Intelligence xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Intelligence

Statistical inference and spatial patterns in correlates of IQ

Christopher Hassall *, Thomas N. Sherratt

Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S 5B6

ARTICLE INFO

Article history:
Received 9 November 2010
Received in revised form 9 May 2011
Accepted 9 May 2011
Available online xxxx

Keywords:
IQ
Intelligence
Spatial autocorrelation
Geography
Disease
Statistics

ABSTRACT

Cross-national comparisons of IQ have become common since the release of a large dataset of international IQ scores. However, these studies have consistently failed to consider the potential lack of independence of these scores based on spatial proximity. To demonstrate the importance of this omission, we present a re-evaluation of several hypotheses put forward to explain variation in mean IQ among nations namely: (i) distance from central Africa, (ii) temperature, (iii) parasites, (iv) nutrition, (v) education, and (vi) GDP. We quantify the strength of spatial autocorrelation (SAC) in the predictors, response variables and the residuals of multiple regression models explaining national mean IQ. We outline a procedure for the control of SAC in such analyses and highlight the differences in the results before and after control for SAC. We find that incorporating additional terms to control for spatial interdependence increases the fit of models with no loss of parsimony. Support is provided for the finding that a national index of parasite burden and national IQ are strongly linked and temperature also features strongly in the models. However, we tentatively recommend a physiological – via impacts on host–parasite interactions – rather than evolutionary explanation for the effect of temperature. We present this study primarily to highlight the danger of ignoring autocorrelation in spatially extended data, and outline an appropriate approach should a spatially explicit analysis be considered necessary.

1. Introduction

The measurement of intelligence is a controversial field (Gould 1981; Jensen 1982), particularly where comparisons are made among races (Hunt & Carlson 2007) or nations (Lynn & Vanhanen 2006). The recent compilation of an international dataset of IQ results from a wide range of countries (Lynn & Vanhanen 2006) has made possible broad comparisons between nations, of which a great many have already been published (see Wicherts, Dolan, & van der Maas 2010 for a review of this literature). While criticisms have been levelled at how this IQ dataset was collated (Wicherts, Dolan, & van der Maas 2010), there are statistical issues with international comparisons even with perfectly-collated data due to the potential lack of independence of individual data points driven by spatial proximity. We first highlight the general nature of this problem and explain why it matters. We then re-evaluate a set of hypotheses that have been put forward to explain variation in national IQ as a case study to provide guidance for future studies. Note that while the global variation in mean national IQ has received considerable recent attention, it remains debateable whether variation in national IQ is a strict reflection of variation in underlying cognitive abilities that they are proposed to measure, since their psychometric properties may also vary across space (Wicherts, Dolan, Carlson, & van der Maas 2010) and time (Wicherts et al. 2004). For example, recent work has indicated that IQ score may vary with individual motivation, and that this simple phenomenon may confound relationships between individual IQ and late-life outcomes (Duckworth et al. in press). Thus, while we have followed others in focusing on national mean IQ as the key dependent variable of interest, we recognise at the outset that it has significant limitations as a measure of latent intelligence.

* Corresponding author. Tel.: +1 613 520 2600x3866; fax: +1 613 520 3539.
E-mail address: chassall@connect.carleton.ca (C. Hassall).

0160-2896/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2011.05.001

2. Why spatial autocorrelation matters

Recently, Gelade (2008) used spatial autocorrelation analysis to show that nations that are geographical neighbours have more similar mean IQs than nations that are far apart. One might equally find positive autocorrelation in candidate predictor variables of national mean IQ such as average temperature, or national *per capita* income, reflecting Tobler's (1970) First Law of Geography: "everything is related to everything else, but near things are more related than distant things".

Acknowledgement of spatial autocorrelation in response variables and/or their potential predictors is extremely important. As an example from the intelligence literature, nearby nations may have similar sized values of a response variable (*e.g.* national IQ) and similar sized values of any given predictor (*e.g.* mean temperature). This association may stem from a causal relationship, *i.e.* the sites share a similar climate regime and this results in a similar national mean IQ. However, it may be that there are one or more underlying factors that drive both variables, resulting in a correlation without a causal relationship. One such example is local movement of peoples between countries that share similar temperature attributes simply through spatial proximity. Thus, the apparent association between the two variables may be due to their proximity rather than independently driven causal relationships. Classical significance testing is based on the assumption of independence and if one cannot be confident that each data point represents an independent realisation of the same causal process, the significance values become unreliable. It seems intuitively unreasonable, for example, to compare data for France, Germany and Belgium with Ghana, Togo and Benin, assuming each to be entirely independent. We have illustrated precisely this problem in Fig. 1. Countries on the same continent are more similar to one another than to countries on different continents both in terms of national mean IQ and any number of potential predictors (*e.g.* disease burden as shown in Fig. 1 and as hypothesised by Eppig, Fincher, & Thornhill 2010). Additional statistical controls must be taken into account to explicitly deal with the spatial relationships among data points. Specifically, without controlling for autocorrelation, tests of association between spatially autocorrelated variables can lead to an inflated proportion of Type I errors (rejection of the null hypothesis when true), since the effective sample size is always smaller than the total number of genuinely independent data points (Clifford, Richardson, & Hemon 1989; Legendre & Fortin 1989; Legendre & Legendre 1998). The problem may also be more severe than simply inflating Type I error rate. In particular, Lennon (2000) argued that correlations between an autocorrelated response variable and a set of candidate predictors will be strongly biassed in favour of identifying autocorrelated predictors as significant over non-autocorrelated predictors.

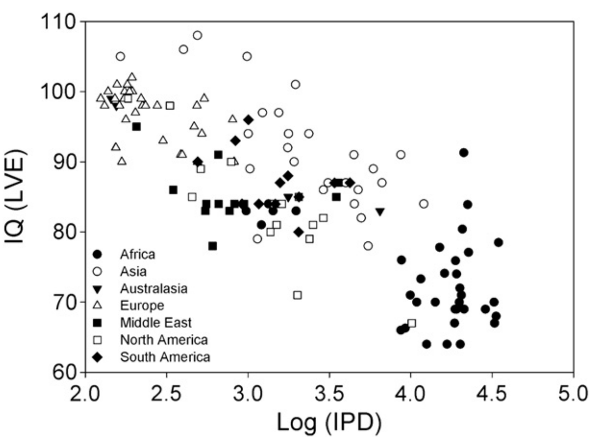

Fig. 1. The relationship between national mean IQ (LVE) and IPD (daily-adjusted life years due to infectious and parasitic diseases) for 137 countries grouped by continent. Note the clear lack of independence of the data, with African countries consistently exhibiting high IPD and low mean IQ, while European countries consistently exhibit low IPD and high mean IQ. It is unlikely that these spatially dependent relationships arise as independent realisations of the same causal process.

While many papers have highlighted the problems posed by spatial autocorrelation in data, far fewer studies have offered a solution (Dale & Fortin 2002). These solutions include discarding data, adjusting the Type I error rate, adjusting the effective sample size to control for lack of independence and accounting for spatial structure directly in the fitted model (Dale & Fortin 2002). Whatever the remedy, one simply cannot ignore spatial autocorrelation and hope for the best (Beale, Lennon, Yearsley, Brewer, & Elston 2010). Of course, it is quite possible for a spatially autocorrelated predictor to generate independent yet spatially autocorrelated responses when the response variable would not otherwise be autocorrelated. Using the example above, a positive correlation between national mean IQ and temperature would, by virtue of the spatial structure in temperature, produce a spatial structure in national IQ. Thus the two variables would be spatially autocorrelated but with an independent relationship. Therefore, conservatively controlling for spatial autocorrelation in predictor and response can "throw the baby out with the bathwater" and leave researchers with little additional variation to explain other than processes operating at different (usually smaller) spatial scales. Arguably therefore, controlling for a lack of spatial independence is only essential when the residuals of fitted models continue to show significant spatial signature (Diniz-Filho, Bini, & Hawkins 2003) above and beyond those accounted for by the predictor, which will arise when the response continues to show a lack of independence even after controlling for the predictor's effect. Here we adopt this conservative approach in re-evaluating competing hypotheses to explain geographical patterns in national mean IQ. We show that spatial autocorrelation is present not only in the predictors of national mean IQ, but also in the residuals of models used to describe national IQ. The best fitting models exhibit greater explanatory power after control for spatial autocorrelation so, rather than obliterate any pattern, they remain capable of yielding insights into the question of how and why IQ varies across nations.

3. Competing hypotheses to explain geographical variation in mean IQ

Since Lynn and Vanhanen (2001) published their monographs on geographical variation in IQ, a number of competing hypotheses have emerged to explain variation between countries. We present a subset of representative hypotheses which can be classified using three broad categories:

Evolutionary hypotheses:

- *Distance from the environment of evolutionary adaptedness* (hereafter, "D_{EEA}") (Kanazawa 2008) — Kanazawa proposed that the human brain was adapted to a particular ancestral environment: the savannah of central Africa. In order to exploit environments that differ from this habitat, the human brain would need to be able to adapt to solve new challenges. Kanazawa proposes that this requirement for greater intelligence is what selected for higher-IQ individuals in locations further from the environment of evolutionary adaptedness (EEA).
- *Temperature* (Kanazawa 2008; Templer & Arikawa 2006) — In a similar hypothesis, a variety of authors have suggested that cold weather and harsh winters select for higher intelligence to be able to cope with the extremes of climate.

Physiological hypotheses:

- *Nutrition* (Lynn 1990) — Lynn observed that changes in height and head size were occurring over time. He hypothesised that this was the result of increasing levels of nutrition, citing evidence that nutritional deficiencies retard growth. Citing correlations between head size, brain size and IQ, Lynn then proposes that increases in nutrition are also increasing national mean IQ.
- *Parasite burden* (Eppig et al. 2010) — Significant international variation in IQ can be explained by variation in the disability-adjusted life years (DALY, a measure of disease burden) due to parasitic and infectious disease. The reasoning behind this hypothesis is that the response to parasites by the immune system requires energy which can then not be used in cognitive development.

Socioeconomic hypotheses:

- *Education* (Barber 2005) — This hypothesis assumes that the amount of time put into education is related to the extent of cognitive development, which then influences IQ. Evidence for such a causal relationship has been presented using longitudinal studies (*e.g.* Richards & Sacker 2003). Marks (2010) has argued that geographical variation in IQ is purely an artefact of literacy levels. However, literacy data are no longer collected in many high-income countries which are typically considered to be 99% literate (*e.g.* United Nations Development Programme 2009). Here we assume that Marks' hypothesis based on literacy can be tested using data on education.
- *Gross domestic product* (GDP) (Lynn & Vanhanen 2002) — GDP per capita is related to development which, in turn, is related to the average amount of education. For reasons described in the previous hypothesis, it might be expected that a higher general level of education would result in higher IQ.

All studies cited above have provided significant statistical results to support their hypotheses. However, none so far has either tested for or controlled for the spatial structure of the data in a rigorous way.

3.1. Outline of the analysis

We begin by describing the sources for our data (which are provided in Appendix 1). We then demonstrate the extent of the spatial autocorrelation in the raw predictor and response variables. We show that strong correlations exist between all six candidate predictors and three measures of national mean IQ, even when spatial autocorrelation is taken into account. We use an exhaustive model selection method to find the most parsimonious model to explain variation in national mean IQ. Next, and most importantly, we show that the residuals of these best-fit multiple regression models exhibit spatial autocorrelation, which even by the least conservative standards necessitates the control of this autocorrelation in the analysis of the model (Diniz-Filho et al. 2003). Finally, we then carry out the model selection procedure, this time including control for SAC.

3.2. Data sources

Data sources were used mostly as specified in Eppig et al. (2010): national IQ data were taken from Lynn and Vanhanen (2006) with 17 alternative values from Wicherts, Dolan, and van der Maas (2010); disability life-adjusted year (DALY) values for infectious and parasitic diseases (hereafter "IPD") and nutritional deficiencies ("Nut") were generated by the WHO (2004); average years in education ("AVED"), % population reaching enrolment in secondary education ("Sec_E") and % population completing secondary education ("Sec_C") from Barro and Lee (2010) and data at http://www.barrolee.com/ for 2010; and GDP per capita ("GDP") from the CIA (2007). Three IQ datasets were defined, as in Eppig et al.: Lynn and Vanhanen's (2006) data based only on censuses ("LVCD"), Lynn and Vanhanen's data with estimates for missing values ("LVE") and LVE with the 17 alternative values from Wicherts, Dolan, and van der Maas (2010) ("WEAM"). Distance from the point 5°S, 25°E (the "environment of evolutionary adaptedness") to the centroid of each country ("D_{EEA}") was calculated in ArcGIS v9.2 (ESRI, 2006). Centroids were also used in subsequent control for SAC. As an index of temperature, we calculated the mean temperature of the coldest quarter ("MTCQ") for each country using the WORLDCLIM dataset (Hijmans, Cameron, Parra, Jones, & Jarvis 2005) in ArcGIS v9.2 (ESRI, 2006). Countries lacking any data were excluded leaving a total of 137 countries for the comparison (Table S1). IPD, Nut, GDP and D_{EEA} were log-transformed for normality. The three education measures were highly collinear (Sec_E *vs.* Sec_C, $r = 0.942$, $p < 0.001$; Sec_E *vs.* AVED, $r = 0.935$, $p < 0.001$; Sec_C *vs.* AVED, $r = 0.892$, $p < 0.001$). Therefore, the three education variables were entered into a principal components analysis to produce a single education measure ("ED") from the first principal component which explained 97.7% of the variance in the three measures.

3.3. Data analysis

3.3.1. SAC in predictors and responses

A statistical measure of spatial autocorrelation, Moran's I, was calculated for each of the three national IQ datasets (the response variables) and the six predictors described above and in Table 1. An alternative measure of SAC is Geary's C, which is approximately inversely related, though not identical,

Please cite this article as: Hassall, C., & Sherratt, T.N., Statistical inference and spatial patterns in correlates of IQ, *Intelligence* (2011), doi:10.1016/j.intell.2011.05.001

Table 1
Three measures of national IQ and six predictor variables with the extent of spatial autocorrelation (global Moran's I). Each of these variables exhibit highly significant (denoted ***) spatial structuring, in that we can readily reject the null hypothesis of no spatial structure ($p<0.001$). $N=137$, except for LVCD where $N=88$.

Variable	Abbreviation	Moran's I
National IQ (Lynn and Vanhanen including estimates)	LVE	0.312***
National IQ (Lynn and Vanhanen with Wicherts et al. (2010) alternative values)	WEAM	0.286***
National IQ (Lynn and Vanhanen's census data)	LVCD	0.253***
Infectious and parasitic disease burden	IPD	0.321***
Nutritional deficiency burden	Nut	0.199***
Mean temperature of the coldest quarter	MTCQ	0.275***
Education	Ed	0.205***
Gross domestic product (*per capita*)	GDP	0.221***
Distance from the environment of evolutionary adaptedness	D_{EEA}	0.359***

to Moran's I (Sokal & Oden 1978). We use Moran's I as it gives a more global indicator of spatial autcorrelation while Geary's C is more sensitive to local differences. Moran's I also tends to perform better in ecological analyses, describing patterns more cleanly and being easier to interpret (Legendre & Fortin 1989).

A distance matrix was first calculated based on great circle distances between each pair of country centroids using the "distCosine" function in the R package geosphere (Hijmans, Williams, & Vennes 2011). Great circle distances take into account the curvature of the earth when calculating distances between two sets of latitude–longitude coordinates. The "Moran.I" function in the R package APE (Paradis, Claude, & Strimmer 2004) was used to calculate the global Moran's I value for each of the nine variables. We have attached the R code for this operation in Appendix 2. To further illustrate the pattern of SAC in the data, the three IQ variables and IPD, highlighted as the most important predictor in a recent analysis (Eppig et al. 2010) were analysed in SAM v4.0 (Rangel, Diniz-Filho, & Bini 2006) over a range of distances. SAM ("Spatial Analysis in Macroecology") is free software available from http://www.ecoevol.ufg.br/sam/. This software provides tools to carry out a variety of analyses including spatial eigenvector mapping, the quantification of SAC using Moran's I, and multimodel inference using Akaike's Information Criteria (AIC).

3.3.2. Correlations between national mean IQ and predictors

Correlations between each of the predictors and the three national IQ indices were assessed using Pearson product–moment correlations (Table 2). Having previously demonstrated the presence of spatial autocorrelation in the predictors and response variables, it was clear that the degrees of freedom in the tests would be artificially inflated due to the lack of independence between data points. The "spatial correlation" function in SAM was used to recalculate the geographically effective degrees of freedom according to the method of Clifford et al. (1989). This allows a more accurate calculation of statistical significance.

3.3.3. First model construction

Having demonstrated that all predictor variables are strongly correlated with all three national IQ indices, even when the lack of independence is controlled for, we were left with all six predictor variables as viable predictors for linear regression. Extensive collinearity exists within the predictors, which poses problems for using stepwise model selection to identify subsets of variables for use in regression models. Wicherts, Borsboom, and Dolan (2010) highlight this collinearity among socioeconomic and health variables – and suggest that national mean IQ is simply another indicator of development – although the same is true for most predictors of national IQ. If left unchanged, multicollinearity (linear relationship between two or more variables) results in an inflation of the variance associated with parameter estimates within multiple regression models. However, cases of multicollinearity can be identified using variance inflation factors (VIFs) to determine the extent to which the variance associated with each term is increased by the collinearity, where $VIF>10$ is considered "high" multicollinearity (Kutner, Nachtsheim, Neter, & Li 2005). We avoid this problem by using an "exhaustive search" method to compare all possible combinations of variables (Graham 2003). The relative performance of the models was then judged using AIC controlling for small sample size (AICc; Kutner et al. 2005). This measure of model performance incorporates goodness-of-fit as well as the number of explanatory variables to rank models relative to one another to indicate the most parsimonious models. Alternative model selection methods using only goodness of fit (*e.g.* R^2 or adjusted R^2) neglect the principle of parsimony, while the Bayesian information criterion (BIC, also known as the Schwartz criterion) rests on assumptions that are rarely met with empirical data (Johnson & Omland 2004). A ΔAICc (the difference between the AICc of a given model and that of the top model) of <2 indicates that there is substantial evidence for the given model above alternative candidate models, $3<\Delta\text{AICc}<7$ indicates considerably less support and $\Delta\text{AICc}>10$ indicates essentially no support (Burnham & Anderson 2002). We also calculate adjusted R^2 (the proportion of overall variance explained by the fitted model) as an absolute measure of goodness-of-fit to complement the relative measure provided by AICc. Six predictors yield a potential 63 models including a null model (with only a floating intercept) and each of these was constructed in R for each of the three IQ variables. The resulting models were compared using the "aictab" function in the AICcmodavg package (Mazerolle 2010) in R. We have provided the R code for this stage of the analysis in Appendix 3.

3.3.4. SAC in model residuals

As stated above, the presence of SAC in model residuals indicates a need to account for SAC in the model itself. We tested for evidence of spatial autocorrelation in the best

Please cite this article as: Hassall, C., & Sherratt, T.N., Statistical inference and spatial patterns in correlates of IQ, *Intelligence* (2011), doi:10.1016/j.intell.2011.05.001

Table 2
Product moment coefficients and significance of correlations between three national IQ measures (see text for details) and eight putative predictors (see text for definitions) before (r and p) and after (p* = corrected p-value, df* = estimated corrected degrees of freedom) control for spatial autocorrelation. Degrees of freedom prior to correlation for autocorrelation are 135 for LVE and WEAM and 85 for LVCD.

	LVE (n = 137)				WEAM (n = 137)				LVCD (n = 88)			
	r	p	p*	df*	r	p	p*	df*	r	p	p*	df*
IPD	−0.854	<0.001	0.002	7.65	−0.812	<0.001	0.003	8.60	−0.855	<0.001	0.003	7.17
Nut	−0.748	<0.001	0.002	12.76	−0.718	<0.001	0.002	14.09	−0.753	<0.001	0.003	10.95
MTCQ	−0.642	<0.001	0.026	9.73	−0.630	<0.001	0.022	10.87	−0.671	<0.001	0.018	9.83
Ed	0.638	<0.001	0.008	13.81	0.606	<0.001	0.009	15.32	0.707	<0.001	0.005	11.96
GDP	0.717	<0.001	0.003	12.76	0.680	<0.001	0.004	14.18	0.795	<0.001	0.002	10.32
D_{EEA}	0.605	<0.001	0.031	10.74	0.531	<0.001	0.049	12.15	0.594	<0.001	0.011	15.29

fitting models (for which ΔAICc<2) for each of the three IQ variables. This was done by calculating global Moran's I in R, as described above, for the residuals of each of the models.

3.3.5. Control for SAC

Having demonstrated that the residuals of the best fitting models exhibited spatial autocorrelation, the model selection procedure was carried out a second time with a control for SAC. The incorporation of SAC into these models was through a technique called "spatial eigenvector mapping" (SEVM) and was carried out in SAM. This method decomposes the spatial relationships between data into explanatory variables which capture spatial effects at different spatial resolutions. The method can be viewed as equivalent to a principal components analysis carried out on the distance matrix of the data (Dormann et al. 2007). Whereas selection of relevant components in PCA hinges on their eigenvalues, we based selection of eigenvectors on the minimisation of Moran's I (to a threshold of 0.05) in the model residuals. The resulting eigenvectors are then included in all models during the model selection procedure. Global Moran's I was calculated for the residuals of each of the best fitting (ΔAICc<2) models to evaluate the success of the method.

4. Results

4.1. SAC in predictors and responses

LVE and WEAM data showed a positive autocorrelation that was significantly ($p<0.001$) different from zero at each distance up to 3500 km then a significant ($p<0.01$) negative autocorrelation up to 16,000 km. LVCD showed a significant ($p<0.001$) positive autocorrelation up to 3,500 km and a significant ($p<0.001$) negative autocorrelation to 10,000 km after which there was no significant spatial structure (Fig. 2). Comparing predictors and response variables, we find that SAC is higher in national IQ than in national temperature (Table 1), as shown by Gelade (2008). As Gelade points out, there is an intuitive spatial autocorrelation involving temperature where two neighbouring nations tend to have a more similar climate than two more-distant nations. That national IQ exhibits stronger SAC than temperature emphasises the strength of the pattern. In fact, the only variable with higher SAC than national IQ was the distance from the environment of evolutionary adaptedness (D_{EEA}), which is itself a distance measure. What this SAC in D_{EEA} tells us is that two points that are closer together are a more-similar distance from another given point. This near-tautological example of SAC is instructive in demonstrating the importance of accounting for lack of independence in analyses.

4.2. Correlations between national IQ and predictors

Before control for SAC, there were strong, significant ($p<0.001$ in all cases) correlations between all six predictor variables and the three national IQ measures (Table 2). The proportion of variance in the national IQ measures that was explained by the individual predictors range from 28% to 73%, with the strongest correlations between national IQ and IPD and the weakest between IQ and D_{EEA}. When SAC was controlled for in these pairwise correlations there were still significant correlations at the reduced degrees of freedom. It is worth noting that the variables with higher SAC in Table 1 (IPD, D_{EEA} and MTCQ) are those which have the greatest reduction in degrees of freedom in Table 2. However, this

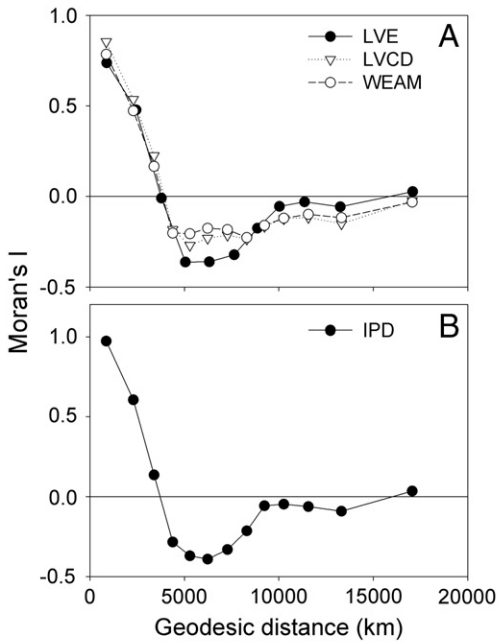

Fig. 2. Spatial autocorrelation in (A) three measures of national IQ, and (B) a proposed explanatory variable, namely the incidence of infectious and parasitic diseases (IPD, see text for details). Moran's I is a measure of spatial clustering. A positive Moran's I indicates that values are more similar at a given distance than would be expected by chance, while a negative Moran's I indicates that values are less similar than would be expected.

method still gives us no reason to choose between the competing hypotheses as all terms remain significant.

4.3. First model construction

An exhaustive search of models prior to control for SAC yielded very similar models for each of the three national IQ measures (Table 3). In each of the LVE, WEAM and LVCD measures, IPD, MTCQ and D_{EEA} formed the top model and were contained in all models where ΔAICc<2. Nut also featured in the second-ranking models in each case, and GDP featured in the third- and fourth-ranking models for LVCD. All models explain a large proportion of the variance in the response variables (between 72.3 and 81.1%).

4.4. SAC in model residuals

Examining the residuals for SAC we see that there is highly significant autocorrelation in the residuals of all the top models (Table 3). While this SAC is not as strong as that present in the raw data (Table 1), it provides strong evidence for a continuing effect of spatial interdependence in the models.

4.5. Control for SAC

The inclusion of spatial eigenvectors in the model selection procedure, results in a change in our interpretation of the results. The first is that the explanatory power of all models increases (note the adjusted R^2 values in Table 3). The lower AICc values demonstrate that this increase in goodness-of-fit does not come at a cost of decreased parsimony. In fact, the model fit according to AIC is substantially better after control for SAC, with ΔAICc values comparing best-fit models before and after SAC of 44.875, 31.286 and 24.185 for LVE, WEAM and LVCD, respectively.

Second, the SAC of the model residuals of two of the three measures was non-significant after control for SAC. SAC in the residuals of LVE was particularly high in the original models (Table 3) and, although the SEVM approach reduced SAC considerably, it was still significant. It is worth noting that the SEVM approach was designed not to render SAC non-significant, but to reduce it below a certain threshold (Moran's I<0.05) where it has a negligible effect. Using this criterion, the procedure was successful.

Third, the composition of the models changes. There is consistent evidence for an effect of IPD and MTCQ in the top models before controlling for SAC and this remains after the control is applied (Table 3). The most noticeable difference in model composition is the omission of D_{EEA} (distance from the environment of evolutionary adaptedness) from most of the models after control for SAC. Having been present in all top models prior to control for SAC, D_{EEA} occurs only once in the second-best fit model for the WEAM IQ measure. Nut also seems to increase in importance but only in the LVCD IQ measure, where GDP also remains in the best-fit models.

5. Discussion

We have highlighted the importance of dealing with spatial autocorrelation when analysing spatial patterns, and re-examined competing hypotheses explaining geographical variation in national IQ to illustrate our case. Cross-national research in mean IQ is a relatively new field but has already produced a number of studies which have sought predictors of variation in IQ. Such putative predictors have included temperature and skin colour (Templer & Arikawa 2006), evolutionary novelty (Kanazawa 2008), irreligion (Lynn, Harvey, & Nyborg 2009), inbreeding (Woodley 2009) and a range of economic factors (*e.g.* Dickerson 2006). While these studies may provide interesting results, none have explicitly considered spatial autocorrelation. It has long been appreciated (*e.g.* Clifford et al. 1989) that not accounting for spatial autocorrelation in the response variable results in inflated significance due to overestimation of the true sample size of data. While this is true for any spatial analysis, different fields have taken different lengths of time to address the problem. Geography was among the first (Cliff & Ord 1970), with ecology following later (Legendre 1993) and other sub-disciplines of biology only now incorporating the issues into their paradigms (Valcu & Kempenaers 2010). In this paper we

Table 3
Model selection table for exploratory analysis before (SAC is "no") and after (SAC is "yes") control for spatial autocorrelation. For definitions of model terms see text and Table 1. Significance of Moran's I is indicated by: $^{***}p<0.001$, $^{NS}p>0.05$. Note that after control for SAC, Moran's I for the model explaining LVE is still significant. This is due to the SEVM routine acting to reduce the magnitude of SAC below a specific threshold (0.05), rather than reducing the significance of the pattern.

Response	SAC	Model	K	AICc	ΔAICc	w_i	R^2 (adj)	Moran's I
LVE	No	IPD + MTCQ + D_{EEA}	5	836.695	0.000	0.368	0.811	0.161^{***}
		IPD + MTCQ + D_{EEA} + Nut	6	838.194	1.499	0.174	0.811	0.164^{***}
	Yes	IPD + MTCQ + SEVM	8	791.820	0.000	0.312	0.868	0.047^{***}
WEAM	No	IPD + MTCQ + D_{EEA}	5	869.189	0.000	0.378	0.724	0.105^{***}
		IPD + MTCQ + D_{EEA} + Nut	6	870.694	1.505	0.178	0.723	0.106^{***}
	Yes	IPD + MTCQ + SEVM	8	837.903	0.000	0.271	0.786	0.012^{NS}
		IPD + MTCQ + D_{EEA} + SEVM	9	838.751	0.848	0.177	0.787	0.006^{NS}
LVCD	No	IPD + MTCQ + D_{EEA}	5	545.176	0.000	0.254	0.787	0.099^{***}
		IPD + MTCQ + Nut + D_{EEA}	6	545.278	0.102	0.241	0.790	0.101^{***}
		IPD + MTCQ + GDP + D_{EEA}	6	545.616	0.441	0.204	0.789	0.100^{***}
		IPD + MTCQ + Nut + GDP + D_{EEA}	7	547.079	1.903	0.098	0.789	0.101^{***}
	Yes	IPD + MTCQ + Nut + SEVM	8	520.991	0.000	0.194	0.845	-0.003^{NS}
		IPD + MTCQ + GDP + SEVM	8	521.294	0.303	0.167	0.845	0.004^{NS}
		IPD + MTCQ + SEVM	7	521.906	0.914	0.123	0.841	0.002^{NS}
		IPD + MTCQ + Nut + GDP + SEVM	9	522.342	1.350	0.099	0.845	-0.001^{NS}

Please cite this article as: Hassall, C., & Sherratt, T.N., Statistical inference and spatial patterns in correlates of IQ, *Intelligence* (2011), doi:10.1016/j.intell.2011.05.001

highlight the issue of spatial autocorrelation in the context of spatial variation in intelligence.

Correcting for SAC in conjunction with exhaustive model selection enables us to circumvent the twin problems of spatial autocorrelation and collinearity among variables. This permits the most comprehensive and statistically rigorous assessment of six potential hypotheses explaining variation in geographical patterns in IQ that has yet been conducted. When a comprehensive model comparison was conducted to analyse national variation in IQ scores, then infectious and parasitic diseases (IPD) and temperature (mean temperature of the coldest quarter) were the only two variables consistently included in models. Mortality and morbidity resulting from nutritional deficiencies (Nut), GDP, and distance from the environment of evolutionary adaptedness (D_{EEA}) also feature in some of the best fitting models. However, it is worth noting that D_{EEA} becomes far less important in models after controlling for SAC. This is not surprising given that the variable itself is, by definition, autocorrelated across space. It seems likely that the distance from the environment of evolutionary adaptedness has no causal link with national mean IQ.

The case for an effect of infectious and parasitic disease burdens influencing national IQ has been made elsewhere (Eppig et al. 2010). Previously, the relationship between temperature and national mean IQ has been explained in terms of the greater cognitive demands of surviving in colder environments (Templer & Arikawa 2006). Given the strength of evidence for the physiological effects of disease, it may be that temperature is acting not through an impact on the environment but through an impact on the interaction between humans and their diseases. Temperature influences a number of disease-related parameters such as disease distribution (Guernier, Hochberg, & Guégan 2004), transmission seasons (*e.g.* malaria, Hay, Guerra, Tatem, Noor, & Snow 2004), the ability of insect vectors to transmit diseases (Cornel, Jupp, & Blackburn 1993) and the development and survival of parasites and host susceptibility (Harvell et al. 2002). It may be that temperature is having an effect on national mean IQ by mediating the response to infectious diseases rather than via environmental complexity.

We have highlighted SAC as a cause for concern in these analyses of geographic variation in IQ and briefly mentioned multicollinearity in the predictor variables as a second issue. While we use exhaustive (or "all-subsets") modelling to avoid issues with collinear predictor variables and model construction, an alternative method would be structural equation modelling (SEM, or "path analysis") (Graham 2003); (van der Maas et al. 2006). SEM involves the explicit, *a priori* statement of causal and correlative relationships between variables and provides estimates of the relative strengths of interactions. Where, for example, changes in sanitation are thought to cause changes in disease, or changes in nutrition cause changes in infant mortality, these effects can be stated and the direct and indirect effects on national IQ can be assessed. While this approach shows promise for testing hypotheses of national IQ variation, there are cases in which the nature of relationships is unclear. For example, does GDP exert a causal relationship on other factors? Does education improve nutrition and/or disease incidence?

Socioeconomic factors do not feature strongly in the analysis when other factors are taken into account. GDP is present in some of the best-fitting models but it is unclear as to how this variable is acting. There has been debate in the literature over the competence of IQ tests to accurately measure intelligence over a range of education or literacy levels (Barber 2005), with some researchers claiming that global variation in IQ is entirely an artefact of varying literacy (Marks 2010). We find no evidence to support this. However, we stress that our measure of education, despite being a composite statistic will not have captured all aspects of educational experience, so as always, alternative measures could have given different results. Intriguingly, cross-fostering studies have demonstrated that socio-economic factors can influence IQ, with children from high socioeconomic status (SES) parents who were subsequently fostered by low SES parents having lower IQ scores than those children from high SES families who were then fostered by other high SES parents. Conversely, children from low SES parents who were fostered by high SES foster parents exhibited higher IQ scores than did children from low SES parents who were fostered by low SES foster parents (Capron & Duyme 1989). It is worth noting that this study was conducted only in France, and so the results may not be applicable to a global study with far greater variations in SES. It may be that SES acts at a smaller scale that is dwarfed by other factors on a global level.

Like all correlative studies, we cannot ascribe causality on the basis of statistical significance and so all potential relationships identified require further investigation. Here is not the place to present any alternative hypotheses in depth, especially on the basis of automated searches for candidate models rather than directed tests. However, it is possible that reduced parasite prevalence may play a role in the generation of the Flynn Effect, the apparent increase in mean IQ over time (but *cf.* Wicherts et al. 2004). Other studies have shown that generational increases in intelligence are focused at the lower end of the IQ distribution (Colom, Lluis-Font, & Andrés-Pueyo 2005). Parasites in host populations commonly exhibit aggregation, with a few individuals carrying large numbers of parasites and most individuals carrying few (Anderson & Gordon 1982). It could be reasoned that either improved hygiene or clinical intervention for diseases and parasites is benefitting those few heavily infected individuals disproportionately and, if those individuals also exhibit low IQ as a result of their disease burden, IQ would also increase to the greatest extent at the lower end of the scale. Thus, a parasite-induced depression in IQ with subsequent improvement due to hygiene and medicine could provide an explanation for the Flynn Effect (Eppig et al. 2010).

Controlling for autocorrelation may remove real biological patterns and this has been offered as an argument against controlling for both spatial (Legendre 1993) and phylogenetic (Ricklefs & Starck 1996) autocorrelation. However, any statistical analysis with an inherent spatial component should consider spatial autocorrelation, if only to demonstrate that its control is not necessary. Failure to account for this lack of independence in data violates statistical assumptions and renders statistical inference invalid. The initial dogmatism with which controls for spatial and phylogenetic autocorrelation were enforced has now given way to an acceptance that such controls are not always necessary. However, with the advent of numerous tools and techniques (such as those presented here) for assessing this need, we encourage researchers to at least give the topic due consideration as it can substantially influence results.

Please cite this article as: Hassall, C., & Sherratt, T.N., Statistical inference and spatial patterns in correlates of IQ, *Intelligence* (2011), doi:10.1016/j.intell.2011.05.001

Supplementary materials related to this article can be foundonline at doi 10.1016/j.intell.2011.05.001.

Acknowledgements

We would like to thank the Carleton University ECO-EVO Group for discussion. Conor Dolan, Douglas Detterman, Jelte Wicherts and an anonymous referee provided comments which greatly improved the manuscript. CH was supported by an Ontario Ministry of Research and Innovation Postdoctoral Fellowship and TNS is funded by NSERC.

References

Anderson, R. M., & Gordon, D. M. (1982). Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology*, *85*, 373–398.

Barber, N. (2005). Educational and ecological correlates of IQ: A cross-national investigation. *Intelligence*, *33*, 273–284.

Barro, R. J., & Lee, J. -W. (2010). *A new data set of educational attainment in the world, 1950–2010 (Vol. 15902)*. Cambridge, USA: National Bureau of Economic Research.

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010). Regression analysis of spatial data. *Ecology Letters*, *13*, 246–264.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer-Verlag.

Capron, C., & Duyme, M. (1989). Assessment of effects of socio-economic status on IQ in a full cross-fostering study. *Nature*, *340*, 552–553.

CIA (2007). CIA 2007 World factbook. https://www.cia.gov/library/publications/the-world-factbook

Cliff, A. D., & Ord, K. (1970). Spatial autocorrelation: A review of existing and new measures with applications. *Economic Geography*, *46*, 269–292.

Clifford, P., Richardson, S., & Hemon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics*, *45*, 123–134.

Colom, R., Lluis-Font, J. M., & Andrés-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence*, *33*, 83–91.

Cornel, A. J., Jupp, P. G., & Blackburn, N. K. (1993). Environmental temperature on the vector competence of *Culex univittatus* (Diptera: Culicidae) for West Nile Virus. *Journal of Medical Entomology*, *30*, 449–456.

Dale, M. R. T., & Fortin, M. -J. (2002). Spatial autocorrelation and statistical tests in ecology. *Ecoscience*, *9*, 162–167.

Dickerson, R. E. (2006). Exponential correlation of IQ and the wealth of nations. *Intelligence*, *34*, 291–295.

Diniz-Filho, J. A. F., Bini, L. M., & Hawkins, B. A. (2003). Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, *12*, 53–64.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, *30*, 609–628.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (in press). Role of test motivation in intelligence testing. Proceedings of the National Academy of Sciences.

Eppig, C., Fincher, C. L., & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society: Series B (Biological Sciences)*, *277*, 3801–3808.

ESRI (2006). *ArcGIS v.9.2*. Redlands: Environmental Systems Research Institute, Inc.

Gelade, G. A. (2008). The geography of IQ. *Intelligence*, *36*, 495–501.

Gould, S. J. (1981). *The mismeasure of man*. New York: W.W. Norton & Co..

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

Guernier, V., Hochberg, M. E., & Guégan, J. -F. (2004). Ecology drives the worldwide distribution of human diseases. *PLoS Biology*, *2*, e141.

Harvell, C. D., Mitchell, C. E., Ward, J. R., Altizer, S., Dobson, A. P., Ostfeld, R. S., et al. (2002). Climate warming and disease risks for terrestrial and marine biota. *Science*, *296*, 2158–2162.

Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., & Snow, R. W. (2004). The global distribution and population at risk of malaria: Past, present and future. *The Lancet Infectious Diseases*, *4*, 327–336.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Hijmans, R. J., Williams, E., & Vennes, C. (2011). Geosphere: Spherical trigonometry. R package version 1.2-19. http://CRAN.R-project.org/package=geosphere

Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, *2*, 194–213.

Jensen, A. R. (1982). The debunking of scientific fossils and straw persons. *Contemporary Education Review*, *1*, 121–135.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*, 101–108.

Kanazawa, S. (2008). Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence*, *36*, 99–108.

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). Irwin, CA: McGraw-Hill.

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, *74*, 1659–1673.

Legendre, P., & Fortin, M. -J. (1989). Spatial pattern and ecological analysis. *Vegetatio*, *80*, 107–138.

Legendre, P., & Legendre, L. (1998). *Numerical ecology*. Amsterdam: Elsevier Science.

Lynn, R. (1990). The role of nutrition in secular increases in intelligence. *Personality and Individual Differences*, *11*, 273–285.

Lynn, R., Harvey, J., & Nyborg, H. (2009). Average intelligence predicts atheism rates across 137 countries. *Intelligence*, *37*, 11–15.

Lynn, R., & Vanhanen, T. (2001). National IQ and economic development: A study of eighty-one nations. *Mankind Quarterly*, *41*, 415–435.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger.

Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality*. Augusta, GA: Washington Summit.

Marks, D. F. (2010). IQ variations across time, race and nationality: An artifact of differences in literary skills. *Psychological Reports*, *106*, 643–664.

Mazerolle, M. J. (2010). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 1.13. http://CRAN.R-project.org/package=AICcmodavg

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290.

Rangel, T. F. L. V. B., Diniz-Filho, J. A. F., & Bini, L. M. (2006). Towards an intergrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology and Biogeography*, *15*, 321–327.

Richards, M., & Sacker, A. (2003). Life course antecedents of cognitive reserve. *Journal of Clinical Experimental Neuropsychology*, *25*, 614–624.

Ricklefs, R. E., & Starck, J. M. (1996). Applications of phylogenetically independent contrasts: A mixed progress report. *Oikos*, *77*, 167–172.

Sokal, R. R., & Oden, N. L. (1978). Spatial autocorrelation in biology 1. *Methodology. Biological Journal of the Linnaean Society*, *10*, 199–228.

Templer, D. I., & Arikawa, H. (2006). Temperature, skin color, per capita income, and IQ: An international perspective. *Intelligence*, *34*, 121–139.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

United Nations Development Programme (2009). *Human development report 2009: Overcoming barriers — Human mobility and development*. New York: Palgrave Macmillan.

Valcu, M., & Kempenaers, B. (2010). Spatial autocorrelation: An overlooked concept in behavioral ecology. *Behavioral Ecology*, *21*, 902–905.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.

WHO (2004). *Global burden of disease: 2004 update*. Geneva, Switzerland: World Health Organisation.

Wicherts, J. M., Borsboom, D., & Dolan, C. V. (2010). Why national IQs do not support evolutionary theories of intelligence. *Personality and Individual Differences*, *48*, 91–96.

Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. J. (2010). Raven's test performance of sub-Saharan Africans: Mean level, psychometric properties and the Flynn Effect. *Learning and Individual Differences*, *20*, 135–151.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., et al. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, *32*, 509–537.

Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (2010). A systematic literature review of the average IQ of sub-Saharan Africans. *Intelligence*, *38*, 1–20.

Woodley, M. A. (2009). Inbreeding depression and IQ in a study of 72 countries. *Intelligence*, *37*, 268–276.

Please cite this article as: Hassall, C., & Sherratt, T.N., Statistical inference and spatial patterns in correlates of IQ, *Intelligence* (2011), doi:10.1016/j.intell.2011.05.001