

REVIEW

Spatial regression techniques for inter-population data: studying the relationships between morphological and environmental variation

S. I. PEREZ*, J. A. F. DINIZ-FILHO†, V. BERNAL* & P. N. GONZALEZ*

*División Antropología, Museo de La Plata, Universidad Nacional de La Plata, La Plata, Argentina

†Departamento de Ecología, ICB, Universidade Federal de Goiás, CP 131, 74001-970, Goiânia, Goiás, Brazil

Keywords:autocorrelation;
evolutionary anthropology;
morphometric techniques;
spatial statistical techniques.**Abstract**

Understanding the importance of environmental dimensions behind the morphological variation among populations has long been a central goal of evolutionary biology. The main objective of this study was to review the spatial regression techniques employed to test the association between morphological and environmental variables. In addition, we show empirically how spatial regression techniques can be used to test the association of cranial form variation among worldwide human populations with a set of ecological variables, taking into account the spatial autocorrelation in data. We suggest that spatial autocorrelation must be studied to explore the spatial structure underlying morphological variation and incorporated in regression models to provide more accurate statistical estimates of the relationships between morphological and ecological variables. Finally, we discuss the statistical properties of these techniques and the underlying reasons for using the spatial approach in population studies.

Introduction

Phenotypic diversification in the intra-specific level results from random and nonrandom factors (Reznick *et al.*, 1997; Hendry & Kinnison, 1999; Carroll *et al.*, 2007). Environmental variation can profoundly affect the phenotypic variation within and among populations – yet the developmental and evolutionary mechanisms behind this correlation are poorly understood (Badyaev, 2005)–, and therefore nonrandom factors such as selection and phenotypic plasticity can be of great importance to account for phenotypic diversity at this taxonomic level (Hendry & Kinnison, 1999; Carroll *et al.*, 2007; Ezard *et al.*, 2009; Perez & Monteiro, 2009). Moreover, it is now widely documented that evolutionary change can occur on ecological timescales. Organisms can undergo adaptive phenotypic evolution over a few generations, leading to a

rapid diversification of populations that are under different environmental conditions (Carroll *et al.*, 2007). Therefore, it is important to consider the importance of the environmental dimensions behind the morphological variation in evolutionary studies of phenotypic diversification among populations (Schluter, 2000; Roseman, 2004; Carroll *et al.*, 2007; Perez & Monteiro, 2009).

A common approach to evaluate the importance of environmental dimensions behind morphological variation is based on testing statistically the association between morphological (e.g. cranial length and body size) and environmental (e.g. climate) variables using a set of natural populations (e.g. Katzmarzyk & Leonard, 1998; Felsenstein, 2002). The main problem with this approach is that geographically mediated gene flow among populations, divergence from a shared population history and/or local environmental conditions can cause close populations to become autocorrelated, i.e. populations that are closer together in geographical space and/or close in phylogeny tend to be more similar to each other than expected by chance alone, for a given phenotypic variable (Barbujani, 1987; Legendre, 1993; Cavalli-Sforza

Correspondence: S. Ivan Perez, División Antropología, Museo de La Plata, Universidad Nacional de La Plata, Paseo del Bosque s/n, La Plata 1900, Argentina. Tel.: +54 221 4215184; fax: +54 221 4257744; e-mail: iperez@fcnym.unlp.edu.ar

et al., 1994; Felsenstein, 2002; Relethford, 2004a; Ives & Zhu, 2006). When the response or dependent variable (e.g. phenotypic data) is modelled as a function of explanatory or independent variables (e.g. environmental data), the existence of autocorrelation perturbs significance tests as well as parameter estimates of the standard statistical techniques, which can lead to a misunderstanding of the relationship between these variables. For example, if a population attained a large body size by climatic factors (e.g. low temperature), the neighbouring populations may have a similar size due to gene flow with the former, even though they are not directly affected by the climate with exactly the same intensity. Therefore, similar size among these populations should not be taken as proof of a response to a local climatic influence (Felsenstein, 2002). In this case, more complex models incorporating the autocorrelation structures based on geography (i.e. spatial regression techniques) and/or phylogenetic relationships (i.e. phylogenetic comparative methods) must be used instead of the standard, well-known regression or correlation techniques (Rohlf, 2001; Garland *et al.*, 2005; Ives & Zhu, 2006; Bini *et al.*, 2009; Freckleton & Jetz, 2009).

The statistical problems generated by autocorrelation in a data set are widely recognized and taken into account in ecological and evolutionary inter-specific studies (Rohlf, 2001; Ives & Zhu, 2006). Moreover, several recent papers review the spatial and phylogenetic statistical techniques used to solve this problem at the inter-specific level (Garland *et al.*, 2005; Dormann *et al.*, 2007; Bini *et al.*, 2009). Conversely, at the intra-specific level the influence of autocorrelation is generally underestimated and the associations between traits and environmental variables are evaluated using standard correlation or regression (Sokal, 1984; Felsenstein, 2002). As a consequence, the main objective of this paper was to review the available spatial regression techniques – which incorporate the autocorrelation structures of data sets based on geography – used to test the association between morphological and environmental variables at the intra-specific level. We argue that any study aimed at evaluating the environmental influence on phenotypic evolution within a species ought to apply an adequate methodology that account for spatial autocorrelation in data. In addition, we empirically illustrate the use of such spatial regression techniques to test the association between cranial form variation among worldwide human populations and a set of environmental variables (i.e. mean annual temperature, average annual rainfall and elevation), using a cranial data set of recent human populations widely employed in biological anthropology (Howells, 1973, 1989). Finally, we discuss the performance of generalized least squares, trend surface, autoregression and spatial eigenvector mapping (SEVM) techniques as well as the conceptual and methodological reasons underlying the use of a spatial approach in population studies.

Spatial and comparative analyses in population biology

Spatial variation among populations is a central research issue in evolutionary biology, particularly within the framework of studies interested in neutral variation (Sokal *et al.*, 1989a; Barbujani, 2000; Relethford, 2008). This is due to the fact that most neutral evolutionary processes occur in a spatial context (Epperson, 2003), where the genetic variation originated by random mutations within local populations will disperse through geographically mediated gene flow. Several approaches can be used to analyse the resulting patterns of spatial variation, that usually involve the estimation of parameters such as the geographical distance at which genetic or phenotypic data can be considered independent (Sokal & Oden, 1978; Barbujani, 2000; Manel *et al.*, 2003).

The magnitude of spatial autocorrelation can be evaluated using autocorrelation coefficients, such as the Moran's I coefficient, which is commonly applied in population studies (Sokal & Oden, 1978; Barbujani, 2000; Diniz-Filho *et al.*, 2009), and given by

$$I = \left(\frac{n}{S} \right) \left[\frac{\sum_i \sum_j (y_i - \bar{y})(y_j - \bar{y}) w_{ij}}{\sum_i (y_i - \bar{y})^2} \right],$$

where n is the number of local populations, y_i and y_j are the values of the biological trait measured in populations i and j , \bar{y} is the average of y , and w_{ij} is an element of a \mathbf{W} or weighting matrix. In this \mathbf{W} matrix, the elements are equal to 1 if the pair i, j of local populations is within a given distance class interval (indicating samples that are 'connected' in this class); otherwise $w_{ij} = 0$. S indicates the number of entries (connections) in the \mathbf{W} matrix. The value expected under the null hypothesis of the absence of autocorrelation is $-1/(n-1)$. Moran I is usually calculated by using several distance classes, and in this case multiple \mathbf{W} matrices are built by connecting pairs of local populations situated at increasing geographical distances. This sequence of coefficients is plotted against geographical distances, generating a correlogram that describes the complexity of spatial patterns, in the original variable as well as in the residuals (see below; Sokal & Oden, 1978; Legendre & Legendre, 2003). These parameters can be linked to evolutionary processes, such as dispersion (Sokal *et al.*, 1989a). More complex micro-evolutionary inferences can be performed by comparing patterns of geographical variation for different alleles and loci using multiple correlograms (Sokal & Oden, 1978; Sokal & Wartenberg, 1983; Sokal *et al.*, 1989a).

Graphic representations and randomization tests of biological and geographical distances among a set of populations are also employed (Smouse *et al.*, 1986; Hutchison & Templeton, 1999; Relethford, 2004b; Ramachandran *et al.*, 2005). Mantel (1967) introduced a method for deciding whether the matrix of biological

distances correlated with the matrix of geographical distances (see Smouse *et al.*, 1986). The basic Mantel Z -statistic is the sum of cross-products of the values in two matrices:

$$Z_{yx} = \sum_{ij} (X_{ij}Y_{ij}),$$

where X and Y are unfolded distance matrices (i.e. the distance matrices are unfolded column by column to form a long vector, excluding the diagonal term) (Smouse *et al.*, 1986; Legendre & Legendre, 2003). The ordinary product-moment correlation coefficient, r , is monotonically related to Z (Smouse *et al.*, 1986). Several other approaches are also available (see Sokal & Oden, 1978; Peres-Neto & Jackson, 2001; Manel *et al.*, 2003; Relethford, 2008).

Although these approaches are slightly different, their ultimate goal is to describe and explore the spatial structure underlying neutral genetic or phenotypic variation. In population studies of several species, the spatial statistics have shown that many genetic and phenotypic variables are spatially correlated, such that geographically close populations tend to be biologically similar (Barbujani, 1987; Cavalli-Sforza *et al.*, 1994; Hutchison & Templeton, 1999; Relethford, 2004a; Manica *et al.*, 2005). Particularly, two endogenous processes have been used to explain the spatial pattern of variation among populations: it could emerge as the result of gene flow restricted by the geographical distance (i.e. model of isolation by distance) or because of the serial founder effect (Cavalli-Sforza *et al.*, 1994; Relethford, 2004a; Ramachandran *et al.*, 2005; Templeton, 2007). As a result of the spatial structure of populations, gene flow will occur more frequently between nearby populations, leading to high genetic affinities between groups in close geographical proximity and the probable genetic differentiation of more distant groups due to the effect of genetic drift (i.e. the IBD model; Wright, 1943; Barbujani, 1987; Cavalli-Sforza *et al.*, 1994; Hutchison & Templeton, 1999; Relethford, 2004a). On the other hand, the increase in the biological distance with geographical distance could be the result of the colonization of an area through multiple and successive dispersion events of groups that have a small number of individuals, a process known as expansion of range (Slatkin, 1993). This expansion of range leads to several events of random sampling – serial founder events, resulting in a gradient of reduction in biological diversity within populations in the direction that the groups are moving away from the centre of expansion, unless rates of migration are extremely high (Ramachandran *et al.*, 2005; Ray *et al.*, 2005; but see Templeton, 2007).

However, when we study the effects of environmental variables over morphology, we should use other approaches that incorporate the spatial autocorrelation of morphological and/or environmental variables directly into the statistical model (Sokal, 1984; Legendre, 1993;

Diniz-Filho *et al.*, 2003, 2009; Dormann *et al.*, 2007). Generally, population studies use the partial Mantel's matrix correlation statistic (Smouse *et al.*, 1986) to remove the effects of spatial and/or phylogenetic variation in the relationship between two sets of data (e.g. Relethford, 2004b; Roseman, 2004). However, partial Mantel's matrix correlation is just a linear correction that removes all morphological variation correlated with space (Oden & Sokal, 1992). Therefore, it does not correspond to what spatial regression techniques (e.g. generalized least squares) do because they correct for the effect of spatial similarity among neighbour populations, i.e. they model local-scale autocorrelations in residuals of the regression model (Dormann *et al.*, 2007; Perez *et al.*, 2009; see below).

Other techniques that directly emerge from the overall linear modelling framework – i.e. linear regression techniques – could be used to test whether a morphological variable is associated with environmental variation, in order to account for spatial structures in data (Dormann *et al.*, 2007; Bini *et al.*, 2009; Diniz-Filho *et al.*, 2009). In the following section we describe generalized least squares, trend surface, autoregression and SEVM techniques.

Spatial regression models

Conventional statistical analysis assumes the independence of all observations (independence entails that no observation in a sample can be predicted by another observation in the same sample and that the best predictor of any observation is the mean; Sokal & Rohlf, 1986; Zar, 1999), frequently overestimating the number of independent observations in spatial studies (Legendre, 1993; Peres-Neto, 2006). Overestimating the number of independent observations could lead to incorrectly refute the null hypothesis of nonassociation between morphological and environmental variables (H_0), i.e. inflating type I error rates. Consequently, in this section we illustrate a set of available techniques that can be used to take into account the problem of nonindependence, or autocorrelation, in the study of morphological variation among populations.

The problem of estimating the level of relationship between morphological and environmental variables has the general structure of a regression model (the ordinary least squares model, OLS; Table 1), where the dependent – or morphological – variable is modelled as a function of the independent – or environmental – variable (Sokal & Rohlf, 1986; Zar, 1999). In this model the error term, or residuals, must be normally distributed, with constant variance and independently distributed among observations, i.e. the covariance matrix among residuals is the identity matrix. In biological studies the residuals are generally independent when the populations are not correlated by geography and/or phylogeny.

When autocorrelation in residuals is detected (e.g. by using autocorrelation analysis such as Moran's I

Table 1 Regression models most frequently used in spatial ecological analysis.

Model	General approach	Formula
Ordinary least squares (OLS)		$y = Xb + \varepsilon$, where y is the vector that describes trait variation, X is the matrix of independent variables, b is the vector of regression coefficients, ε is the error term, and the covariance matrix C among residuals is $C = \sigma^2 I$, where σ^2 is the variance of the residuals, and I is an identity matrix
Simultaneous spatial autoregressive (SAR)	Model residuals	$y = Xb + \varepsilon$ and $C_{\text{SAR}} = \sigma^2 [(I - \rho W)^T]^{-1} [I - \rho W]^{-1}$, where W is the weighting matrix and ρ is an autoregressive coefficient for response variable
Conditional spatial autoregressive (CAR)	Model residuals	$y = Xb + \varepsilon$ and $C_{\text{SAR}} = [(\sigma^2 W_{i+})] [I - \rho W]^{-1}$
Moving average (MA)	Model residuals	$y = Xb + \varepsilon$ and $C_{\text{MA}} = \sigma^2 [(I + \rho W)(I - \rho W)]$
Trend surface analysis (TSA)	Model structure	$y = Xb + G + \varepsilon$, where $G = LB_L$, where L is a matrix with the spatial coordinates of local populations and B_L are the slopes of these coordinates
Lagged-response autoregressive (ARM-response)	Model structure	$y = Xb + G + \varepsilon$ where $G = \rho W y$
Lagged-predictor or mixed autoregressive (ARM-mixed)	Model structure	$y = Xb + G + \varepsilon$ where $G = \rho W y + \gamma W X$, where γ is the autoregressive coefficients for each explanatory variable
Spatial eigenvector mapping (SEVM)	Model structure	$y = Xb + G + \varepsilon$ where $G = PC$, where PC are the principal coordinates

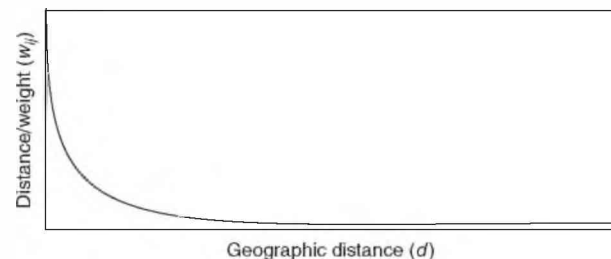
The problem of estimating the level of relationship between morphological and ecological variables has the general structure of a regression model. Here, we show the different regression models most frequently used in spatial ecological analysis: ordinary least squares, regression techniques that incorporate autocorrelation into residuals (model residuals) and regression techniques that incorporate autocorrelation into the structure of the regression model (model structure). All spatial analyses described in this paper can be performed using the SAM software (spatial analysis in macroecology) version 3.1 (Rangel *et al.*, 2006), which is freely available at <http://www.ecoevol.ufg.br/sam>. In addition, the spatial and phylogenetic regression analyses can be made using several R packages (e.g. APE), which are freely available at <http://www.r-project.org/>. Finally, NTSYS 2.2, available at <http://www.exetersoftware.com>, perform many regression techniques that consider the autocorrelation of data.

coefficient), there is a clear violation of the assumptions for the standard regression model. Therefore, the residual variation must be modified in order to improve our understanding of morphological variation, as well as to achieve a better parameter's estimation and to test the statistical model. In this scenario, *spatial regression models* have been proposed to solve this problem. These models can be grouped into two classes (Table 1) based on the idea of incorporating autocorrelation either into the residuals of the regression model (model residual approach) correcting their covariance matrix, or into the structure of the model (model structure approach) including a new term (Diniz-Filho *et al.*, 2003, 2007, 2009; Legendre & Legendre, 2003; Dormann *et al.*, 2007; Kissling & Carl, 2008; Bini *et al.*, 2009).

In the model residual approach, known as *generalized least squares model*, the error structure in covariance matrix among residuals is designed to incorporate the expected lack of independence of the observations due to the spatial distribution of the populations. In this model the covariance matrix among residuals is based on the W matrix, 'expected relationship matrix' or weighting matrix, which contains the correlation structure among the populations. The elements of W can be estimated by different and complex inverse functions of geographical distance (d_{ij}) between populations, given by inverse distance-powered functions of the form $w_{ij} = 1/d_{ij}^\alpha$, where α is the parameter that regulates the model. With $\alpha = 1$ this formula generates a large decline in distance, with a geographical distance between 0 and a given

distance, and shows a plateau with little change in distance after this value (Fig. 1), such as it was shown for biological distance among populations (Relethford, 2004a). Several techniques, such as SEVM (see below), truncate the W matrix in a specific distance, being equal to 0 the distances greater than such specific distance. This procedure gives greater importance to small geographical distances. There are several generalized least squares techniques that can be found in the literature related to spatial analyses (Wall, 2004; Rangel *et al.*, 2006; Dormann *et al.*, 2007; Diniz-Filho *et al.*, 2009) and they are named after the different ways of defining the covariance matrix among residuals (simultaneous spatial autoregressive, conditional spatial autoregressive and moving average; Table 1).

Instead of modifying the error term, the model structure approach introduces new explanatory variables in

**Fig. 1** Plot of geographical distance (d) vs. distance/weight (w_{ij}).

the model that 'capture' the spatial variation, thereby minimizing the autocorrelation in the residuals. There are several ways of incorporating spatial variables into the model structure to eliminate or minimize residual autocorrelation (Table 1). The simplest way of defining space is by using the spatial coordinates of local populations (i.e. latitude and longitude), which can be added as spatial independent variables in the model. This technique is known as trend surface analysis (TSA1; Legendre & Legendre, 2003), and is better suited to model broad-scale trends and not local autocorrelation in residuals. The simplest equation expresses part of the morphological variation as a plane in geographical space. The spatial component in this equation can be changed by adding polynomial expansions, thereby adjusting to quadratic (TSA2) or cubic functions of spatial coordinates. Another way to take into account spatial patterns into the model structure is by using an autoregressive model.

There are several forms used to express autoregressive models, but the main idea is the pure autoregression model (Diniz-Filho *et al.*, 2009), which estimates the variation in a trait that can be explained by space. In spatial analysis it is possible to incorporate autoregressive terms for the response variable (lagged-response autoregressive model) and for both, response and explanatory variables (lagged-predictor or mixed autoregressive model) (Table 1).

Finally, another approach to incorporate space into models structure is to extract principal coordinates (i.e. eigenvectors) from the weighting matrix – i.e. the matrix expressing the spatial relationship among local populations – and to use part of these vectors to establish the regression model (Table 1). This approach is called SEVM (Griffith, 2003; Griffith & Peres-Neto, 2006). The basic difference between the various applications of this approach lies on the principal coordinates that are extracted to represent geographical space. The principal coordinates of a spatial matrix express the relationships among local populations at decreasing spatial scales, so that first principal coordinates with large eigenvalues tend to express broad-scale structures, whereas principal coordinates with small eigenvalues tend to express local patterns. Thus, the advantage of eigenvector mapping is its flexibility in dealing with patterns at multiple scales, and the possibility of iteratively improving the modelling process by adding or removing these principal components (PCs) (Diniz-Filho & Bini, 2005; Griffith & Peres-Neto, 2006).

An example of spatial regression techniques in human population analyses

Understanding the importance of nonrandom factors and environmental dimensions in the origin of the worldwide pattern of morphological variation among human populations has long been a central goal of evolutionary

anthropology (Roberts, 1953; Howells, 1973, 1989; Beals *et al.*, 1984; Relethford, 1994, 2004a; Ruff, 1994; Katzmarzyk & Leonard, 1998; Roseman, 2004; Harvati & Weaver, 2006). Craniofacial form and shape variation has been widely investigated across modern human populations (Beals *et al.*, 1984; Relethford, 1994, 2004a; Roseman, 2004; Harvati & Weaver, 2006). These studies point out that cranial shape variation is mainly influenced by neutral evolutionary processes, such as mutation, gene flow and genetic drift (Relethford, 1994, 2004a). Conversely, variation in craniofacial size and form (i.e. shape plus size) has been related to nonrandom factors, like natural selection (Beals *et al.*, 1984; Roseman, 2004; Harvati & Weaver, 2006). Specifically, several works pointed out that temperature could be the principal environmental dimension shaping the worldwide pattern of form and size variation among populations. However, some investigators suggested the possibility that the observed association between craniofacial form and temperature could be due to a spurious correlation of each with the neutral patterns of inter-regional difference generated by spatial structure of the populations (i.e. autocorrelation; Sokal, 1984; Relethford, 1994). Here, we employ spatial regression techniques in order to establish whether craniofacial form is significantly associated with climatic variables (i.e. mean annual temperature, average annual rainfall and elevation), independent of the spatial structure. The existence of a significant correlation between these variables could be used to support the importance of nonrandom factors, such as natural selection, driving the morphological divergence among human populations (e.g. Roseman, 2004; Harvati & Weaver, 2006; Perez & Monteiro, 2009).

We analysed 45 linear cranial measurements collected from a sample of 1367 male individuals from 30 populations distributed worldwide (Fig. 2; Howells, 1973, 1989). All the samples belong to recent modern human populations that inhabited different geographical and ecological regions around the world (Howells, 1989); distributed from 70°N latitude to 45°S latitude, and from 30 to –8 °C of mean annual temperature (Fig. 2). The geographical locations of the samples (local populations) were reported by Howells (1989). The geographical coordinates of each local population were transformed to a geodesic system and used to compute a matrix of great circle geographical distances between them. The mean annual temperature, average annual rainfall and elevation at each local population were obtained and used as estimators of climate variation across the globe (Beals *et al.*, 1984; Katzmarzyk & Leonard, 1998; Harvati & Weaver, 2006). These variables were obtained for each of the 30 populations (i.e. geographical localization or close to) using Internet climatic databases (i.e. <http://www.worldclimate.com>; Relethford, 2004b) and geographical maps.

Rather than performing a separate analysis on each of the 45 craniometric variables, we used the original

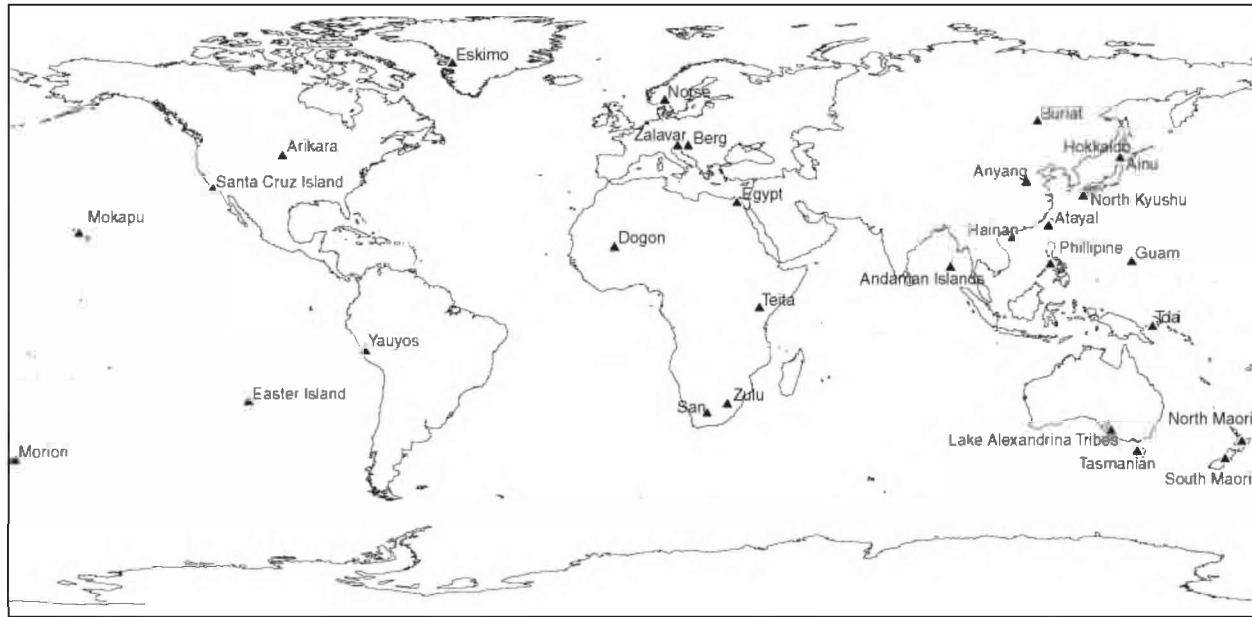


Fig. 2 Geographical location of the 30 samples used in this study.

variables to perform a PC analysis of a covariance matrix using mean values; and the resulting first PC score was used as the general form vector. The calculation of PC score generates a data reduction and avoids redundancy (Marcus, 1990; Thalib *et al.*, 1999). This first PC score, accounting for 45% of the total among mean samples variation, has strong correlations with the size measurement, the arithmetic mean of all variables ($r = 0.982$). In addition, this procedure is essentially the same as the one used in others works of spatial techniques applied to population morphometrics (e.g. Sokal & Uytterschaut, 1987; Relethford, 2008). Although the other PC scores represent important shape variation among human populations, and because the main objective of this paper was to review the statistics of spatial regression techniques, in the following analyses we restrict the tests to the first PC score to simplify the explanation. Although we used a univariate approach to study variation among human populations, the spatial regression techniques can be generalized to use multivariate multiple regression models (Rohlf, 2001; Perez *et al.*, 2009).

We first generated a spatial correlogram (Sokal & Oden, 1978; Barbuji, 2000) to explore the spatial autocorrelation of form variation. Although there are alternative approaches to describe spatial patterns (e.g. semi-variograms; Relethford, 2008), correlograms have been repeatedly used in previous exploratory autocorrelation analyses of inter-population variation, mainly based on genetic data (e.g. Sokal & Oden, 1978; Sokal *et al.*, 1989b; Barbuji, 2000). Here, Moran's I coefficients were calculated for five geographical distance classes, whose intervals were defined such that each class

contains approximately the same number of connections among local populations. The statistical significance of the autocorrelation coefficients, Moran's I , was calculated with 4999 randomizations (for details, see Legendre & Legendre, 2003).

The spatial correlograms of form variation (i.e. PC1 score) are shown in the Figure 3. These correlograms show a cline in the PC1 score affecting the entire worldwide distribution, starting from about 6000–7000 km (Fig. 3a). Perhaps because of the relatively large and irregular distances among close populations, Moran's I in the first distance class is not very high, as is usually observed for clinal patterns. The cline observed in the PC1 score can be explained by several processes, such as migration from a single direction or one side, gene flow among populations or environmental influence acting in geographically close and similar environments (see Sokal *et al.*, 1989a,b; Legendre & Legendre, 2003). Anyway, the most important issue is that a similar cline is also observed in the residuals of morphometric against climate variation obtained with the OLS techniques (Fig. 3b). Therefore, the residuals of neighbour populations are similar, and that suggests the importance of spatial endogenous processes such as gene flow to explain the PC1 variation. Consequently, evolutionary spatial factors, local environmental conditions or historical factors are important in accounting for craniofacial variation among worldwide populations (Cavalli-Sforza *et al.*, 1994; Eller, 1999; Relethford, 2004a; Manica *et al.*, 2005).

We then regressed the PC1 score against climate (i.e. mean annual temperature, average annual rainfall and

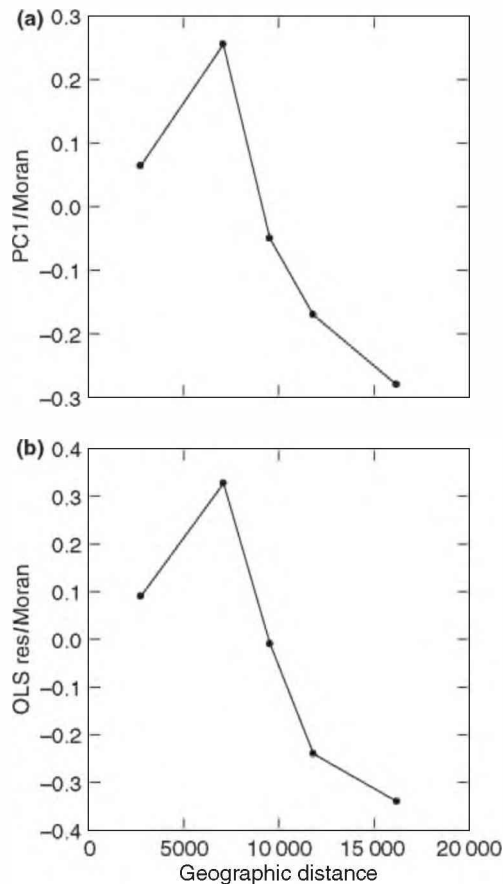


Fig. 3 Autocorrelogram of (a) principal component 1 (PC1; form cranial variation) and (b) OLS residuals.

elevation) using three forms of generalized least squares models based on autoregressive processes (SAR, CAR and MA), first- and second-order trend surface (TSA1 and TSA2), lagged-response and lagged-predictor autoregressive models (ARM-response and ARM-mixed), and SEVM techniques. To define the spatial structures to be used in these spatial regression models, we employed one weighting matrix (\mathbf{W}) estimated assuming an inverse relationship between craniofacial variation and geographical distances among populations (e.g. isolation-by-distance model; Relethford, 2004a). This \mathbf{W} matrix was calculated as the inverse function of great circle geographical distances between populations, $w_{ij} = 1/d_{ij}^1$, generating a large decline in distance with a geographical distance between 0 and 6000 km, and showing a plateau with little change in distance after ca. 8000–10 000 km (see Relethford, 2004a). We estimated the r^2 and the standardized regression slopes of the spatial models and assessed their significance by using the t -statistic (Akaike information criterion could be an alternative measure to r^2 for comparing model fit; Freckleton, 2009).

The success of these techniques for eliminating residual autocorrelation is not always guaranteed, because of

model-fit problems and variation in the robustness of each technique against violations in some of their assumptions. For example, if the \mathbf{W} matrix (i.e. the expected spatial structure) does not capture the true spatial processes underlying genetic variation, then the residual can still possess spatial autocorrelation (Diniz-Filho *et al.*, 2003). Therefore, it is important to use some exploratory autocorrelation coefficient, such as Moran's I , to test whether the assumption of the spatial independence of the residuals of each spatial regression is still being violated or not (see Gittleman & Kot, 1990). For SEVM, the matrix was truncated based on the \mathbf{W} matrix – i.e. the distances greater than 6092 km were equal to 0 – and the selection of the principal coordinates to be used in the model was based on minimizing residual Moran's I (see Griffith & Peres-Neto, 2006). We tested Moran's I in regression residuals at the five geographical distance classes and also computed the Euclidian distances between each residual correlogram and the null expectation, as a measure of the amount of autocorrelation still present in model residuals (so that a better technique will have a relatively small distance between the residual and null correlograms, indicating minimization of the autocorrelation).

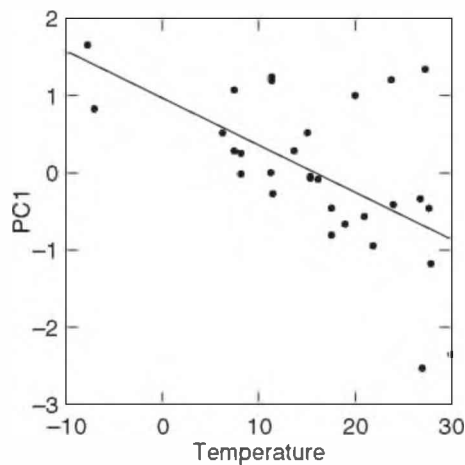
The OLS analysis suggests that climate has a significant effect on patterns of form variation calculated with the first PC for cranial measurements (Table 2). The slope value of temperature is the largest one, followed by elevation and rainfall (although these last two are not statistically significant). The temperature shows a clear negative association with the PC1, with larger crania found in cooler regions, although some populations from Oceania are outliers in this relationship (Fig. 4). This is shown by the correlogram, which detected autocorrelation in residuals, showing a clear violation in the assumptions of a standard OLS (Fig. 3b; Table 2).

Results from spatial regression techniques are reported in Table 2. In general, all techniques show qualitatively the same result, in which the most important variable driving cranial variation is temperature; with partial standardized slopes ranging from -0.549 to -0.642 . In all cases, these coefficients were highly statistically significant ($P < 0.001$, but see below). The regression slopes of model residual approaches (SAR, CAR and MA) are very similar to the OLS results, and the correlograms revealed similar levels of (high) autocorrelation in residuals. Conversely, the model structure approaches, i.e. TSA1, TSA2, ARM-response, ARM-mixed and SEVM, were more effective, on average, in minimizing residual spatial autocorrelation (Table 2). Unlike OLS, these techniques generate residuals with normal distribution.

Our results pointed out that although random factors are important to explain spatial inter-population differentiation in craniofacial characteristics in modern humans (supporting recent studies, e.g. Relethford, 1994, 2004a; Roseman, 2004), there is a significant correlation between craniofacial form and climate independent of spatial

Table 2 Results of the regression analyses performed between PC1 score and the environmental variables.

Regression models	Technique	Slopes					
		R^2	Elevation	Temperature	Rainfall	Moran $I < 0.05$	Distance from H_0
OLS		0.398	-0.231	-0.612*	0.058	3	0.288
Model residuals	SAR	0.434	-0.220	-0.613*	0.060	3	0.258
	CAR	0.468	-0.264	-0.635*	0.061	3	0.251
	MA	0.432	-0.222	-0.614*	0.060	3	0.263
Model structure	TSA1	0.467	-0.265	-0.784*	0.225	1	0.096
	TSA2	0.663	-0.104	-0.613	0.023	0	0.031
	ARM-response	0.313	-0.173	-0.549*	0.033	1	0.089
	ARM-mixed	0.350	-0.241	-0.594*	0.038	1	0.081
	SEVM	0.474	-0.203	-0.642*	0.104	0	0.065

* $P < 0.001$.**Fig. 4** Plot of PC1 vs. mean annual temperature among male samples.

structure. These results also refuted the possibility that the observed correlation between craniofacial form and temperature could be due to a spurious correlation of each with the patterns of inter-regional difference generated by spatial structure. The large-scale pattern of Howells (1989) data set is mainly related to climate (Fig. 4), suggesting the importance of nonrandom factors to explain cranial diversification among human populations.

Performance of spatial regression models

Ordinary least squares technique, which does not incorporate spatial information, makes the tacit assumption that all the populations studied are equally related to each other. In human population analyses there is a large amount of information that suggests the importance of geography in morphological variation, particularly in a worldwide scale (e.g. Relethford, 2004a; Roseman, 2004), and independently of other climate and ecological variation. Therefore, the assumptions of OLS are not achieved by our data set. Nevertheless, under different

circumstances these assumptions might not be completely rejected. For example, if morphological traits evolve very rapidly in response to environmental fluctuations, we would never know the relationships among populations just by looking at the traits under study because spatial autocorrelation is absent. This could be true for some geographical regions with broad ecological variation and recent peopling (see Perez & Monteiro, 2009). Some authors have suggested that spatial statistical techniques, as well as the phylogenetic comparative analysis, should only be used when there is spatial or phylogenetic autocorrelation in the morphological variable (see Garland *et al.*, 2005); however, Rohlf (2006) pointed out that this introduces a conditional test, affecting the type I error.

Our example suggests that model residual approaches cannot adequately incorporate the spatial autocorrelation structure present in data set, using the weighting matrix. This is probably not due to problems with techniques *per se*, but to the difficulty in expressing complex spatial patterns in residuals in the weighting matrix employed by GLS techniques. In addition, these results stress the necessity to assume a more realistic model of spatial structuring (e.g. migration patterns and/or shared evolutionary history) for a better understanding of the relationship between morphological and ecological variation among human populations.

TSA2 and SEVM are the spatial regression techniques that were better capable of incorporating the spatial autocorrelation structures in our example, minimizing residual autocorrelation. However, TSA2 incorporates the geographical coordinates in the model structure and adjusts the quadratic function, with a total of five predictors (latitude and longitude and their quadratic expansions), thereby greatly affecting the statistical power of the regression model (inflating the type II error; Table 2). This technique can be useful to incorporate broad-scale effects, but it is not usually very successful in incorporating local autocorrelation in residuals. In our example, the simultaneous incorporation of geography as a broad-scale quadratic trend, plus the

temperature (another broad-scale effect), generates a loss of statistical power and, consequently, the partial slope for temperature is not statistically significant (the opposite of what was found using every other techniques).

On the other hand, SEVM is the most flexible technique for dealing with patterns at multiple scales, and can add principal coordinates to minimize the spatial autocorrelation using explicitly the minimum Moran's I coefficient (Griffith & Peres-Neto, 2006; Peres-Neto, 2006). The SEVM does not present the same problems as its phylogenetic version (phylogenetic eigenvector method; Diniz-Filho *et al.*, 1998) where the fit of morphological and phylogenetic variation will always be perfect ($r^2 = 1$) and there will be no residual variation left to investigate association with ecological variables when we incorporate more principal coordinates to the regression model (Diniz-Filho *et al.*, 1998; Rohlf, 2001). This is because the phylogenetic eigenvector method uses path length distances (patristic distances) on the tree to define the \mathbf{W} matrix, which have properties very different from that of the Euclidean distance matrices usually used in spatial analyses (Rohlf, 2001). Conversely, in the spatial version of SEVM the distance between points in space has a Euclidean metric and is truncated to account for short distance effects only (Griffith, 2003; Griffith & Peres-Neto, 2006); therefore, the fit of morphological and spatial variation will not always be perfect.

The result of our example agrees with the recent comparative evaluation by Bini *et al.* (2009) and Diniz-Filho *et al.* (2009), in the sense that the performance of spatial regression models is quite idiosyncratic and data dependent. From our analyses, it is evident that model structure approaches (especially SEVM) seem to work better for our data set than those incorporating autocorrelation in model residuals (see also Diniz-Filho *et al.*, 2009), a result which is opposed to those found by Bini *et al.* (2009) when analysing 99 macroecological data sets. This may be due to the strong endogeneous component in our data set (also found in the simulated data set used by Diniz-Filho *et al.*, 2009), whereas, in macroecological data, exogenous components are usually dominant (Hawkins *et al.*, 2007; Bini *et al.*, 2009).

Thus, in general, the results showed here are in agreement with previous studies in suggesting that although model structure regression techniques are useful in our evolutionary and ecological scenario, model residuals could be useful in different ecological scenarios where exogenous components are dominant.

Intra-specific spatial regression models and inter-specific comparative phylogenetic methods

Autocorrelation is common in nature and it mainly occurs along three dimensions: spatial, temporal and phylogenetic variation (Ives & Zhu, 2006; Peres-Neto, 2006). Therefore, the regression techniques have been

generalized to incorporate these different sources of autocorrelation into the residuals or the structure of the regression model, such as in the comparative phylogenetic methods (Cheverud *et al.*, 1985; Grafen, 1989; Martins & Hansen, 1997; Diniz-Filho *et al.*, 1998; Rohlf, 2001; Garland *et al.*, 2005). In comparative phylogenetic methods, the generalized least squares technique was proposed by Grafen (1989) and Martins & Hansen (1997) and is now the current standard comparative tool (Garland *et al.*, 2005; Ives & Zhu, 2006; Rohlf, 2006; Freckleton, 2009). On the other hand, applications of autoregressive methods in phylogenetic comparative analyses, starting with studies by Cheverud *et al.* (1985) and Gittleman & Kot (1990), are based on the pure autoregression model (i.e. $y = \rho W y + \epsilon$). Finally, SEVM method is called eigenvector method (PVR; Diniz-Filho *et al.*, 1998) in phylogenetic comparative analysis, and it employs principal coordinates or eigenvectors from a phylogenetic distance matrix or from the weighting matrix in the regression model.

Martins & Hansen (1997) and Rohlf (2001) showed how a phylogenetic tree can be used to construct the expected covariance matrix or weighting matrix for taxa, when different models of evolutionary divergence are assumed, by means of an algorithm similar to the one used to compute a matrix of cophenetic values (Sokal & Rohlf, 1986; Rohlf, 2001). Assuming the Brownian motion model, the \mathbf{W} matrix for the tree in Fig 5 is

$$\begin{pmatrix} w_1 + w_{1+2} & w_{1+2} & 0 \\ w_{1+2} & w_2 + w_{1+2} & 0 \\ 0 & 0 & w_3 \end{pmatrix}.$$

Although we stress the use of spatial regression techniques, these phylogenetic approaches could be used to incorporate phylogenetic autocorrelation in inter-populations studies.

Concluding remarks

Eco-evolutionary studies at the intra-specific level have been recently revitalized (Carroll *et al.*, 2007; Ezard *et al.*, 2009; Pelletier *et al.*, 2009) as a consequence of

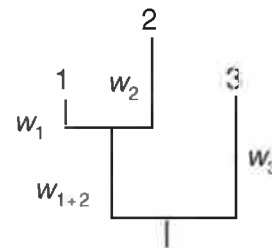


Fig. 5 Phylogenetic tree with three terminal populations. The quantities w_1 , w_2 , w_3 and w_{1+2} are the lengths of the branches supporting the populations indicated by their subscripts (modified after Rohlf, 2001).

recognizing that environment-related morphological changes accompany most evolutionary changes (Badyaev, 2005). Here, we show that morphological diversification of *Homo sapiens* could be explained as the result of nonrandom factors acting closely related to climatic variation (also see Beals *et al.*, 1984; Roseman, 2004; Harvati & Weaver, 2006; Perez & Monteiro, 2009). In population studies, Sokal (1984) stressed that conventional association analyses of morphometric and environmental data sets must be corroborated by incorporating spatial autocorrelation in regression models. However, to date no systematic approaches have been used to solve this problem at the intra-specific level. In this paper, we illustrate several regression techniques which take into account spatial autocorrelation.

Several works have pointed out that although autocorrelation can introduce bias in regression models, the processes that generate spatial autocorrelation can also be interesting on their own (Legendre, 1993; Peres-Neto, 2006). For instance, gene flow restricted by the geographical distance, which may cause spatial autocorrelation in form variation among populations, is interesting as an evolutionary process (Sokal & Oden, 1978; Sokal & Wartenberg, 1983; Sokal *et al.*, 1989b; Relethford, 2004a, 2008); although it causes bias in a model that tests for relationships between morphological and environmental variables. Therefore, spatial autocorrelation must be studied to explore the spatial structure underlying human genetic or phenotypic variation (Sokal & Oden, 1978; Barbujani, 2000; Relethford, 2008) and incorporated in regression models to provide more accurate statistical estimates of the relationships between morphological and environmental variables (Rohlf, 2006; Dormann *et al.*, 2007).

The regression techniques used in our example provided qualitatively similar results, but this does not necessarily indicate that all techniques are absolutely equivalent in any situation (Legendre, 1993; Legendre & Legendre, 2003). Under certain circumstances, the slopes can be qualitatively affected and the relative order of importance of the explanatory variables may shift between methods (see Lennon, 2000; Kühn, 2007), although it is still difficult to predict the situation in which this occurs (Bini *et al.*, 2009).

This revision highlights some methodological and conceptual topics in regression statistical techniques that need more study. Particularly, we need more realistic computer simulations to determine the performance of these statistical techniques in relation to type I and II errors (Rohlf, 2001; Diniz-Filho *et al.*, 2009). In addition, as all techniques assume spatial stationarity (i.e. spatial autocorrelation and effects of ecological correlates are constant across regions; Dormann *et al.*, 2007), it is necessary to develop techniques that consider the spatial variation in autocorrelation. Finally, we require expanding the discussion regarding alternative approaches to explore the underlying environmental variables and

nonrandom factors that generate morphological variation (e.g. Desdevises *et al.*, 2003; Peres-Neto, 2006).

The spatial regression techniques described and applied here are uncommon in population morphometric studies (but see Cheverud & Dow, 1985) and promise a new avenue for understanding the origin of morphological variation among populations. However, we remark that the change in statistical methodology should be followed by several conceptual advances. It must be clear that spatial regression techniques are correlational, and the cause of the relationship between morphology and ecology from comparative data can only be suggested (Pucciarelli, 1974; Garland *et al.*, 2005). Although non-random factors could be the probable cause of morphological divergence among populations, it is difficult to know the specific ecological factor shaping inter-population morphological variation. This is mainly because of the conceptual problems underlying correlation and causation (Shipley, 2000), and not necessarily due to problems of statistical techniques. Spatial regressions are mainly designed to deal with inflated type I errors due to spatial autocorrelation, and cannot solve the problem of broad-scale and direct-indirect associations. For this reason, understanding the causes of the relationship between morphology and environment requires the use of both comparative and experimental approaches.

Acknowledgments

We thank W. W. Howells for making publicly available the human morphometric data set. We are sincerely grateful to S. F. dos Reis for discussions and comments about phylogenetic and spatial comparative techniques. We also thank EditMyEnglish editors and Amelia Barreiro for help with the English version of the manuscript and D. Gobbo for help with Fig 2. We are deeply indebted to one anonymous reviewer who contributed greatly to improve the clarity of the manuscript. S. I. Perez, V. Bernal and P. N. Gonzalez are supported by research and postdoctoral fellowship from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). J. A. F. Diniz-Filho is partially supported by research fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico.

References

- Badyaev, A.V. 2005. Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation. *Proc. R. Soc. B* **272**: 877–886.
- Barbujani, G. 1987. Autocorrelation of gene frequencies under isolation by distance. *Genetics* **117**: 777–782.
- Barbujani, G. 2000. Geographic patterns: how to identify them and why. *Hum. Biol.* **72**: 133–153.
- Beals, K.L., Smith, C.L. & Dodd, S.M. 1984. Brain size, cranial morphology, climate, and time machines. *Curr. Anthropol.* **25**: 301–330.

- Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F.L.V.B., Akre, T.S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araújo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Böhning-Gaese, K., Borges, P.A.V., Castro-Parga, I., Chey, V.K., Chown, S.L., de Marco, P. Jr, Dobkin, D.S., Ferrer-Castán, D., Field, R., Filloy, J., Fleishman, E., Gómez, J.F., Hortal, J., Iverson, J.B., Kerr, J.T., Kissling, W.D., Kitching, I.J., León-Cortés, J.L., Lobo, J.M., Montoya, M., Morales-Castilla, I., Moreno, J.C., Oberdorff, T., Olalla-Tárraga, M.Á., Pausas, J.G., Qian, H., Rahbek, C., Rodríguez, M.A., Rueda, M., Ruggiero, A., Sackmann, P., Sanders, N.J., Terribile, L.C., Vetaas, O.R. & Hawkins, B.A. 2009. Coefficients ships in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography* **32**: 1–12.
- Carroll, S.P., Hendry, A.P., Reznick, D.N. & Fox, C.W. 2007. Evolution on ecological time-scales. *Funct. Ecol.* **21**: 387–393.
- Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. 1994. *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.
- Cheverud, J.M. & Dow, M.M. 1985. An autocorrelation analysis of genetic variation due to lineal fission in social groups of rhesus macaques. *Am. J. Phys. Anthropol.* **67**: 113–121.
- Cheverud, J.M., Dow, M.M. & Leutenegger, W. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* **39**: 1335–1351.
- Desdevises, Y., Legendre, P., Azouzi, L. & Morand, S. 2003. Quantifying phylogenetically structured environmental variation. *Evolution* **57**: 2647–2652.
- Diniz-Filho, J.A.F. & Bini, L.M. 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecol. Biogeogr.* **14**: 177–185.
- Diniz-Filho, J.A.F., Sant’Ana, C.E.R. & Bini, L.M. 1998. An eigenvector method for estimating phylogenetic inertia. *Evolution* **5**: 1247–1262.
- Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecol. Biogeogr.* **12**: 53–64.
- Diniz-Filho, J.A.F., Hawkins, B.A., Bini, L.M., Marco, J.R.P. & Blackburn, T. 2007. Are spatial regression methods a panacea or a Pandora’s box? A reply to Beale *et al.* (2007). *Ecography* **30**: 848–851.
- Diniz-Filho, J.A.F., Nabout, J.C., Campos Telles, M.P., Soares, T.N. & Rangel, T.F.L.V.B. 2009. A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genet. Mol. Biol.* **32**: 203–211.
- Dormann, C.F., McPherson, J., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D. *et al.* 2007. Methods to account for spatial autocorrelation in the analysis of distributional species data: a review. *Ecography* **30**: 609–628.
- Eller, E. 1999. Population substructure and isolation by distance in three continental regions. *Am. J. Phys. Anthropol.* **108**: 147–159.
- Epperson, B.K. 2003. *Geographical Genetics*. Princeton University Press, Princeton, NJ.
- Ezard, T.H.G., Côté, S.D. & Pelletier, F. 2009. Ecoevolutionary dynamics: disentangling phenotypic, environmental and population fluctuations. *Phil. Trans. R. Soc. B* **364**: 1491–1498.
- Felsenstein, J. 2002. Contrasts for a within-species comparative method. In: *Modern Developments in Theoretical Population Genetics* (M. Slatkin & M. Veuille, eds), pp. 118–129. Oxford University Press, Oxford.
- Freckleton, R.P. 2009. The seven deadly sins of comparative analysis. *J. Evol. Biol.* **22**: 1367–1375.
- Freckleton, R.P. & Jetz, W. 2009. Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proc. R. Soc. B* **276**: 21–30.
- Garland, T. Jr, Bennett, A.F. & Rezende, L. 2005. Phylogenetic approaches in comparative physiology. *J. Exp. Biol.* **208**: 3015–3035.
- Gittleman, J.L. & Kot, M. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst. Zool.* **39**: 227–241.
- Grafen, A. 1989. The phylogenetic regression. *Phil. Trans. R. Soc. Lond. B* **326**: 119–157.
- Griffith, D.A. 2003. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Visualization*. Springer-Verlag, New York.
- Griffith, D.A. & Peres-Neto, P. 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* **87**: 2603–2613.
- Harvati, K. & Weaver, T.D. 2006. Human cranial anatomy and the differential preservation of population history and climate signatures. *Anat. Rec. A* **288A**: 1225–1233.
- Hawkins, B.A., Diniz-Filho, J.A.F., Bini, L.M., De Marco, P. & Blackburn, T.M. 2007. Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography* **30**: 375–384.
- Hendry, A.P. & Kinnison, M.T. 1999. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53**: 1637–1653.
- Howells, W.W. 1973. *Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference among Recent Human Populations*, Papers of the Peabody Museum No. 67. Harvard University Press, Cambridge, MA.
- Howells, W.W. 1989. *Skull Shapes and the Map: Craniometric Analyses in the Dispersion of Modern Homo*, Papers of the Peabody Museum No. 79. Harvard University Press, Cambridge, MA.
- Hutchison, D.W. & Templeton, A.R. 1999. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**: 1898–1914.
- Ives, A.R. & Zhu, J. 2006. Statistics for correlated data: phylogenies, space, and time. *Ecol. Appl.* **16**: 20–32.
- Katzmarzyk, P.T. & Leonard, W.R. 1998. Climatic influences on human body size and proportions: ecological adaptations and secular trends. *Am. J. Phys. Anthropol.* **106**: 483–503.
- Kissling, W.D. & Carl, G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecol. Biogeogr.* **17**: 59–71.
- Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. *Divers. Distrib.* **13**: 66–69.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**: 1659–1673.
- Legendre, P. & Legendre, L. 2003. *Numerical Ecology*. Elsevier, Amsterdam.
- Lennon, J.J. 2000. Red-shifts and red herrings in geographical ecology. *Ecography* **23**: 101–113.
- Manel, S., Schwartz, M.K., Luikart, G. & Taberlet, P. 2003. Landscape genetics: combining landscape ecology and population genetics. *TREE* **18**: 189–197.
- Manica, A., Prugnolle, F. & Balloux, F. 2005. Geography is a better determinant of human genetic differentiation than ethnicity. *Hum. Genet.* **118**: 366–371.

- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**: 209–220.
- Marcus, L.F. 1990. Traditional morphometrics. In: *Proceedings of the Michigan Morphometrics Workshop* (F.J. Rohlf & F.L. Bookstein, eds), pp. 77–122. Special Publication, Number 2. The University of Michigan Museum of Zoology, Ann Arbor, MI.
- Martins, E.P. & Hansen, T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**: 646–667.
- Oden, N.L. & Sokal, R.R. 1992. An investigation of 3-matrix permutation tests. *J. Classif.* **9**: 275–290.
- Pelletier, F., Garant, D. & Hendry, A.P. 2009. Eco-evolutionary dynamics. *Phil. Trans. R. Soc. B* **364**: 1483–1489.
- Peres-Neto, P.R. 2006. A unified strategy for estimating and controlling spatial, temporal and phylogenetic autocorrelation in ecological models. *Oecol. Bras.* **10**: 105–119.
- Peres-Neto, P.R. & Jackson, D.A. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**: 169–178.
- Perez, S.I. & Monteiro, L.M. 2009. Nonrandom factors in modern human morphological diversification: a study of craniofacial variation in southern South American populations. *Evolution* **63**: 978–993.
- Perez, S.I., Diniz-Filho, J.A.F., Rohlf, F.J. & dos Reis, S.F. 2009. Ecological and evolutionary factors in the morphological diversification of South American spiny rats. *Biol. J. Linn. Soc.* **98**: 646–660.
- Pucciarelli, H.M. 1974. The influence of experimental deformation on neurocranial wormian bones in rats. *Am. J. Phys. Anthropol.* **41**: 29–38.
- Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W. & Cavalli-Sforza, L.L. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* **102**: 15942–15947.
- Rangel, T.F.L.V.B., Diniz-Filho, J.A.F. & Bini, L.M. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecol. Biogeogr.* **15**: 321–327.
- Ray, N., Currat, M., Berthier, P. & Excoffier, L. 2005. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res.* **15**: 1161–1167.
- Relethford, J.H. 1994. Craniometric variation among modern human populations. *Am. J. Phys. Anthropol.* **95**: 53–62.
- Relethford, J.H. 2004a. Global patterns of isolation by distance based on genetic and morphological data. *Hum. Biol.* **76**: 499–513.
- Relethford, J.H. 2004b. Boas and beyond: Migration and craniometric variation. *Am. J. Hum. Biol.* **16**: 379–386.
- Relethford, J.H. 2008. Geostatistics and spatial analysis in biological anthropology. *Am. J. Phys. Anthropol.* **136**: 1–10.
- Reznick, D.N., Shaw, F.H., Rodd, F.H. & Shaw, R.G. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**: 1934–1937.
- Roberts, D.F. 1953. Body weight, race and climate. *Am. J. Phys. Anthropol.* **11**: 533–558.
- Rohlf, F.J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* **55**: 2143–2160.
- Rohlf, F.J. 2006. A comment on phylogenetic correction. *Evolution* **60**: 1509–1515.
- Roseman, C.C. 2004. Detection of interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proc. Natl Acad. Sci. USA* **101**: 12824–12829.
- Ruff, C.B. 1994. Morphological adaptation to climate in modern and fossil hominids. *Yearb. Phys. Anthropol.* **37**: 65–107.
- Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.
- Shipley, B. 2000. *Cause and Correlation in Biology: A User's Manual to Path Analysis, Structural Equations and Causal Inference*. Cambridge University Press, Cambridge.
- Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279.
- Smouse, P.E., Long, J.C. & Sokal, R.R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* **35**: 627–632.
- Sokal, R.R. 1984. Comment to “Beals, K.L., Smith, C.L. & Dodd, S.M. 1984. Brain size, cranial morphology, climate, and time machines.”. *Curr. Anthropol.* **25**: 322–323.
- Sokal, R.R. & Oden, N.L. 1978. Spatial autocorrelation in biology. 1. Methodology. *Biol. J. Linn. Soc.* **10**: 199–228.
- Sokal, R.R. & Rohlf, F.J. 1986. *Biometry*. W. H. Freeman and Company, San Francisco, CA.
- Sokal, R.R. & Uytterschaut, H. 1987. Cranial variation in European populations: a spatial autocorrelation study at three time periods. *Am. J. Phys. Anthropol.* **74**: 21–38.
- Sokal, R.R. & Wartenberg, D.E. 1983. A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics* **105**: 219–237.
- Sokal, R.R., Jacquez, G.M. & Wooten, M.C. 1989a. Spatial autocorrelation analysis of migration and selection. *Genetics* **121**: 845–855.
- Sokal, R.R., Harding, R.M. & Oden, N.L. 1989b. Spatial patterns of human gene frequencies in Europe. *Am. J. Phys. Anthropol.* **80**: 267–294.
- Templeton, A.R. 2007. Genetics and recent human evolution. *Evolution* **61**: 1507–1519.
- Thalib, L., Kitching, R.L. & Bhatti, M.I. 1999. Principal component analysis for grouped data: a case study. *Environmetrics* **10**: 565–574.
- Wall, M.M. 2004. A close look at the spatial structure implied by the CAR and SAR models. *J. Stat. Plan. Inference* **121**: 311–324.
- Wright, S. 1943. Isolated by distance. *Genetics* **28**: 114–138.
- Zar, J.H. 1999. *Biostatistical Analysis*. Prentice-Hall, Nueva York.

Received 24 June 2009; accepted 23 October 2009