MJS, Vol. 6, No. 1, January - June 2007, pp. 27-35
ISSN 0975-3311
https:/ / doi.org/ 10.12725/ mjs.10.2

BIODIVERSITY STUDIES AND MULTICOLLINEARITY IN MULTIVARIATE DATA ANALYSIS

Sabu K. Thomas[*] *& K.T. Thomachen*[**]

ABSTRACT

Multicollinearity of explanatory variables often threatens statistical interpretation of ecological data analysis in biodiversity studies. Using litter ants as an example, the impact of mulitcollinearity on ecological multiple regression and complications arising from collinearity is explained. We list the various statistical techniques available for enhancing the reliability and interpretation of ecological multiple regressions in the presence of mulitcollinearity.

Key words: *biodiversity studies; mulitcollinearity; multiple regression; litter ants.*

Introduction

Biodiversity studies require analysis of the regulatory effect of various environmental and biological factors (independent or explanatory variables) on the population, and community ecology (dependent or response variables) of various organisms. Following the convention of Biological Diversity (CBD) signed

* *Litter Entomology Research Unit, PG & Research Department of Zoology; St. Joseph's College. Devagiri, Calicut 673008*

** *Department of Economics, St. Joseph's College. Devagiri, Calicut 673008, India. Corresponding author:e-mail: sabukthomas@gmail.com*

at the UNCED "Earth Summit" in Rio de Janeiro in June 1992 and the later declaration of Western as biodiversity hot spot efforts are being going on to document the biodiversity and to analyze the regulatory effect of various environmental and biological factors (independent or explanatory variables) on the population, and community ecology (dependent or response variables) of various flora and fauna in the region. Often multivariate analysis of the relationship or influence of various environmental and biological factors (independent or explanatory variables) on the population and community of both flora and fauna is necessary to determine the causative factors and often it will be hindered by the complex nature of ecological data, in which targeted ecological responses are linked to many explanatory variables that are often correlated among each other (multicollinear). Ecologists recognize that the complex nature of ecological data or macroecological and biodiversity data are linked to many explanatory variables that are often correlated among each other (multicollinear) and leads to erroneous interpretations. For example rainfall may lead to rise in moisture levels and humidity and reduction in temperature in forest litter habitats; but if a forest region is having even terrain with low slope, more litter will accumulate and loss of litter by way of rainwater run off will be less leading to higher moisture than in a uneven region with very high rainfall. Similarly if a region is geographically located in the windward mountains gaps or in tops there will be more wind mediated evapotranspiration leading to less moisture and humidity facing with high wind velocity at least during the rainless periods and even if the region gets high annual rainfall. It is well known that rainfall and moisture influences the ground dwelling fauna in a complex way. This sort of complex nature of macro-ecological and biodiversity data linked to many explanatory variables that are often correlated among each other (multicollinear) will lead to erroneous interpretations (Graham, 2003). We experienced these difficulties while analyzing role of ecological variables in determining litter ant abundance in the shola and deciduous forests of Western Ghats and found that multicollinear explanatory variables are difficult to analyze because their effects on the response can be due to either true synergistic relationships among the variables or spurious correlations... The statistical and inferential problems of multicollinearity in multiple regressions have been well established in the theoretical statistical literature and straightforward techniques exist for diagnosing and remediating the effects of multicollinearity in multiple regression. Yet, despite previous warnings by statisticians, very few published ecological papers at national levels that used multiple regressions for data analysis even discussed the potential presence of multicollinearity in ecological studies and very few papers tried to detect the presence of collinearity and follow the remedial measures.

Given the increasing importance of multivariate analysis in ecology and biodiversity studies, it is important to detect the presence of multicollineariy in ecological studies and in the present work we explain it with reference to our experience on the community ecology of forest litter ants. We selected litter ants as they dominate the macro-arthropod ground fauna of tropical forests (Levings & Windsor 1985; Hölldobler & Wilson1990). Even seemingly minor changes in environmental conditions are important to individual ants and these differences may explain some of the ant-community structure variations (Greenslade & Greenslade, 1977). Their relative stability, moderate diversity and sensitivity to climatic variables make them a suitable group for measurement of biodiversity assessments and indicating ecosystem modifications in various habitats (Agosti *et al.*, 2000; Brühl *et al.*, 2003). In the present study we explain to analyse how to detect the potential presence of multicollinearity in ecological studies employing multiple regression and what are alternate statistical techniques or multivariate analysis techniques that should be used for enhancing the reliability and interpretation of ecological data in the presence of multicollinearity.

Materials and Methods

Site: Study sites are located at Thirunelli in Wayanad region of southern Western Ghats in the state of Kerala (11°58', 11°3'N; 75° 45', 76° 28'E). Thirunelli forests spread over an area of 20.55 km^2 and occur at an elevation of 900 m amsl and above. No area records about the evergreen forests commonly referred as shola forests exist. Our estimates indicate that shola forests occupy an area of 6+2 km^2 at 1300 m amsl. Northeast monsoon from October to November supplements the June-September southwest monsoon in this region. But because of the deeply dissected topography with sheltering high ridges on the windward western and the eastern sides, the deciduous forests at Thirunelli receive less than 2000 mm/a rainfall whereas Periya and Vythiri regions on the western and shola forests on the eastern reaches receive more than 3000 mm/a. Rainfall data from the shola region is not available. Shola forests in the region are locally recognized as a high-rainfall region with greater wind velocity and higher evapo-transpiration rate compared with the deciduous forests.

Sampling protocol: A preliminary transect, following the standardized protocol abbreviated "A.L.L." (Ants of the Leaf Litter) (Agosti et al. 2000), was conducted in the evergreen forest at Thirunelly in January 2004. Fauna was extracted with a mini-Winkler apparatus (Fisher 1998; Bestelmeyer et al. 2000) for 24 h. Ants were hand picked, and transferred to labelled containers of 70% alcohol. Collected ant species samples were identified primarily based on Fauna of British India, Bingham (1903).

Rainfall data was collected from the records of State Electricity Board office at Thirunelli. Humidity and litter temperature was assessed with thermo-hygrometer (Barigo; Barigo Barometer fabric GmbH; Villingen-Schwenningen; Germany), moisture with moisture analyzer (Advance; M-3A deluxe; Advance Research Instrument Co; Ambala; India.), and litter depth with a standard 30 cm ruler. Slope of the terrain was calculated using the trigonometric formula 'tanø (where 'ø' is the angle of inclination (Jacobs & Meyer, 1972). Litter humidity, litter moisture, litter temperature, slope of the terrain and litter depth were calculated as the mean (arithmetic mean) sampling spots at each site (Table 1).

Table 1. Incidence based abundance of litter ants in deciduous forests at Thirunelli in Wayanad.

Subfamily	Species	Incidence data
Formicinae	Camponotus radiatus	2
Myrmecinae	Monomorium sp	2
Myrmecinae	Crematogaster sp	10
Myrmecinae	Solenopsis sp1	3
Myrmecinae	Solenopsis sp2	8
Myrmecinae	Cardiocondyla sp	7
Myrmecinae	Cardiocondyla sp1	4
Myrmecinae	Strumigenys sp	7
Myrmecinae	Myrmicaria brunnea	24
Myrmecinae	Myrmicaria sp1	2
Myrmecinae	Myrmicaria sp2	5
Myrmecinae	Phidole sp	2
Myrmecinae	Tetramorium sp1	5
Dolichoderinae	Tapinoma sp	30
Ponerinae	Leptogenys sp	2
Ponerinae	Pachycondyla sp	1
Ponerinae	Cryptopone sp	10
Ponerinae	Odontomachus sp	3
Ponerinae	Diacamma sp1	7
Ponerinae	Diacamma sp2	3
Dorylinae	Aenictus sp1	2
Dorylinae	Aenictus sp2	1
Dorylinae	Dorylus sp	5

Data Analysis: Effects of physical characteristics of the forest vegetation types on ant abundance was tested with multiple regressions (Zar, 2003) and the inherent collinearity (multicollinearity) of explanatory variables was analyzed with auxiliary regression (Neter & Kutner, 1996). Variations in physio-climatic factors, á-diversity and functional and trophic guild wise â-diversity between the forests types were analyzed with one way ANOVA (Zar, 2003). The Gretl (Version 1.1; Cottrell, 2006) open-source software for Windows was used for all statistical analysis.

Results & Discussion

Twenty three ant species in 17 genera from deciduous forests were collected. *Tapinoma* (21%) and *Myrmicaria brunnea* (17%) were the dominant species in deciduous forests. Multiple regression with physical factors as independent variable and ant abundance as dependant variable indicated significant overall relation in both the forests (R^2=0.93, $F_{5,24}$=59.54, $P < 0.05$. for shola forest: R^2=0.85, $F_{5,24}$=26.19, $P < 0.05$ for deciduous forest). However partial regression coefficients showed significant negative influence of temperature, moisture and humidity, and insignificant effect of litter depth and slope on ant abundance contradictory to what we got from the overall analysis. Negative influence of moisture and humidity on litter ants is not unexpected as it widely known that excessive wetness and coldness associated with the litter habitat in forests reduces the foraging area and time available for foraging ants (Olson, 1994; Brühl *et al.*, 1999). But the significant negative relation with temperature and negative though insignificant relationship with litter depth and is contradictory and theoretically wrong as low-litter cover reduces the foraging area and prey resources availability for litter ants and ants prefer open and dry environments in general (Vineesh *et al.*, 2007; Brühl *et al.*, 1999). This along with the high R^2 value (0.8) indicate the presence of collinearity among regressors and experienced statistician knows that it's an instance of high collinearity among regressors seriously impairing the validity of regression results. It will not only make the regression coefficients of many variables statistically insignificant but make the signs of many coefficients contrary to the theoretical expectations. This study has quantitatively showed that statistical and inferential problems created by multicollinearity can be extremely severe under realistic ecological conditions and any statistical investigation involving multiple regressors requires a careful examination of the likely presence of collinearity among regressors. Yet, very few articles at national level that used multiple regression for data analysis in ecological studies even discussed the potential presence of multicollinearity.

Table 2: Linear regression analysis of the relationship between physical characteristics and ant abundance in deciduous forests at Thirunelli in Wayanad.

Physical factors	Deciduous forest		
	B	t	Probability
(Constant)	25.14	7.78	
Humidity	-0.08	-2.30	*
Litter depth	-0.10	-0.96	n.s
Moisture	-0.18	-5.69	***
Slope	-0.04	-0.89	n.s
Temperature	-0.17	-2.55	*
R^2 values : 0.8			

Table 3: Auxilary regression analysis for the multicollinearity among physical characteristics in deciduous forests at Thirunelli in Wayanad.

Physical factors	R^2	$F_{4,25}$	p-value
Humidity	0.07	0.47	n.s
Litter depth	0.22	1.77	n.s
Moisture	0.56	7.91	***
Slope	0.14	0.99	n.s
Temperature	0.62	10.04	***

*p d", 0.05; ** p d" 0.01; *** p d", 0.001; n.s, not significant*

A large number of testing procedures are available in the literature (Draper & Smith, 1998; Netter *et al.*, 1996) to detect the pattern of mulitcollinearity. The classic symptoms of multicollinearity as the above example suggests is high R^2 and few significant *t* ratios. We outline below a few procedures suggested in the literature to detect the presence of multicollinearity.

1. Pair wise correlation among regressors
2. Partial correlations analysis
3. Running auxiliary regression
4. Calculation of Eigenvalnes and condition index and
5. Examination of the value of variance inflation factor

Our subsequent analysis of collinearity with auxiliary regression, (rated as the best among the above; Gjuarati, 2003, Table 3) indicated significant collinear influence of moisture and temperature and indicates the need to go for the following statistical procedures viz., Apriori information, Combining cross sectional and home series data, dropping of variables, transformation of variables, additional or new data, ridge regression, factor analysis and the more widely used principal component analysis. Most of these procedures are "rules of thumb procedures" depending on the type of data only help to stabilize the statistical analyses, making them less biased, less subjective, and more repeatable, but only the statistical collinearity will have been removed from the data (Netter *et al.*, 1996). The explanatory variables are still, by nature and in nature, correlated, whether or not functionally. And it should be emphasized that multicollinearity is a data deficiency problem aside from designing manipulative experiments to break correlations among explanatory variables, (Graham, 2003) no technique exists that allows researchers to infer the different functional relationships between the response and explanatory variables and sometimes we have no choice over the data made available for empirical analysis as in biological sciences.

Acknowledgements

We thank Ministry of Environment and Forests (MoEF), Govt. of India, for financial support; the Chief Wildlife Warden, Kerala, for permission; Vineesh PJ for critical reading of a final draft of the manuscript and incisive comments and T.N. Ananthakrishnan for literature.

References

1. AGOSTI, D., MAJER, J. D., ALONSO, L. E. & SCHULTZ, T. R., (Eds.), 2000. Ants- Standard Methods for Measuring and Monitoring Biodiversity. Smithsonian Institution Press, Washington. 280 pages.
2. BESTELMEYER, B. T., AGOSTI, D., LEEANNE, F., ALONSO, T., BRANDÃO, C. R. F., BROWN, W. L., DELABIE, J. H. C. & SILVESTRE, R. 2000 Field techniques for the study of ground-living

ants: An Overview, description, and evaluation. In *Ants: standard methods for measuring and monitoring biodiversity* (eds. Agosti, D., Majer, J. D., Alonso, L. E., and Schultz, T. R.,), Smithsonian Institution Press, pp.122-144..

3. BINGHAM, C. T., 1903. The Fauna of British India, Hymenoptera, Taylor and Francis, London, Volume 2, pages 506.

4. BRÜHL, C. A, ELTZ, T. & LINSENMAIR, K. E., 2003. Size does matter - effects of tropical rainforest fragmentation on the leaf litter ant community in Sabah, Malaysia. *Biodiversity. Conserv.*, 12: 13711389.

5. BRÜHL, C. A., MOHAMED, M. & LINSENMAIR, K. E., 1999. Altitudinal Distribution of leaf litter ant along a transect in primary forest on mount Kinabalu, Sabah, Malaysia. *J. Trop. Ecol.*, 16: 265267.

6. COTTRELL, A., 2006. Version 1.1 Gnu general public license department of Economics Wake Forest University. http://WWW.ecn.wfu.edu. /gretl. Accessed on Jan, 2006.

7. DRAPER, N. R. & H. SMITH. 1998. Applied regression analysis. John Wiley and Sons, New York, New York, USA. Using multivariate statistics. HarperCollins College Publishers

8. FISHER, B. L 1998. Ant diversity patterns along an elevational gradient in the Reserve Speciale d' Anjanaharibe- Sud and on the western Masoala Peninsula, *Madagascar. Fieldana Zool.* New Ser. 90: 3967.

9. FORESTS & WILDLIFE DEPARTMENT WORKING PLAN., 2001. North Wayanad Division, Forests and Wildlife Department, Govt. of Kerala, India. 250 pages

10. GRAHAM, M.H., 2003 Confronting multicollinearity in ecological multiple regression. *Ecology*, 84, 28092815.

11. GREENSLADE, P. J., & GREENSLADE, M. P., 1977. Some effects of vegetation cover and disturbance on a tropical ant fauna. *Insect Soc.* 24: 163182.

12. GUJARATI DN. 2003. Basic econometrics. Mc Graw- Hill, Inc.

13. HÖLLDOBLER, B. & WILSON, E. O., 1990. The ants. Belknap Press of Harvard University Press, Cambridge. 752 pages.

14. JACOBS, R. F & MEYER, R. A., 1972. Discovering Geometry. Harcourt Brace Jovanovich, Inc. Press. New York. 704 pages.

15. LEVINGS, S. C. & WINDSOR, D. M., 1985. Litter Arthropod Populations in a Tropical Deciduous Forest: Relationships between Years and Arthropod Groups. *The Journal of Animal Ecology*. 54: 61-69

16. MAYERS, N., MITTERMEIER, R. A., MITTERMEIER, C. G., DA FONSECA, G. A .B. & KENT, J., 2000. Biodiversity hotspots for conservation priorities. *Nature*. 403: 853858.

17. NETER, J., KUNTER, M. H., NACHTSHEIM, C. J., & WASSERMAN, W., 1996. Applied linear statistical models. Irwin, Chicago, Illinois, USA.

18. OLSON, D. M., 1994. The distribution of leaf litter invertebrates along a Neotropical altitudinal gradient. *J. Trop. Ecol.* 10: 129150.

19. VINEESH, P. J, SABU, K. THOMAS. & KARMALY, K. A., 2007: Community structure and functional group classification of litter ants in the montane evergreen and deciduous forests of wayanad region of western ghats, southern india : *Oriental insects*, vol. 41: 185200, (in press).

20. ZAR J. H, 2003. Biostatistical Analysis Fourth Edition. Pearson Education Press, Singapore. 663 pages.