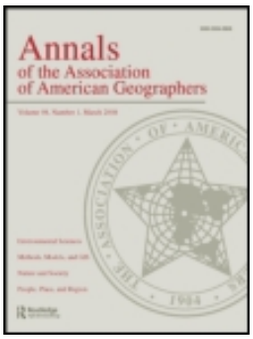

Routledge
Taylor & Francis Group

Annals of the Association of American Geographers

ISSN: 0004-5608 (Print) 1467-8306 (Online) Journal homepage: www.tandfonline.com/journals/raag20

Principal Component Analysis on Spatial Data: An Overview

Urška Demšar, Paul Harris, Chris Brunsdon, A. Stewart Fotheringham & Sean McLoone

To cite this article: Urška Demšar, Paul Harris, Chris Brunsdon, A. Stewart Fotheringham & Sean McLoone (2013) Principal Component Analysis on Spatial Data: An Overview, Annals of the Association of American Geographers, 103:1, 106-128, DOI: 10.1080/00045608.2012.689236

To link to this article: https://doi.org/10.1080/00045608.2012.689236

Principal Component Analysis on Spatial Data: An Overview

Urška Demšar,* Paul Harris,† Chris Brunsdon,‡ A. Stewart Fotheringham,* and Sean McLoone§

*Centre for GeoInformatics, School of Geography and Geosciences, University of St. Andrews
†National Centre for Geocomputation, National University of Ireland Maynooth
‡School of Environmental Sciences, University of Liverpool
§Department of Electronic Engineering, National University of Ireland Maynooth

This article considers critically how one of the oldest and most widely applied statistical methods, principal components analysis (PCA), is employed with spatial data. We first provide a brief guide to how PCA works: This includes robust and compositional PCA variants, links to factor analysis, latent variable modeling, and multilevel PCA. We then present two different approaches to using PCA with spatial data. First we look at the nonspatial approach, which avoids challenges posed by spatial data by using a standard PCA on attribute space only. Within this approach we identify four main methodologies, which we define as (1) PCA applied to spatial objects, (2) PCA applied to raster data, (3) atmospheric science PCA, and (4) PCA on flows. In the second approach, we look at PCA adapted for effects in geographical space by looking at PCA methods adapted for first-order nonstationary effects (spatial heterogeneity) and second-order stationary effects (spatial autocorrelation). We also describe how PCA can be used to investigate multiple scales of spatial autocorrelation. Furthermore, we attempt to disambiguate a terminology confusion by clarifying which methods are specifically termed "spatial PCA" in the literature and how this term has different meanings in different areas. Finally, we look at a further three variations of PCA that have not been used in a spatial context but show considerable potential in this respect: simple PCA, sparse PCA, and multilinear PCA. *Key Words: dimensionality reduction, multivariate statistics, principal components analysis, spatial analysis and mathematical modeling, spatial data.*

这篇文章批判性地探讨一个最古老和最广泛应用的统计方法之一，即主成分分析（PCA），是如何被应用到空间数据的。我们首先提供一个有关PCA如何工作的简要指南：它包括完整的和成分性的PCA变种，与因素分析，潜变量模型，多层次的PCA相关联。然后，我们提出了两种不同的把PCA应用到空间数据的方法。首先我们查看了非空间的方法，该方法通过只在属性空间使用一个标准的PCA，避免了空间数据带来的挑战。在这类方法中，我们确定了四个主要方法，并把它们定义为（1）应用于空间对象的PCA，（2）应用于栅格数据的PCA，（3）大气科学PCA，（4）流动科学PCA。在第二种方法中，我们通过查看适应第一阶非稳效应（空间异质性）和二阶固定效果（空间自相关）的PCA, 测试了适应地理空间影响的PCA。我们还描述了如何可以用PCA来研究多尺度的空间自相关。此外，我们试图通过澄清哪些方法是专门在文献中被称为"空间PCA"的，以及这个术语在不同的领域有怎样不同的含义，来消除该术语的歧义。最后，我们期待进一步观察PCA的三个变种，它们还没被应用到空间范围内，但是在这方面已显示了相当大的潜力：简单的PCA，稀疏的PCA，和多重线性的PCA。*关键词：降维，多元统计分析，主成分分析，空间分析和数学模型，空间数据。*

Este artículo considera críticamente la manera de utilizar con datos espaciales uno de los métodos estadísticos más viejos y de aplicación generalizada, el análisis de componentes principales (ACP). Antes de todo, suministramos una breve guía sobre cómo trabaja el ACP: Esto incluye variantes del ACP robustas y composicionales, vínculos con el análisis factorial, modelización de variable latente, y ACP de nivel múltiple. Luego presentamos dos enfoques diferentes para utilizar el ACP con datos espaciales. Primero, dirigimos nuestra atención al enfoque no espacial, que evita los problemas que surgen cuando los datos espaciales se utilizan con un ACP estándar de solo el espacio como atributo. Dentro de este enfoque identificamos cuatro metodologías principales, las cuales definimos como (1) el ACP aplicado a objetos espaciales, (2) el ACP aplicado a datos raster, (3) el ACP para ciencia atmosférica, y (4) el ACP para flujos. En el segundo enfoque, tratamos al ACP adaptado para efectos en el espacio geográfico, examinando métodos de ACP adaptados para efectos no estacionarios de primer orden (heterogeneidad espacial) y efectos estacionarios de segundo orden (autocorrelación espacial). También describimos la manera de utilizar el ACP para investigar múltiples escalas de autocorrelación espacial. Adicionalmente, intentamos desambiguar una confusión de terminología aclarando qué métodos son específicamente denominados "ACP

Annals of the Association of American Geographers, 103(1) 2013, pp. 106–128
Initial submission, November 2010; revised submissions, May and September 2011; final acceptance, September 2011
Published by Taylor & Francis, LLC.

espacial" en la literatura y cómo esta expresión tiene significados diferentes en áreas distintas. Por último, dirigimos nuestra atención a tres variaciones adicionales del ACP que no han sido usadas en un contexto espacial pero que muestran considerable potencial en este respecto: ACP simple, ACP ralo y ACP multilineal. *Palabras clave: reducción de dimensionalidad, estadísticas multivariadas, análisis de componentes principales, análisis espacial y modelización matemática, datos espaciales.*

Following its introduction at the beginning of the twentieth century by Pearson (1901) and Hotelling (1933), principal components analysis (PCA) has been used in many different disciplines, including agriculture, biology, chemistry, climatology, demography, ecology, economics, genetics, geography, geology, meteorology, oceanography, and psychology. The purpose of this article is not to give a full historic overview of its use (see Jolliffe [2002] for an extensive review) but to highlight the need for special types of PCA for use with spatial data and to investigate how different versions of PCA have been and should be used on spatial data.

This overview is intended for geographers who might want to use PCA in some way for their particular problems and data. The hope is that this overview might help them select an appropriate version of the method or suggest an improvement to their existing methodology. Therefore, we attempt to present the material in a very general form without going into details, to try to make it accessible to the widest possible audience. For a more expert reader, the underlying theory can be found in the method-specific references cited and in the comprehensive book by Jolliffe (2002). Key historical literature on the use of PCA from a geographer's perspective includes the work of Berry (1964, 1966, 1968a, 1971), Gould (1967), Hägerstrand (1967), Tinkler (1972), Mather and Openshaw (1974), Goddard and Kirby (1976), Daultrey (1976), and Johnston (1978). In many of these articles, there is much interchange between the use of PCA and factor analysis (FA), where for applications in urban geography their use came under a general umbrella term of *factorial ecology*.

Spatial data contain geographic as well as attribute information. Thus, whereas typical data sets only contain measurements of variables or attributes, spatial data sets are characterized by having a location associated with each measurement; that is, the geographic location within the basic three-dimensional framework of our physical world, where the measurement was taken. In contrast with nonspatial data, the data space can therefore be separated into two distinct components: geographic space and attribute space. Occasionally, if temporal information is also present, then time forms a third component, the temporal space. As such, the data space can consist of n-dimensional attribute space, three-dimensional geographic space, and one-dimensional temporal space, where the space–time components provide the framework for attribute space.

Two properties that can make spatial data special and different from nonspatial data are spatial heterogeneity and spatial autocorrelation. *Spatial heterogeneity* refers to the nonstationarity of geographic processes, meaning that processes can vary locally and are not necessarily the same at each spatial location. Commonly, this nonstationarity is modeled as a first-order (mean response) or second-order (variance) effect. With respect to spatial heterogeneity, in this article we limit ourselves to *nonstationary* first-order effects only, where such effects change across space. *Spatial autocorrelation* is the tendency of attributes at some location in space to be related. Spatial autocorrelation is a second-order effect and we limit ourselves to *stationary* second-order effects only (noting that nonstationary second-order effects are possible). The presence of spatial heterogeneity and spatial autocorrelation invalidates two basic assumptions of many standard statistical analyses: that data are independently generated and identically distributed. As a consequence, using a standard statistical methodology, including PCA, on spatial data poses particular challenges. Analogous effects are possible in temporal space (and spatiotemporal space combined), but we do not discuss them here.

In this article, we distinguish between two different approaches to using PCA on spatial data: (1) those that avoid spatial challenges altogether by using a standard nonspatial PCA and (2) those that adapt PCA for spatial effects with respect to spatial heterogeneity or autocorrelation. Although most spatial applications of PCA stem from the geosciences (physical geography, geology, geochemistry, atmospheric sciences, environmental sciences, etc.) and, to a lesser extent, the social sciences (human, social, economic geography), we focus on the manner in which PCA is applied to spatial data, rather than on the discipline-specific topics themselves.

Principal Components Analysis

Data dimension is the number of variables measured at each observation. Many spatial data sets are highly

dimensional and as such can be difficult to visualize and interpret. However, there often exists a smaller intrinsic dimensionality in the data set, where not all of the variables are needed to convey the information relevant to an understanding of the underlying process. Therefore, it is often of interest to reduce the dimensionality of the data. Methods for dimensionality reduction attempt to capture the maximum information present in the original data, at the same time minimizing the error between the original data and the new lower dimensional representation (Donoho 2000; Fodor 2002; Afifi, Clark, and May 2004).

PCA is one of the most popular dimensionality reduction methods. It is a linear method, meaning that the transformation between the original data and the new lower dimensional representation is a linear projection. Its main purpose is dimensionality reduction, but it can also be used to explore relationships between variables. Often it is used as a preprocessing method either for data orthogonalization and eliminating redundancy caused by variable correlation or for dimensionality reduction, before employing another statistical method, such as regression or clustering (Fodor 2002; Jolliffe 2002). As principal components (PCs) are orthogonal, regression and clustering methods can proceed with data independence assured.

PCA maps the original n dimensions (variables) of the data matrix $\mathbf{X}$ onto a new orthogonal space, such that the new axes are oriented in directions of largest variance in the data. The new dimensions are called the PCs and are mathematically defined as follows.

PCA is a factorization or decomposition of an $m \times n$ matrix $\mathbf{X}$, with m measurements and n variables, such that

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}}, \tag{1}$$

where $\mathbf{P}$ is an orthonormal projection matrix (i.e., $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}$) and $\mathbf{T}$ is the projection of n-dimensional $\mathbf{X}$ onto the new r-dimensional space defined by $\mathbf{P}$; that is,

$$\mathbf{T} = \mathbf{X}\mathbf{P}. \tag{2}$$

Matrix $\mathbf{P} \in \Re^{n \times r}$ is referred to as the *loading matrix* and $\mathbf{T} \in \Re^{m \times r}$ is referred to as the *score matrix*. The dimension r is the number of independent columns in $\mathbf{X}$ (i.e., the rank of $\mathbf{X}$) and is bounded by the minimum of m and n. $\mathbf{P}$ is computed so that its columns are the directions of maximum variance in the data, with the first column (or PC1) representing the direction of maximum variance, the second column (PC2) the direction of the next largest variance, and so on. These directions correspond to the eigenvectors of either the data covariance or correlation matrix, Σ, where

$$\Sigma = \frac{\breve{\mathbf{X}}^{\mathrm{T}}\breve{\mathbf{X}}}{m-1}. \tag{3}$$

Here, $\breve{\mathbf{X}}$ denotes $\mathbf{X}$ with the mean removed from each column in the case of covariance and $\mathbf{X}$ with each column standardized to have zero mean and unit variance in the case of correlation. By definition $\mathbf{\Sigma}$ is a positive semidefinite matrix and therefore its eigenvalues are greater than or equal to zero. Hence, ordering the eigendecomposition of $\mathbf{\Sigma}$ so that the eigenvalues are in descending amplitude order gives

$$\boldsymbol{P}\Lambda\boldsymbol{P}^{\mathrm{T}} = \Sigma \tag{4}$$

where $\mathbf{P}$ is the score matrix and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues; that is,

$$\begin{aligned}\mathbf{\Lambda} &= diag(\lambda_1, \lambda_2, \ldots, \lambda_r, 0, \ldots 0), \text{ with} \\ &\lambda_1 \geq \lambda_2 \cdots \geq \lambda_r \geq 0\end{aligned} \tag{5}$$

In many situations $\mathbf{X}$ can be approximated by a small number of PCs, k, where $k << r \leq n$, while still explaining most of the variance in the data; that is, where $\mathbf{\Lambda}$ only has a small number of large eigenvalues and many small ones. Denoting $\mathbf{P}_k$ as the matrix containing the first k columns of $\mathbf{P}$ (i.e., the most significant PCs) then the corresponding scores matrix is given by

$$\mathbf{T}_k = \mathbf{X}\mathbf{P}_k \tag{6}$$

and the proportion of the total variance explained by $\mathbf{T}_k$ is given by $\frac{v_k}{v_r} \times 100$, where

$$v_k = \sum_{i=1}^{k} \lambda_i \text{ and } v_r = \sum_{i=1}^{r} \lambda_i = trace(\mathbf{\Lambda}) = trace(\mathbf{\Sigma}). \tag{7}$$

From this description it follows that each PC corresponds to the direction of one eigenvector and is a linear combination of the original variables. Because PCs are ordered according to the size of their respective eigenvalues, starting with the largest, this means that the new space of the PCs is oriented so that the first few PCs are aligned with the directions of the largest variance in the data; that is, the first PC represents the direction in which the variance of data is the largest,

the second PC the direction of the next greatest variance, and so on. If the first few dimensions (k) explain most of the variance in the data, the rest can usually be disregarded with minimal loss of information. Dimensionality reduction is performed by taking the first k PCs, where $k << n$ such that the new k-dimensional space contains the majority of the information according to some criterion. A key decision is the size of k; that is, how many PCs should be retained? The answer is study dependent and has to be determined by examination of the data. A number of heuristic methods exist for this purpose. The k selected PCs are sometimes referred to as *unobserved latent variables*.

PCs will differ by the choice of the matrix $\mathbf{\Sigma}$ used for their calculation. The covariance matrix is scale dependent and should only be used when all variables have the same measurement units. If the measurement units of variables differ in size and type, then the scale-independent correlation matrix should be used instead to standardize the original variables. This avoids a domination of variables with the largest measurement units in the first few PCs (Maćkiewicz and Ratajczak 1993; Jolliffe 2002).

With respect to data set structure, PCA can be run in two ways: either in the so-called R-mode or Q-mode. In R-mode, the goal is to identify combinations of variables that explain the pattern of variation among the objects—this is the standard way of running PCA. Q-mode PCA, which is sometimes referred to as *inverted PCA*, focuses on combinations of samples that explain variation among variables. That is, the PCA is run on a data set where the matrix of samples and attributes is transposed so that the roles of the variables and measurements are reversed (Tanaka and Zhang 1999; Choulakian 2001; Schuenemeyer and Drew 2011). If time is added as one of the measurements, resulting in space–time series data, then there are in total six different modes—O, P, Q, R, S, and T—each of which addresses a different combination of time, objects, and attributes (Richman 1986). We explain these modes more fully later, as we specifically deal with space–time series data.

Statistical inference for PCA deals with estimating characteristics of the PCs defined by the entire population given the PCs derived from a data sample. The key limitation is that inference should only ever be attempted when the data are (at least approximately) multivariate normal (Jolliffe 2002). If this requirement were to be imposed every time, it would limit the use of PCA to only a very small number of cases, because in reality, true multivariate normality is rarely the case. In our context, PCA should be seen as a descriptive methodology rather than an inferential one, as it can produce valuable information regardless if the attributes are normally distributed or not. That is, it provides a view on the structure of data as they are within the sample, rather than attempting to infer characteristics of the entire population. This is the perspective that the majority of applications in this overview would take. Details on the inferential side of PCA can be found in Jolliffe (2002).

Robust PCA and Outlier Detection

When there are outliers in the sample data, basic statistical methods often produce unreliable results, as the presence of outliers violates basic assumptions of the methods. This is usually prevented by using a robust version of the same method. By construction, robust methods also detect outliers and a robust version of PCA can be used for multivariate outlier detection via dimensionality reduction, so that in the resultant transformed (PC) space, outliers are more readily observable. PCA in a basic form is not very robust to outlying observations (i.e., its covariance estimates are nonrobust) and, as such, is not ideally suited to their detection. In this respect, numerous robust PCA-based techniques (together with their associated outlier detection tools) have been proposed (see Jackson and Chen 2004; Rousseeuw et al. 2006; Daszykowski et al. 2007; Stanimirova, Daszykowski, and Walczak 2007). Here PCA can be made robust to outliers by using (1) some robust covariance estimator (such as a reweighted minimum covariance determinant [MCD]; Croux and Haesbroeck 2000), (2) a projection pursuit (PP) technique where projections of the data are searched for outliers (Hubert, Rousseeuw, and Verboven 2002), and (3) a mixture of both MCD and a PP technique (Hubert, Rousseeuw, and Vanden Branden 2005). Many robust PCA techniques are computationally intensive and, as such, computationally fast algorithms are required for analyzing large high-dimensional data sets (e.g., see Filzmoser, Maronna, and Werner 2008).

Compositional PCA

In some cases, data have the property that all attributes sum to a constant; that is, they are descriptions of a part of some whole and give only relative information. An example of this is proportional values that sum to one—this routinely occurs in geochemistry when water or soil samples are taken and the proportions

in each sample of a number of chemical elements are measured (e.g., Filzmoser, Hron, and Reiman 2009). This constraint means that the attribute values of such data occur in a space limited by a simplex and are therefore closed in space, which affects the structure of the correlations. A standard multivariate method such as PCA is therefore not applicable to such data. Instead, a common way to model this type of data is to use a technique from compositional data analysis (Aitchison 1982; Aitchison and Egozcue 2005), which replaces variables with logarithmic ratios of variables (or some related transformation). This removes the constraint of the simplex and transforms the variables into an unconstrained multivariate space and consequently allows a standard statistical method to be applied. Specifically, compositional PCA (Aitchison 1983) calculates PCs of log ratio transformations of the raw data. Many compositional data sets are inherently curved; that is, the largest variance is distributed along a curved line and not a new straight line dimension. Compositional PCA is able to correctly pick up this curvature, whereas a standard PCA, which only produces linear combinations of variables, is powerless (Jolliffe 2002). However, the log ratio transformation in compositional PCA further complicates the already difficult interpretation of PCs (Aitchison and Egozcue 2005).

PCA and Links with Factor Analysis

As suggested, a known disadvantage of PCA is that the PCs do not always correspond to meaningful physical variables. Indeed, there is no reason why a purely mathematically calculated linear combination of variables should have a physical meaning (Jolliffe 2002). PCs are therefore not always easy to interpret. One attempt to solve this problem is to rotate the PCs post-analysis into new dimensions that might have an easier-to-interpret connection with the original variables. This is commonly done using varimax, covarimax, or similar rotations that originate in factor analysis (FA) and were designed to maximize the variance of the factors (Kaiser 1958). Although this is a relatively common approach (e.g., Widmann and Schär 1997; Frank and Esper 2005; Esteban, Martin-Vide, and Mases 2006), the usefulness of postrotation of PCs is debatable, as PCs already maximize the variance, and any further rotation of the axes in the PC space, while preserving the amount of the variance, might change the ordering according to variance size (Daultrey 1976; Richman 1986; Jolliffe 2002).

PCA is sometimes considered to be a special form of FA. This, however, is not quite correct, because although both methods have a similar goal of dimensionality reduction to a number of latent variables, their postulations are very different. As described earlier, PCA can be defined as a technique that identifies a set of linear combinations of original variables with no preassumed models: PCs are defined simply as directions of largest variance. FA, on the other hand, attempts to achieve dimensionality reduction by assuming the existence of k latent variables or factors (where $k << n$ and n is the number of variables), such that each original variable is a linear combination of factors. Factors are separated into common factors, which contribute to all variables, and specific factors, each of which contributes to only one particular variable and describes the variable-specific model error (which can be either an observational or measurement error). Thus, FA is defined in the familiar format of a "deterministic term + random error term," used, for example, in regression models. The factors define a model and depend on known parameters such as k defined in the text immediately preceding and unknown parameters to be estimated, such as the component loadings. There is no explicit model in the classic derivation of PCA whose results are dependent on data only (although it is possible to consider model-based PCA, as one solution of the maximum likelihood calibration of FA where k factors coincide with the first k PCs; see next subsection on latent variable modeling and Jolliffe [2002]). Standard PCA finds a data-defined linear transformation from an n-dimensional space to another n-dimensional space and no additional parameters have to be specified. To summarize, FA provides us with a model of the lower dimensional space, whereas PCA produces a unique data-driven projection. This is the most fundamental difference between the two methods. An interested reader can find details on FA in any textbook on multivariate statistics (e.g., Afifi, Clark, and May 2004). Detailed differences between PCA and FA are discussed in Jolliffe (2002).

Latent Variable Modeling and Multilevel PCA

Some recent developments in PCA are latent variable modeling and multilevel PCA. From a theoretical perspective, these can be best understood by adopting a model-based interpretation of PCA—see, for example, Anderson (1984) or Whittle (1953). Although often presented as an exploratory algorithm, PCA can be viewed as the solution to a maximum-likelihood model

calibration problem, where the data matrix is represented as a matrix of lower rank than its number of columns. In standard PCA, the data are assumed to be such a matrix plus an independent Gaussian error term for each element—but the model can be modified to represent a number of alternative situations. One such example is that of a spatial PCA, where the statistical distribution of the error terms reflects the spatial structure of the observations. This approach, in which the data might be thought of as a linear combination of a number of unobserved variables (PCs) plus a random error, is sometimes referred to as *latent models* or *latent variable models* (see subsection on regionalized PCA).

Another modification can be used to reflect a hierarchical structure in the data. This means that the data set consists of individual data records for which factors are calculated at different levels of aggregation. An example would be data collected on the level of individual students and the two higher aggregation levels were schools and the aeral units to which each school belongs (Goldstein and Browne 2005). In such cases, the variability in the data can be represented by random terms at different levels in the hierarchy. In turn, this implies that the models can be thought of as having components at different levels of the hierarchy. Goldstein and Browne (2005) applied this approach to investigate Organisation for Economic Co-operation and Development (OECD) data recorded for thirty-two industrialized countries (OECD 1999) including tests of reading, mathematics, and science. Factors at country, school, and individual student levels were considered and component loadings related to responses to individual questions in the tests were estimated at school and country levels.

In this model-based approach, Bayesian estimation of the loadings can also be used and through techniques such as Markov chain Monte Carlo (MCMC) estimation, a very wide portfolio of models and modifications to existing models can be considered. For example, the variables in the data matrix might be categorical and the link between the component scores and the observed data might take the form of a logistic regression.

Standard Nonspatial PCA on Spatial Data

We have identified four main methodologies for using standard nonspatial PCA (as described in previous section) on spatial data, which we refer to as follows:

- Spatial objects PCA
- Raster data PCA
- Atmospheric science PCA
- PCA on flows

In the following, we give a description of each of these approaches and list some examples from various academic disciplines. Table 1 provides a summary.

Spatial Objects PCA

One of the most common uses of PCA on spatial data is in studies where spatial data consist of spatial objects. These are typically either irregularly spaced points (e.g., sampling sites of environmental measurements) or areas (e.g., watersheds or administrative districts). Variables are measurements of several different properties (characteristics) at each point or area location. In these studies, PCA is run on the entire data set, statistical software is (commonly) used for processing rather than a geographical information system (GIS), and geographical effects do not play any role in the PCA itself. Results give a global summary of the data and are presented nonspatially using tables and statistical summaries. Figure 1 shows the schematic flow of this methodology. Observe that it is possible to map the PC scores (i.e., the transformed data values after applying PCA) for each PC, as they correspond to each vector of observations at each spatial location of the data set. Regionalization studies in the social sciences provide numerous examples of this practice (see the text immediately following).

Spatial objects PCA is commonly used for dimensionality reduction or as a data preprocessing method, as with any nonspatial application. Examples can be found in many of the geosciences: in environmental sciences for environmental indices (Tran et al. 2002; Parinet, Lhote, and Legube 2004); for atmospheric, soil, and water pollution (Hernández, Adarve Alcazar, and Pastor 1998; Felipe-Sotelo et al. 2006; Zhang 2006); in environmental geochemistry (Zhang and Selinus 1998; Reid and Spencer 2009); in environmental management (Bastianoni et al. 2008); and in biogeography for wildlife and vegetation distribution studies (Antunes et al. 2008). There are also numerous instances of such simple PCA applications in the social sciences; for example, as a data preprocessor prior to statistical modeling—see the regression modeling of fire and rescue incidents (Corcoran et al. 2007) and the nonstationary regression modeling of hedonic house price data (Bitter, Mulligan, and Dall'erba 2007).

A common application in the social sciences is to develop composite social indices for development, health, or quality of life from the first few PCs using

Table 1. Use of four different standard principal components analysis methodologies on spatial data

Discipline	Spatial objects PCA	Raster data PCA	Atmospheric science PCA	PCA on flows
Atmospheric science, climatology, meteorology			X	
Biogeography: Ecology, vegetation, and wildlife	X	X		
Dendrochronology			X	
Environment: Atmospheric pollution		X	X	
Environment: Soil and groundwater pollution	X	X		
Environmental geochemistry	X			
Environmental management	X			
Remote sensing		X		
Geology, sedimentology	X	X		
Historical geography: Regionalization	X			
Human geography: Migration and spatial interaction				X
Human geography: Regionalization	X			
Seismology			X	
Social indicators research: Composite indices	X	X		
Social sciences: Preprocessing/orthogonalization	X			
Transportation				X
Underwater acoustics	X			

Note: PCA = principal component analysis.

sociodemographic input variables. Here PCA is used on administrative units, both for dimensionality reduction and to explore relationships between variables (Boelhouwer and Stoop 1999; Booysen 2002; Fotso and Kuate-Defo 2005; Anselin, Srihdaran, and Gholston 2007; Lengen and Blasius 2007; Kelly and Teljeur 2007). Spatial objects PCA has also been used to link remotely sensed (RS) data with area census data for quality of life indices (Lo 1997).

In human (social and economic) geography, spatial objects PCA has been used for regionalization (Gould 1967; Daultrey 1976; Hall 1977). Here PCs are calculated on areal data with the goal being to aggregate similar areal units into regions (internally cohesive larger spatial units) through the first few PCs (via their respective PC scores). Identified regions can be environmental, geographical, or social, depending on the study. Examples of this (more spatially oriented) use of standard PCA can be found in Skånes and Bunce (1997) with respect to landscape dynamics, Horner and Grubesic (2001) with respect to transportation planning, and Campbell and Power (1989) with respect to historical geography.

Finally, spatial objects PCA is routinely applied in geological studies (Davis 1986). Examples include the study of marine mineral properties (Andrews 2008) and seabed classification from acoustic data (Preston 2009). Commonly in geological (and geochemical) studies, sample data are compositional and, as such, compositional PCA is applied (Thomas and Aitchison 2005; Thió-Henestrosa and Martín-Fernández 2005; Reyment 2006; van den Boogaart and Tolosana-Delgado 2008). These same disciplines are also at the forefront in the application of robust forms of PCA and outlier detection (Filzmoser 1999; Filzmoser, Garrett, and Reimann 2005; Filzmoser, Hron, and Reiman 2009).

Raster Data PCA

Our second identified methodology for applying PCA occurs in the analysis of raster data. Here PCs are calculated for a data set where data elements are cells of raster surfaces (or locations in the center of raster cells) with measurements of several variables at each location. This type of PCA focuses on the creation of PCA maps—new rasters, where each pixel is assigned a value or score in each new PC dimension (Eastman 2003). As with the analogous methods in the previous section, geographical effects are not accounted for in the calculation of the PCs, as the analysis is run strictly on attribute space only. It should be noted that, mathematically, this methodology is a variation of spatial objects PCA from the previous section; the only difference is that in this case "spatial objects" correspond to regular grid cells or locations of their centers. Based on the studies that we found, however, we decided that it warrants a separate description because, in contrast with

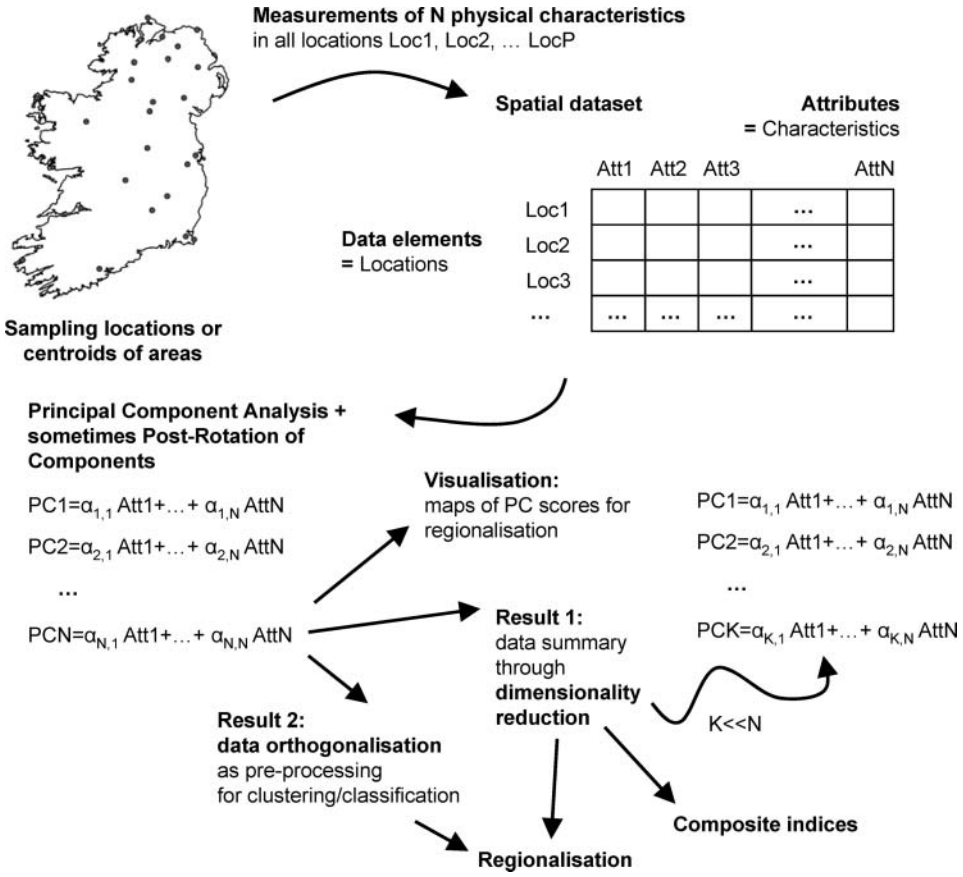

Figure 1. Spatial objects principal component analysis (PCA).

the previous methodology, there is a focus on the spatial distribution of results. The new PCs are almost always displayed in map form and often used in subsequent spatial analysis based on map algebra. The schematic flow of this methodology is shown in Figure 2.

This methodology is commonly used when satellite or other RS data need to be combined with other types of raster data; for example, with interpolated surfaces of some meteorological or soil variables, with interpolated animal counts in zoology, and with rasters of socioeconomic and census data in the social sciences. Analysis is usually performed using the PCA functionality provided in a GIS, typically Idrisi Kilimanjaro or ArcGIS Spatial Analyst. Statistical software is rarely used.

Raster PC maps are often used to produce composite indices that describe a certain subset of data with particular properties. Again, indices are derived using the first few PCs and describe inter-variable relationships. Examples include ecological susceptibility (Hoersch, Braun, and Schmidt 2002), wildlife distribution patterns (Khaemba and Stein 2000; Brito et al. 2008; Ngene et al. 2009), water pollution (Satapathy, Salve, and Katpatal 2009), the ecogeographical analysis of underwater acoustic bathymetry data (Verfaillie et al. 2009), and other topics in the environmental sciences (e.g., Arbia, Griffith, and Haining 2003; Li et al. 2006; Maina et al. 2008; Shi et al. 2009).

Again with RS data, raster data PCA can be used for change detection, feature detection of natural and man-made features, and classification of spectral classes. This has been done for a variety of RS source data; for example, Landsat TM bands (Collins and Woodcock 1996; Floras and Sgouras 1999; Aminzadeh and Samani 2006), multispectral and hyperspectral imagery (Goovaerts, Jacquez, and Marcus 2005; Panda, Hoogenboomb, and Pazb 2009), and near-infrared astronomical imagery (Klassen 2009). An example in the social sciences can be found in Lo and Faber (1997), where RS data are linked with census data to develop composite social indices.

In some studies, input rasters are weighted by multiplying them with another raster that describes a particular spatial relationship or distribution, resulting in the so-called spatially weighted PCA method (W. Wang and Cheng 2008; this is not to be confused with the geographically weighted PCA method described in the next section, which is an entirely different method with

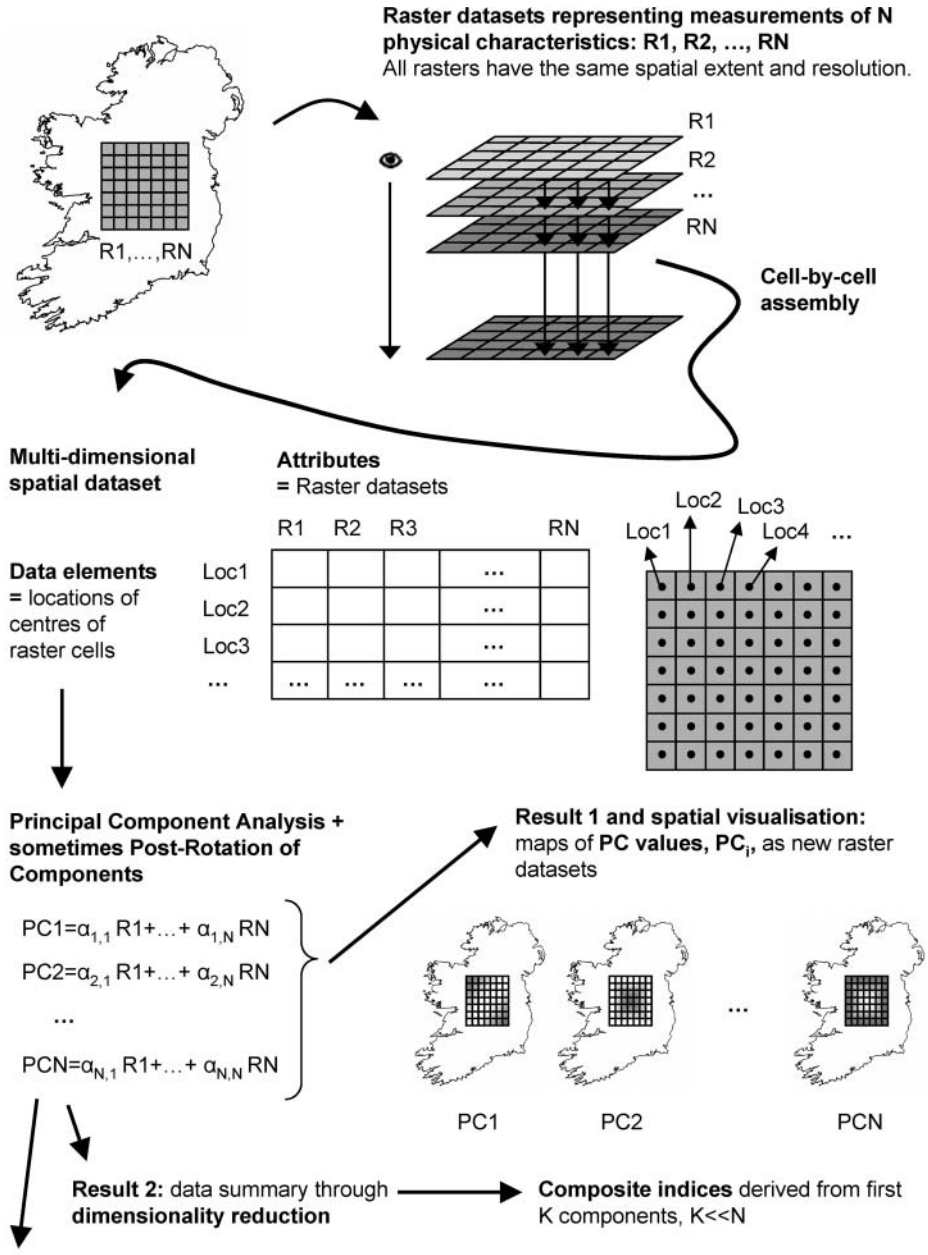

Figure 2. Raster data principal component analysis (PCA) and its spatial visualization—raster principal component (PC) maps. Location is only relevant for visualization, not computation, as PCA is run on attribute space only.

different objectives; see the final subsection of the next section for disambiguation).

Atmospheric Science PCA

Our third methodology comes from the atmospheric sciences, where standard PCA is applied to spatio-temporal data; that is, a time series of measurements collected at specific spatial locations. It is considered a standard analytical tool in climatology and meteorology and has been in use since the 1950s.

Data for atmospheric science PCA consist of time series of measurements of one particular meteorological field variable (this can be air temperature, sea-level pressure, or similar), measured at equidistant time intervals at each sampling location. Observe that only one variable is measured in this approach, as opposed to the previous two methodologies, in which several variables are measured. Sampling locations are either

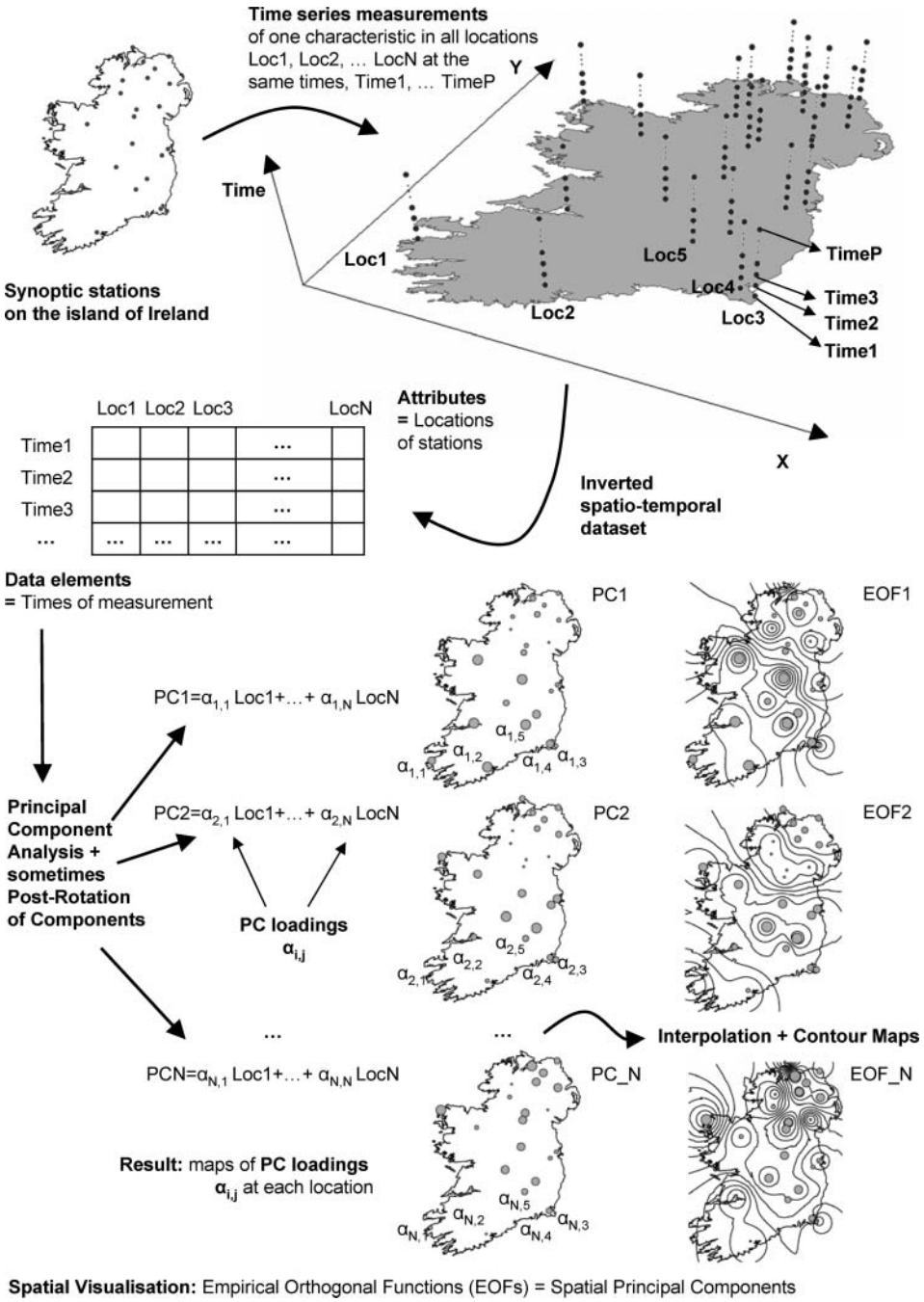

Figure 3. Atmospheric science principal component analysis (PCA) and its spatial visualization—empirical orthogonal functions (EOFs)/spatial principal components. Location is implicitly involved in the process through inversion of the spatio-temporal data set, where sampling locations are considered attributes and sampling times data elements of the data set on which the PCA is run.

locations of meteorological stations or centers of grid cells of meteorological raster data. In the previous two approaches, sampling locations represent data elements and field measurements variables. Atmospheric science PCA inverts this concept and considers sampling locations as variables and sampling times as data elements. Thus, PCs are calculated from the transposed data set, where the covariance/correlation matrix is $n \times m$ (as opposed to $m \times n$ in the nontransposed case). Again the number of PCs is bounded by $\min(n, m)$. Note that, because of the conceptual inversion of the data set, covariance/correlation is calculated between each pair of sampling locations and not between two measured variables as before. Therefore, location does play a role in calculation of the PCs, albeit implicitly. Figure 3 schematically shows this process.

As mentioned in the previous section, when dealing with space–time series data there are six different operational modes of use of PCA: O, P, Q, R, S, and T. These modes are defined by having the data

matrix for PCA defined by two out of the three subspaces of the space–time data set. The three subspaces of the space–time data set are geographic space, temporal space, and attribute space. The six modes are then defined as follows (Richman 1986):

- O-mode data matrix is between attributes and time (i.e., attributes are considered data elements and sampling times variables for PCA).
- P-mode: time vs. attributes (i.e., sampling times are data elements and attributes variables).
- Q-mode: attributes vs. locations.
- R-mode: locations vs. attributes (spatial objects PCA and raster PCA).
- S-mode: time vs. locations (atmospheric science PCA).
- T-mode: locations vs. time.

Different modes provide different insights into data—details of which can be found in Richman (1986). The S-mode is the most common one in atmospheric science, however, and is the one described in the preceding text.

The resulting PCs are sometimes referred to as empirical orthogonal functions (EOFs, termed *empirical* because they originate from observed values of the meteorological field; North, Bell, and Cahalan 1982; Jolliffe 2002). Because each PC/EOF is a linear combination (i.e., a weighted sum) of all locations, a map can be produced for each PC/EOF. Here the calculated weights of the respective PC/EOF at each sampling location are spatially interpolated to form a contour map, which is then inspected for spatial patterns. Often patterns from several of the first few PC/EOF maps correspond to typical situations in the atmosphere at particular times of the temporal period studied (Jolliffe 2002). Furthermore, the PCs/EOFs are often rotated postanalysis (as in FA) to facilitate the interpretation of the resulting component maps. In atmospheric science, there are several types of rotations, some of which are orthogonal (e.g., varimax, quartimax, and equimax rotations, which preserve the orthogonality of the PCs) and others oblique (these produce correlated rotated PCs). The aim of all of these rotations is to discover a transformation of the PCs that results in a so-called simple structure—a description of the data set with the smallest necessary number of rotated PCs, which are oblique yet still constitute a set of linearly independent vectors in original PC space (Richman 1986). This structure was defined by Thurstone (1947) for FA (cited in Richman 1986), but simple structure rotation is often applied in atmospheric PCA, in spite of the controversy over post-PCA rotation being necessary or not (see subsection on PCA and factor analysis). In meteorology it has been observed that such rotated PC/EOF maps are more similar to particular observed weather situations (Yarnal et al. 2001; Jolliffe 2002; Zwiers and Von Storch 2004).

It is important to note the difference in the information represented by this approach compared to raster data PCA and the difference in their spatial visualizations. Raster PC maps show values (scores) of PCs (latent variables) at each location (represented as a grid cell), as each location is considered as a data element in the spatial data set. This is in contrast to atmospheric science PC/EOF maps, which show PC loadings at each location (because each location is considered to be a variable and not a data element). To emphasize the difference between the two, the PC/EOF maps present a spatial distribution of the importance of one single meteorological variable at each particular location, whereas raster PC maps show the distribution of values of each latent variable (PC) at each location, where each PC is a linear combination of the original variables, thus reflecting several variables, not just one.

Detailed historical reviews of atmospheric science PCA can be found in Jolliffe (2002) and Esteban, Martin-Vide, and Mases (2006). Some noteworthy uses include investigating sea-level pressure (Esteban, Martin-Vide, and Mases 2006; Lopez-Bustins et al. 2007), precipitation (Widmann and Schär 1997; Krepper and García 2004), and synoptic climatology (Yarnal et al. 2001). Again, this use of PCA often precedes the use of some clustering or classification algorithm; for example, circulation pattern classification, classification of weather types, climate regionalization, atmospheric circulation reconstructions, circulation anomalies associated with natural climatic variability, and climate series reconstruction (see Esteban, Martin-Vide, and Mases [2006], for a review). There is also a linkage between atmospheric science PCA and the modeling of nonstationary spatial autocorrelation structures in spatiotemporal data sets (Obled and Creutin 1986; Sampson, Damian, and Guttorp 2001). This topic directly relates to the methods of the next section, but its description is beyond the scope of this study.

Finally, the use of atmospheric science (i.e., S-mode) PCA can also be found in other disciplines that similarly collect long-term time series data at spatial locations—one example is the dendrochronological study of Frank and Esper (2005), where time series

of tree ring width and wood density were converted into PC/EOF maps, which were then compared with climate maps to identify similarities between properties of tree growth and climatic conditions. Other examples can be found in environmental pollution analysis (Ibarra-Berastegi et al. 2009) and in seismology for earthquake series data (Savage 1988; Holliday et al. 2006).

PCA on Flows

Our fourth methodology is very different than the previous three and concerns the use of PCA in the identification of structure in flow matrices (Berry 1966, 1968a, 1968b). Flow matrices are spatial in that they can be, for example, a data matrix of the flow of migrants between countries (Magee 1971; Hay and Raihan Sharif 1986). In this particular example (both references analyzed the same data), after conducting a standard PCA on the raw data flow matrix, one can relate each PC to a subsystem of flows emanating from a country (or countries) and terminating in another, where the first PC has the strongest subsystem of flows. Again, in this example, flows originating from Spain and Portugal and terminating in France were found to be the strongest, reflecting their position as neighboring countries and their political and economic status at the time of the study. Other examples of this use of PCA can be found in Goddard (1970) with taxi data, Black (1973) with transportation of commodities, and both Goddard (1973) and Clark (1973) with phone call data. These studies belong to geography's historical literature and are often labeled as factorial ecology (see introduction). A more recent application of the same underlying methodology can be found in Reades, Calabrese, and Ratti (2009), where the spatiotemporal structure of a rasterized representation of a mobile phone network in Rome, Italy, is characterized using PCA.

PCA Adapted for Spatial Effects

In this section, we describe three truly spatial PCA techniques that are specifically designed for, or account for, spatial effects in spatial data. First, we look at two closely related PCA techniques that are adapted locally in attribute and in geographical space, respectively. These techniques adopt a nonparametric, kernel-based approach and are termed *locally weighted PCA* (LWPCA) and *geographically weighted PCA* (GWPCA). These can be considered nonstationary forms of PCA, where only the second technique is spatial. Second, we describe a PCA technique that directly accounts for spatial autocorrelation in the data via Moran's I statistic and, in doing so, has links to related models concerning eigenfunction spatial filtering. Third, we look at the use of PCA for exploring different structures of the variogram, a measure of spatial autocorrelation commonly used in geostatistics, and, in turn, link these established techniques to more contemporary and sophisticated spatial FA models. Finally, as a caveat to this section, we discuss the ambiguities that are often found in the literature when labeling a PCA technique as spatial.

LWPCA and GWPCA: Moving from Global PCA to Local PCA in Attribute or Geographic Space

LWPCA (Tipping and Bishop 1999; Skočaj, Leonardis, and Bischof 2007; Hoffmann, Schaal, and Vijayakumar 2009; Charlton et al. 2010) is applied to the situation when the data are not described well by a universal set of PCs but where there are localized regions in attribute data space where a suitably localized set of PCs provide a better description. That is, in different parts of the data space, a different set of PCs is needed. This technique uses a moving window weighting approach in the data space where PCs are found in the locality of some point **x** in the data space. For each individual LWPCA around **x**, neighboring data points are first weighted according to some distance-decay kernel function (e.g., bi-square, Gaussian, etc.) where Mahalanobis (attribute space) distances of the neighboring points to **x** are used. Each observation is then multiplied by its respective weight and a standard PCA algorithm is (locally) applied to this weighted data. As a different PCA is computed for every point **x**, the results vary continuously through the data space. The size of the window over which a local PCA might apply is controlled by the bandwidth. Small bandwidth values lead to more rapid variation in the results, whereas very large bandwidths give subspaces increasingly close to the universal (global) PCA solution.

GWPCA models (Fotheringham, Brunsdon, and Charlton 2002; Charlton et al. 2010; Lloyd 2010; Kumar, Lal, and Lloyd in press) are similar to LWPCA models, but in this case it is assumed that there are regions of geographical space in which distinct PCA models apply (i.e., the study data set or process is spatially heterogeneous and should be modeled as such). The technique is identical to LWPCA except that now the distance-decay weights are based on geographical (usually Euclidean) distances between some point **z** and

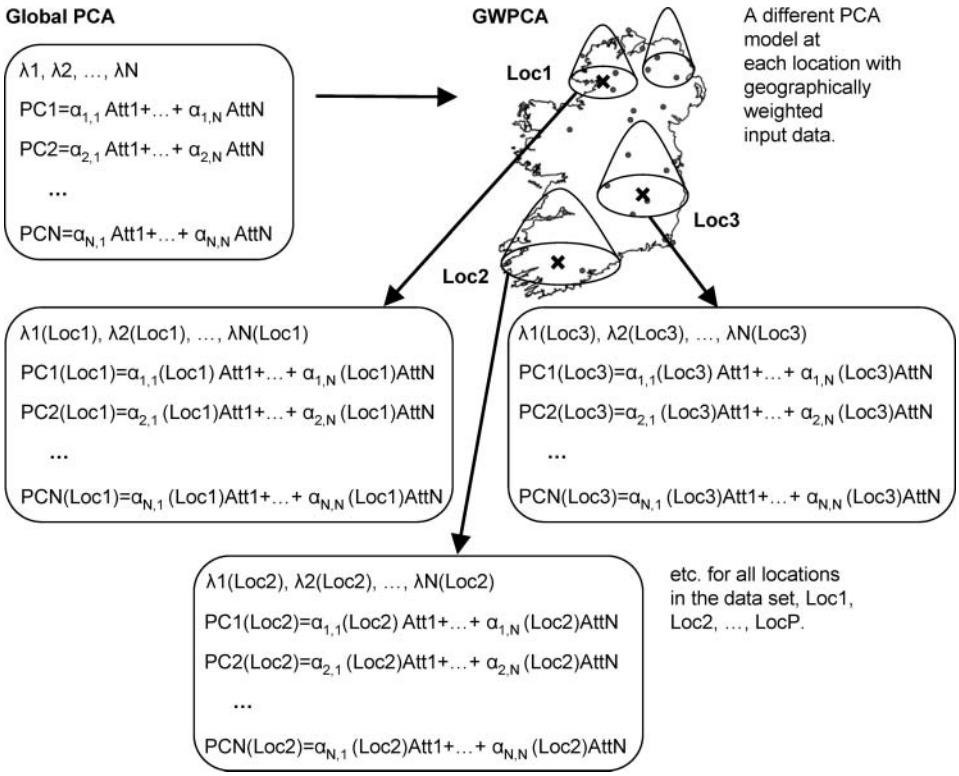

Figure 4. Geographically weighted principal component analysis (GWPCA). A local model is calculated at each location based on a geographically weighted subset of neighboring data points. λ_1 to λ_N are eigenvalues, ordered from largest to smallest. PC1 to PCN are principal components, defined as respective eigenvectors. In a global principal component analysis model, eigenvalues and eigenvectors are constant, whereas in GWPCA, they become dependent on geographic location (Loc1, . . . LocP).

its neighboring data points. As for LWPCA, a different PCA is computed for every $\mathbf{z}$, but now the results vary continuously over geographic space and, as such, they can be mapped (Figure 4).

Spatial patterns in the behavior of local eigenvalues from GWPCA inform on the complexity and local intrinsic dimensionality of the data and could be used for local dimensionality reduction. Local PCs describe local relationships between original variables at each location and could be used to derive local composite indices that depend on local environmental circumstances. GWPCA could also serve as a locally defined orthogonalization prior to the application of some other local statistical method, such as geographically weighted regression (GWR; Fotheringham, Brunsdon, and Charlton 2002), as an alternative to a global PCA orthogonalization in combination with GWR (Bitter, Mulligan, and Dall'erba 2007). Currently, and unlike GWR, GWPCA is rather limited in that there are no associated diagnostics to indicate whether it provides an advantage over its global counterpart. Preliminary research in addressing this drawback is reported in Charlton et al. (2010).

PCA with Spatial Autocorrelation

Spatial effects can also be taken into account when PCA is combined with a measure of spatial autocorrelation. Jombart et al. (2008) presented such a spatial modification of PCA (termed sPCA) to investigate the spatial pattern of genetic variability with respect to the (multivariate) genetic characteristics (termed *alleles*) of a set of individuals or populations under study. Spatial autocorrelation is measured using Moran's I (Moran 1950) and incorporated within the sPCA algorithm. The sPCA technique provides PC scores that summarize both the aspatial genetic variability in attribute space and the spatial autocorrelation structure in geographical space among the individuals or populations. Here statistical (Monte Carlo) tests are used to partition the spatial structure into random, local, and global variance patterns, where local patterns are taken to relate to highly negative spatial autocorrelation and global patterns are taken to relate to highly positive spatial autocorrelation. Observe that it is unlikely that this should be viewed as clear-cut separation of spatial structures, as vectors with relatively small positive eigenvalues produce local patterns of positive spatial autocorrelation. The technique can be implemented and its output visualized using functions provided in the R (Ihaka and Gentlemen 1996) adeganet package (Jombart 2008), where its application in related research areas, such as those found in ecology, should be straightforward. Applications of sPCA should only be viewed as explorative, especially as its output depends strongly on the particular (often arbitrary) connection

network that needs to be specified when computing Moran's I.

The sPCA technique can be viewed as a direct alternative to GWPCA for incorporating spatial effects into a PCA, but whereas GWPCA accounts for first-order (nonstationary) spatial effects, sPCA accounts for second-order (stationary) spatial effects. Such methodological differences are analogous to the use of a GWR or a regression with a spatially autocorrelated error term when choosing a regression model to study spatially referenced data. A natural extension would be to adapt sPCA locally to provide a GWsPCA hybrid.

Linkages and Related Techniques. The sPCA technique has strong conceptual links to the multivariate spatial correlation technique of Wartenberg (1985a, 1985b) where both techniques require the computation of a spatial weighting matrix **W** to account for spatial autocorrelation between spatial units. Furthermore, Dray (2011) provided useful linkages concerning how the spectral decomposition of **W** has been used in different contexts. For example, in quantitative geography, the eigenvectors of **W** are used in spatial filtering where spatial autocorrelation is removed from the residuals of a statistical model and, in turn, can be used for spatial prediction (e.g., see Griffith 1996, 2000; Griffith and Amrhein 1997; Getis and Griffith 2002). The same eigenvectors are also used in ecology for multivariate spatial exploration and prediction (Dray, Legendre, and Peres-Neto 2006; Griffith and Peres-Neto 2006; Jombart, Dray, and Dufour 2009). Useful linkages can also be made between spatial filtering and GWR (Griffith 2008), and Chun (2008) adapted a spatial interaction model using spatial filtering to improve its parameter estimates when modeling migration flows.

Regionalized PCA: A Geostatistical Methodology

In the classical geostatistics framework, a univariate spatial prediction algorithm such as ordinary kriging (OK) and its multivariate extension, ordinary cokriging (OCoK), can each be adapted to decompose the data into spatial components. This results in the (univariate) factorial kriging (FK) and multivariate factorial kriging (MFK) algorithms, respectively (Matheron 1982; Goovaerts 1997; Wackernagel 2003). Here the geostatistical objective is no longer spatial prediction, but an exploration of the origins of the data, where spatial variability (and covariability for MFK) is investigated at different spatial scales. Such scale-dependent variation is typically revealed by nested structures in the empirical variogram(s) and cross-variogram(s) for MFK (i.e., our measures of spatial autocorrelation). FK and MFK estimate and map the different sources of variation suggested by the nested variography and in doing so can provide a greater understanding of the process under investigation. Similarly, FK and MFK can be used as a data filter, where one (or more) of the spatial components are filtered from the data so that the primary analysis can focus on the behavior of the residual process, which in this case is considered more important.

It is only for MFK that applications of PCA are needed. Here PCA is used to decompose the variance–covariance (i.e., coregionalization) matrices that describe the correlation structure of multiple variables at characteristic spatial scales. Outputs from such matrix decompositions are commonly visualized using a circle of correlation plots, one for each spatial scale of interest. Here any significant change in the relationships between variables at the different spatial scales should become immediately apparent. Software to implement MFK can be found in Pardo-Igùzquiza and Dowd (2002); as with any kriging method, MFK can be embedded with a conditional simulation algorithm to provide an assessment of spatial uncertainty for variables with coregionalized components (Larocque et al. 2006).

Numerous applications of FK and MFK can be found in the geosciences and include Galli, Gerdill-Neuillet, and Dadou (1984) in geophysics (FK); Bourgault and Marcotte (1991), Lin et al. (2006), and Imrie et al. (2008) in geochemistry (all MFK); Goovaerts (1992) (FK and MFK) and Castrignanò, Buttafouco, and Puddu (2008; MFK) in soil science; Goovaerts, Sonnet, and Navarre (1993) in hydrogeology (FK and MFK); and Ma and Royer (1988) and Rodgers and Oliver (2007) in image analysis and remote sensing (both FK). Applications outside of the geosciences are rarer, and more recent and examples include (FK only) Goovaerts, Jacquez, and Greiling (2005) and Goovaerts (2010) for health data; Kerry et al. (2010) for crime data; and Nagle (2010) for employment data.

The key drawback to any application of FK and MFK is that the output depends wholly on the form of the nested variogram model(s), which tend to be arbitrarily fitted to the empirical variography. In this respect, any nested behavior observed in the empirical variography should always be expertly related to any physical knowledge of the given process. Appropriate techniques to minimize the effects of outlying data on the FK and MFK analysis are also recommended. Furthermore, each

spatial component that is identified is assumed stationary (or constant) across space. For processes where such an assumption is unrealistic, a wavelet approach (which identifies local changes in variation across a range of spatial scales) can provide a useful alternative (e.g., see Oliver, Bosch, and Slocum 2002).

Linkages and Related Techniques. Conceptual relationships between MFK and similar multivariate techniques (commonly termed *spatial FA*) proposed outside of the geostatistical paradigm are given in Bailey and Krzanowski (2000). Moving on from these predominantly exploratory techniques to spatial FA models (or spatial latent variable models, see also subsection on latent variable modeling), where estimation, inference, and spatial prediction procedures are formally developed, can be found in Christensen and Amemiya (2001, 2002, 2003). Here the models are demonstrated using simulated soil geochemical data. F. Wang and Wall (2001, 2003) used a related spatial FA model to analyze health data, where their second article extends to a Bayesian methodology. Another Bayesian spatial FA model was described in Hogan and Tchernis (2004) and was used to model social and health (material) deprivation indices. A spatiotemporal FA model, also within a Bayesian framework, can be found in Lopes, Gamerman, and Salazar (2011) and was demonstrated using meteorological data (and links to the use of EOFs described earlier). Folmer and Oud (2008) presented a structural equation model with spatial latent variables and in doing so provided an alternative to spatial regressions that require the spatial weighting matrix **W** from previous subsection (linkages).

Disambiguation of the Term *Spatial PCA* and Its Variations

We identified a number of studies that claim to use a method termed "spatial PCA". In a geographical context, the term seems to have originated in the first half of the last decade (we were not able to find any explicit reference to "spatial PCA" pre-2002 in geography, although there are earlier studies using "spatial," "temporal," and "spatiotemporal PCA" in, for example, neuroscience; see Spencer, Dien, and Donchin [1999, 2001], but we found that its meaning differs). The term can refer either to one of the first three methodologies for standard PCA on spatial data (identified in the previous section) or else (and more appropriately) to the spatial adaptation of PCA or sPCA (from two subsections ago). To address this confusion in terminology, we list studies in our literature search, aside from sPCA, that explicitly use this term.

In most cases "spatial PCA" refers to raster data PCA. Examples are mainly found in the environmental sciences: to define indices of environmental vulnerability or quality (Li et al. 2006; Ruimin and Zhenyao 2007; Shi et al. 2009), to analyze susceptibility to coral bleaching (Maina et al. 2008), in environmental geology (Satapathy, Salve, and Katpatal 2009), and also in biogeography (Brito et al. 2008).

The second explicit use of the term refers to spatial objects PCA. Two examples that we found were PCA on areas—watersheds to calculate environmental indicators (Tran et al. 2002) and PCA on point measurements at sampling locations in lakes (Parinet, Lhote, and Legube 2004).

The third explicit use of the term refers to atmospheric science PCA, where EOFs are called spatial PCs (Krepper and García 2004).

We would also like to clarify the difference between *spatially weighted PCA* (SWPCA) and *geographically weighted PCA* (GWPCA). SWPCA applies to raster PCA, where each of the input rasters (e.g., bands of RS images) is weighted by another weight raster. This weight raster is a surface that represents some type of spatial relationship that is important in the context of the respective input raster. For example, it can be a distance surface from each pixel to a certain object (e.g., ore deposits) or to areas with extremely high or low values in the input raster (W. Wang and Cheng 2008). This is not to be confused with GWPCA (see relevant subsection three subsections ago), where geographically local PCA models are calculated at the location of each spatial object in the data set. The two methods, SWPCA and GWPCA, are therefore completely different from one another and are meant for different types of data and for different purposes and should not be confused.

Further Topics

In this section we look at three recent variants of PCA from the statistics, machine learning, and pattern recognition communities and discuss their potential usefulness in a spatial context, either in a basic form or some spatially adapted form. In particular, we look at simple PCA, sparse PCA, and multilinear PCA.

Simple PCA

Spatial data sets are often highly dimensional (sometimes containing several hundreds of dimensions) and

very large; for example, acoustic marine data sets contain hundreds of millions of data points and hundreds of attributes represented by statistical features calculated from the sonar backscatter (Preston 2009). Basic PCA methods, which are often used on such data sets as discussed in the previous sections, are matrix based, which means that they require an explicit calculation and diagonalization of the variance–covariance matrix. This is computationally a very demanding process that in practice is often not viable for such large or highly dimensional data sets. There exist several recent developments, however, that calculate approximations of mathematical PCs in a way that is computationally fast, but they have not been used widely in a spatial context. One such development is simple PCA, which is an approximation of the traditional PCA algorithm, such that the PCs are calculated by an iterative calculation of one approximated component at a time. It therefore does not require the explicit calculation of the variance–covariance matrix.

Simple PCA works by using a training procedure, similar to those in neural networks, to calculate the first PC from the data. This is done by iteratively computing a series of linear transformations to a set of orthogonal axes (approximations of PCs), one axis at a time. In each step a linear transformation is found such that the variance of the data with respect to only one axis is maximized—this is the PC that is being sought in this particular iteration. This component is then removed to ensure that it is not found again in the next step. The process of removing the effect of one component from the data is called *deflation* and in the next step the search for the next component is rerun on the deflated data. This iterative process is repeated until the desired number of components is reached. There are two main algorithms for simple PCA, those by Partridge and Calvo (1998) and Vines (2000).

Simple PCA is a data-oriented method (as opposed to matrix-based methods such as traditional PCA). It has been shown to be more efficient for highly dimensional data sets than basic matrix-based PCA methods (Partridge and Calvo 1998). There also exists an even more efficient variation of simple PCA (Oyama et al. 2008) that deals with data incrementally; that is, by adding one data point at a time in a process called *incremental learning*. This procedure enables the calculations to be performed on extremely large data sets that are fed to the algorithm sequentially instead of all at once. Therefore, simple PCA offers a promising practical alternative for large and highly dimensional spatial data sets.

Sparse PCA

Although PCA is a very powerful tool for dimensionality reduction, it is often difficult to relate patterns in the resulting latent variables back to physical quantities or determine which variables are significant contributors to the patterns. This is because the loadings obtained by PCA are linear combinations of all variables in the data set. Furthermore, if a group of highly correlated (or collinear) variables contributes to a latent variable, their contribution is distributed evenly across all variables in the group. This so-called grouping effect is a property of linear least squares regression and by extension PCA (Zou and Hastie 2005). Although this is a desirable property in terms of averaging out noise, it masks the significance of variables, making the identification of key variables difficult. This issue has motivated the development of extensions to PCA that result in PCs that are sparse (i.e., with many zero coefficients). The methods are based on the assumption that many real-life data sets exhibit a low-dimensional structure in a sparse form.

As discussed previously, an early method developed to improve the interpretability of PCA is varimax rotation (Kaiser 1958), which involves rotating the subspace defined by selected PCs so that a small number of the coefficients in the loading vectors have much greater values than the remaining coefficients. To obtain sparse components, the smaller coefficients are then simply set to zero. Jeffers (1967) proposed setting small coefficients of the original PCs to zero as a means of obtaining sparse components, although this can lead to a selection deficiency when the variables have high mutual correlations (Cadima and Jolliffe 1995). It also invalidates the orthogonality of the resulting components.

The first true algorithmic method for achieving sparse loadings was proposed by Jolliffe, Trendafilov and Uddin (2003) and is known as *Simplified Component Technique for Least Absolute Shrinkage and Selection* (SCOTLASS). This employs a penalty term referred to as the *Least Absolute Shrinkage and Selection Operator* (LASSO; Tibshirani 1996) to force loadings to be sparse. Unlike a ridge penalty, which encourages parameters to be small, the LASSO penalty has the attractive property that it forces parameters to be exactly zero. SCOTLASS has a relatively high computational cost with the result that several researchers have developed alternative implementations that are substantially more efficient (Zou, Hastie, and Tibshirani 2006; D'Aspremont et al. 2007; Shen and Huang 2008).

In the context of spatial data analysis, sparse PCA has not often been used but seems promising for problems where there are a large number of attributes (variables) and there is a need to determine the key factors contributing to the underlying spatial patterns being investigated.

Multilinear PCA of Tensor Objects

With the unprecedented advances in data collection, many disciplines collect spatial data that fill a certain spatially constrained area and where the phenomenon under observation is continuously distributed in this area. These are the so-called tensor data objects. Examples include 2D tensors, such as gray-level images in computer vision and pattern recognition, or 3D tensors, such as MRI scans in medical imaging. In terms that are familiar to geographers, these tensors could be seen as 2D rasters and 3D volumes, examples of which include hyperspectral satellite imagery for the former and geological volumetric data for the latter.

Tensors can be understood as n-dimensional bounded areas, separated into regular meshes, where each mesh element (pixel in 2D or voxel in 3D) represents a measurement of an attribute. Tensor data are therefore highly dimensional: a $100 \times 100 \times 100$ volume has a million voxels and each of these voxels is a separate attribute. However, these attributes are not independent. The spatial proximity of their position in the volume and continuity of the phenomenon observed or measured entails that there is a high level of spatial correlation present (Lu, Plataniotis, and Venetsanopoulos 2011).

Feature extraction and pattern recognition in tensor data (e.g., patterns in MRI scans) are usually performed through dimensionality reduction, where the goal is to map the tensor space onto a lower dimensional subspace that captures most of the signal variation present in the original tensorial representation. This is often done with PCA in several different ways (Lu, Plataniotis, and Venetsanopoulos 2011).

The standard way is to employ basic linear PCA, where the tensors are represented as highly dimensional vectors of attributes. Imagine a 3D volume being "unwound" into a long row vector by taking rows of voxels one by one off the 3D volume. Each tensor object is then represented as one such vector. This results in a vector with as many dimensions as there are voxels but breaks the natural structure in the data in that it completely ignores any spatial correlation and proximity of voxels (Lu, Plataniotis, and Venetsanopoulos 2011).

To counter this problem, tensor data analysis suggests that PCA is run in either 2D mode (Yang et al. 2004; Ye, Janardan, and Li 2004) or 3D mode (Lu, Plataniotis, and Venetsanopoulos 2008). In both these approaches, input data are represented in their natural multidimensional form as tensors, bound by a 2D or 3D spatial area. This keeps the basic tensor elements together with spatial correlation and proximity taken into consideration.

Such approaches are increasingly common in areas such as face recognition, gait recognition, medical imaging, and shape analysis (Aguirre et al. 2007; Lu, Plataniotis, and Venetsanopoulos 2011). Typical tasks include 3D object recognition tasks, content-based retrieval of patterns, gait or gesture recognition, and activity recognition in data of various very complex types, such as medical images, spatial video sequences, and space–time series (Lu, Plataniotis, and Venetsanopoulos 2008; Leibovici 2010).

Given the similarity of data type, these approaches could also be of interest to geographers. In particular, 2D mode PCA could replace raster PCA, whereas 3D mode PCA would be welcome, for example, in geology and seismology, which often deal with pattern recognition in 3D volumetric data (Gao 2009; Hsieh, Chen, and Ma 2010). An example of a tensor mode PCA applied in spatial context was given in Leibovici (2010), where it was used to solve a spatiotemporal ecoclimatic delineation problem.

Conclusions

In this article we surveyed the use of PCA on spatial data in an attempt to identify methodological characteristics of its use and also to investigate uses that take into account particular characteristics of spatial data (spatial heterogeneity and spatial autocorrelation). Studies reviewed were primarily from geography and the geosciences, where we found that standard nonspatial PCA is commonly applied in one of four methodological ways: on spatial objects, on raster data, on meteorological space–time series data, and on flow matrices. Although such applications of PCA have merit, they overlook spatial effects that could furnish a greater understanding of a given process. In this respect, we reported on adapted PCA methods that account for spatial heterogeneity (with respect to nonstationary mean response effects) and spatial autocorrelation (with respect to stationary variance effects). For the former we described a geographically weighted PCA method,

whereas for the latter we used a PCA method that could account for spatial autocorrelation via the calculation of Moran's I. Furthermore, we described how PCA can be used to investigate multiple scales of spatial autocorrelation (from the geostatistical literature) and also attempted to provide links to more sophisticated techniques and models from the statistical literature.

Perhaps rather surprisingly, in surveyed literature we found proportionally few studies that use spatially adapted versions of PCA to analyze their data. There seems to be a need to promote such spatially aware techniques and this is something that we hope that this article will achieve—that the reader will consider the fact that geographic space can often matter and therefore look into alternatives that account for this.

We also attempted to bridge the gap between recent developments in the statistics, machine learning, and pattern recognition communities with geography and the geosciences by suggesting alternative algorithms for direct application on spatial data or eventual spatial adaptation. Here we reviewed methods that are computationally faster (simple PCA), facilitate easier interpretation of PCs (sparse PCA) than standard PCA, and those that might have use in 2D and 3D raster-type data sets (multilinear PCA). These all seem promising for use in a spatial context.

Consequently, this article serves as a useful catalyst to increased recognition of the issues involved in using PCA with spatial data and the potential uses for alternative PCA methodologies on spatial data that perhaps could be explored by geographers and that to date remain relatively underused.

Acknowledgments

The authors would like to thank the three anonymous reviewers whose comments helped to significantly improve this article. Urška Demšar's work on this topic is supported by a Research Frontiers Programme Grant (09/RFP/CMS2250) by Science Foundation Ireland under the National Development Plan. Paul Harris and Sean McLoone are funded by a Strategic Research Cluster Grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan.

References

Afifi, A., V. A. Clark, and S. May. 2004. *Computer-aided multivariate analysis*. London and New York: Chapman & Hall/CRC.

Aguirre, M. R., M. G. Linguraru, K. Marias, N. Ayache, L.-P. Nolte, and M. A. Gonzalez-Ballester. 2007. Statistical shape analysis via principal factor analysis. In *Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1216–19. doi: 10.1109/ISBI.2007.357077

Aitchison, J. 1982. The statistical analysis of compositional data. *Journal of the Royal Statistical Society Series B (Methodological)* 44 (2): 129–77.

———. 1983. Principal component analysis of compositional data. *Biometrika* 70 (1): 57–65.

Aitchison, J., and J. J. Egozcue. 2005. Compositional data analysis: Where are we and where should we be heading? *Mathematical Geology* 37 (7): 829–50.

Aminzadeh, B., and F. Samani. 2006. Identifying the boundaries of the historical site of Persepolis using remote sensing. *Remote Sensing of Environment* 102:52–62.

Anderson, T. W. 1984. Estimating linear statistical relationships. *Annals of Statistics* 12 (1): 1–45.

Andrews, J. T. 2008. The role of the Iceland Ice Sheet in the North Atlantic during the late Quaternary: A review and evidence from Denmark Strait. *Journal of Quaternary Science* 23 (1): 3–20.

Anselin, L., S. Srihdaran, and S. Gholston. 2007. Using exploratory spatial data analysis to leverage social indicator databases: The discovery of interesting patterns. *Social Indicators Research* 82:287–309.

Antunes, S. C., E. Pereira, J. P. Sousa, M. C. Santos, and F. Gonçales. 2008. Spatial and temporal distribution of litter arthropods in different vegetation covers of Porto Santo Island. *European Journal of Soil Biology* 44:45–56.

Arbia, G., D. A. Griffith, and R. P. Haining. 2003. Spatial error propagation when computing linear combinations of spectral bands: The case of vegetation indices. *Environmental and Ecological Statistics* 10:375–96.

Bailey, T. C., and W. J. Krzanowski. 2000. Extensions to spatial factor methods with illustrations in geochemistry. *Mathematical Geology* 32:657–82.

Bastianoni, S., F. M. Pulselli, S. Focardi, E. B. P. Tiezzi, and P. Gramatica. 2008. Correlations and complementarities in data and methods through principal components analysis (PCA) applied to the results of the SPIn-Eco Project. *Journal of Environmental Management* 86:419–26.

Berry, B. J. L. 1964. Approaches to regional analysis: A synthesis. *Annals of the Association of American Geographers* 54:2–11.

———. 1966. Essays on commodity flows and the spatial structure of the Indian economy. Research Paper No. 111, Department of Geography, University of Chicago.

———. 1968a. Interdependency of spatial structure and spatial behavior: A general field theory formulation. *Papers, Regional Science Association* 21:205–27.

———. 1968b. A synthesis of formal and functional regions using a general field theory of spatial behavior. In *Spatial analysis*, ed. B. J. L. Berry and D. F. Marble, 419–30. Englewood Cliffs NJ: Prentice-Hall.

———. 1971. Comparative factorial ecology. *Economic Geography* 47 (Suppl.).

Bitter, C., G. F. Mulligan, and S. Dall'erba. 2007. Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems* 9:7–27.

Black, W. R. 1973. Toward a factorial ecology of flows. *Economic Geography* 49:59–67.

Boelhouwer, J., and I. Stoop. 1999. Measuring well-being in the Netherlands: The SCP index from 1974 to 1997. *Social Indicators Research* 48:51–75.

Booysen, F. 2002. An overview and evaluation of composite indexes of development. *Social Indicators Research* 59:115–51.

Bourgault, G., and D. Marcotte. 1991. Multivariate variogram and its applications to the linear model of coregionalization. *Mathematical Geology* 23:899–928.

Brito, J. C., X. Santos, J. M. Pleguezelos, and N. Sillero. 2008. Inferring evolutionary scenarios with geostatistics and geographical information systems for the viperid snakes *Vipera latastei* and *Vipera monticola*. *Biological Journal of the Linnean Society* 95:790–806.

Cadima, J., and I. P. Jolliffe. 1995. Loadings and correlations in the interpretation of principal components. *Journal of Applied Statistics* 22 (2): 203–14.

Campbell, B. M. S., and J. P. Power. 1989. Mapping the agricultural geography of medieval England. *Journal of Historical Geography* 15 (1): 24–39.

Castrignanò, A., G. Buttafouco, and R. Puddu. 2008. Multiscale assessment of the risk of soil salinization in an area of south-eastern Sardinia (Italy). *Precision Agriculture* 9:17–31.

Charlton, M., C. Brunsdon, U. Demšar, P. Harris, and A. S. Fotheringham. 2010. Principal components analysis: From global to local. Paper presented at the 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal.

Choulakian, V. 2001. Robust Q-mode principal component analysis in L_1. *Computational Statistics & Data Analysis* 37:135–50.

Christensen, W. F., and Y. Amemiya. 2001. Generalized shifted-factor analysis method for multivariate geo-referenced data. *Mathematical Geology* 33:801–24.

———. 2002. Latent variable analysis of multivariate spatial data. *Journal of American Statistical Association* 97 (457): 302–17.

———. 2003. Modeling and prediction for multivariate spatial factor analysis. *Journal of Statistical Planning and Inference* 115:543–64.

Chun, Y. 2008. Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems* 10 (4): 317–44.

Clark, D. 1973. Normality, transformation and the principal component solution: An empirical note. *Area* 5: 110–13.

Collins, J. B., and C. E. Woodcock. 1996. An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing of Environment* 56:66–77.

Corcoran, J., G. Higgs, C. Brunsdon, A. Ware, and P. Norman. 2007. The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study. *Computers, Environment and Urban Systems* 31:623–47.

Croux, C., and G. Haesbroeck. 2000. Principal components analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika* 87:603–18.

D'Aspremont, M., L. Ghaoui, M. I. Jordan, and G. Lanckriet. 2007. A direct formulation for sparse PCA using semidefinite programming. *SIAM Review* 49 (3): 434–48.

Daszykowski, M., K. Kaczmarek, Y. Vander Heyden, and B. Walczak. 2007. Robust statistics in data analysis—A review of basic concepts. *Chemometrics and Intelligent Laboratory Systems* 85:203–19.

Daultrey, S. 1976. *Principal component analysis*. Concepts and techniques in modern geography, GeoAbstracts. Norwich, UK: University of East Anglia.

Davis, J. C. 1986. *Statistics and data analysis in geology*. New York: Wiley.

Donoho, D. L. 2000. High-dimensional data analysis: The curses and blessings of dimensionality. Keynote address at the conference of The American Mathematical Society, Los Angeles. http: //www- stat.stanford.edu/~donoho/Lectures/AMS2000/AMS2000.html (last accessed 4 September 2009).

Dray, S. 2011. A new perspective about Moran's coefficient: Spatial autocorrelation as a linear regression problem. *Geographical Analysis* 43:127–41.

Dray, S., P. Legendre, and P. Peres-Neto. 2006. Spatial modelling: A comprehensive framework for principle coordinate analysis of neighbours matrices (PCNM). *Ecological Modelling* 196:483–93.

Eastman, J. R. 2003. *IDRISI Kilimanjaro tutorial*. Worcester, MA: Clark Labs, Clark University.

Esteban, P., J. Martin-Vide, and M. Mases. 2006. Daily atmospheric circulation catalogue for western Europe using multivariate techniques. *International Journal of Climatology* 26:1501–15.

Felipe-Sotelo, M., M. Gustems, I. Hernàndez, M. Terrado, and R. Tauler. 2006. Investigation of geographical and temporal distribution of tropospheric ozone in Catalonia (North-East Spain) during the period 2000–2004 using multivariate data analysis methods. *Atmospheric Environment* 40:7421–36.

Filzmoser, P. 1999. Robust principal component and factor analysis in the geostatistical treatment of environmental data. *Environmetrics* 10:363–75.

Filzmoser, P., R. G. Garrett, and C. Reimann. 2005. Multivariate outlier detection in exploration geochemistry. *Computers & Geosciences* 31:579–87.

Filzmoser, P., K. Hron, and C. Reiman. 2009. Principal component analysis for compositional data with outliers. *Environmetrics* 20:621–32.

Filzmoser, P., R. Maronna, and M. Werner. 2008. Outlier identification in high dimensions. *Computational Statistics and Data Analysis* 52:1694–1711.

Floras, S. A., and I. D. Sgouras. 1999. Use of geoinformation techniques in identifying and mapping areas of erosion in a hilly landscape of central Greece. *International Journal of Applied Earth Observation and Geoinformation* 1 (1): 68–77.

Fodor, I. K. 2002. *A survey of dimension reduction techniques*. LLNL Technical Report, Lawrence Livermore National Laboratory, Livermore, CA. http: //www.llnl.gov/CASC/sapphire/pubs/148494.pdf (last accessed 4 September 2009).

Folmer, H., and J. Oud. 2008. How to get rid of W: A latent variables approach to modelling spatially lagged variables. *Environment & Planning A* 40:2526–38.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, UK: Wiley.

Fotso, J.-C., and B. Kuate-Defo. 2005. Measuring socioeconomic status in health research in developing countries: Should we be focusing on households, communities or both? *Social Indicators Research* 72:189–237.

Frank, D., and J. Esper. 2005. Characterization and climate response patterns of a high-elevation multi-species tree-ring networks in the European Alps. *Dendrochronologia* 22:107–21.

Galli, A., F. Gerdill-Neuillet, and C. Dadou. 1984. Factorial kriging analysis: A substitute to spectral analysis of magnetic data. In *Geostatistics for natural resources characterization*, ed. G. Verly, M. David, A. G. Journel, and A. Marachal, 543–57. Dordrecht, The Netherlands: Reidel.

Gao, D. 2009. 3D seismic volume visualization and interpretation: An integrated workflow with case studies. *Geophysics* 74 (1): 1–12.

Getis, A., and D. A. Griffith. 2002. Comparative spatial filtering in regression analysis. *Geographical Analysis* 34:130–40.

Goddard, J. 1970. Functional regions within the city centre: A study by factor analysis of taxi flows in central London. *Transactions of the Institute of British Geographers* 49:161–82.

———. 1973. Office linkages and location: A study of communications and spatial patterns in Central London. *Progress in Planning* 1:109–232.

Goddard, J., and A. Kirby. 1976. *An introduction to factor analysis*. Concepts and techniques in modern geography, GeoAbstracts. Norwich, UK: University of East Anglia.

Goldstein, H., and W. J. Browne. 2005. Multilevel factor analysis models for continuous and discrete data. In *Contemporary psychometrics: A festschrift for Roderick P. McDonald*, ed. A. Maydeu-Olivares and J. J. McArdle, 453–75. Mahwah, NJ: Erlbaum.

Goovaerts, P. 1992. Factorial kriging analysis: A useful tool for exploring the structure of multivariate spatial information. *Journal of Soil Science* 43:597–619.

———. 1997. *Geostatistics for natural resources evaluation*. New York: Oxford University Press.

———. 2010. Geostatistical analysis of county-level lung cancer mortality rates in the southeastern United States. *Geographical Analysis* 42:32–52.

Goovaerts, P., G. M. Jacquez, and D. Greiling. 2005. Exploring scale-dependent correlations between cancer mortality rates using factorial kriging and population weighted semivariograms. *Geographical Analysis* 37:152–82.

Goovaerts, P., G. M. Jacquez, and A. Marcus. 2005. Geostatistical and local cluster analysis of high resolution hyperspectral imagery for detection of anomalies. *Remote Sensing of Environment* 95:351–67.

Goovaerts, P., P. Sonnet, and A. Navarre. 1993. Factorial kriging analysis of spring-water contents in the Dyle River basin, Belgium. *Water Resources Research* 29:2115–25.

Gould, P. R. 1967. On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers* 42:53–86.

Griffith, D. A. 1996. Spatial autocorrelation and eigenfunctions of the geographical weights matrix accompanying geo-referenced data. *Canadian Geographer* 40:351–67.

———. 2000. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* 2:141–56.

———. 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment & Planning A* 40:2751–69.

Griffith, D. A., and C. G. Amrhein. 1997. *Multivariate statistical analysis for geographers*. Upper Saddle River, NJ: Prentice-Hall.

Griffith, D. A., and P. Peres-Neto. 2006. Spatial modelling in ecology: The flexibility of eigenfunction spatial analyses. *Ecology* 87:2603–13.

Hägerstrand, T. 1967. *Innovation diffusion as a spatial process*. Chicago: University of Chicago Press.

Hall, D. R. 1977. Applied social area analysis: Defining and evaluating areas for urban neighbourhood councils. *Geoforum* 8:277–310.

Hay, A. M., and A. H. M. Raihan Sharif. 1986. The use of unstandardised data in components analysis of flow matrices: A comment. *Area* 18:35–38.

Hernández, A. J., M. J. Adarve Alcazar, and J. Pastor. 1998. Some impacts of urban waste landfills on Mediterranean soils. *Land Degradation and Development* 9:21–33.

Hoersch, B., G. Braun, and U. Schmidt. 2002. Relation between landform and vegetation in Alpine regions of Wallis, Switzerland: A multiscale remote sensing and GIS approach. *Computers, Environment and Urban Systems* 26:113–39.

Hoffmann, H., S. Schaal, and S. Vijayakumar. 2009. Local dimensionality reduction for non-parametric regression, *Neural Processing Letters* 29:109–31.

Hogan, J. W., and R. Tchernis. 2004. Bayesian factor analysis for spatially-correlated data, with application to summarizing area-level material deprivation from census data. *Journal of American Statistical Association* 99 (466): 314–24.

Holliday, J. R., J. B. Rundle, K. F. Tiampo, W. Klein, and A. Donnellan. 2006. Modification of the pattern informatics method for forecasting large earthquake events using complex eigenfactors. *Tectonophysics* 413:87–91.

Horner, M. W., and T. W. Grubesic. 2001. A GIS-based planning approach to locating urban rail terminals. *Transportation* 28:55–77.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24 (6): 417–41, 498–520.

Hsieh, T.-J., C.-K. Chen, and K.-L. Ma. 2010. Visualizing field-measured seismic data. In *Proceedings of the IEEE Pacific Visualization Symposium 2010*, 65–72. doi: 10.1109/PACIFICVIS.2010.5429610

Hubert, M., P. J. Rousseeuw, and K. Vanden Branden. 2005. ROBPCA: A new approach to robust principal components analysis. *Technometrics* 47:64–79.

Hubert, M., P. J. Rousseeuw, and S. Verboven. 2002. A fast robust method for principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems* 60:101–11.

Ibarra-Berastegi, G., J. Sáenz, A. Ezcurra, U. Ganzedo, J. Díaz de Argandoña, I. Errasti, A. Fernandez-Ferrero, and J. Polanco-Martí. 2009. Assessing spatial variability of SO_2 field as detected by an air quality network using self-organizing maps, cluster, and principal component analysis. *Atmospheric Environment* 43:3829–36.

Ihaka, R., and R. Gentleman. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5:299–314.

Imrie, C. E., A. Korre, G. Munoz-Melendez, I. Thornton, and S. Durucan. 2008. Application of factorial kriging analysis to the FOREGS European topsoil geochemistry database. *Science of the Total Environment* 393:96–110.

Jackson, D. A., and Y. Chen. 2004. Robust principal component analysis and outlier detection with ecological data. *Environmetrics* 15:129–39.

Jeffers, N. R. 1967. Two case studies in the application of principal component analysis. *Applied Statistics* 16 (3): 225–36.

Johnston, R. J. 1978. *Multivariate statistical analysis in geography*. London: Longman.

Jolliffe, I. T. 2002. *Principal component analysis*. 2nd ed. Berlin, Germany: Springer Verlag.

Jolliffe, I. T., N. T. Trendafilov, and M. Uddin. 2003. A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics* 12 (3): 531–47.

Jombart, T. 2008. Adeganet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–05.

Jombart, T., S. Devillard, A.-B. Dufour, and D. Pontier. 2008. Revealing cryptic patterns in genetic variability by a new multivariate method. *Heredity* 101:92–103.

Jombart, T., S. Dray, and A. Dufour. 2009. Finding essential scales of spatial variation in ecological data: A multivariate approach. *Ecography* 32:161–68.

Kaiser, F. H. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23:187–200.

Kelly, A., and C. Teljeur. 2007. The national deprivation index for health and health services research. SAHRU Technical Report, Department of Public Health and Primary Care, Trinity College, Dublin, Ireland.

Kerry, R., P. Goovaerts, R. P. Haining, and V. Ceccato. 2010. Applying geostatistical analysis to crime data: Car-related thefts in the Baltic states. *Geographical Analysis* 42:53–77.

Khaemba, W. M., and A. Stein. 2000. Use of GIS for a spatial and temporal analysis of Kenyan wildlife with generalised linear modelling. *International Journal of Geographical Information Science* 14 (8): 833–53.

Klassen, D. R. 2009. Principal components analysis of Mars in the near-infrared. *Icarus* 204 (1): 32–47.

Krepper, C. M., and N. O. García. 2004. Spatial and temporal structures of trends and interannual variability of precipitation over the La Plata Basin. *Quaternary International* 114:11–21.

Kumar, S., R. Lal, and C. D. Lloyd. In press. Assessing spatial variability in soil characteristics with geographically weighted principal component analysis. *Computational Geosciences*.

Larocque, G., P. Dutilleul, B. Pelletier, and J. W. Fyles. 2006. Conditional Gaussian co-simulation of regionalised components of soil variation. *Geoderma* 134:1–16.

Leibovici, D. G. 2010. Spatio-temporal multiway decompositions using principal tensor analysis on *k*-modes: The R package PTAk. *Journal of Statistical Software* 34 (10): 1–34.

Lengen, C., and J. Blasius. 2007. Constructing a Swiss health space model of self-perceived health. *Social Science & Medicine* 65:80–94.

Li, A., A. Wang, S. Liang, and W. Zhou. 2006. Eco-environmental vulnerability evaluation in mountainous region using remote sensing and GIS—A case study in the upper reaches of Minjiang River, China. *Ecological Modelling* 192:175–87.

Lin, Y.-B., Y.-P. Lin, C.-W. Lui, and Y.-C. Tan. 2006. Mapping of spatial multi-scale sources of arsenic variation in groundwater on ChiaNan floodplain of Taiwan. *Science of the Total Environment* 370:168–81.

Lloyd, C. D. 2010. Analysing population characteristics using geographically weighted principal components analysis: A case study of Northern Ireland in 2001. *Computers, Environment and Urban Systems* 34:389–99.

Lo, C. P. 1997. Application of Landsat TM data for quality of life assessment in an urban environment. *Computers, Environment and Urban Systems* 21 (3–4): 259–76.

Lo, C. P., and B. J. Faber. 1997. Integration of Landsat TM and census data for quality of life assessment. *Remote Sensing of Environment* 62:142–57.

Lopes, H. F., D. Gamerman, and E. Salazar. 2011. Generalized spatial dynamic factor models. *Computational Statistics and Data Analysis* 55:1319–30.

Lopez-Bustins, J.-A., P. Esteban, K. Labitzke, and U. Langematz. 2007. The role of the stratosphere in Iberian Peninsula rainfall, a preliminary approach in February. *Journal of Atmospheric and Solar-Terrestrial Physics* 69:1471–84.

Lu, H., K. N. K. Plataniotis, and A. N. Venetsanopoulos. 2008. MPCA: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks* 19 (1): 18–39.

———. 2011. A survey of multilinear subspace learning for tensor data. *Pattern Recognition* 44:1540–51.

Ma, Y. Z., and J. J. Royer. 1988. Local geostatistical filtering: Application to remote sensing. *Sciences de la Terre, Série Informatique* 27:17–36.

Maćkiewicz, A., and W. Ratajczak. 1993. Principal component analysis (PCA). *Computers & Geosciences* 19 (3): 303–42.

Magee, A. 1971. Problems of economic development and migration in Southern Europe with special reference to Spain. In *Proceedings of the Sixth New Zealand Geography Conference* (Christchurch), ed. R. J. Johnston and J. M. Soons, 179–85. Christchurch, NZ: New Zealand Geographical Society.

Maina, J., V. Venus, T. R. McClanahan, and M. Ateweberhan. 2008. Modelling susceptibility of coral reefs to environmental stress using remote sensing data and GIS models. *Ecological Modelling* 212:180–99.

Mather, P. M., and S. Openshaw. 1974. Multivariate methods and geographical data. *The Statistician* 23:283–308.

Matheron, G. 1982. Pour une analyse krigeante de données régionalisées [Kriging analysis for regional data]. Technical Report N-732, Centre de Géostatistique, Ecoles des Mines de Paris, Fontainebleau.

Moran, P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37:17–23.

Nagle, N. N. 2010. Geostatistical smoothing of areal data: Mapping employment density with factorial kriging. *Geographical Analysis* 42:99–117.

Ngene, S. M., A. K. Skidmore, H. Van Gils, I. Douglas-Hamilton, and P. Omondi. 2009. Elephant distribution around a volcanic shield dominated by a mosaic of

forest and savanna (Marsabit, Kenya). *African Journal of Ecology* 47:234–45.
North, G. R., T. L. Bell, and R. F. Cahalan. 1982. Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review* 110:699–706.
Obled, C., and J. D. Creutin. 1986. Some developments in the use of empirical orthogonal functions for mapping meteorological fields. *Journal of Applied Meteorology* 25:1189–1204.
Oliver, M. A., E. Bosch, and K. Slocum. 2002. Wavelets and kriging for filtering and data reconstruction. In *Geostatistics 2000*, ed. W. J. Kleingold and D. G. Krige, 571–80. Cape Town, South Africa: Geostatistical Association of Southern Africa.
Organisation for Economic Co-operation and Development (OECD). 1999. *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD.
Oyama, T., S. Karungaru, S. Tsuge, Y. Mitsukura, and M. Fukumi. 2008. Incremental learning method of simple-PCA. In *Knowledge-based intelligent information and engineering systems*, ed. I. Lovrek, R. J. Howlett, and L. C. Jain, 403–10. Berlin, Germany: Springer-Verlag.
Panda, S. S., G. Hoogenboomb, and J. Pazb. 2009. Distinguishing blueberry bushes from mixed vegetation land use using high resolution satellite imagery and geospatial techniques. *Computers and Electronics in Agriculture* 67:51–58.
Pardo-Igùzquiza, E., and P. Dowd. 2002. FACTOR2D: A computer program for factorial cokriging. *Computers & Geosciences* 28:857–75.
Parinet, B., A. Lhote, and B. Legube. 2004. Principal component analysis: An appropriate tool for water quality evaluation and management—Application to a tropical lake system. *Ecological Modelling* 178:295–311.
Partridge, M., and R. A. Calvo. 1998. Fast dimensionality reduction and simple PCA. *Intelligent Data Analysis* 2:203–14.
Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series* 6 2 (11): 559–72.
Preston, J. 2009. Automated acoustic seabed classification of multibeam images of Stanton Banks. *Applied Acoustics* 70 (10): 1277–87.
Reades, J., F. Calabrese, and C. Ratti. 2009. Eigenplaces: Analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B* 36:824–36.
Reid, M. K., and K. L. Spencer. 2009. Use of principal components analysis (PCA) on estuarine sediment datasets: The effect of data pre-treatment. *Environmental Pollution* 157:2275–81.
Reyment, R. A. 2006. On stability of compositional canonical variate vector components. In *Compositional data analysis in the geosciences: From theory to practice*, ed. A. Buccianti, G. Mateu-Figueras, and V. Pawlowsky-Glahn, 59–66. London: Geological Society.
Richman, M. B. 1986. Rotation of principal components. *Journal of Climatology* 6:293–335.
Rodgers, S. E., and M. A. Oliver. 2007. A geostatistical analysis of soil, vegetation, and image data characterizing land surface variation. *Geographical Analysis* 39:195–216.
Rousseeuw, P. J., M. Debruyne, S. Engelen, and M. Hubert. 2006. Robust and outlier detection in chemometrics. *Critical Reviews in Analytical Chemistry* 36:221–42.
Ruimin, L., and S. Zhenyao. 2007. Integrated assessment and changes of ecological environment in the Daning River watershed. *Frontiers of Biology in China* 2 (4): 474–78.
Sampson, P. D., D. Damian, and P. Guttorp. 2001. Advances in modeling and inference for environmental processes with nonstationary spatial covariance. NRCSE-TRS No. 61, National Research Center for Statistics and the Environment, Seattle, Washington.
Satapathy, D. R., P. R. Salve, and Y. B. Katpatal. 2009. Spatial distribution of metals in ground/surface waters in the Chandrapur district (Central India) and their plausible sources. *Environmental Geology* 56:1323–52.
Savage, J. C. 1988. Principal component analysis of geodetically measured deformation in Long Valley caldera, eastern California, 1983–1987. *Journal of Geophysical Research, Solid Earth* 93:13297–305.
Schuenemeyer, J. H., and L. J. Drew. 2011. *Statistics for earth and environmental scientists*. Hoboken, NJ: Wiley.
Shen, H., and J. Z. Huang. 2008. Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis* 99 (6): 1015–34.
Shi, Z. H., L. D. Chen, J. P. Hao, T. W. Wang, and C. F. Cai. 2009. The effects of land use change on environmental quality in the red soil hilly region, China: A case study in Xianning county. *Environmental Monitoring Assessment* 150:295–306.
Skånes, H. M., and R. G. H. Bunce. 1997. Directions of landscape change (1741–1993) in Virestad, Sweden—Characterized by multivariate analysis. *Landscape and Urban Planning* 38:61–75.
Skočaj, D., A. Leonardis, and H. Bischof. 2007. Weighted and robust learning of subspace representations. *Pattern Recognition* 40 (5): 1556–69.
Spencer, K. M., J. Dien, and E. Donchin. 1999. A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology* 36:409–14.
———. 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology* 38:343–58.
Stanimirova, I., M. Daszykowski, and B. Walczak. 2007. Dealing with missing values and outliers in principal component analysis. *Talanta* 72:172–78.
Tanaka, Y., and F. Zhang. 1999. R-mode and Q-mode influence analyses in statistical modelling: Relationship between influence function approach and local influence approach. *Computational Statistics and Data Analysis* 32:197–218.
Thió-Henestrosa, S., and J. A. Martín-Fernández. 2005. Dealing with compositional data: The freeware CoDaPack. *Mathematical Geology* 37 (7): 773–93.
Thomas, C. W., and J. Aitchison. 2005. Compositional data analysis of geological variability and process: A case study. *Mathematical Geology* 37 (7): 753–72.
Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *Journal of Royal Statistical Society B* 58 (1): 267–88.
Tinkler, K. J. 1972. The physical interpretation of eigenfunctions of dichotomous matrices. *Transactions of the Institute of British Geographers* 55:17–46.

Tipping, M., and C. Bishop. 1999. Probabilistic principal components analysis. *Journal of the Royal Statistical Society (Series B)* 61 (3): 611–22.

Tran, L. T., C. G. Knight, R. V. O'Neill, E. R. Smith, K. H. Ritters, and J. Wickham. 2002. Fuzzy decision analysis for integrated environmental vulnerability assessment of the Mid-Atlantic region. *Environmental Management* 29 (6): 845–59.

van den Boogaart, K. G., and R. Tolosana-Delgado. 2008. "Compositions": A unified R package to analyze compositional data. *Computers & Geosciences* 34:320–38.

Verfaillie, E., I. Du Four, M. Van Meirvenne, and V. Van Lancker. 2009. Geostatistical modeling of sedimentological parameters using multi-scale terrain variables: Application along the Belgian part of the North Sea. *International Journal of Geographical Information Science* 23 (2): 135–50.

Vines, S. K. 2000. Simple principal components. *Applied Statistics* 49 (4): 441–51.

Wackernagel, H. 2003. *Multivariate geostatistics*. 3rd ed. Berlin, Germany: Springer-Verlag.

Wang, F., and M. M. Wall. 2001. *Modelling multivariate data with a common spatial factor. Report 2001–008*, University of Minnesota, Division of Biostatistics, St. Paul, MN.

———. 2003. Generalized common spatial factor model. *Biostatistics* 4:569–82.

Wang, W., and Q. Cheng. 2008. Mapping mineral potential by combining multi-scale and multi-source geoinformation. Paper presented at the IEEE International Geoscience & Remote Sensing Symposium 2008, Boston.

Wartenberg, D. 1985a. Multivariate spatial correlations: A method for exploratory geographical analysis. *Geographical Analysis* 17:263–83.

———. 1985b. Spatial autocorrelation as a criterion for retaining factors in ordinations of geographic data. *Mathematical Geology* 17:665–82.

Whittle, P. 1953. On principal components and least square methods of factor analysis. *Scandinavisk Aktuarietidskrift* 36:223–39.

Widmann, M., and C. Schär. 1997. A principal component and long-term analysis of daily precipitation in Switzerland. *International Journal of Climatology* 17:1333–56.

Yang, J., D. Zhang, A. F. Frangi, and J. Yang. 2004. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions in Pattern Analysis and Machine Intelligence* 26 (1): 131–37.

Yarnal, B., A. C. Comrie, B. Frakes, and D. P. Brown. 2001. Developments and prospects in synoptic climatology. *International Journal of Climatology* 21:1923–50.

Ye, J., R. Janardan, and Q. Li. 2004. GPCA: An efficient dimension reduction scheme for image compression and retrieval. In *Proceedings of the 10th ACM SIGKDD International Conference in Knowledge Discovery and Data Mining*, 354–63. doi: 10.1145/1014052.1014092

Zhang, C. 2006. Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland. *Environmental Pollution* 142:501–11.

Zhang, C., and O. Selinus. 1998. Statistics and GIS in environmental geochemistry—Some problems and solutions. *Journal of Geochemical Exploration* 64:339–54.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B* 67 (2): 301–20.

Zou, H., T. Hastie, and R. Tibshirani. 2006. Sparse principal component analysis. *Journal of Computational and Graphical Statistics* 15 (2): 265–86.

Zwiers, F. W., and H. Von Storch. 2004. On the role of statistics in climate research. *International Journal of Climatology* 24:665–80.

Correspondence: Centre for GeoInformatics, School of Geography and Geosciences, University of St. Andrews, St. Andrews, Fife KY16 9AL, Scotland, UK, e-mail: urska.demsar@st-andrews.ac.uk (Demšar); stewart.fotheringham@st-andrews.ac.uk (Fotheringham); National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland, e-mail: paul.harris@nuim.ie (Harris); School of Environmental Sciences, University of Liverpool, Liverpool L69 3BX, UK, e-mail: christopher.brunsdon@liverpool.ac.uk (Brunsdon); Department of Electronic Engineering, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland, e-mail: sean.mcloone@eeng.nuim.ie (McLoone).