Geographically Weighted Non-negative Principal Components Analysis for Exploring Spatial Variation in Multidimensional Composite Index

N. Tsutsumida[*,1], D. Murakami[2], T. Yoshida[3], T. Nakaya[4], B. Lu[5], and P. Harris[6]

[1,*] Graduate School of Global Environmental Studies, Kyoto University, Japan
[2] Department of Data Science, Institute of Mathematical Statistics, Japan
[3] Center for Global Environmental Studies, National Institute for Environmental Studies, Japan
[4] Graduate School of Environmental Studies, Tohoku University, Japan
[5] School of Remoting sensing and information engineering, Wuhan University, China
[6] Sustainable Agriculture Systems North Wyke, Rothamsted Research, UK
*Email: naru@kais.kyoto-u.ac.jp

Abstracts

The objectives of this study are to develop geographically weighted (GW) non-negative principal components analysis (PCA) and to explore spatial variations of contributions to a multidimensional composite index (MCI) from spatial multidimensional data. As a case study, we produced a MCI for earthquake risk in Tokyo, Japan, in 2018 from the collapse risk of buildings, the fire risk, and the evacuation risk associated with earthquake. GW non-negative PCA was applied to these data to uncover spatial variation of non-negative weightings (eigenvectors). This study demonstrates GW non-negative PCA provides more informative outputs when considering local differences of contributions of multidimensional data to the composite index.

Keywords: non-negative PCA, geographically weighted model, multivariable spatial data, multidimensional composite index

1. Introduction

Principal components analysis (PCA) is a well-known dimension reduction technique used in multidimensional data. The PCA allows the extraction of several orthogonal principal components (eigenvectors) which consist of linear combinations of multidimensional data by accounting for the variance in a data set. The eigenvector with the largest variation of input data is allocated as the first eigenvector, providing a means to build a multidimensional composite index (MCI) in the case where the multidimensional data correlate with each other as a whole. Eigenvectors, which are coefficients or weights of the linear combination of the input data, would be feasibly interpretable if such eigenvectors are all positive, however eigenvectors calculated by conventional PCA can be negative. Constraints of non-negative eigenvectors are important to be additive, not subtractive, combinations of multidimensional data. To deal with this issue, non-negative PCA has been developed to coerce eigenvectors to be non-negative (Sigg and Buhmann 2008). Another issue with conventional PCA is that it does not incorporate spatial effects for the use of geospatial data. To incorporate spatial heterogeneity into the analysis (Demšar et al. 2013), geographically weighted (GW) PCA has been proposed and applied to geographical studies (Harris, Brunsdon, and Charlton 2011; Lu et al. 2014; Gollini et al. 2015; Tsutsumida, Harris, and Comber 2017). GWPCA is locally weighted PCAs in geographic space and is applied at each point at the centre of a distance-decayed moving window or kernel. The GWPCA eigenvectors can vary spatially, however again such eigenvectors can be negative, resulting in the difficulty of building a MCI straightforwardly. To address this, this study proposes GW non-negative PCA which combines non-negative PCA and GWPCA to make local eigenvectors non-negative at any locations with accounting for a spatial heterogeneity. We applied GW non-negative PCA to three risks of earthquake in Tokyo metropolitan areas, Japan, to build a MCI as a case study.

2. Methods

2.1. Data

Our study data consist of 3138 administrative units in 23 wards in Tokyo, Japan. We chose three earthquake risks: the collapse risk of buildings; the earthquake-related fire risk; and the evacuation risk due to insufficient infrastructures (Figure 1). With considerations for summaries of building types/characteristics, and landform/geological types, the collapse risk of buildings was investigated based on the degree of liquefaction and subsidence of the ground, and the fire risk was estimated by a potential of fire occurrence and fire extensions caused by the effects of earthquake. The evacuation risk due to insufficient infrastructures was estimated based on the lack of sufficient open space and road networks for evacuations.

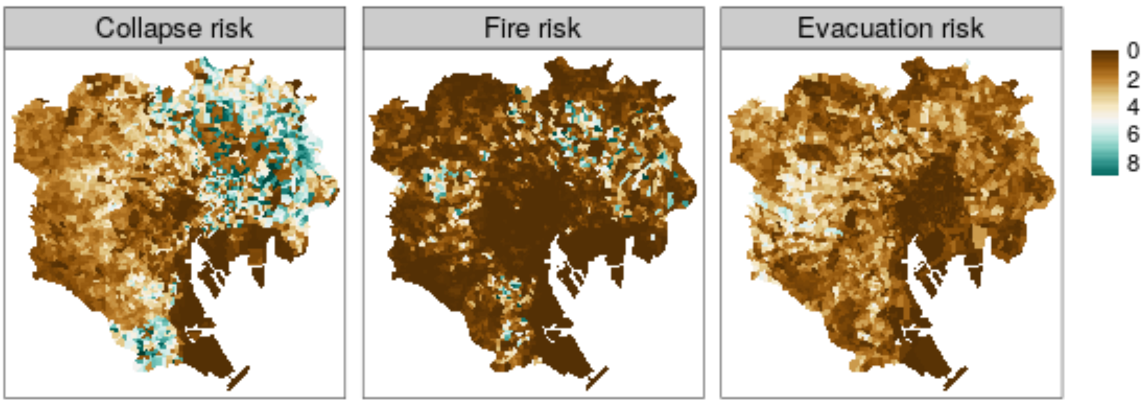

Figure 1. Three variables used in this study.

2.2. Methods

Logarithm transformations were firstly applied to these data. Napier's constant value was added to three risks to avoid negative values prior to the transformation. GW non-negative PCA integrates the essence of conventional PCA, non-negative PCA, and GW model. Given a $n \times m$ matrix X consists of m objective variables at n observation sites, GWPCA at the i-th location with coordinates (u_i, v_i) on the geographic space decomposes the GW variance-covariance matrix of X, which is defined by $\Sigma_i = X^T W_i X$, as follows:

$$L_i V_i L_i^T = \Sigma_i,$$

Equation 1

where L_i is a GW matrix of eigenvectors and V_i is a GW diagonal matrix of eigenvalues. W_i is a diagonal matrix of geographic weights that can be generated using a kernel function. In the case study, we used a bi-squared function for the j-th diagonal:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 & \text{if } |d_{ij}| < b, \\ 0 & \text{otherwise} \end{cases}$$

Equation 2

where the bandwidth b is the geographic distance b and d_{ij} is the distance between spatial locations of the i-th and j-th locations in the data. b was arbitrary determined as 1200 (38.2% of n in this study). Optimization approaches such as Leave-one-out residual (LOOR) found in Harris, Brunsdon, and Charlton (2011) for GWPCA is under development.

The first eigenvector l_i^1 for GWPCA at the location i is applied so that:

$$l_i^1 = argmax\ l_i^T \Sigma_i l_i,\ \ subject\ to\ ||l_i|| = 1$$

Equation 3

where l_i^1 is the m-length first column of the GW eigenvector matrix L_i at the location i. Finally, the GW non-negative PCA uses an additional restriction:

$$subject\ to\ \ l_i >= 0,$$

Equation 4

so that all eigenvectors at any locations are non-negative.

The MCI was calculated by a weighted geometric mean with the use of the result of GW non-negative PCA written as:

$$MCI_i = \left(\prod_{k=1}^{m} x_{i,k}{}^{l^1 ik}\right)^{\frac{1}{m}},$$

Equation 5

where $x_{i,k}$ and l_{ik}^1 are the k-th variable and its fitst eigenvector at the location i, respectively, and are both non-negative.

The GW non-negative PCA function was developed based on the combination of the *nsprcomp* function in the *nsprcomp* package (Sigg and Buhmann 2008) and the *gwpca* function in the *GWmodel* package (Gollini et al. 2015) in the R environment.

3. Results

Figure 2 indicates the MCI calculated from GW non-negative PCA and finds many patches with relatively higher risks surrounding the centre of the study area. We would like to know how the degree of each risk contributes to this MCI locally.

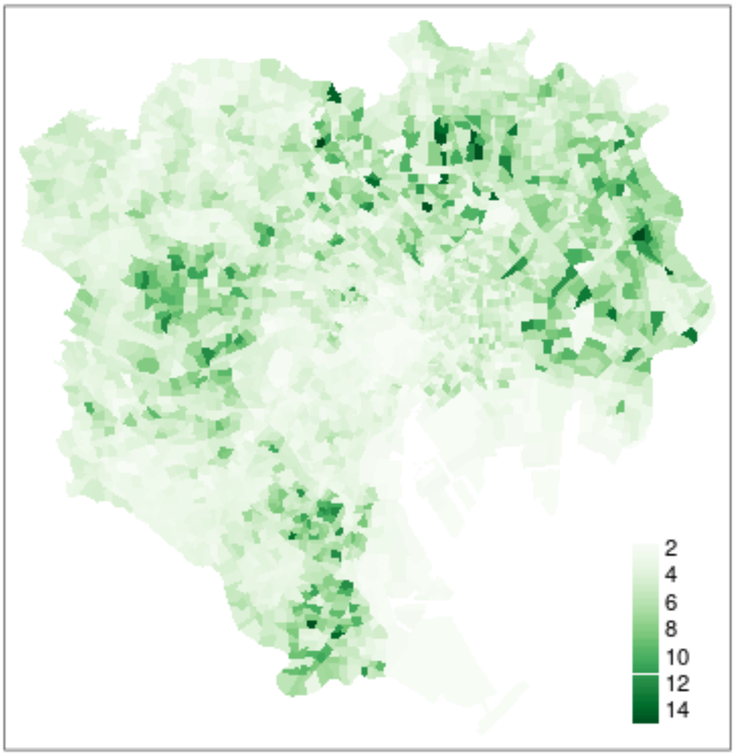

Figure 2. Multidimensional composite index built by GW non-negative PCA.

GW non-negative PCA gives spatial surfaces of the first eigenvector shown in Figure 3. The first eigenvector map (not shown) of GWPCA indicates many aggregated areas with negative values, suggesting the difficulty of building a MCI. Spatial surfaces of the first eigenvector, used as weightings of the geometric mean for the MCI, demonstrates how the degree of contributions to the MCI from variables vary spatially. The collapse risk variable is likely to contribute to the MCI relatively high in the middle-east part of the area with a high density of buildings on reclaimed land or an alluvial plain. The fire risk variable is likely to contribute to the MCI being relatively high in the north-east part of the area and surrounding the centre of the Tokyo. The evacuation risk variable tends to contribute to the MCI being relatively high in the western side of Tokyo.

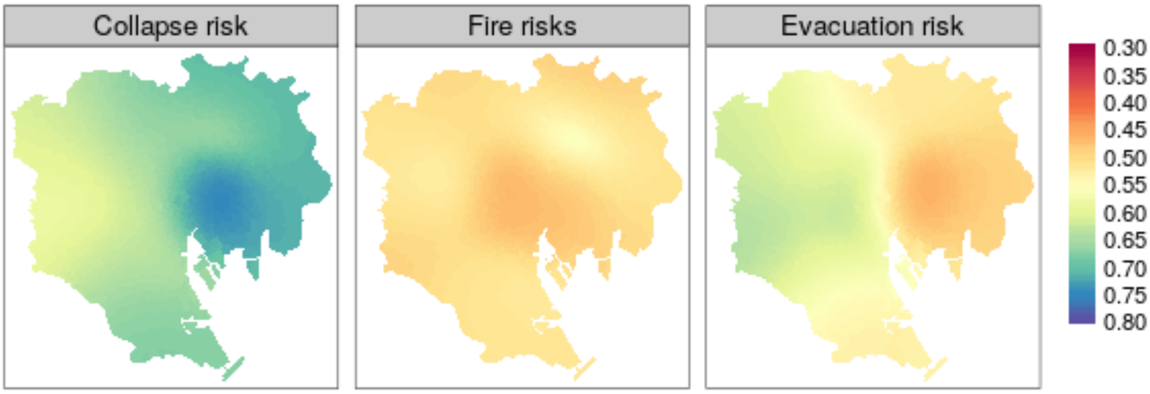

Figure 3. Spatial distribution of eigenvectors for GW non-negative PCA.

4. Conclusions

In this study, we developed and proposed GW non-negative PCA for building a MCI while allowing for spatial variations in the first eigenvector, which suggests the degree of contributions to the MCI. GW non-negative PCA adapts standard GWPCA for this purpose, as standard GWPCA may load negatively in some areas, causing difficulties in MCI construction. Our demonstrated approach provides detailed local information of earthquake risks hidden in any global measure of MCI, and is transferable to many other domains.

Acknowledgement

This research was funded by ROIS-DS-JOINT (006RP2018 and 004RP2019) and the Kyoto University Foundation. We acknowledge Bureau of Urban development Tokyo Metropolitan Government for providing data.

Reference

Demšar, Urška, Paul Harris, Chris Brunsdon, A. Stewart Fotheringham, and Sean McLoone. 2013. “Principal Component Analysis on Spatial Data: An Overview.” *Annals of the Association of American Geographers* 103 (1): 106–28.

Gollini, Isabella, Binbin Lu, Martin Charlton, Christopher Brunsdon, and Paul Harris. 2015. “GWmodel : An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models.” *Journal of Statistical Software* 63 (17): 85–101.

Harris, Paul, Chris Brunsdon, and Martin Charlton. 2011. “Geographically weighted principal components analysis.” *International Journal of Geographical Information Science* 25 (10): 1717–36.

Lu, Binbin, Paul Harris, Martin Charlton, and Chris Brunsdon. 2014. “The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models.” *Geo-Spatial Information Science* 17 (2): 85–101.

Sigg, Christian D., and Joachim M. Buhmann. 2008. “Expectation-maximization for sparse and non-negative PCA.” In *Proceedings of the 25th International Conference on Machine Learning - Icml ’08*, 960–67. New York, New York, USA: ACM Press.

Tsutsumida, Narumasa, Paul Harris, and Alexis Comber. 2017. “The Application of a Geographically Weighted Principal Component Analysis for Exploring Twenty-three Years of Goat Population

Change across Mongolia." *Annals of the American Association of Geographers* 107 (5): 1060–74.