Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity

Jayanti Ray-Mukherjee[1]*, Kim Nimon[2], Shomen Mukherjee[1], Douglas W. Morris[3], Rob Slotow[1] and Michelle Hamer[1,4]

[1]*The School of Life Sciences, University of KwaZulu-Natal, Westville Campus, Private Bag: X54001, Durban 4000, South Africa;* [2]*Learning Technologies and College of Information, University of North Texas, 3940 N. Elm, Rm G150 Denton, TX, 76207, USA;* [3]*Department of Biology, Lakehead University, Thunder Bay, ON, P7B 5E1, Canada; and* [4]*South African National Biodiversity Institute, Private Bag X101, Pretoria 0001, South Africa*

Summary

1. In the face of natural complexities and multicollinearity, model selection and predictions using multiple regression may be ambiguous and risky. Confounding effects of predictors often cloud researchers' assessment and interpretation of the single best 'magic model'. The shortcomings of stepwise regression have been extensively described in statistical literature, yet it is still widely used in ecological literature. Similarly, hierarchical regression which is thought to be an improvement of the stepwise procedure, fails to address multicollinearity.

2. We propose that regression commonality analysis (CA), a technique more commonly used in psychology and education research will be helpful in interpreting the typical multiple regression analyses conducted on ecological data.

3. CA decomposes the variance of R^2 into *unique* and *common* (or shared) variance (or effects) of predictors, and hence, it can significantly improve exploratory capabilities in studies where multiple regressions are widely used, particularly when predictors are correlated. CA can explicitly identify the magnitude and location of multicollinearity and suppression in a regression model.

In this paper, using a simulated (from a correlation matrix) and an empirical dataset (human habitat selection, migration of Canadians across cities), we demonstrate how CA can be used with correlated predictors in multiple regression to improve our understanding and interpretation of data. We strongly encourage the use of CA in ecological research as a follow-on analysis from multiple regressions.

Key-words: stepwise regression, hierarchical regression, structure coefficients, standardized partial regression coefficient, suppressor variable, habitat selection

Introduction

Multiple linear regression (MR) is widely used to identify models that capture the essence of ecological systems (Whittingham *et al.* 2006). MR extends simple linear regression to model the relationship between a dependent variable (Y) and more than one independent (also known as 'predictor' and so-called henceforth) variables (Sokal & Rohlf 1995). Owing to the complexity of ecological systems, multicollinearity among predictor variables frequently poses serious problems for researchers using MR (Graham 2003).

Quite often, ecologists pose the question, to what extent is variation in each predictor variable associated (linearly) with variation in the dependent variable? To answer this, in MR, there are three main effects that need to be assessed: (i) *total effects* – total contribution of each predictor variable to the regression when the variance of other predictors are accounted for; (ii) *direct effects* –contribution of a predictor, independent of all other predictors; and (iii) *partial effects* –contribution of a predictor when accounting for variance of a specific subset or subsets of remaining predictors (LeBreton, Ployhart & Ladd 2004). However, conventionally, one reports the coefficient of determination (R^2) and regression coefficients, where the P-values of the regression coefficients are the information one mostly relies upon to answer the fairy-tale question, 'Mirror, mirror on the wall, what is the best predictor of them all?' (Nathans, Oswald & Nimon 2012; p. 1).

The R^2 quantifies the extent to which the variance of the dependent variable is explained by variance in the predictor

*Correspondence author. E-mail: pollen.jrm@gmail.com

set, and regression coefficients help rank the predictor variables according to their contributions in the regression equation (Nimon *et al.* 2008). However, relying only on the unstandardized slope of regression (also known as partial regression coefficients, Sokal & Rohlf 1995) or standardized (partial) regression coefficients (also known as beta coefficients and so-called henceforth, Sokal & Rohlf 1995) may generate erroneous interpretations, particularly when multicollinearity is involved (Courville & Thompson 2001; Kraha *et al.* 2012). It is not uncommon for researchers to erroneously refer to beta coefficients as measures of relationship between predictors and the dependent variable (Courville & Thompson 2001). Hence, it is often advisable to use multiple methods while interpreting MR results, especially when the intentions are not strictly predictive (Nimon & Reio 2011). In this paper, we use simulated and empirical data to demonstrate how a relatively similar approach, regression commonality analysis (CA), can be used along with beta coefficients and structure coefficients to better interpret MR results in the face of multicollinearity.

Why are stepwise or hierarchical regressions not enough?

Complexities in most systems warrant that ecologists collect data on variables of importance and of interest in concert with data on factors that might affect a response (Whittingham *et al.* 2006). When a researcher is interested in determining predictors' contribution, MR is used to either choose predictor variables based on statistics (stepwise regression) or identify the predictor variables that support a theory (hierarchical regression; Lewis 2007).

Stepwise regressions, which are widely used in ecological studies, provide a methodology to determine a predictor variable's meaningfulness as it is introduced into a regression model (Pedhazur 1997; Graham 2003). Though widely used in fields such as animal behavior and species distribution modelling (Araújo & Guisan 2006; Smith *et al.* 2009), stepwise regression is often discouraged owing to its approach of 'data dredging' and lack of theory (Burnham & Anderson 2002; Whittingham *et al.* 2006). This procedure cause biases and inconsistencies in parameter estimation, model selection algorithms, selection order of different predictors and selection of the single best *magic model* (see Whittingham *et al.* 2006; Zientek & Thompson 2006; Nathans, Oswald & Nimon 2012; Suppl mat.). Stepwise regression relies largely on the first predictor entering the model, which determines the variance of other predictors in the model, posing serious Type I errors associated with inflated *F*-values (Thompson 1995; Zientek & Thompson 2006; Nimon *et al.* 2008; Nathans, Oswald & Nimon 2012). Hence, the use of stepwise regression has been discouraged for assessing the contributions of predictor variables in MR (Nathans, Oswald & Nimon 2012).

Hierarchical regression is more replicable and reliable for evaluating the contribution of predictors and is thus an improvement over stepwise regression (Thompson 1995; Pedhazur 1997; Lewis 2007). Rather than using the familiar stepwise procedure, the choice and order of variables in hierar-

chical regression is based on *a priori* knowledge of theory (Thompson 1995; Pedhazur 1997; Lewis 2007; Nathans, Oswald & Nimon 2012) that help researchers to more effectively choose the best predictor set (Henderson & Velleman 1981; Lewis 2007), for example, by entering distal or lateral variables (generally variables we are less interested in and would want to control for) in early steps and prime or proximal variables later (variables we are truly interested in). However, hierarchical regression ignores the relative importance of certain predictor variables and fails to address multicollinearity (Petrocelli 2003). Furthermore, the misuse of hierarchical regression in one step leads to additional errors in subsequent steps, with compounding as well as errors in interpretation (Cohen & Cohen 1983).

Can beta and structure coefficients address multicollinearity?

In MR, the beta coefficient (b'_Y) of a predictor indicates the expected increase (or decrease) in standard deviation units of the dependent variable, with one standard deviation increase in the predictor holding all other predictors constant (Nimon & Reio 2011; Nathans, Oswald & Nimon 2012). Hence, beta coefficients account for a predictor's total contribution to the regression equation and have a simple relationship with the more conventional partial regression coefficient (b_Y) as follows:

$$b'_Y = b_Y \cdot \frac{S_x}{S_Y}$$

where S_x and S_Y are the standard deviations of the predictor and the dependent variables, respectively (Sokal & Rohlf 1995). When research intentions are strictly predictive, the use of partial regression coefficients and beta coefficients is appropriate. However, relying only on beta coefficients may lead to misinterpretations while informing theory or explaining the predictive powers of a variable of interest (Nimon & Reio 2011; Kraha *et al.* 2012). This is because the accuracy of these regression coefficients depends on a fully and perfectly specified model (Kraha *et al.* 2012), as adding or removing of variables may change these values.

When predictor variables are correlated, beta coefficients may mislead the interpretation of how different predictors influence the dependent variable, because, they are based on a predictor's relation with Y, as well as with all other predictors in the model (Kraha *et al.* 2012). In multicollinear data, risks of misinterpretation are enhanced, because even though a predictor's contribution to the regression equation may be negligible, it might still be highly correlated with the dependent variable Y (Courville & Thompson 2001; Nimon 2010; Nimon & Reio 2011). Still in other cases, a predictor's contribution to the regression equation may be negative even when its relationship to the dependent variable is positive, which some researchers may inappropriately interpret as a negative relationship with the dependent variable (Courville & Thompson 2001). Similar to regression coefficients, confidence intervals (Nimon & Oswald 2013) and standard error can be generated for beta coefficients.

Interpretation of MR results can also be improved with the structure coefficients. Structure coefficients (r_s) are the bivariate (Pearson's) correlations between a predictor and the predicted dependent variable's score resulting from the regression model ($\hat{Y}$; Nathans, Oswald & Nimon 2012). Additionally, the squared structure coefficients (r_s^2) identify how much variance is common between a predictor and $\hat{Y}$:

$$r_s^2 = \frac{r_{xy}^2}{R^2}, \text{ or}$$

$$r_s^2 = r_{x\hat{y}}^2$$

where $r^2_{x.y}$ is the squared bivariate correlation between a given predictor x and y and $r^2_{x\hat{y}}$ is the squared bivariate correlation between a given predictor x and $\hat{Y}$ values (Nimon & Reio 2011). Structure coefficients are thus independent of collinearity among variables and have the additional property of ranking independent variables based on their contribution to the regression effect (Kraha *et al.* 2012). Hence, when predictors are correlated, interpretation of both beta coefficients and structure coefficients should be considered when attempting to understand the essence of the relationships among the variables (Nimon *et al.* 2008; Kraha *et al.* 2012). Structure coefficients should not, however, be confused with partial correlation coefficients, which is a measure of linear dependence of two random variables when the influence of other variables are eliminated, and can actually be obtained by correlating two sets of residuals (Kerlinger & Pedhazur 1973). However, neither of these measures can inform us about both the 'unique' variance (*unique effects* or *direct effects*) of a predictor along with 'common' variance (*common effects* or *partial effects*) that is shared by two or more predictors.

What does commonality analysis do?

Commonality analysis (also called element analysis and/or components analysis) was developed in the 1960s (Newton & Spurrell 1967; Mood 1969) and is frequently used in social sciences, psychology, behavioural sciences and education research (Siebold & McPhee 1979; Nimon *et al.* 2008; Nimon 2010; Kraha *et al.* 2012; Nathans, Oswald & Nimon 2012), yet CA has rarely been used in ecological research (but see Raffel *et al.* 2010; Sorice & Conner 2010). Regression CA improves the ability to understand complex models because it decomposes regression R^2 into its unique and common effects. *Unique* effects indicate how much variance is uniquely accounted for by a single predictor. *Common* effects indicate how much variance is common to a predictor set (Pedhazur 1997; Thompson 2006; Nimon 2010; Nathans, Oswald & Nimon 2012).

To understand the value of CA, consider a hypothetical example adapted from Nimon *et al.* (2008). Imagine that an independent variable 'y' is explained by two predictors 'm' and 'n'. The total variance explained by both of these variables is $R^2_{y.m.n}$. The unique contribution (U) of a variable is the proportion of variance assigned to it when it is entered last in the MR equation. In this case, the *unique* effects of m and n will be

$$U_m = R^2_{y.m.n} - R^2_{y.n},$$

$$U_n = R^2_{y.m.n} - R^2_{y.m}$$

and the *common* contribution (C) will be

$$C_{m.n} = R^2_{y.m.n} - U_m - U_n$$

Substituting the solutions for U_m and U_n

$$C_{m.n} = R^2_{y.m} + R^2_{y.n} - R^2_{y.m.n}$$

Thus, U explains the unique (minimum) explanatory power of a predictor variable, while $U+C$ explains the total (maximum) explanatory power of a predictor. Understanding these effects becomes more interesting when C (the sum of all commonalities associated with a predictor) is substantively larger than U, indicating greater collinearity among variables and making it harder to interpret how the predictor contributes to the regression effect (Nathans, Oswald & Nimon 2012). In a regression model with k predictor variables, CA decomposes the explained variance into 2^k-1 independent effects (Siebold & McPhee 1979). With two predictors, there are three readily interpretable commonalities. However, interpretation becomes more difficult in higher-order models because the number of commonalities expands exponentially with the number of predictors (for 6, 7 and 8 predictors, the number increases to 63, 127 and 255, respectively).

Commonality effects should not be confused with interaction effects (e.g. in regression or ANOVA). Common effects identify how much variance a set of variables has in common with a dependent variable, while interaction effect models the contrasts that exist between different levels or values of at least two independent variables (Supporting information). Any interaction effect thus should be considered as an additional predictor in the regression model. As indicated by Siebold & McPhee (1979); 'the function of commonality analysis is to ferret out these common effects so that they may be interpreted' (p. 365).

Commonalities can be either positive or negative. Negative commonalities can occur in the presence of suppression or when some of the correlations among predictor variables have opposite signs (Pedhazur 1997). A particularly interesting case, likely to emerge in ecological research, is given by variables that suppress or remove irrelevant variance in another predictor and thus increase the predictive ability of that predictor (or a set of predictors) and R^2 by its inclusion in a regression equation (Cohen & Cohen 1983; MacKinnon, Krull & Lockwood 2000; Capraro & Capraro 2001; Zientek & Thompson 2006). Irrelevant variance is the variance shared with another predictor and not with the dependent variable, and hence, it does not directly affect R^2 (Pedhazur 1997). A suppressor (say $X1$) has zero or almost zero (*classic* suppression) or small positive (*negative* suppression) correlation with the dependent variable (Y) but is correlated with one or more predictor variables (say $X2$), generating negative regression weights in the equation (Pedhazur 1997; Thompson 2006; Beckstead 2012). When a predictor and a suppres-

Kraha *et al.* 2012), indicating that 34·9% of the variation in the dependent variable is accounted for by the three independent predictor variables. Please note that, although the predictor variable $X2$ had zero correlation with Y (Table 1a), its beta coefficient is relatively high in the regression equation. Such a result alerts one to the possible suppression by $X2$ on one or more of the remaining variables and reinforces that beta coefficients are not direct measures of the relationships in a regression unless the predictors are perfectly uncorrelated (Courville & Thompson 2001; Kraha *et al.* 2012).

Recall that the structure coefficients are the bivariate correlation between $\hat{Y}$ and each of the predictors $X1$, $X2$ and $X3$; thus, the squared structure coefficients (r_s^2) represent the proportion of variance in the regression effect explained by each predictor alone irrespective of collinearity with other predictors (Kraha *et al.* 2012). For instance, the r_s^2 of $X1$ (0·715) shows that $X1$ was able to account for 71·5% of the regression effect given by the R^2 (=0·349, Table 1b), which with rounding error yields the corresponding r^2 for $X1$ (0·715 × 0·349 = 0·250). Similarly, examining $X2$ ($r_s^2 = 0·00$) shows that $X2$ shared no variance of $\hat{Y}$, while $X3$ ($r_s^2 = 0·179$) was able to explain 17·9% of the variation in $\hat{Y}$.

Looking at *unique* effects of the predictors (Table 1b) shows that predictor $X1$ uniquely explained 25·2% variation in the dependent variable, followed by 9·4% by $X3$, while $X2$ uniquely explained 5·9% of the variation in Y (Table 1b, Fig. 1). This seems contradictory, as although $X2$'s contribution to Y is zero ($r^2 = 0·00$), it receives credit in the regression equation. This is because $X2$'s *unique* effect ($U = 0·059$) is offset by its *common* effect ($C = -0·059$; Table 1b), identifying $X2$ as a suppressor variable. $X2$ thus helps other predictors to better predict the regression, although it itself is unrelated to the equation.

Table 2 decomposes the regression effect into first-order (variance unique to each X), second-order effects (variance common to pairs of variables), and third-order effects (variance common to $X1$, $X2$ and $X3$), and it also includes the per-
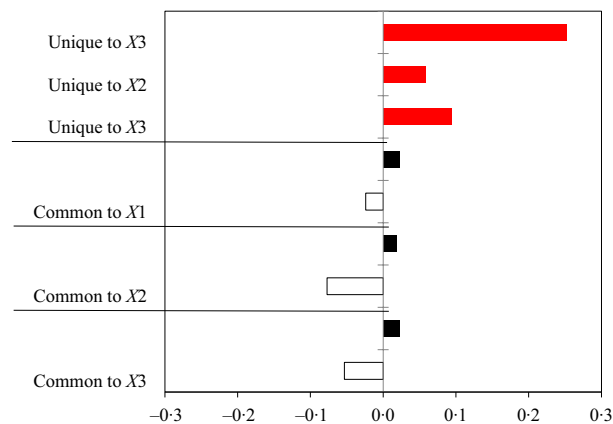
Table 2. Commonality analysis output representing commonality coefficients, both unique (also presented in Table 1a) and common, along with % total contribution of each predictor variable or sets of predictor variables to the regression effect

	$X1$	$X2$	$X3$	% Total
Unique to $X1$	0·252			72·06
Unique to $X2$		0·059		16·90
Unique to $X3$			0·094	26·82
Common to $X1$, $X2$	−0·024	−0·024		−6·84
Common to $X1$, $X3$	0·004		0·004	1·12
Common to $X2$, $X3$		−0·053	−0·053	−15·26
Common to $X1$, $X2$, and $X3$	0·018	0·018	0·018	5·20
Total				100·00

centage of contributions (% Total) of these coefficients to the regression effect.

If we sum the negative contributions to the total regression effect in Table 2, we find that 22·1% of the regression effect is caused by suppression. Table 2 also reveals that each of the negative commonality coefficients involves $X2$ (Fig. 1). Although the suppression effect of $X2$ on $X1$ is lower than that with $X3$, collectively they add to a moderate amount of variance. In this simulated dataset, the remaining second-order common effects account only for 1·12 % of the regression effect, whereas 5·2% of the regression effect was uniquely common to $X1$, $X2$ and $X3$.

EXAMPLE 2: AN ECOLOGICAL EXAMPLE

Morris and Mukherjee (2006) used MR as part of their assessment of the role that habitat selection plays in metapopulation dynamics. In particular, they used MR to evaluate whether the *proportion of immigrants* (*PropIm or Y*) moving into a city depended on the median household income (*Income*) of the city, the number of people employed in the city (*Employ*), the proportion of people emigrating (*PropEmm*) out of the city, the weighted distance of immigration (*DistIm*), and the weighted distance of emigration (*DistEm*) (Appendix S1, Supporting information). Although a stepwise regression may have seemed appropriate to address Morris & Mukherjee's (2006) intent to identify whether humans choose habitats in a way consistent with theory, a more careful assessment of each variable's contribution to the model is necessary if the intent is to understand how the different variables interact to influence human dispersal. Such an understanding could be crucially important in developing economic and social policies for humans. Similar understanding of dispersal by other animals may be important and necessary to develop viable conservation strategies. For more details on this study and its conceptual theory, see Morris and Mukherjee (2006).

Table 3a represents the pairwise Pearson's correlation coefficients of the predictor variables and the dependent variable. Among these, it is notable that the dependent variable, *PropIm,* has the highest correlation with *Income* ($r = 0·58$), while its correlation with *DistEm* is close to zero ($r = -0·01$).

Fig. 1. *Unique* (red bars) and *common* regression effects of three predictor variables ($X1$, $X2$, and $X3$), on the dependent variable (Y). Note the presence of both positive (black bars) and negative (empty bars) commonality coefficient values for each predictor variable.

Table 3. Human habitat selection represented by immigration of Canadians across Canadian cities was tested using commonality analysis. (a) Pearson's correlation coefficients of the predictor variables with *Y*, which in this case is proportion of immigrants (or *PropIm*). (b) In addition to typical regression results, which includes multiple R^2, adjusted R^2 (R^2_{adj}), beta coefficients[1] (β), standard error or beta coefficients (SE), 95.0% confidence interval (lower) for β (CIL), 95.0% confidence interval (upper) for β (CIU), structure coefficients[1] (r_s), squared structure coefficient (r_s^2), each predictor's total unique (U), total common (C) and total variance (r^2) in the regression equation

	Income	*Employ*	*PropEmm*	*DistIm*	*DistEm*
(a)					
Y	0·58	−0·20	0·20	−0·29	−0·01
Income		0·31	−0·32	0·05	0·07
Employ			−0·62	0·05	0·19
PropEmm				−0·08	−0·30
DistIm					0·42

Predictors (*x*)	Multiple R^2	R^2_{adj}	β	SE	CIL	CIU	r_s	r_s^2	U	C	r^2
(b)											
	0·675***	0·594									
Income			0·766***	0·136	0·482	1·050	0·704	0·495	0·513	−0·179	0·334
Employ			−0·272	0·165	−0·616	0·073	−0·245	0·060	0·044	−0·004	0·040
PropEmm			0·318†	0·170	−0·380	0·673	0·240	0·058	0·056	−0·017	0·039
DistIm			−0·397*	0·141	−0·691	−0·104	−0·356	0·127	0·130	−0·044	0·086
DistEm			0·252	0·147	−0·055	0·558	−0·007	0·000	0·048	−0·048	0·000

[1]The 'regr' function in yhat also computes beta coefficients and structure coefficients.
Significance codes: $0 < {***}\, 0.001 < {**}\, 0.01 < {*}\, 0.05 < {\dagger}\, 0.1$.

Table 3b provides an overview of the human migration model, where in addition to multiple R^2, beta coefficients, structure coefficients and squared structure coefficients are provided to help interpret the regression effect, along with the total variance explained by each predictor (r^2), partitioned into its 'unique' (U) and total 'common' (C) effects.

The predictor variables in the regression explained 67·5% ($R^2 = 0.675$, $R^2_{adj.} = 0.594$, $P < 0.001$) of the variance in *PropIm* (Table 3b). The standardized regression equation from the beta coefficients (Table 3b) yields the prediction equation:

$$\hat{Y}(PropIm') = (0\cdot766 * Income) + (-0\cdot272 * Employ) + \\ (0\cdot318 * PropEmm) + (-0\cdot397 * DistIm) + \\ (0\cdot252 * DistEm).$$

We note that although *DistEm* had negligible correlation with *PropIm* (Table 3a), yet its beta coefficient was relatively high in the regression equation. As well, the zero-order correlation between *DistEm* and *PropIm* was negative (Table 3a), while the beta coefficient of *DistEm* was high and positive (Table 3b). The combined pattern indicated the presence of multicollinearity among variables as well as suppression caused by *DistEm*.

Examining the unique variance, we find that, similar to Morris and Mukherjee (2006), *Income* was indeed the best unique predictor of *PropIm* (Table 3b, Fig. 2), uniquely explaining 51·3% of variation in Y. *DistIm* (i.e. how far people are moving in from) was the second best unique predictor, uniquely explaining 13·0% of variance in Y (Table 3b, Fig. 2). Although *PropEmm* uniquely explained 5·6% of variance in Y, its contribution to the regression equation was weak (Table 3b, Fig. 2).

If we now decompose each predictor's variance with the dependent variable into unique and the sum of common effects, we find that each predictor's total *common* effect was
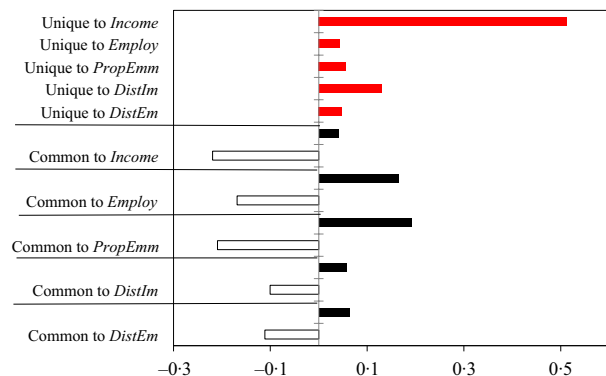
Fig. 2. *Unique* (red bars) and *common* regression effects of *Income*, *Employ*, *PropEmm*, *DistIm*, and *DistEm* on proportion of immigrants. Note the presence of both positive (black bars) and negative (empty bars) commonality coefficient values for each predictor variable.

negative (Table 3b, Fig. 2). Table 4 displays the decomposition of the regression effect for this data set. Among the 26 second-, third- and fourth-order commonalities, 15 were negative and contributed to nearly half of the regression effect. A substantial amount of the regression effect (25·0%) involved negative commonalities associated with *Income*, *Employ* and/ or *PropEmm* (Table 4, Fig. 2). Furthermore, the majority of the negative commonality coefficients stemmed from the high correlation between *Employ* and *PropEmm* (Table 3a), which suggested that in cities with high employment, emigration is reduced. This correlation was in conjunction with the following conditions: First, *Income* increased with *Employ* ($r = 0.31$), while it decreased with *PropEmm* ($r = -0.32$, Table 3a). This yielded a substantial negative commonality coefficient among the three predictors (−.0925, Table 4). Secondly, *Employ* increased with *DistEm* ($r = 0.19$), while *PropEmm* decreased with *DistEm* ($r = -0.30$; Table 3a). Thirdly, *Employ* increased

Table 4. Contribution of predictors and predictor sets on proportion of Canadians immigrating into cities to address human habitat selection. Table shows commonality coefficients (unique and common) and % total contribution of each predictor or predictor sets to the regression effect

	Commonality coefficients	% Total
Unique to *Income*	0·5131	75·97
Unique to *Employ*	0·0439	6·50
Unique to *PropEmm*	0·0564	8·36
Unique to *DistIm*	0·1295	19·17
Unique to *DistEm*	0·0476	7·04
Common to *Income*, and *Employ*	−0·0334	−4·94
Common to *Income*, and *PropEmm*	−0·0433	−6·41
Common to *Employ*, and *PropEmm*	0·1286	19·05
Common to *Income*, and *DistIm*	−0·0193	−2·86
Common to *Employ*, and *DistIm*	−0·0021	−0·32
Common to *PropEmm*, and *DistIm*	−0·0070	−1·03
Common to *Income*, and *DistEm*	−0·0116	−1·72
Common to *Employ*, and *DistEm*	−0·0014	−0·21
Common to *PropEmm*, and *DistEm*	−0·0193	−2·86
Common to *DistIm*, and *DistEm*	−0·0420	−6·22
Common to *Income, Employ,* and *PropEmm*	−0·0925	−13·69
Common to *Income, Employ,* and *DistIm*	0·0016	0·23
Common to *Income, PropEmm,* and *DistIm*	0·0044	0·64
Common to *Employ, PropEmm,* and *DistIm*	−0·0108	−1·59
Common to *Income, Employ,* and *DistEm*	0·0010	0·15
Common to *Income, PropEmm,* and *DistEm*	0·0116	1·72
Common to *Employ, PropEmm,* and *DistEm*	−0·0178	−2·64
Common to *Income, DistIm,* and *DistEm*	0·0092	1·37
Common to *Employ, DistIm,* and *DistEm*	0·0011	0·16
Common to *PropEmm, DistIm,* and *DistEm*	0·0143	2·12
Common to *Income, Employ, PropEmm,* and *DistIm*	0·0061	0·91
Common to *Income, Employ, PropEmm,* and *DistEm*	0·0067	1·00
Common to *Income, Employ, DistIm,* and *DistEm*	−0·0008	−0·12
Common to *Income, PropEmm, DistIm,* and *DistEm*	−0·0086	−1·28
Common to *Employ, PropEmm, DistIm,* and *DistEm*	0·0200	2·95
Common to *Income, Employ, PropEmm, DistIm,* and *DistEm*	−0·0099	−1·46
Total	0·6753	100·00

with *DistIm* (*r* = 0·05), while *PropEmm* decreased with *DistIm* (*r* = −0·08; Table 3a). A substantial amount (16·5%) of the regression effect involving *DistEm* generated negative regression effect (Table 4). *DistEm* thus satisfies the classic definition of suppression, having virtually no shared variance with the dependent variable (Table 3b. Fig. 2), yet making a noteworthy contribution to the regression equation. *DistEm*'s unique variance was offset by its common variance (Table 3b, Fig. 2),

suggesting that *DistEm*'s major contribution to the regression equation was to suppress variance in the remaining predictors (most notably *DistIm*) that was irrelevant to predict variance in *PropIm*.

Our analysis confirms Morris and Mukherjee's (2006) conclusion that immigration (*PropIm*) is mostly driven by *Income*. But we also learned that immigration is driven by the distance from which Canadians are immigrating (*DistIm*) and, to a lesser degree, by *DistEm*, *Employ* and *PropEmm*. The substantial negative commonalities associated with, and common to, *Income, Employ* and *PropEmm* reveal that *Income*'s predictive power was increased after controlling for *Employ* and *PropEmm* in the model. It thus appears that decisions by Canadians to permanently move from one city to another are not a simple weighting of expected income and distance from 'current home' (as implied in the MR results from Morris & Mukherjee 2006), but that individuals may also be assessing emigration patterns correlated with employment opportunities. Although Morris and Mukherjee (2006) correctly interpret the data and their model as consistent with habitat-selection theory, application of CA sharpen our insights into the causes of human dispersal and deepen our understanding of underlying behaviours influencing habitat (city) selection. Recognizing that, *DistEm* suppresses other predictor's irrelevant variance provide deeper insights into how humans make the decision to move from one city to another. From here, we can say that humans tend to discount income and other factors influencing dispersal as the distance between cities increased. This insight may explain why cities (and countries) hoping to increase immigration often provide incentives additional to employment opportunities. Such incentives may be unnecessary if immigrants are targeted from shorter distances. To better understand the relationships among these predictors and to further understand the increase in predictor power associated with suppression and inconsistent signs in predictor correlations, path analysis and/or structural equation modelling approach may be warranted.

How CA does compare with other analyses?

To some degree CA competes with other analyses in providing information that researchers can garner. However, there are subtle and useful differences in the information that CA generates compared with other analyses. CA differs from other variance partitioning algorithms (Pratt's measure, relative weights, general dominance weights) in the way in which CA partitions regression effect into $2^k − 1$ effects, as opposed to k partitions for other measures. Although k partitioning is easier to interpret, such analyses do not necessarily indicate how much and where multicollinearity exist.

Commonality analysis is the process of decomposing the regression effect from all-possible-subsets of regression and a set of formulae that depend on the number of predictors. CA is independent of variable order and hence does not suffer from the problems associated with stepwise regression (wrong degrees of freedom, type 1 errors, inflated *F*-value), thus yielding results that are replicable (Pedhazur 1997; Zientek &

Thompson 2006; Nathans, Oswald & Nimon 2012). CA may also subsume some of the information gained through a hierarchical linear regression. For example, in the case of a regression model with three predictors (*X1*, *X2* and *X3*) where *X1* and *X2* were entered in the first block of a hierarchical regression and *X3* was added in the second block, the unique effect for *X1* will be the same as the ΔR^2 between the regression model with only *X1* and *X2* and the full regression model with *X1*, *X2* and *X3*. Here, it should be kept in mind that hierarchical regressions are different from hierarchical linear models.

Hierarchical linear models (also known as multilevel models, mixed models, nested models or random-effects models) are commonly used in biological or ecological data analysis (Nakagawa & Schielzeth 2013). CA can also be applied to these hierarchical linear (mixed) models using the methods described in this paper. However, instead of using R^2 analogues (Luo & Azen 2013; Nakagawa & Schielzeth 2013) due to lack of monotonicity (R^2 increasing with added predictors), fit indices resulting from an all-possible-subsets analysis are used as input to the commonality analysis algorithms described in this paper. For a description of this solution and the statistical software, see Nimon, Henson & Roberts (2013), Nimon, Oswald & Roberts (2013).

Identifying the magnitude and location of multicollinearity using CA may be useful in 'fully conceptualizing and representing the regression dynamics' (Zientek & Thompson 2006, p. 306). Structural equation models are a class of sophisticated multivariate techniques that are largely confirmatory (based on prior knowledge of theory) and test the validity of a model as opposed to identifying the best model (Tomer 2003). Structural equation models explicitly address the underlying causality patterns among variables (Tomer 2003). Although CA is applied in the same way for a given set of dependent and independent variables regardless of one's causal model (Pedhazur 1997), its results may be useful in identifying what causal models should be explored. In the presence of suppression, it may be particularly useful to employ structural equation models to assess the regression model with and without the suppressor effects (Beckstead 2012). CA may also be used as a technique to identify potential general factors (Hale *et al.* 2001); however, such analyses have been criticized and require further research (Schneider 2008).

Limitations, conclusions and future applications

We caution readers to keep in mind that 'methods are not a panacea' and require prudent use and diligent interpretation (Kraha *et al.* 2012, p. 9). In regression analysis, the uniqueness of variables depends on a specific set of predictors under study, and the addition or deletion of predictors may change the uniqueness attributed to some or all of the variables (Kerlinger & Pedhazur 1973). As CA decomposes the regression effect resulting from a MR, it is possible that failure to meet the statistical assumptions of MR may also impact the bias and precision of the resulting commonality coefficients; however, research has yet to be conducted to determine how such violations of the statistical assumptions of MR affect commonality

coefficients. Understanding the complexities of ecological systems often demands collection and analysis of elaborate data sets with confounding factors and increases in the number of predictor variables exponentially increases the complexity of computation and interpretation of these predictors, when considering analyses based on all possible subsets regression (e.g. commonality analysis, dominance analysis). In such cases, it might be advisable to conduct variance partitioning using sets of predictors and interpret the unique and common effects of a set of predictors (generally highly correlated family of predictors), rather than single predictor variables (Zientek & Thompson 2006).

Traditional MR analysis is inefficient when multicollinear variables arise by either true synergistic association, through spurious correlations, or by improperly specified models (Graham 2003; Whittingham *et al.* 2006). CA models multicollinear data explicitly (Kraha *et al.* 2012); nevertheless, difficulties in understanding such models should not be attributed to CA, but to our ignorance of variables (or the model) in question. Hence, CA should be used to identify indicators that are failing badly with respect to the specificity of the model (Mood 1969; Kerlinger & Pedhazur 1973). Thus, in exploratory stages of ecological research, CA may help researchers select variables in a predictive framework (Kerlinger & Pedhazur 1973; Pedhazur 1997), by revealing the confounding relationships of multicollinear predictor variables. Ecologists should nevertheless use and interpret CA with the same caution as any other analysis, because it does not differentiate theoretical relationships from other causal relations (Kerlinger & Pedhazur 1973).

Commonality analysis has rarely been used in ecological studies (but see Raffel *et al.* 2010; Sorice & Conner 2010), and we highly recommend the use of CA when one wishes to gain a better understanding of exploratory data. CA can often yield additional insights into ecological studies that have used (or plan to use) MR. Ecologists often attempt to reduce problems of collinear variation with principal components regression, canonical correspondence analysis, cluster analysis and stepwise regression (Graham 2003; Araújo & Guisan 2006; Whittingham *et al.* 2006). However, interpretations using these techniques have not always been satisfactory (Graham 2003). Regression CA can provide an improved solution to this problem. We suggest that, along with beta coefficients and structure coefficients (which can also be generated using *yhat* package in R), CA should be used, especially when the aim of research is exploratory. Other approaches, such as path analysis and/or structural equation modelling, are more appropriate when the researcher has an prior theoretical understanding of how different variables and combinations of variables, are likely to influence a particular ecological process (Morris, Dupuch & Halliday 2012). The ability to conduct CA for a large number of variables might allow the discussion on how CA might form the 'missing link' between exploratory and theory-driven research frameworks. CA thus deserves strong consideration as a tool in ecological research, where it can improve model selection and predictor contribution in regression analysis.

Acknowledgements

References

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modeling. *Journal of Biogeography*, **33**, 1677–1688.

Beckstead, J.W. (2012) Isolating and examining sources of suppression and multicollinearity in multiple linear regression. *Multivariate Behavioral Research*, **47**, 224–246.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practice Information-Theoretic Approach*, 1st edn. Springer-Verlag, New York.

Capraro, R.M. & Capraro, M.M. (2001) Commonality analysis: understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. *Multiple Linear Regression Viewpoints*, **27**, 16–23.

Cohen, J. & Cohen, P. (1983) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd edn. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Conger, A.J. (1974) A revised definition for suppressor variables: a guide to their identification and interpretation. *Educational and Psychological Measurement*, **34**, 35–46.

Courville, T. & Thompson, B. (2001) Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, **61**, 229–248.

Graham, M.H. (2003) Multicollinearity in ecological multiple regression. *Ecology*, **84**, 2809–2815.

Hale, J.B., Fiorello, C.A., Kavanagh, J.A., Hoeppner, J.B. & Gaither, R.A. (2001) WISC-III predictors of academic achievement for children with learning disabilities: are global and factor scores comparable? *School Psychology Quarterly*, **16**, 31–55.

Henderson, H.V. & Velleman, P.F. (1981) Building multiple regression models interactively. *Biometrics*, **37**, 391–411.

Kerlinger, F.N. & Pedhazur, E.J. (1973) *Multiple Regression in Behavioral Research*, 1st edn. Holt, Rinehart and Winston Inc., New York.

Kraha, A., Turner, H., Nimon, K., Zientek, L.R. & Henson, R.K. (2012) Interpreting multiple regression in the face of multicollinearity. *Frontiers in Psychology*, **3**, 1–10.

LeBreton, J.M., Ployhart, R.E. & Ladd, R.T. (2004) A Monte Carlo comparison of relative importance methodologies. *Organizational Research Methods*, **7**, 258–282.

Lewis, M. (2007) Stepwise versus hierarchical regression: pros and cons. Paper presented at the Annual Meeting of the Southwest Educational Research Association, San Antonio, TX.

Luo, W. & Azen, R. (2013) Determining predictor importance in hierarchical linear models using dominance analysis. *Journal of Educational and Behavioral Statistics*, **38**, 3–31.

MacKinnon, D.P., Krull, J.L. & Lockwood, C.M. (2000) Equivalence of the mediation, confounding, and suppression effect. *Prevention Research*, **4**, 173–185.

Mood, A.M. (1969) Macro-analysis of the American educational system. *Operations Research*, **17**, 770–784.

Morris, D.W., Dupuch, A. & Halliday, W.D. (2012) Climate-induced habitat selection predicts future evolutionary strategies of lemmings. *Evolutionary Ecology Research*, **14**, 689–705.

Morris, D.W. & Mukherjee, S. (2006) Simulated and human metapopulations created by habitat selection. *Evolutionary Ecology Research*, **8**, 1263–1275.

Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.

Nathans, L., Oswald, F.L. & Nimon, K. (2012) Interpreting multiple linear regression: a guidebook of variable importance. *Practical Assessment Research & Evaluation*, **17**, 1–19.

Newton, R.G. & Spurrell, D. (1967) A development of multiple regression for the analysis of routine data. *Applied Statistics*, **16**, 51–64.

Nimon, K. (2010) Regression commonality analysis: demonstration of an SPSS solution. *Multiple Linear Regression Viewpoints*, **36**, 10–17.

Nimon, K., Henson, R. & Roberts, K. (2013) Using fit indices to holistically assess forms of predictor importance in multilevel models. Paper presented at the Annual Meeting of the American Psychological Association, Honolulu, HI.

Nimon, K. & Oswald, F.L. (2013) Understanding the results of multiple linear regression: beyond standardized regression coefficients. *Organizational Research Methods*, **16**, 650–674.

Nimon, K., Oswald, F.L. & Roberts, J.K. (2013) Interpreting regression effects. R package version 2.0-0.URL: http://cran.r-project.org/web/packages/yhat/index.html [accessed on 17th September, 2013].

Nimon, K. & Reio, T. (2011) Regression commonality analysis: a technique for quantitative theory building. *Human Resource Development*, **10**, 329–340.

Nimon, K., Lewis, M., Kane, R. & Haynes, R.M. (2008) An R package to compute commonality coefficients in multiple regression case: an introduction to the package and a practical example. *Behavior Research Methods*, **40**, 457–466.

Pedhazur, E.J. (1997) *Multiple Regression in Behavioral Research: Explanation and Prediction*, 3rd edn. Harcourt Brace, Orlando, FL.

Petrocelli, J.V. (2003) Hierarchical multiple regression in counseling research: common problems and possible remedies. *Measurement and Evolution in Counseling and Development*, **36**, 9–22.

R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL: http://www.R-project.org.

Raffel, T.R., Michael, P.J., Sites, E.W. & Rohr, J.R. (2010) What drives chytrid infections in newt populations? Associations with substrate, temperature, and shade. *EcoHealth*, **7**, 526–536.

Schneider, W.J. (2008) Playing statistical Ouija board with commonality analysis: good questions, wrong assumptions. *Applied Neuropsychology*, **15**, 44–53.

Siebold, D.R. & McPhee, R.D. (1979) Commonality analysis: a method for decomposing explained variance in multiple regression analysis. *Human Communication Research*, **5**, 355–363.

Smith, A.C., Koper, N., Francis, C.M. & Fahrig, L. (2009) Confronting collinearity: comparing methods for disentangling the effects of habitat loss and fragmentation. *Landscape Ecology*, **24**, 1271–1285.

Sokal, R.F. & Rohlf, F.J. (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd edn. W.H. Freeman, New York.

Sorice, M.G. & Conner, J.R. (2010) Predicting private landowner intensions to enroll in an incentive program to protect endangered species. *Human Dimensions of Wildlife*, **15**, 77–89.

Thomas, R.D., Hughes, E. & Zumbo, B.D. (1998) On variable importance in linear regression. *Social Indicators Research*, **45**, 253–275.

Thompson, B. (1995) Stepwise regression and stepwise discriminant analysis need not apply here: a guidelines editorial. *Educational and Psychological Measurement*, **55**, 525–534.

Thompson, B. (2006) *Foundations of Behavioral Statistics: An Insight-Based Approach*, 1st edn. The Guilford Press, New York.

Tomer, P. (2003) A short history of structural equation models. *Structural Equation Modeling: Applications in Ecological and Evolutionary Biology* (eds B.H. Pugesek, A. Tomer & A. von Eye), pp. 85–124. Cambridge University Press, New York.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B. & Freckleton, R.P. (2006) Why do we still use stepwise modeling in ecology and behaviour? *Journal of Animal Ecology*, **75**, 1182–1189.

Zientek, L.R. & Thompson, B. (2006) Commonality analysis: partitioning variance to facilitate better understanding of data. *Journal of Early Intervention*, **4**, 299–307.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. R codes for example 1 & 2 on how to perform commonality analysis.

Appendix S2. Supporting materials.

Appendix S3. Human.csv

APPENDIX – 1

R codes for example 1 (simulated dataset)

Step 1: Codes for loading appropriate packages and generate data from given correlation matrix

```
library(yhat)
library(MASS)
library(corpcor)

covm<-c(1.00,0.50, 0.00, 0.25,
        0.50, 1.00, 0.15, 0.10,
        0.00, 0.15, 1.00, 0.60,
        0.25, 0.10, 0.60, 1.00)

covm<-matrix(covm,4,4)
covm<-make.positive.definite(covm)
varlist<-c("Y","X1","X2","X3")
dimnames(covm)<-list(varlist,varlist)
data1<-mvrnorm(n=200,rep(0,4), covm, empirical=TRUE)
data1<-data.frame(data1)
```

Step 2: Codes to generate beta coefficients, structure coefficients, and commonality coefficients

```
lmOut<-lm(Y~X1+X2+X3,data1)
regrOut<-regr(lmOut)
regrOut
```

R codes for example 2 (Human migration and habitat selection)

```
library(yhat)

popcomm <- read.csv("Humans.csv", header = TRUE, sep = ",", quote="\"",
dec=".",fill=TRUE) ## Humans.csv file is available online

lmout<-lm(PropIm ~ Income + Employ + PropEmm + DistIm + DistEm, data=popcomm)

regr(lmout)
```