Summarizing large spatial datasets: spatial principal components and spatial canonical correlation

Working Paper Series

#2013-011

Summarizing large spatial datasets: Spatial principal components and spatial canonical correlation

Samyukta Bhupathiraju, Bart Verspagen and Thomas Ziesemer

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://mgsog.merit.unu.edu

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

UNU-MERIT Working Papers

ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology, UNU-MERIT

Maastricht Graduate School of Governance MGSoG

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSoG to stimulate discussion on the issues raised.

Summarizing Large Spatial Datasets: Spatial Principal Components and Spatial Canonical Correlation

Samyukta Bhupathiraju,

Bart Verspagen &

Thomas Ziesemer

UNU-MERIT

6 February 2013

We thank Denis de Crombrugghe for useful comments and discussions.

Abstract

We propose a method for spatial principal components analysis that has two important advantages over the method that Wartenberg (1985) proposed. The first advantage is that, contrary to Wartenberg's method, our method has a clear and exact interpretation: it produces a summary measure (component) that itself has maximum spatial correlation. Second, an easy and intuitive link can be made to canonical correlation analysis. Our spatial canonical correlation analysis produces summary measures of two datasets (e.g., each measuring a different phenomenon), and these summary measures maximize the spatial correlation between themselves. This provides an alternative weighting scheme as compared to spatial principal components analysis. We provide example applications of the methods and show that our variant of spatial canonical correlation analysis may produce rather different results than spatial principal components analysis using Wartenberg's method. We also illustrate how spatial canonical correlation analysis may produce different results than spatial principal components analysis.

Keywords: spatial principal components analysis; spatial canonical correlation analysis; spatial econometrics; Moran coefficients; spatial concentration

JEL Classification: R10, R15, C10

1. Introduction

"Principal components" is an often used technique used to summarize data. It is based on correlation analysis and can roughly be seen as a way to summarize a dataset of many variables into only a few dimensions. Spatial analysis has a counterpart of correlation analysis, e.g., in the form of the Moran coefficient. The Moran coefficient measures the extent to which a phenomenon, measured by an indicator, is spatially concentrated (a positive spatial correlation), or spread-out (a negative correlation).

As spatial datasets often consist of a large number of variables, it was only natural that Wartenberg (1985) turned to the principal components method in order to summarize spatial correlations. He proposed a method that was closely analogous to principal components. As "regular" principal components essentially means to undertake a spectral decomposition (obtaining eigenvalues and eigenvectors) of the correlation matrix of a dataset, Wartenberg (1985) simply proposed to spectrally decompose the matrix of Moran coefficients of a spatially organized dataset.

While the spectral decomposition has a clear interpretation in the case of regular principal components, it does not have the same, nor an alternative, clear-cut interpretation in the case of Wartenberg's proposed method. His justification of the method was purely at the intuitive level, both by the analogy of taken eigenvalues and eigenvectors, and by the results that it produced for a number of constructed datasets that show different kinds of spatial dependence.

Below, we will first propose an alternative method for undertaking spatial principal components. Like Wartenberg's method, it is based on spectral decomposition. However, rather than taking eigenvalues and eigenvectors of the raw matrix of Moran coefficients in a dataset, we propose to spectrally decompose a slightly different matrix. Our method results from a clear objective, i.e., that the Moran coefficient of the resulting component is maximized. Thus, our summary measure is aimed at showing maximal spatial correlation itself.

The spatial principal components method that we propose can also be extended in a direction that is similar to canonical correlation analysis. Canonical correlation analysis is a way of summarizing two separate datasets, each into one or a few components, in such a way that the correlation between these components, across the datasets, is maximized. Since our spatial principal components method already maximizes spatial correlation within a single dataset, it can easily be extended to maximize spatial correlation between the summary measures of two datasets.

The rest of our paper is organized as follows. The new method for spatial principal components analysis is explained in the next Section (2). Section 3 presents the spatial canonical correlation method. Section 4 provides a few applications of the two methods, illustrating their use, and comparing the weighting schemes that they produce to Wartenberg's method, and regular principal components.

2. Spatial Principal Components: A New Method

Like Wartenberg (1985), our method starts from the Moran coefficient (e.g., Cliff and Ord, 1981). We will denote our data matrix of n observations (spatial units) and k variables by the symbol X (i.e., X is an nxk matrix of observations). Throughout the paper, we will assume that the k variables are z-scores, i.e., that for each variable the mean is zero and the standard deviation is one (this simplifies the notation). The n x n matrix of spatial weights is denoted by W, and we assume that the sum of elements of this matrix is equal to n (i.e., the weights matrix is standardized). Then, the n x n matrix of Moran coefficients between the k variables is calculated as

$$M = X^T WX. \quad \textbf{(1)}$$

The superscript T denotes the matrix transpose.

In case of no spatial weighting (W is the unity matrix), equation (1) would yield a matrix of normal (Pearson) correlation coefficients. In that case, an eigenvector of the correlation matrix, denoted by u, would yield a set of weights that could be used to calculate a composite measure Xu. This composite measure is actually a projection of the data matrix X on to the vector u, and it can be shown that the eigenvector that belongs to the largest eigenvalue is associated with the projection that minimizes the residual variance between X itself and the projection. In other words, that eigenvector maximizes the "fit" between the data and the lower-dimensional projection. The eigenvector associated with every next largest eigenvalue correspondingly maximizes the fit of the remaining residual variance. This procedure is called principal components.

Note that the length of the vector u is constrained to one, so that the projection is identified, and hence the procedure of minimizing residual variance (or maximizing fit) is a problem of constrained optimization. In particular, the principal components procedures maximizes $u^T X^T Xu$ subject to $u^T u = 1$, which yields the first-order condition and eigenvalue problem $X^T Xu = \lambda u$.

Wartenberg's spatial principal components analysis takes eigenvalues of the matrix M, which can be seen as the outcome of maximizing $u^T X^T WXu$ subject to $u^T u = 1$, which yields the first-order condition and eigenvalue problem $X^T WXu=\lambda u$. Note that this can also be written as $Mu=\lambda u$, which brings out clearly that what we are in fact doing in this case is taking the eigenvalues of the matrix of Moran coefficients M. Thus, Wartenberg maximizes $u^T X^T WXu$, which can be seen as the covariance of the non-spatially weighted factor (Xu) and the spatially weighted factor (WXu). While this is mathematically sound, we argue that this maximization does not have a clear intuitive interpretation, and hence the procedure is in need of an objective that is more directly related to the basics of spatial analysis.

Our proposal is to calculate a set of weights v in such a way that the Moran coefficient of the weighted summary variable Xv (which, as in the case of principal components, is the projection of the original data on the vector v) is maximized. This idea is the intuition that we want to put behind the spatial principal components procedure, which then becomes aimed at finding a summary measure that yields maximum spatial correlation. Mathematically, it can be implemented as follows. As long as the vector Xv is a z-score, the Moran coefficient that we want to maximize can be written as

$$(Xv)^T W(Xv) \tag{2}$$

The mean of Xv is zero because X is already expressed as z-scores. However, in order for Xv to be a z-score, we also need to choose the weights v such that the variance of Xv is one. Hence, given our objective of maximizing the Moran coefficient of Xv, we have to maximize equation (2) subject to the condition $(Xv)^T(Xv) = 1$. Note that in comparison to Wartenberg's procedure (as outlined above), we only propose to change the constraint, not the objective function. Wartenberg (implicitly) uses the constraint $v^T v = 1$, whereas we use $(Xv)^T(Xv) =1$.

It can relatively easily be shown that this constrained maximization problem leads to the following first-order condition:

$$(X^T X)^{-1}(X^T WX)v = \lambda v, \tag{3}$$

where λ is the Lagrange multiplier. This is clearly an eigenvalue problem. On the left hand side of (3), we find the matrix of Moran coefficients (1), pre-multiplied with the inverse of the covariance matrix of the non-spatially weighted data. The solution of our constrained maximization problem is an eigenvector of this left hand side matrix. The eigenvector that belongs to the largest eigenvalue will maximize the Moran coefficient of the projection Xv, and the corresponding eigenvalue will be equal to the largest possible value of the Moran coefficient, given the constraint of unit variance.[1]

Summarizing, our modification of Wartenberg's procedure is to pre-multiply the Moran matrix M of the variables in the dataset by the inverse of its (non-spatially weighted) covariance matrix. Doing so provides a clear interpretation of the eigenvectors of the matrix: this eigenvector is the set of weights that maximizes the spatial correlation of the projection of the raw data on to this eigenvector.

3. Spatial Canonical Correlation

The method of spatial principal components as outlined in the previous section has an intuitive extension into the direction of canonical correlation analysis. Suppose that instead of just one dataset, we have two, i.e., a matrix X and a matrix Y. Both X and Y have n rows (observations), but their number of variables (columns, denoted by k_X and k_Y) may differ. As before, we assume that X and Y are z-scores.

The aim of our spatial canonical correlation analysis will be to find two sets of weights, vectors v_X and v_Y, such that the weighted summary variables Xv_X and Yv_Y are maximally spatially correlated to each other. This can be achieved by maximizing the Moran coefficient

$$(Xv_X)^T W(Yv_Y) \tag{4}$$

[1] Note that we can also obtain (3) through maximizing $v^T(X^T X)^{-1}(X^T WX)v$ subject to $v'v=1$, which would be a more general form that does not require the data to be standardized.

subject to two conditions: $(Xv_X)^T(Xv_X) = 1$ and $(Yv_Y)^T(Yv_Y) = 1$ (the rationale for these conditions is the same as in the previous section). After re-arranging, this problem yields two first-order conditions:

$$(X^TX)^{-1}X^TWY(Y^TY)^{-1}(WY)^TXv_X = \lambda_1\lambda_2 v_X \qquad \textbf{(5)}$$

$$(Y^TY)^{-1}Y^TW^TX(X^TX)^{-1}X^TWYv_Y = \lambda_1\lambda_2 v_Y \qquad \textbf{(6)}$$

where λ_1 and λ_2 are Lagrange multipliers. Each of these conditions is an eigenvalue problem. In fact, the first m (where m is the minimum of by k_X and k_Y) eigenvalues of the two problems will be identical, and they will be equal to the square root of the (maximized) Moran coefficient that is associated to the eigenvectors in each of the two conditions. Defining $A = (X^TX)^{-1}X^TWY(Y^TY)^{-1}$ and $M=X^TWY$, we get

$$AM^Tv_X = \lambda_1\lambda_2 v_X$$

$$A^TMv_Y = \lambda_1\lambda_2 v_Y$$

The solution then is

$$\lambda_1\lambda_2 = v_X{}^TAM^Tv_X = v_Y{}^TA^TMv_Y$$

Thus, the two eigenvalue problems (5) and (6) provide a different perspective on summarizing information in a spatial way: for each of the two datasets X and Y they provide a summary measure (Xv_X or Yv_Y) that maximizes spatial correlation to the other dataset. The eigenvectors belonging to the largest eigenvalue in either (5) or (6) provide this summary measure, while the square of the largest eigenvalue is equal to the Moran coefficient between the two summary measures.

4. Applications

We now proceed to illustrate the proposed procedures to some real-world data. The datasets that we use are taken from Gallup et al. (1999), who provide a wealth of information on geography related phenomena on a global, country-level scale. We use data from two of their subsets of data: the physical characteristics of countries (mainly access to waterways and climate), and agriculture-related indicators (soil condition and climate zone). The data are available for 162 countries, and we use data from the CEPII dataset on distances between countries (Mayer and Zignano, 2011) in order to construct our spatial weights (we use the distance between the largest cities in a pair of countries).

These spatial weights are constructed as binary weights, where each (row) country has a weight of one for each of its closest 10 neighbour countries, and zero for other countries. This matrix is symmetrized by taking the average of cells (*i*,*j*) and (*j*,*i*) and assigning this average to both cells. We use such a symmetric spatial weights matrix because it makes the interpretation of the various Moran coefficients somewhat easier (e.g., it produces a symmetric Moran matrix M), but we have also applied the method to cases with a non-symmetric matrix W.[2] Finally, the spatial weights matrix is standardized by dividing each cell by the matrix sum (which obviously is 162*10).

4.1. Spatial Principal Components

We start by applying spatial principal components analysis to the dataset on physical characteristics of countries. This dataset has 13 variables, which are presented in Table 1. Nine of the 13 variables relate to access to waterways, either navigable rivers, or sea. The emphasis on this characteristic comes from the attention to landlocked countries in the debate on economic development (e.g., Gallup et al., 1999). The other variables relate to elevation, climate, and population density.

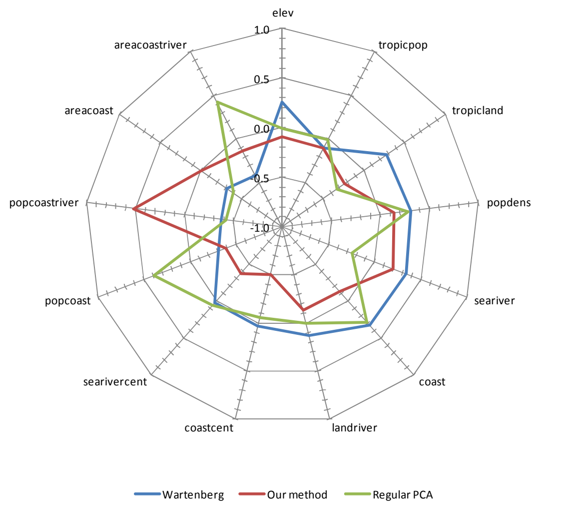
We apply regular (non-spatially weighted) principal components analysis, Wartenberg's spatial principal components analysis, and our own spatial principal components analysis to this dataset.[3] We start by looking at what would be called loadings in regular principal components analysis, i.e., the eigenvectors that were extracted in each of the three procedures. These are documented for the first two components (two largest eigenvalues) in graphical format in Figure 1a and 1b.

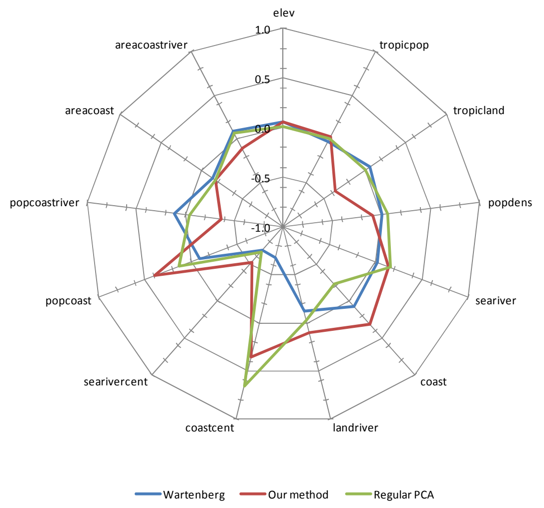
Table 1. Variables in the physical characteristics dataset

Variable description	Variable code
mean elevation (meters above sea level)	*elev*
mean distance to nearest coastline or sea-navigable river (km)	*seariver*
mean distance to nearest coastline (km)	*coast*
mean distance to nearest inland navigable river (km)	*landriver*
distance from centroid of country to nearest coast (km)	*coastcent*
distance from centroid of country to nearest coast or sea-navigable river (km)	*searivercent*
Ratio of population within 100 km of ice-free coast to total population	*popcoast*
Ratio of population within 100 km of ice-free coast/navigable river to total population	*popcoastriver*
population in the geographical tropics (%)	*tropicpop*
% Land area within 100 km of ice-free coast	*areacoast*
% Land area within 100 km of ice-free coast/navigable river	*areacoastriver*
Typical population density experienced by an individual (persons/km2)	*popdens*
% Land area in geographical tropics	*tropicland*

[2] If the spatial weights matrix (W) is non-symmetric, the matrix of Moran coefficients between the 13 variables will also be non-symmetric. This means that that the eigenvalues in the two spatial variants of principal components (and spatial canonical correlation analysis) potentially have imaginary parts. When we worked with non-symmetric W matrices (i.e., not in this paper), we always obtained large purely real eigenvalues.

[3] All our calculations are done in Matlab. The functions that were written to perform the analysis are available on request.





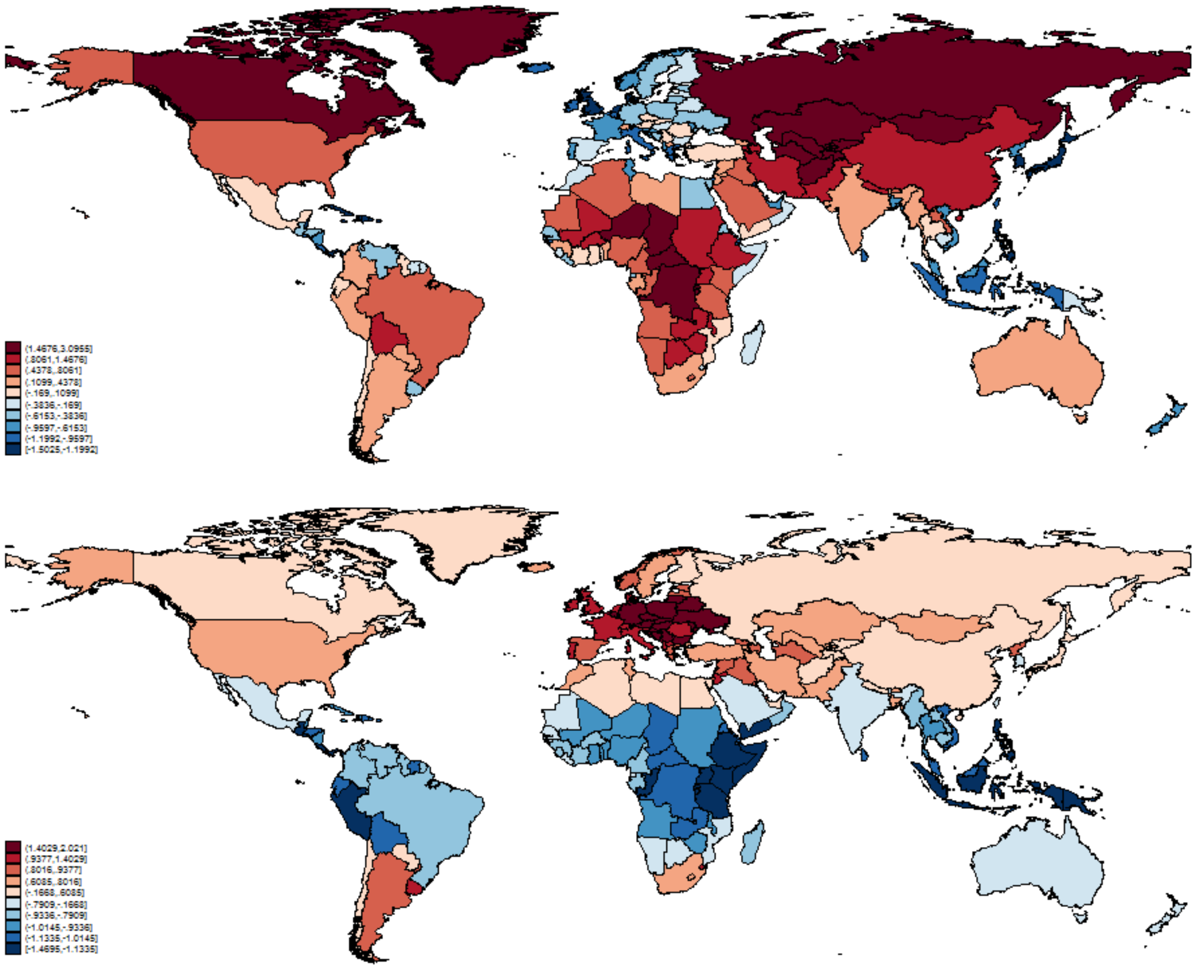
Figures 1a (left) and 1b (right): Eigenvectors for physical characteristics dataset, largest eigenvalue (left) and second-largest eigenvalue (right)

The average spatial correlation between the 13 variables in the dataset is modest to low (mean Moran coefficient of 0.04, with a standard deviation of 0.22). The maximum Moran coefficient observed between two variables in the dataset is 0.77. In terms of the Moran coefficients between the resulting summary variables (what would be called “scores” in regular principal components analysis), the three methods differ by rather much. The regular principal components analysis produces no spatial correlation at all (Moran coefficients -0.01 and 0.01). Wartenberg’s method produces Moran coefficients of 0.38 and 0.78, respectively for the first and second largest eigenvalues. Interestingly, the Moran coefficient associated with the second-largest eigenvalue is larger than that of the largest eigenvalue. In other words, if one would be looking for a high degree of spatial correlation, using the second eigenvector would be better than using the first eigenvector. For our version of spatial principal components, the Moran coefficients are 0.83 and 0.50, respectively for the largest and second largest eigenvalue.

The profiles of the eigenvectors in Figures 1a and 1b show clear differences between the three methods. For the largest eigenvalue (left figure), regular principal components analysis puts most emphasis (highest “loadings”) on *popdens*, *coast*, *popcoast* and *areacoastriver* (positive loadings) and *popcoastriver*, *areacoast*, *tropicland* and *seariver* (negative loadings). Wartenberg’s method stresses *elev*, *tropicland*, *popdens*, *seariver* and *coast* (positive loadings) and *popcoast*, *popcoastriver*, *areacoast* and *areacoastriver* (negative loadings). Our method emphasizes *popdens*, *seariver*, *popcoast* and *popcoastriver* (positive loadings) and *tropicland*, *coastcent*, *searivercent* and *areacoastriver* (negative loadings). All in all, it is clear that the three methods are far from equivalent in terms of which variables should be weighted strongest.

This leads to a rather different picture of which countries are behind the spatial correlation that is observed in the dataset as a whole. Maps 1a and 1b provide an overview of these differences. The colours on the maps indicate the "component scores" (data vectors multiplied by the eigenvectors) for Wartenberg's method and our method for spatial principal components. In order to save space, we do not document maps for any second largest eigenvalues, nor for regular principal components (which produces no spatial correlation).

As could already be suspected from the loadings in Figure 1, the two maps produce rather different pictures. Wartenberg's method shows high scores around the globe, but especially so in the two large countries that border on the North polar area (Canada and Russia). Large parts of Africa, the Americas and Asia also show high scores. Europe is the exception: here Wartenberg's method finds mostly lower values. This is exactly opposite for our method, which produces very high values for a cluster of West- and Central-European countries, stretching into the Caspian Sea area. This method produces also consistently lower values around the equator (on all continents).



Map 1a (top) and 1b (bottom): Component scores for Wartenberg's method (top) and our method (bottom), only largest eigenvalue

4.2. Spatial Canonical Correlation Analysis

In order to apply the spatial canonical correlation analysis, we introduce a second dataset from Gallup et al. (1999). This refers to agriculture, and consists of two main categories of variables: suitability of land (for irrigation and general soil quality) and the percentage of cultivated land in a particular climate zone. The variables are listed in Table 2.

We undertake the spatial canonical correlation analysis for the physical characteristics dataset and the agriculture dataset. This means that we will be looking, for each of the two datasets, for weights that summarize the dataset in such a way that the spatial correlation (Moran coefficient) with the summary variable from the other dataset is maximized. We start by looking at the loadings (eigenvectors), and concentrate only on the eigenvectors that belong to the highest eigenvalue. The Moran coefficient that belongs to this is 0.79 (i.e., the eigenvalue is the square of this value, 0.62).

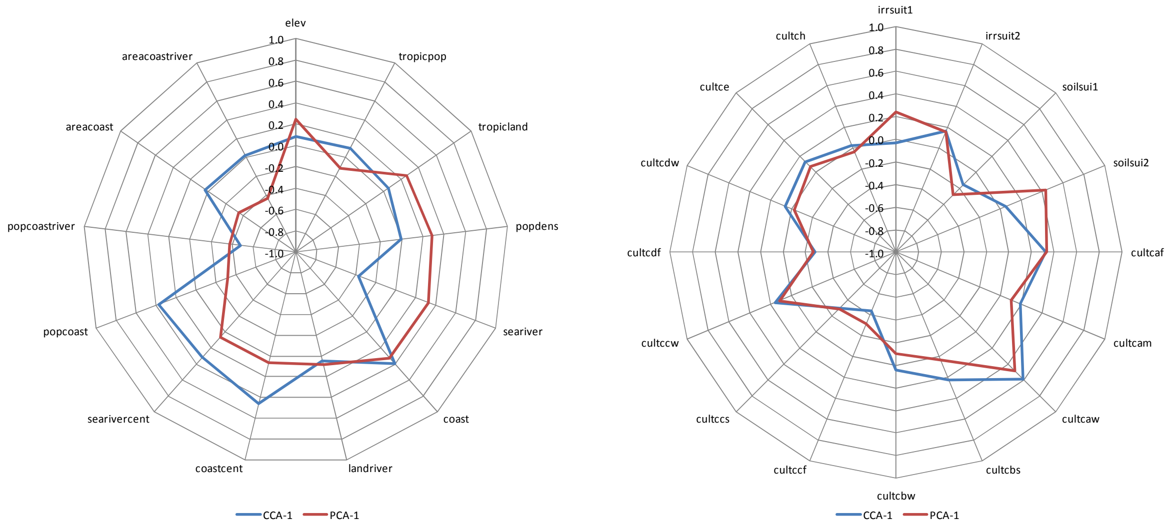
The loadings are in Figures 2a and 2b. We compare the loadings from the spatial canonical correlation analysis with those of the spatial principal components analysis (our method), to see whether any notable differences arise between the cases where we aim for maximizing spatial correlation within the dataset, or between the two datasets. For the spatial principal components analysis for the agriculture dataset, we do not document further details of this procedure, while for the physical characteristics dataset, these are the results from the previous section.

Table 2. Variables in the agriculture dataset

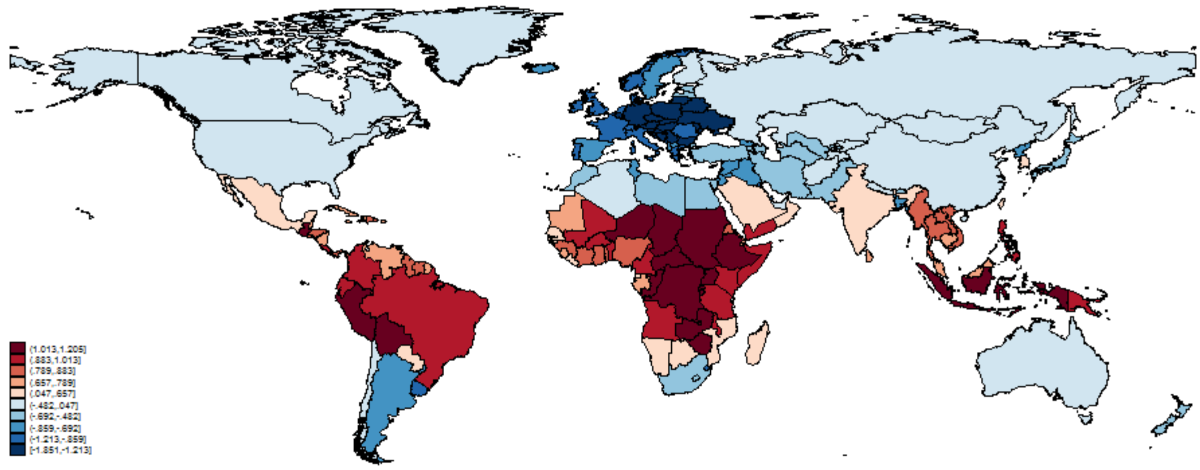
Description	Variable code
Mean irrigation suitability, very suitable (%)	*irrsuit1*
Mean irrigation suitability, moderately suitable (%)	*irrsuit2*
mean soil suitability 1, very suitable (%)	*soilsui1*
mean soil suitability 2, moderately suitable (%)	*soilsui2*
% cultivated land in Köppen-Geiger Af zone	*cultcaf*
% cultivated land in Köppen-Geiger Am zone	*cultcam*
% cultivated land in Köppen-Geiger Aw zone	*cultcaw*
% cultivated land in Köppen-Geiger Bs zone	*cultcbs*
% cultivated land in Köppen-Geiger Bw zone	*cultcbw*
% cultivated land in Köppen-Geiger Cf zone	*cultccf*
% cultivated land in Köppen-Geiger Cs zone	*cultccs*
% cultivated land in Köppen-Geiger Cw zone	*cultccw*
% cultivated land in Köppen-Geiger Df zone	*cultcdf*
% cultivated land in Köppen-Geiger Dw zone	*cultcdw*
% cultivated land in Köppen-Geiger E zone	*cultce*
% cultivated land in Köppen-Geiger H zone	*cultch*

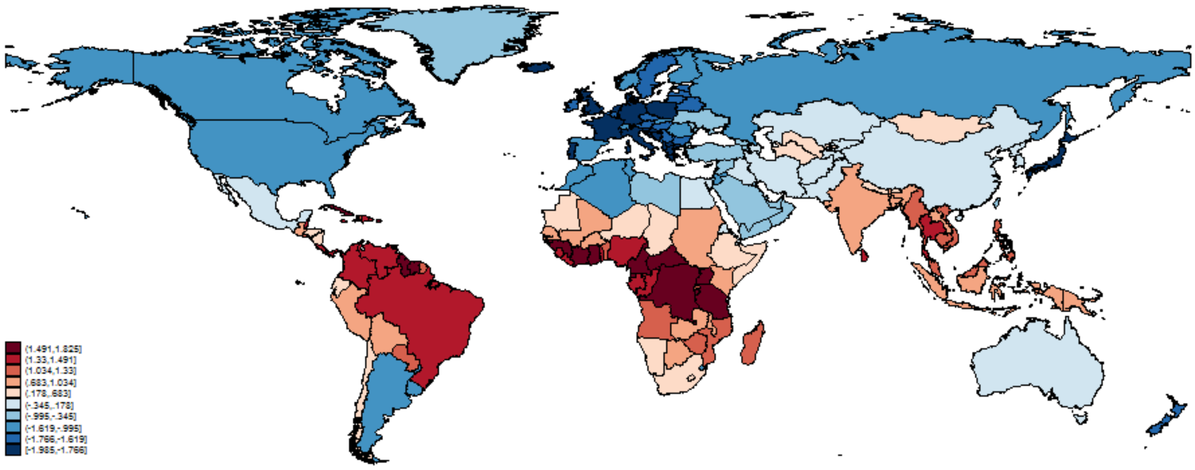
For the physical characteristics dataset, we clearly see changes compared to the previous section. Compared to the principal components analysis, the canonical correlation analysis loads much higher on *coastcent*, *searivercent*, *popcoast*, *areacoast* and *areacoastriver*, and much lower on *seariver*. Note that these are all variables related to waterways. The differences for the other variables are less strong, but still there are differences for these as well (canonical correlation analysis is lower on *elev*, *tropicland* and *popdens*, but higher on *tropicpop*).

The differences between spatial canonical correlation analysis and spatial principal components seem to be somewhat less for the agriculture dataset. Here, many variables are remarkably close in Figure 1b, with *soilsui2* as a main exception, and *irrsuit1*, *cultcbs* and *cultcbw* as more minor exceptions.



Figures 2a (left) and 2b (right): Eigenvectors for physical characteristics dataset (left) and agriculture dataset (right), largest eigenvalue only





Map 2a (top) and 2b (bottom): Component scores for spatial canonical correlation analysis between physical characteristics dataset (top) and agriculture dataset (bottom), only largest eigenvalue

Next, we look at the maps for the component scores for the two datasets, again only for the scores belonging to the largest eigenvalue. These maps are displayed in Maps 2a and 2b. The map at the top (2a) is for physical characteristics, and can therefore be compared to Map 1b above. The first thing that catches the eye in this comparison is that the current map (2a) is almost a mirror image of the previous one (1b). Here (before), we have high (low) values around the equator, and low (high) values in West- and Central-Europe. There are small deviations from this comparison (e.g., Kazakhstan, the USA), but overall, this is a rather strong similarity, despite the differences in loadings in Figure 2a.

Naturally (because of the strong spatial correlation as indicated by the eigenvalue), the map for the agriculture dataset looks similar. Here, however, the areas with high (low) values are somewhat more concentrated within the larger area around the equator (e.g., the Congo and Uganda area rather than all of sub-Saharan Africa), or within Europe.

5. Conclusions

We have presented a method for spatial principal components analysis that has two important advantages over the method that Wartenberg (1985) proposed, and which has found its place in the toolbox of computational methods for spatial analysis (e.g., the adegenet R package, see Jombart, undated). The first advantage is that, contrary to Wartenberg's method, our method has a clear and exact interpretation: it produces a summary measure of a dataset that itself has maximum spatial correlation. Thus, rather than working at an intuitive level, our method specifies exactly what is the goal of the weighting procedure that is derived using the eigenvalue decomposition of the Moran matrix.

Second, by this goal of the analysis, an easy and intuitive link can be made to canonical correlation analysis. Our spatial canonical correlation analysis produces summary measures of two datasets (e.g., each measuring a different phenomenon), and these summary measures produce the maximum spatial correlation between them. This provides an alternative weighting scheme as compared to spatial principal components analysis.

The methods that we propose are computationally easy (with modern computers), and can easily be implemented in a variety of software packages. We have Matlab routines available on request.

We provided example applications of the methods and showed that the spatial principal components analysis may produce rather different results than Wartenberg's method. We also illustrated how spatial canonical correlation analysis may produce different results than spatial canonical correlation analysis. We hope that practitioners in the field of spatial statistical analysis will apply our methods to different problems, and thus show their usefulness.

References

Cliff, A.D. and J.K. Ord, 1981, Spatial Processes: Models & Applications, Taylor & Francis

Gallup, J.L., Sachs, J.D. and A.D. Mellinger (1999), Geography and Economic Development, *International Regional Science Review*, vol. 22, pp. 179-232.

Johnson, R.A. and D.W. Wichern (2007), *Applied Multivariate Statistical Analysis*, 6th edition, Pearson Prentice-Hall, Upper Saddle River, New Jersey.

Jombart, T. (undated), A tutorial for the spatial Principal Component Analysis using the R package adegenet 1.2-2, mimeo.

Mayer, Th. & S. Zignano, 2011, Notes on CEPII's distances measures: The GeoDist database, CEPII Working Paper 2011-25, Paris: CEPII

Wartenberg, D. (1985) Multivariate spatial correlation: a method for exploratory geographical analysis, *Geographical Analysis*, vol. 17, pp. 263–283

Appendix. Our approach of a spatial canonical correlation analysis compared to standard CCA

We want to show in this appendix how our broad approach of maximizing equation (4) in the main text subject to the unitary variance of Xu and Yv relates to standard CCA. There is one main difference between standard CCA and our spatial variant: we introduce the spatial weights matrix W, (in the maximand, not in the constraints). In order to show the equivalence, we will first drop the spatial weights W, and show that in this case our approach and CCA are identical.

The starting point is our maximization problem, which can be written as the Lagrangean

$$L = (Xu)^T W(Yv) - \lambda_1((Xu)^T(Xu) - 1) - \lambda_2((Yv)^T(Yv) - 1). \qquad \textbf{(A1)}$$

The non-spatial version of our approach

Dropping W from the Lagrange function above yields the following alternative Lagrangean:

$$L = (Xu)^T(Yv) - \lambda_1 (u^T X^T Xu - 1) - \lambda_2 (v^T Y^T Yv - 1) \qquad \textbf{(A2)}$$

The maximand is $\mathrm{Cov}(Xu,Yv)$, where X and Y have the same number of rows but potentially a different number of columns. The constraints are identical to the Lagrangean A1.

The first-order conditions are

$$L_u = X^T Yv - \lambda_1 X^T Xu = 0,$$

$$L_v = Y^T Xu - \lambda_2 Y^T Yv = 0.$$

From the first-order conditions we get $u = (X^T X)^{-1} X^T Yv/ \lambda_1$ and $v = (Y^T Y)^{-1} Y^T Xu/ \lambda_2$. Insertion into the first-order conditions yields

$$L_u = X^T Y(Y^T Y)^{-1} Y^T Xu/ \lambda_2 - \lambda_1 X^T Xu = 0,$$

$$L_v = Y^T X(X^T X)^{-1} X^T Yv/ \lambda_1 - \lambda_2 Y^T Yv = 0.$$

Multiplying through by the Lagrange multiplier and $(X^T X)^{-1}$ and $(Y^T Y)^{-1}$ respectively from the left yields

$$L_u = (X^T X)^{-1} X^T Y(Y^T Y)^{-1} Y^T Xu - \lambda_2\lambda_1 u = 0,$$

$$L_v = (Y^T Y)^{-1} Y^T X(X^T X)^{-1} X^T Yv - \lambda_1\lambda_2 v = 0.$$

For the eigenvectors belonging to the highest eigenvalues then the solution is

$$\lambda_2\lambda_1 = u^T(X^T X)^{-1} X^T Y(Y^T Y)^{-1} Y^T Xu = v^T(Y^T Y)^{-1} Y^T X(X^T X)^{-1} X^T Yv.$$

Equivalence with CCA

Going back to the first set of first-order conditions above, we can show that the result is equivalent to CCA, but the above version is much simpler because it goes more directly to the eigenvectors.

Define $e = (X^TX)^{1/2}u$ and $f = (Y^TY)^{1/2}$v. Then, $u = (X^TX)^{-1/2}e$ and $v = (Y^TY)^{-1/2}f$.

Replacing u and v in the first set of first-order conditions yields

$L_u = X^TY(Y^TY)^{-1/2}f - \lambda_1(X^TX)^{1/2}e = 0,$

$L_v = Y^TX(X^TX)^{-1/2}e - \lambda_2(Y^TY)^{1/2}f = 0.$

From these first-order conditions we get $e = (X^TX)^{-1/2}X^TY(Y^TY)^{-1/2}f/\lambda_1$ and

$f = (Y^TY)^{-1/2}Y^TX(X^TX)^{-1/2}e/\lambda_2$. Using these latter equations to replace f in the first and e in the second equation yields

$L_u = X^TY(Y^TY)^{-1/2}(Y^TY)^{-1/2}Y^TX(X^TX)^{-1/2}e/\lambda_2 - \lambda_1(X^TX)^{1/2}e = 0,$

$L_v = Y^TX(X^TX)^{-1/2}(X^TX)^{-1/2}X^TY(Y^TY)^{-1/2}f/\lambda_1 - \lambda_2(Y^TY)^{1/2}f = 0.$

Multiplying through by the Lagrange multiplier and $(X^TX)^{-1/2}$ and $(Y^TY)^{-1/2}$ respectively from the left yields

$L_u = (X^TX)^{-1/2}X^TY(Y^TY)^{-1/2}(Y^TY)^{-1/2}Y^TX(X^TX)^{-1/2}e - \lambda_2\lambda_1 e = 0$

$L_v = (Y^TY)^{-1/2}Y^TX(X^TX)^{-1/2}(X^TX)^{-1/2}X^TY(Y^TY)^{-1/2}f - \lambda_1\lambda_2 f = 0$

Multiplying with e^T and f^T from the left yields and solving for $\lambda_2\lambda_1$ yields:

$\lambda_2\lambda_1 = e^T(X^TX)^{-1/2}X^TY(Y^TY)^{-1/2}(Y^TY)^{-1/2}Y^TX(X^TX)^{-1/2}e =$

$f^T(Y^TY)^{-1/2}Y^TX(X^TX)^{-1/2}(X^TX)^{-1/2}X^TY(Y^TY)^{-1/2}f$

Taking the eigenvectors and values and calculating u and v is exactly what CCA does (see Johnson and Wichern 2007, chap. 10), but here it has been obtained from our method. By implication, our method is equivalent to CCA if W is dropped.

The UNU-MERIT WORKING Paper Series

2013-01 *Effects of innovation on employment in Latin America* by Gustavo Crespi and Ezequiel Tacsir

2013-02 *Revisiting the porter hypothesis: An empirical analysis of green innovation for the Netherlands* George van Leeuwen and Pierre Mohnen

2013-03 *Impact of external knowledge acquisition strategies on innovation - A comparative study based on Dutch and Swiss panel data* by Spyros Arvanitis, Boris Lokshin, Pierre Mohnen and Martin Wörter

2013-04 *Interactive knowledge exchanges under complex social relations: A simulation modelRobin* by Cowan and Anant Kamath

2013-05 *Innovation systems framework: still useful in the new global context?* by Michiko Iizuka

2013-06 *The importance of intrinsic and extrinsic motivation for measuring IQ* by Lex Borghans, Huub Meijers and Bas ter Weel

2013-07 *Firms' innovation capability-building paths and the nature of changes in learning mechanisms: Multiple case-study evidence from an emerging economy* by Paulo N. Figueiredo , Marcela Cohen and Saulo Gomes

2013-08 *A set of time series data labour market stocks and flows for the Netherlands 1980 to 2010* by Manuel Müllers, Joan Muysken and Erik de Regt

2013-09 *Designing an optimal 'tech fix' path to global climate stability: R&D in a multi-phase climate policy framework* by Adriaan van Zon and Paul A. David

2013-10 *Complementarity between internal knowledge creation and external knowledge sourcing in developing countries* by Jun Hou and Pierre Mohnen

2013-11 *Summarizing large spatial datasets: Spatial principal components and spatial canonical correlation* by Samyukta Bhupathiraju, Bart Verspagen and Thomas Ziesemer