Impact – a FORTRAN program for gradient analysis

Version 1.0

Werner Ulrich

Nicolaus Copernicus University in Toruń

Department of Animal Ecology

Gagarina 9, 87-100 Toruń; Poland

e-mail: ulrichw @ umk.pl

Latest update: 06.09.2010

1. Introduction

In ecological research it is often necessary to assess the impact of environmental variables on species occurrences. In multispecies communities this impact might differ from species to species and a community wide approach seems preferable. Frequently impacts are assessed by ordination techniques like correspondence analysis, CANOCO, PCA (Legendre and Legendre 1998), or nestedness (Ulrich et al. 2009), and indicator species analysis (Dufrene and Legendre 1997).

The present program Impact uses a null model approach to infer the response of species occurrences and/or abundances in multispecies communities on environmental gradients defined by an impact variable similar to the random skewer method of Pielou (1984).

2. The metric

The basic calculations are shown in the first Table. You need two input variables, one containing the occurrences of m species at n sites and one containing the values of environmental variables at these sites. In the simplest case of presence—absence data you calculate for a given species i the average sum P_i ($= \Sigma V_{i,presnt}/n_{i,presnt}$) of the environmental variable V for each site where the species is present ($n_{i,present}$) and the average sum A_i ($= \Sigma V_{i,absent}/n_{i,absent}$) where the species is absent ($n_{i,absent}$). In many cases it will be preferable to use the ranks of the environmental data to correct for outliers and non-linear behaviour in the environmental variable. The difference $D_i = P_i - A_i$ is now a test metric that describes the effect of the environmental variable on species i.

Statistical inference is now done by a randomization procedure. Reshuffle the values of the environmental variable randomly and calculate the expected difference $D_{Exp,i}$. Repeat this procedure 1000 times to obtain the null distribution of expectation. The comparison of the observed difference with this null distribution gives the probability of deviation from null expectation.

This simple test can be extended to species abundance data. In this case appropriate metrics are either the coefficient of correlation between impact (environmental) and response (species) variable or the slope of an ordinary least squares (OLS) regression

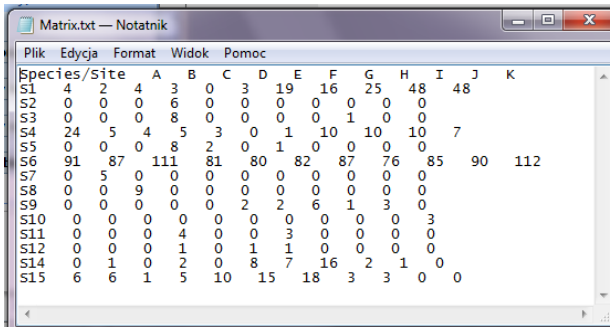
Species/Sample	A	B	C	D	E	F	G	H	I	J	K
S1	0	2	0	3	0	0	19	16	25	48	48
S2	0	0	0	6	0	0	0	0	0	0	0
S3	0	0	0	8	0	0	0	0	1	0	0
S4	24	5	4	5	3	0	1	10	10	10	7

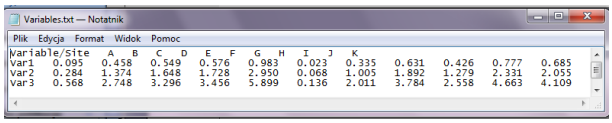
Variable/	A	B	C	D	E	F	G	H	I	J	K
Var1	0.095	0.485	0.049	0.376	0.183	0.023	0.935	0.631	0.426	0.777	0.685
Var2	0.284	1.374	1.648	1.728	2.95	0.068	1.005	1.892	1.279	2.331	2.055
Var3	0.568	2.748	3.296	3.456	5.899	0.136	2.011	3.784	2.558	4.663	4.109

	Presence - absence		Abundances	
		Var1		Var1
S1	Sum of presences/$n_{present}$	0.616	Regression slope	43.8
	Sum of absences/n_{absent}	0.088		
	Difference	0.528		

between both variables. Again statistical inferences comes from the above described randomization procedure.

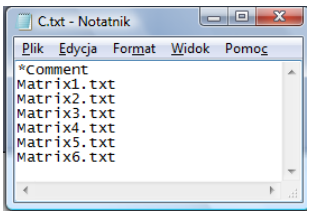
The test values D_i for all species can now be used for a test for heterogeneity in species responses. A ranking of test values shows the contribution of the impact variable on each species. In the case of differential response of species some species will have very large or very small values of D compared to a random expectation. Hence, the community wide variance Var_D of D is a measure of response heterogeneity. Again statistical inference comes from the comparison of the observed Var_D with the expected Var_D obtained from 1000 null distributions where for each reshuffling the community wide Var_D was calculated.



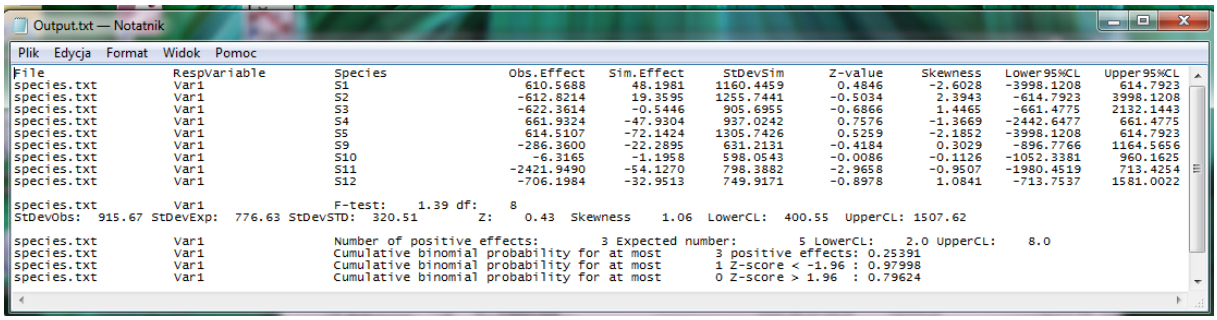


3. Program description

The present FORTRAN 95 software *Impact* calculates the abovementioned metrics for presence—absence and abundance data. The program accepts standard space delimitated text file matrices (cf. the example files below), with sites in columns and species in rows (the Ecosim format, Gotelli and Entsminger 2005). Tab-delimited files are not accepted. Species and site names must not contain spaces. The maximum number of sites is 5000 and the maximum number of signs within rows (incl. spaces) is 30000. Multiple analyses using many matrices are possible, and need an additional text file containing the file names as shown beside. The first line of this file has to be a comment line.



The single output file *Output.txt* gives file names and the environmental variable analyzed. For each species it contains observed and expected metrics as well as the standard deviation of expectation. Expected values are obtained from 1000 reshuffled environmental variables. The out put contains also Z-transformed scores (Z = (observed score – expected score)/$StdDev_{exp}$). Also given are the skewness and the



File	RespVariable	Species	Obs.Effect	Sim.Effect	StDevSim	Z-value	Skewness	Lower95%CL	Upper95%CL
species.txt	Var1	S1	610.5688	48.1981	1160.4459	0.4846	-2.6028	-3998.1208	614.7923
species.txt	Var1	S2	-612.8214	19.3595	1255.7441	-0.5034	2.3943	-614.7923	3998.1208
species.txt	Var1	S3	-622.3614	-0.5446	905.6955	-0.6866	1.4465	-661.4775	2132.1443
species.txt	Var1	S4	661.9324	-47.9304	937.0242	0.7576	-1.3669	-2442.6477	661.4775
species.txt	Var1	S5	614.5107	-72.1424	1305.7426	0.5259	-2.1852	-3998.1208	614.7923
species.txt	Var1	S9	-286.3600	-22.2895	631.2131	-0.4184	0.3029	-896.7766	1164.5656
species.txt	Var1	S10	-6.3165	-1.1958	598.0543	-0.0086	-0.1126	-1052.3381	960.1625
species.txt	Var1	S11	-2421.9490	-54.1270	798.3882	-2.9658	-0.9507	-1980.4519	713.4254
species.txt	Var1	S12	-706.1984	-32.9513	749.9171	-0.8978	1.0841	-713.7537	1581.0022

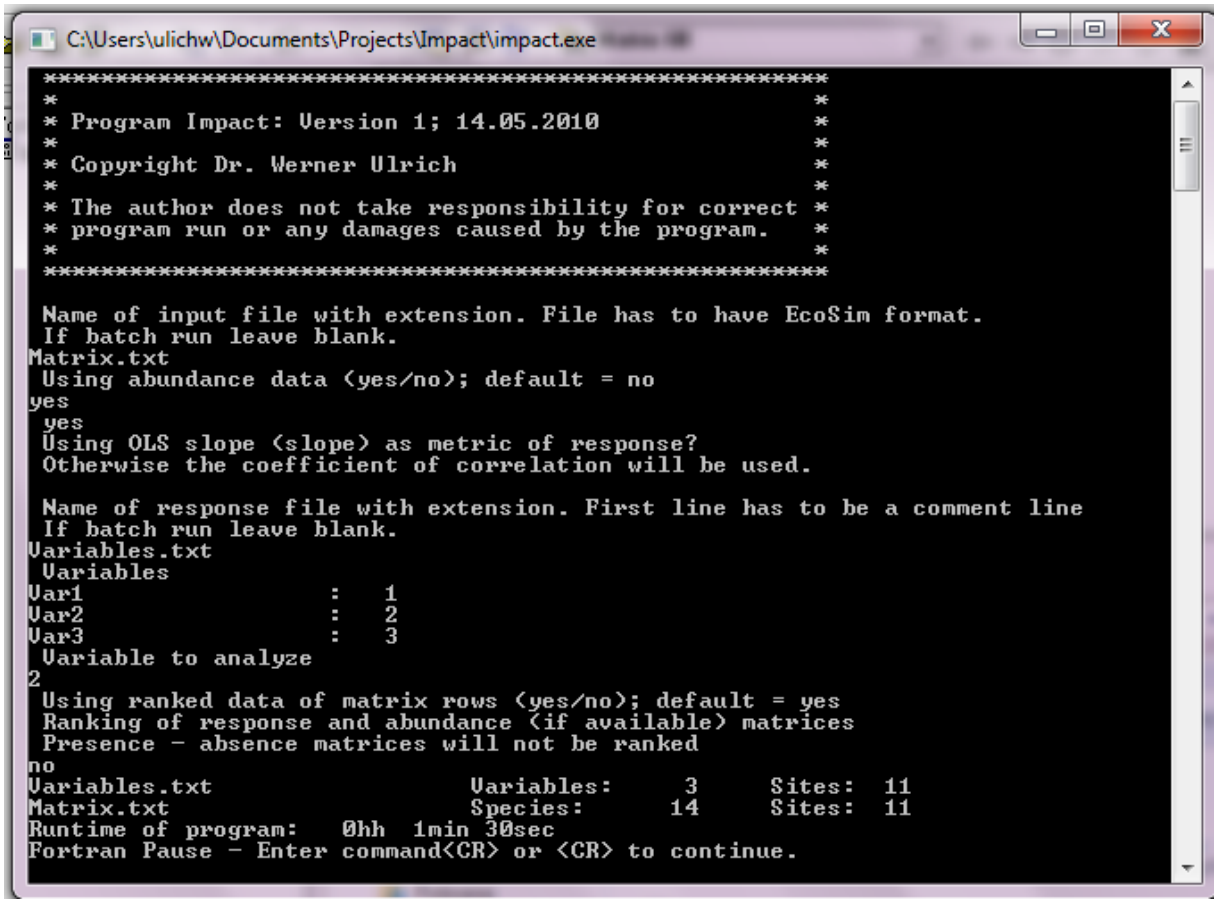
species.txt Var1 F-test: 1.39 df: 8
StDevObs: 915.67 StDevExp: 776.63 StDevSTD: 320.51 Z: 0.43 Skewness 1.06 LowerCL: 400.55 UpperCL: 1507.62

species.txt Var1 Number of positive effects: 3 Expected number: 5 LowerCL: 2.0 UpperCL: 8.0
species.txt Var1 Cumulative binomial probability for at most 3 positive effects: 0.25391
species.txt Var1 Cumulative binomial probability for at most 1 Z-score < -1.96 : 0.97998
species.txt Var1 Cumulative binomial probability for at most 0 Z-score > 1.96 : 0.79624

upper and lower 95% confidence limits of the null distribution (two-sided test).

Lastly *Impact* provides the observed and expected standard deviations of species responses. For statistical inference the program gives the quotient of observed and expected standard deviation (the F-test value) together with the respective degree of freedom (number of species –1). Additionally it provides Z-scores, skewness of the null distribution of variances and upper and lower two tailed 95% confidence limits.

Finally the program provides four tests of deviations of species effects from random expectation. First, the software gives lower and upper two tailed 95% confidence limits of the distribution of species Z-cores. If these do not include zero a significant deviation from random expectation appears at the five percent error level. Of course this test has poor power at low species richness. Second, the program calculates the number of positive effects and gives the cumulative binomial probability for this number. If no positive effect was found it give the cumulative probability to find at least one positive effect. Due to symmetry this is also a test for numbers of negative effects. Lastly, *Impact* provides the cumulative binomial probabilities to find at least the observed number of significant positive and negative Z-scores. If no significant Z-



score was found it give the cumulative probability to find at least one significant score.

4. Program run

Impact first asks for the input file. Give it with extension (example: file.txt). In the case of multiple runs a carriage return results in the question for the name of the file that contains the matrix file names for multiple analysis (cf. the example above). All of the files have to be in the same directory.

Now the program asks whether the species x site file contains abundance or presence absence data. In the case of abundance data you also have to decide whether to use the coefficient of correlation or the OLS slope as impact metric.

Next the program asks for the name of the file containing the environmental data. Again you have to provide the name with extension. The program analysis the file and shows the variables. You have to choose the variable you are interested in.

The next question concerns the use of ranked or raw data. If you choose the default option (ranked data) the program will rank the environmental variable and the abundances of each species. Of course, presence - absence matrices will not be ranked.

5. Citing *Impact*

Impact is freeware but nevertheless if you use *Impact* in scientific work you should cite *Impact* as follows:

Ulrich W. 2010. *Impact* – a FORTRAN program for gradient analysis. Version 1.0. http::www.umk.pl/~ulrichw.

You can also cite the associated article in *Methods in Ecology and Evolution*:

Gotelli, N. J., Ulrich W., Maestre, F. T. 2011. Randomization tests for quantifying species importance to ecosystem function. Meth. Ecol. Evol.: in press.

6. System requirements

Impact is written in FORTRAN 95, has been compiled under a 64 bit architecture, and runs under Windows 7, XP, and Vista. The maximum number of sites is 5000 and the maximum number of signs within rows (incl. spaces) is 30000. Otherwise computation abilities are only limited by the computer's memory.

7. Acknowledgements

Development of this program was initiated by an idea of Nicholas Gotelli. Its development was supported by grants from the Polish Science Committee (KBN, 3 P04F 034 22, KBN 2 P04F 039 29).

8. References

Dufrene M., Legendre P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol. Monogr. 67: 345-366.

Gotelli N.J., Entsminger G.L. 2005. EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear. Jericho, http://garyentsminger.com/ecosim.htm.

Gotelli N. J., Ulrich W., Maestre F. T. 2011. Randomization tests for quantifying species importance to ecosystem function. Meth. Ecol. Evol.: in press.

Legendre P, Legendre L. 1998. Numerical ecology. 2nd edition., Elsevier.

Pielou E. C. 1984. Probing multivariate data with random skewers: a preliminary to direct gradient analysis. Oikos 42: 161-165

Ulrich W., Almeida-Neto M., Gotelli N. G. 2009. A consumer's guide to nestedness analysis. Oikos 118: 3-17.