  -2.3679  0.0179  -0.5329  -0.0502
## indexAR                    -0.2735  0.1482  -1.8461  0.0649  -0.5639   0.0169
## indexBIO                   -0.1449  0.1203  -1.2038  0.2287  -0.3807   0.0910
## indexH                      0.1977  0.1092   1.8111  0.0701  -0.0162   0.4117
## indexNDSI                   0.0840  0.1260   0.6667  0.5050  -0.1630   0.3310
## bioabundance               -0.0815  0.1589  -0.5133  0.6078  -0.3929   0.2298
## biodiversity               -0.0420  0.0950  -0.4423  0.6583  -0.2283   0.1442
## biosound_abundance          0.2600  0.1470   1.7690  0.0769  -0.0281   0.5480
## environA                   -0.0656  0.1484  -0.4422  0.6583  -0.3565   0.2252
## diversity_sourceacoustic   -0.0091  0.1464  -0.0623  0.9504  -0.2962   0.2779
##
## intrcpt                   *
## indexADI
## indexAEI                  *
## indexAR                   .
## indexBIO
## indexH                    .
## indexNDSI
## bioabundance
## biodiversity
## biosound_abundance        .
## environA
## diversity_sourceacoustic
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Collinearity

Supplementary Table 9: VIF values obtained for each moderator level.

Moderators	VIF
ADI	1.497
AEI	1.450
AR	1.250
BIO	1.478
H	1.472
NDSI	1.437
Species abundance	1.244
Species diversity	1.156
Abundance of sounds	1.454
Aquatic	1.393
Acoustic	1.623

Test of moderators

Supplementary Table 10: Wald-type tests for all moderators (first row), and for each moderator separately (remaining rows). 'Q' is the Wald statistic; 'df' are the degrees of freedom; and 'p' is the probability that moderator estimates came from a chi-square distribution, where all estimates are equal to zero. Thus a p-value < 0.05 gives support against the null hypothesis that moderator levels estimates are equal to zero (i.e. they do not explain variation in effect sizes).

Moderator	Q	df	p
All moderators	27.428	11	0.004
Acoustic indices	22.353	6	0.001
Diversity metrics	3.561	3	0.313
Environment	0.196	1	0.658
Diversity source	0.004	1	0.950

Contrasts between moderator levels

Contrasts between H and other acoustic indices

Supplementary Table 11: Wald-type tests for the contrasts between acoustic index H with all other acoustic indices. The column 'Compared' expresses the comparison, in this cases it is the difference between the estimate H and the estimate of each of the other acoustic indices. The column 'Estimate' is the estimate obtained from the difference expressed in the previous column. 'SE' is the standard error of the difference, and CI.lb, CI.up the confidence interval lower and upper bound, respectively. 'QM' is the Wald statistic; 'p' is the probability that the difference between estimates is equal to zero. Thus, a p-value < 0.05 gives support against the null hypothesis of no difference between the estimate of the H index and the estimate of the other index.

Compared	Estimate	SE	CI.lb	CI.up	QM	p
H - ADI	0.327	0.122	0.089	0.566	7.223	0.007
H - AEI	0.489	0.127	0.241	0.738	14.901	0.000
H - AR	0.471	0.152	0.173	0.769	9.623	0.002
H - BIO	0.343	0.124	0.099	0.586	7.591	0.006
H - ACI	0.198	0.109	-0.016	0.412	3.280	0.070
H - NDSI	0.114	0.130	-0.141	0.369	0.765	0.382

Contrasts between NDSI and other acoustic indices

Supplementary Table 12: Wald-type tests for the contrasts between acoustic index NDSI with all other acoustic indices. The column 'Compared' expresses the difference between the estimate NDSI and the estimate of each of the other acoustic indices. The column 'Estimate' is the estimate obtained from the difference expressed in the previous column; 'SE' is the standard error of the difference, and CI.lb, CI.up the confidence interval lower and upper bound, respectively; 'QM' is the Wald statistic; 'p' is the probability that the difference between estimates is equal to zero. Thus, a p-value < 0.05 gives support against the null hypothesis of no difference between the estimate of the NDSI index and the estimate of the other index.

Compared	Estimate	SE	CI.lb	CI.up	QM	p
NDSI - ADI	0.213	0.133	-0.047	0.474	2.586	0.108
NDSI - AEI	0.376	0.138	0.106	0.645	7.442	0.006
NDSI - AR	0.358	0.161	0.041	0.674	4.914	0.027
NDSI - BIO	0.229	0.135	-0.036	0.494	2.869	0.090
NDSI - H	-0.114	0.130	-0.369	0.141	0.765	0.382
NDSI - ACI	0.084	0.126	-0.163	0.331	0.444	0.505

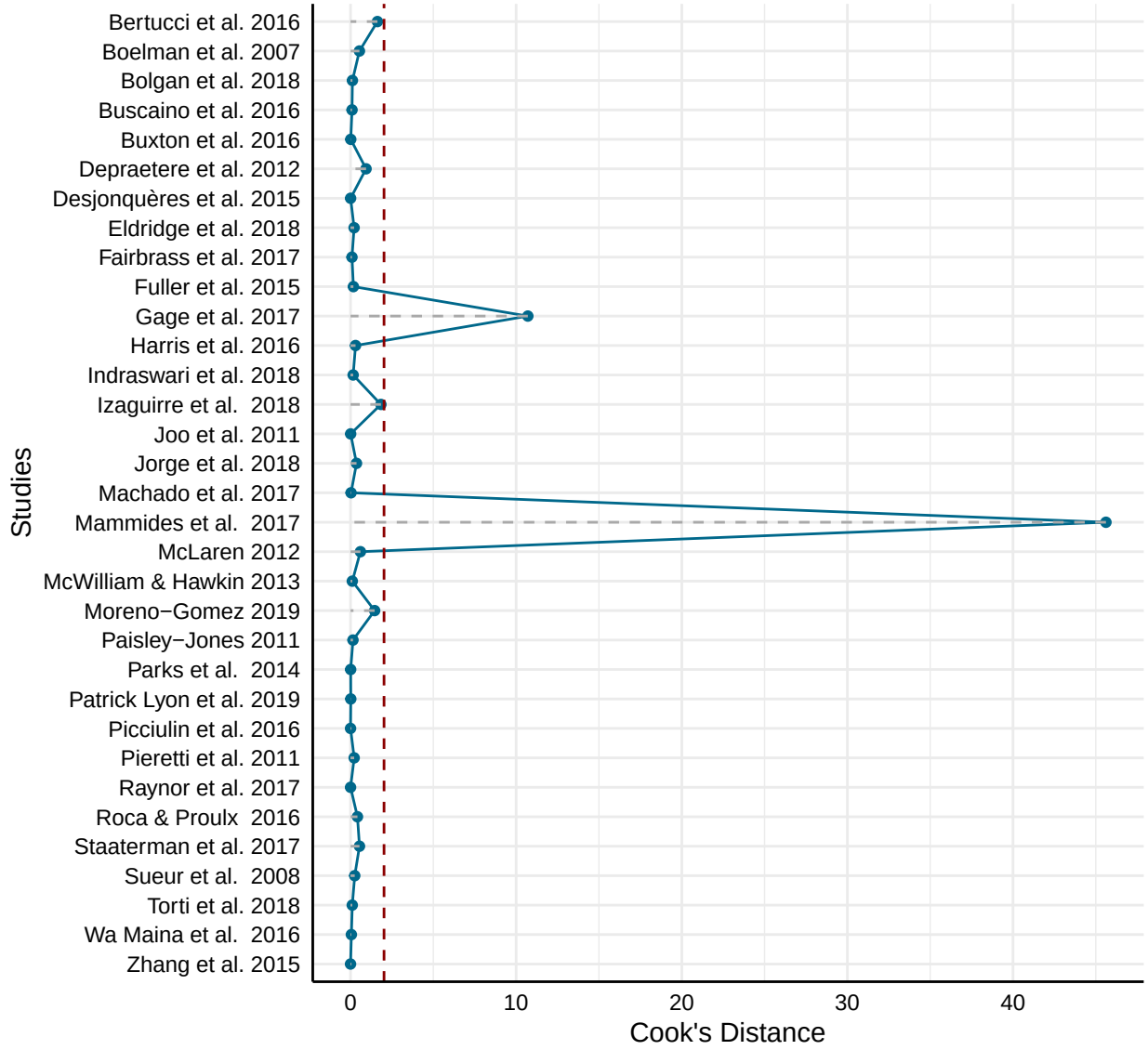
Contrasts between sound abundance and other biodiversity parameters

Supplementary Table 13: Wald-type tests for the contrasts between the diversity metric 'abundance of sounds' with all other diversity metrics. The column 'Compared' expresses the difference between the estimate 'abundance of sounds' and the estimate of each of the other diversity metrics. The column 'Estimate' is the estimate obtained from the difference expressed in the previous column; 'SE' is the standard error of the difference, and CI.lb, CI.up the confidence interval lower and upper bound, respectively; 'QM' is the Wald statistic; 'p' is the probability that the difference between estimates is equal to zero. Thus, a p-value < 0.05 gives support against the null hypothesis of no difference between the estimate of the sound abundance metric and the estimate of the other metric.

Compared	Estimate	SE	CI.lb	CI.up	QM	p
Abundance of sounds - Species abundance	0.342	0.215	-0.079	0.762	2.530	0.112
Abundance of sounds - Species diversity	0.302	0.172	-0.035	0.639	3.087	0.079
Abundance of sounds - Species richness	0.260	0.147	-0.028	0.548	3.129	0.077

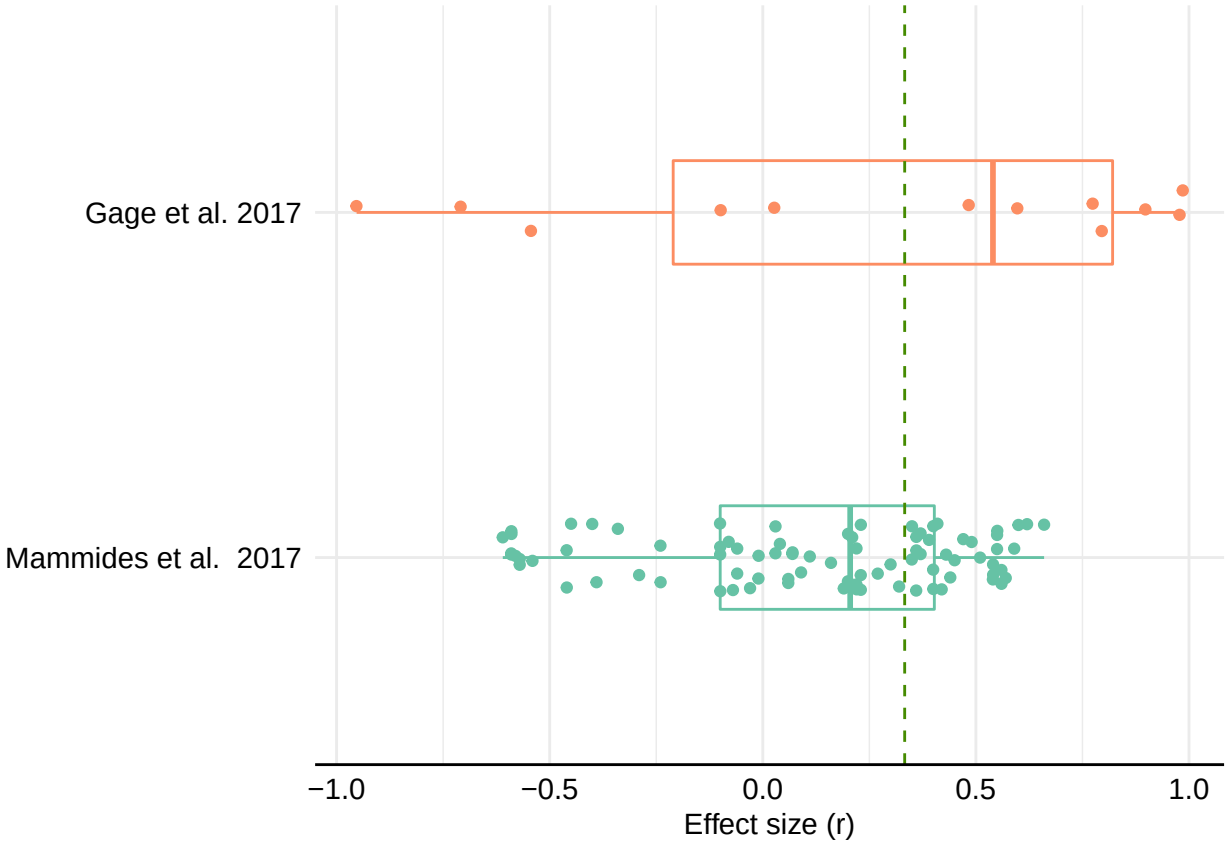
Sensibility analysis

Cook's distance



Supplementary Figure 4: Cook's distance values for each study (blue dots) and average Cook's distance over all studies (dashed vertical red line). The Cook's distance for a given study can be interpreted as the distance between the entire set of predicted values once with this study included and once with this study excluded from the model fitting procedure. On the y-axis are the studies identified by first author and year. The x-axis corresponds to the Cook's distance values.

Inspection of outliers



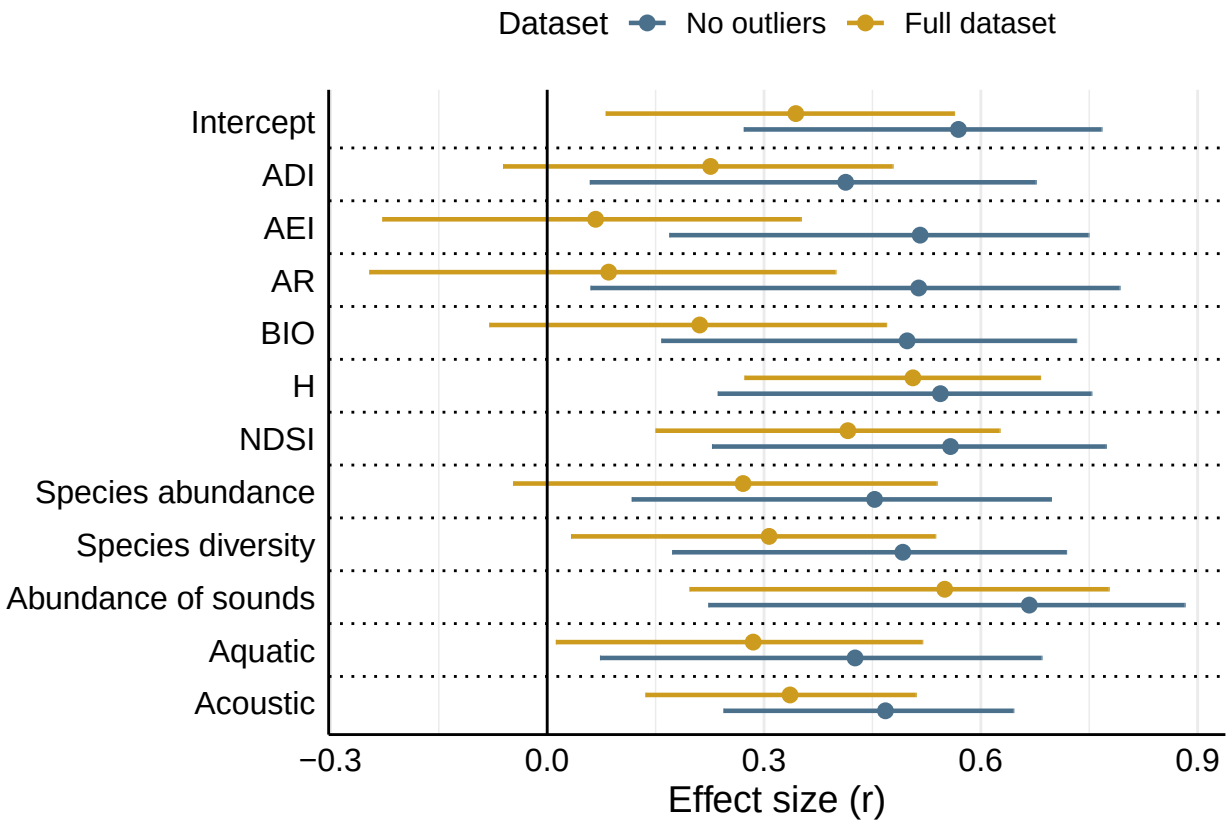
Supplementary Figure 5: Boxplot and distribution of effect size values (dots) of the two studies identified as outliers. The y-axis identifies the study, and the x-axis corresponds to the Pearson r effect size. The green vertical dashed line is the summary effect obtained in the intercept-only model.

Model output without outliers

```
res_no_outliers <- rma.mv(z, var, random = ~1 | id/entry,
                  mods = ~ index + bio + environ + diversity_source,
                  data = df_no_outliers)
res_no_outliers
```

```
##
## Multivariate Meta-Analysis Model (k = 200; method: REML)
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed    factor
## sigma^2.1  0.1225  0.3500     31     no        id
## sigma^2.2  0.0557  0.2360    200     no  id/entry
##
## Test for Residual Heterogeneity:
## QE(df = 188) = 472.4368, p-val < .0001
##
## Test of Moderators (coefficients 2:12):
## QM(df = 11) = 6.3074, p-val = 0.8521
##
## Model Results:
##
##                           estimate      se     zval    pval    ci.lb   ci.ub
## intrcpt                     0.6457  0.1870   3.4522  0.0006   0.2791  1.0123
## indexADI                   -0.2059  0.1118  -1.8412  0.0656  -0.4251  0.0133
## indexAEI                   -0.0753  0.1260  -0.5977  0.5500  -0.3224  0.1717
## indexAR                    -0.0770  0.2082  -0.3700  0.7114  -0.4850  0.3310
## indexBIO                   -0.0992  0.1192  -0.8319  0.4054  -0.3328  0.1345
## indexH                     -0.0357  0.1058  -0.3372  0.7360  -0.2430  0.1716
## indexNDSI                  -0.0164  0.1406  -0.1165  0.9073  -0.2919  0.2591
## bioabundance               -0.1566  0.1512  -1.0355  0.3005  -0.4530  0.1398
## biodiversity               -0.1064  0.1298  -0.8199  0.4122  -0.3607  0.1479
## biosound_abundance          0.1597  0.2105   0.7587  0.4480  -0.2528  0.5722
## environA                   -0.1907  0.2057  -0.9271  0.3539  -0.5939  0.2125
## diversity_sourceacoustic   -0.1379  0.1909  -0.7226  0.4700  -0.5120  0.2362
##
## intrcpt                   ***
## indexADI                    .
## indexAEI
## indexAR
## indexBIO
## indexH
## indexNDSI
## bioabundance
## biodiversity
## biosound_abundance
## environA
## diversity_sourceacoustic
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

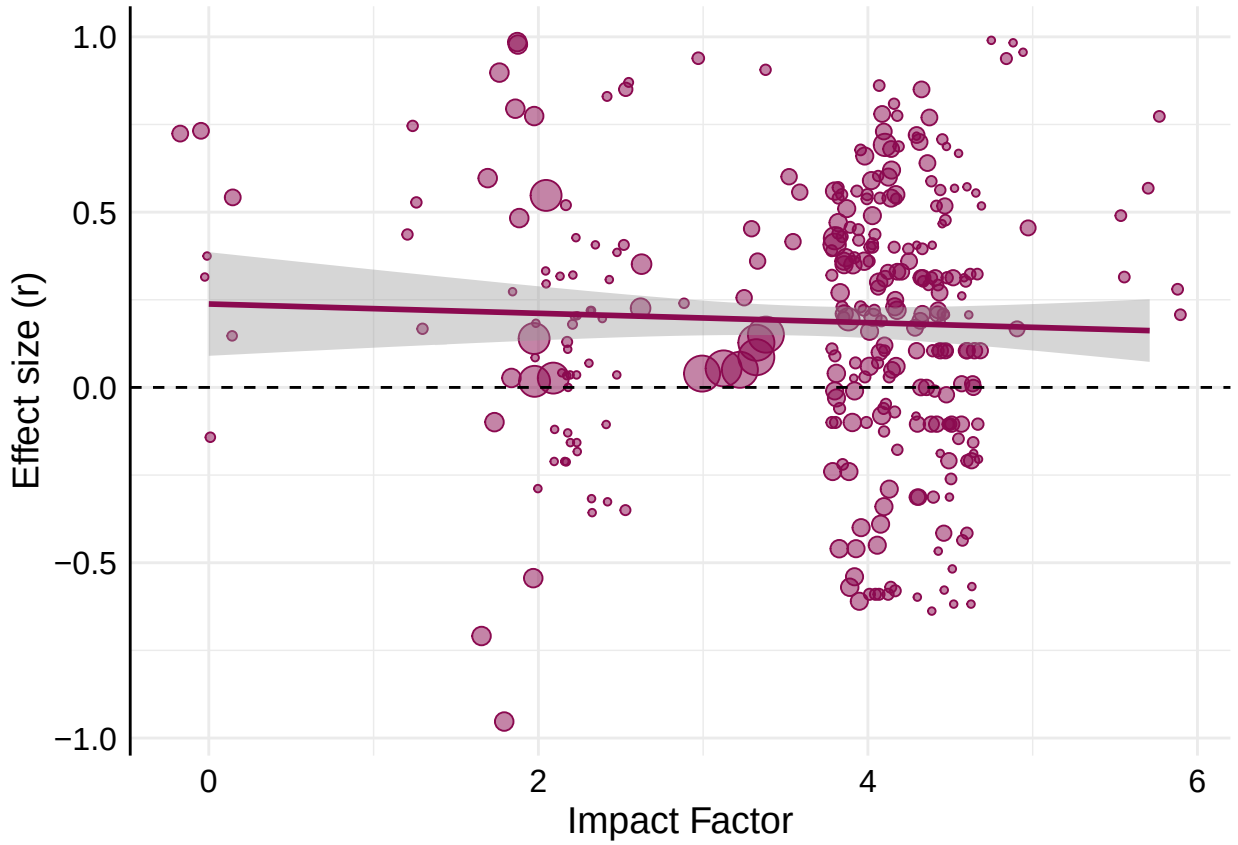
Visualization of model estimates with and without outliers



Supplementary Figure 6: Contrast of model estimates obtained with meta-regression analysis over the full dataset (yellow) and over the dataset with outliers removed (blue). Estimates are represented with circles and corresponding 95% confidence intervals with horizontal lines. Each estimate corresponds to the additive effect when a moderator level is replaced in the intercept (e.g. ADI is the additive effect of ADI when ADI is put as intercept instead of ACI). Estimated effect sizes whose 95% confidence intervals do not overlap zero (black vertical line) indicate a positive correlation between acoustic indices and diversity if they are to the right of zero, or a negative correlation if they are to the left of zero. We considered outliers every study that had a Cook's distance value higher than the mean of all Cook distances. Model moderators were acoustic indices (ADI, AEI, AR, BIO, H, NDSI, with ACI as intercept), diversity metric (Species abundance, Species diversity, Abundance of sounds, with Species richness as intercept), environment (Aquatic, with Terrestrial as intercept), diversity source (Acoustic, with Non-Acoustic as intercept). The solid vertical black line represents a null effect size.

Effect size tendencies

Relationship between effect size and journal impact factor



Supplementary Figure 7: Relation between effect size mean values (circles) and journal impact factor. Circle size indicates the relative sample size of each effect size. The fitted line is a simple least squares regression with the corresponding 95% confidence interval region in grey. The dashed horizontal line represents an effect size of 0. Effect size mean values are positioned along the impact factor axis with minor random noise to reduce overlapping.