Article:

Tsutsumida, N, Harris, P and Comber, A orcid.org/0000-0002-3652-7846 (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). pp. 1060-1074. ISSN 2469-4452

https://doi.org/10.1080/24694452.2017.1309968

Title: The application of a geographically weighted principal components analysis for exploring 23 years of goat population change across Mongolia

Author names/Affiliations:

Narumasa Tsutsumida[1], Paul Harris[2], Alexis Comber[3]

[1] Graduate School of Global Environmental Studies, Kyoto University, Yoshida-honmachi, Sakyo, Kyoto, Kyoto, JP 606-8501

[2] Rothamsted Research, North Wyke, Okehampton, Devon, UK EX20 2SB

[3] School of Geography, University of Leeds, Leeds, UK LS2 9JT

Abstract

The dzud are extreme weather events in Mongolia of deep snow, severe cold, or other conditions that render forage unavailable or inaccessible, which in turn, result in extensive livestock deaths. Mongolia is economically vulnerable to extreme events due to an increase in non-professional herders and the livestock population, that a de-regularised industry has brought about. Thus it is hugely informative to try to understand the spatial and temporal trends of livestock population change. To this end annual livestock census data are exploited and a geographically weighted principal components analysis (GWPCA) is applied to goat data recorded from 1990 to 2012 in 341 regions. This application of GWPCA to temporal data is novel and is able to account for both temporal and spatial patterns in goat population change. Furthermore, the GWPCA methodology is extended to simultaneously optimise the number of components to retain and the kernel bandwidth. In doing so, this study not only advances the GWPCA method but also provides a useful insight into the spatio-temporal variations of the Mongolian goat population.

Keywords

Spatio-temporal; GWmodel; Livestock; Grasslands; Sustainability

Introduction

It is important to evaluate the impacts of disasters to improve and support agricultural planning. In Mongolia, deep snow, severe cold and associated conditions, called *dzud*, occur repeatedly and make forage unavailable or inaccessible to livestock. This results in high livestock mortality (Fernandez-Gimenez, Batbuyan, and Baival 2012; Fernández-Giménez et al. 2015) and huge economic losses, as livestock in Mongolia represents 16% of national GDP (UNDP and NEMA 2010). Traditional nomadic pastoralism is one of the most sustainable ways of life on grasslands and sparsely vegetated lands, as are commonly found in Mongolia (Millennium Ecosystem Assessment 2005; Research Institute for Humanity and Nature 2012). Vegetation availability depends on the impacts of livestock grazing which has been well managed by nomadic herders over thousands of years (Research Institute for Humanity and Nature 2012), and is not suited to intensive livestock and crop production. In particular, excessive livestock populations, whether managed commercially or traditionally, endangers sustainability (Geist and Lambin 2004; Suttie, Reynolds, and Batello 2005). Recent changes to the Mongolian livestock industry, which has become swamped with non-professional herders due to de-regularisation, has made the grasslands vulnerable to environmental change and to extreme weather events. Thus there is a clear need to understand the spatio-temporal trends in Mongolia's livestock populations, accounting for the impacts of the dzuds.

Data on livestock populations (sheep, goat, horse, cattle and camel) are collected for 341 regions (a second administrative subdivision level, called *soum*) in Mongolia by the official statistics service. For this study, goat data for a 23 year period 1990-2012, covering two devastating duzds during 2001-2 and 2009-10, was analysed. A geographically weighted principal components analysis (GWPCA) was used with the aim of generating spatio-temporal insights about goat populations, particularly for abrupt changes caused by dzuds. A standard principal components analysis (PCA) provides a useful starting point to reduce the dimensionality of the temporally-indexed goat data and to observe major trends. However, PCA ignores any spatial structure in the data (Demšar et al. 2013), whilst GWPCA is explicitly designed to do so (Fotheringham, Brunsdon, and Charlton 2002; Lloyd 2010; Harris, Brunsdon, and Charlton 2011; Harris et al. 2015).

GWPCA constructs local PCAs from subsets of the data under a moving window or kernel where the data are weighted by their distance to the kernel centre. Critical factors in the operation of GWPCA are the specification of the kernel bandwidth, which controls the degree of localness, and choosing the number of components to retain (NCR). Bandwidth optimization routines exist, but are dependent on the NCR value, that has to be pre-specified (Harris et al. 2011; 2015). This paper addresses this technical limitation of GWPCA and proposes two novel methods to determine the bandwidth and NCR value simultaneously. In doing so, a better understanding of the spatio-temporal dynamics of the Mongolian goat populations in relation to the duzds is provided.

This article is organised as follows. Firstly, background information on Mongolian livestock populations is presented, together with introductions to PCA and GWPCA. Secondly, the study data is described. Thirdly, PCA and the GWPCA methodology are formally presented. Fourthly, the results of applying PCA and GWPCA to the goat population data are given, including the outcomes of the dual bandwidth and NCR optimisations for GWPCA. Finally, a summary, discussion and concluding remarks section is given.

Background

Livestock populations in Mongolia

Nomadic pastoralism has provided a sustainable way of life for thousands of years in Mongolia (Research Institute for Humanity and Nature 2012). Although Mongolian grasslands have been well-managed, there are concerns about the impacts of increases in livestock populations. The lives of nomadic pastoralists have been strongly influenced by political changes, especially the move from a planned economy to a free-market economy in 1992 (Fernandez-Gimenez 2006). Prior to this, livestock production was managed centrally and nomadic herders raised state-owned livestock, restricting excessive livestock production. The government encouraged herders to organize their collectives locally, and gave salaried (professional) herders the responsibility of breeding livestock. Collectives were self-regulated in their land use and their seasonal long-distance travel, resulting in good pasture maintenance with advance preparedness for keeping livestock secure from extreme events (Fernandez-Gimenez 2006). Since the transition to a free-market economy, pastures have been managed by individual herders, leading to serious sustainability and land management issues, as herders are now focussed on profit and their number has more than doubled (Togtokh 2008) – all of which makes the livestock industry more vulnerable. Five main livestock types are found in Mongolia (sheep, goat, horse, cattle and camel), and the country-wide goat population has rapidly increased since the government policy change in 1992 (Figure 1). This increase is primarily due to the strong demand for goat cashmere (Saizen, Maekawa, and Yamamura 2010), but unfortunately, the rate of increase threatens livestock sustainability and the nomadic lives of herders.

Livestock losses occur during periods of the dzud as a result of deep snow and severe cold (Fernandez-Gimenez, Batbuyan, and Baival 2012; Tsutsumida and Saizen 2014). Additional pressure is also placed on herders as the dzud directly results in reduced opportunities for grazing in the summer that follows, as a result of droughts. Effects of this combination of winter dzuds and summer droughts can be seen in Figure 1 for the years 2001-2 and 2009-10, where declines in the sheep and goat populations are clearly evident. As a result of the 2009-10 dzud, approximately 20% of the country's livestock population were killed, affecting 28% of Mongolia's human population (Fernandez-Gimenez, Batbuyan, and Baival 2012;

Fernández-Giménez, Batkhishig, and Batbuyan 2012). The increase in non-professional herders, with limited knowledge in traditional herding, has compounded this livestock loss in the dzud years (UNDP and NEMA 2010).

Little attention has been paid into the geographical dynamics of the Mongolian livestock population, over this 23-year period of change. Although some research has been conducted, notably by Saizen, Maekawa, and Yamamura (2010) who found areas of goat population increase to be independent of land cover. Saizen, Maekawa, and Yamamura (2010) also noted that in more severe conditions, goat herders were not restricted to the grazing pastures close to Ulaanbaatar, as goats are more resilient to severe conditions, and the fact that a key goat product, cashmere, is relatively portable. Liu et al. (2013) investigated the relationship between goat population density and various climatic factors and suggested that the marked increase in goat population density was a key non-climatic factor affecting grassland degradation. Hilker et al. (2014) observed that livestock population increases, associated with vegetation greenness, were primarily in the western part of Mongolia. Thus previous research has tended to focus on environmental issues and not the vulnerability of the livestock populations due to dzuds, even though they are relatively common. This study seeks to address this oversight by investigating the spatio-temporal pattern of goat population change in relation to the varying impacts of dzuds, via a GWPCA approach.

PCA and geographically weighted PCA

PCA is standard information reduction technique, commonly employed in many areas of data analysis. It transforms a set of m correlated variables into a new set of m uncorrelated variables called components. The components are linear combinations of the original variables and can allow for a better understanding of differing sources of variation and key trends in data. Its use as a dimension reduction technique is viable if the first few components account for most (say, 80 to 90%) of the variation in the original data. Component scores and component loadings data are produced, where the latter display how much each of the original variables attribute to the dimensional variance of the overall data. For details, see Jolliffe (2002).

There are a number of ways that a PCA can be usefully applied to multivariate spatio-temporal data sets, such as the livestock data sets for this study (when all five livestock types are considered). Demšar et al. (2013) provide a review in this respect, where the many dimensional groups can be treated in a variety of ways. This study applies a PCA to the goat population data, collected over a 23-year time period. Thus the application of PCA is to a set of 23 time-stamped geographic variables, where each variable measures goat population for a different year. This means that the PCA only accounts for the temporal correlations in the data.

PCAs have been used to identify spatio-temporal data characteristics in many scientific fields (e.g. Felipe-Sotelo et al. 2006; Lasaponara 2006; and see Demsar et al. 2013 therein). For example, Lasaponara (2006) applied PCA for the evaluation of vegetation anomalies from multi-temporal remote sensing data; and found that the first principal component (PC1) related to a general vegetation distribution pattern, while the second (PC2) indicated a decreasing trend of vegetation amount. In the atmospheric sciences, PCAs are commonly applied to spatio-temporal (univariate) data, and is referred to as an empirical orthogonal function (EOF) analysis (e.g. Obled and Creutin (1986)). However for EOFs, the time series data is sufficiently long enough to consider PCA in Q-mode (rather than the usual R-mode), thus spatial correlations are captured as the data matrix is transposed. If the livestock population data of this study was considered temporally long enough (i.e. collected over 100 years, say), then such an application of Q-mode PCA could also have been considered. Instead, an R-mode PCA is applied and thus only temporal correlations in the goat data are captured. Note that applications of PCA to spatio-temporal data entails that Q-mode PCA is often referred to as S-mode PCA, where "S" denotes spatial, and R-mode PCA is often referred to as T-mode PCA, where "T" denotes temporal. The idea being that Q-mode and R-mode PCAs are reserved for attribute sub-space applications with no spatio-temporal context.

However, a standard (R-mode) PCA application to this study's goat data does not account for any spatial effects, because it only ensures a non-spatial linear transform (Demšar et al. 2013). In order to deal with such a naïve application, but from a spatial effects point of view only, GWPCA can be used. This adaptation of PCA provides a better description of any spatial phenomenon in the structure of the data. It uses a moving window weighting technique and constructs a localized PCA at all target locations (e.g. a grid, such as the application by Comber, Harris, and Tsutsumida (2016)). It is important to note, that although spatio-temporal correlations in the goat population data are captured via GWPCA, only spatial dependencies in the data are fully captured. Temporal dependencies such as those between neighbouring years, are not fully captured nor are true spatio-temporal dependencies. That requires a further extension to GWPCA to a full spatio-temporal approach, similar that proposed for GW regression by Huang, Wu, and Barry (2010). Thus in this study, both PCA and GWPCA are applied in order to provide a better understanding of the dynamics of the Mongolian goat population data, at a soum-level scale, across the period 1990–2012.

Study data

Annual livestock population data were obtained from the National Statistical Office (NSO) of Mongolia for the period 1990–2012. Populations were summarized per soum, an administrative sub-division area. Since local governments collect taxes from herders according to herd size, the data are assumed to reflect livestock numbers reasonably well (Saizen, Maekawa, and Yamamura 2010). Administrative boundaries slightly changed during the 23-year study period. To cater for this, the data were merged accounting for all 341 soums, using the most recent boundaries. Thus all data are taken into account when a soum changed or was incorporated into a neighbour. Missing data that arose because of these changes, were infilled using a probabilistic PCA method provided in the pcaMethod R package (Stacklies et al. 2007). This infilling was fairly minor and was not considered an issue for subsequent analyses.

As would be expected, the goats data are highly correlated, especially across adjacent years as shown in Figure 2, with the weakest correlations between the dzud year of 2002 and all others, and the dzud year of 2010 and all others. Intuitively, this correlation analysis for the temporally-indexed goats data, directly implies that goat population change does not increase or decrease at the same rate across all 341 soums. This in turn, provides some insight into the expected value of a spatial analysis of the goats data, via a GWPCA.

Methods

Principal components analysis (PCA)

Given an $n \times m$ dimensional data matrix X, a PCA to this data consists of conducting this transformation:

$$L\,V\,L^T = S \quad (1)$$

where L is the matrix of eigenvectors with $n \times m$ dimension, V is the diagonal matrix of eigenvalues, and S is the variance–covariance matrix with $m \times m$ dimension. V indicates the eigenvalues of the PCs, representing the axes of a new dimension. Each column of L represents the loadings corresponding to a PC. The PCs are ordered according to the size of eigenvalues, meaning that PC1 corresponds to the largest eigenvalue, and PC2 corresponds to the second largest, and so on. Transformed component scores in matrix T is represented by

$$T = XL \tag{2}$$

where T consists of a linear combination of the original values, which in this study is the multi-temporal goat population data with n = 341 and m = 23.

<u>Geographically weighted principal components analysis (GWPCA)</u>

A GWPCA utilises a kernel weighting approach where localised PCs are found at target locations. At a target location, neighbouring observations are weighted by a distance-decay weighting function, and then a standard PCA is locally applied to its own specific weighted data subset. The size of the window over which this localised PCA might apply is controlled by the kernel's bandwidth. Small bandwidths lead to more rapid spatial variation in the results whereas large bandwidths yield results increasingly close to the global PCA solution. This study identifies an adaptive bandwidth corresponding to a bi-square kernel, a discontinuous function that generates distance-decaying weights data points within the set bandwidth. Observations outside of the bandwidth's range receive weights of zero, and hence the discontinuity. For details, see Gollini et al. (2015).

Thus for coordinates (u, v) at spatial location i, GWPCA involves the conception that the goat population time series variables x_i have a certain dependence on their locality where $\mu_{(u,v)}$ and $\Sigma_{(u,v)}$, are the GW mean vector and the GW variance–covariance matrix, respectively. This GW variance–covariance matrix is calculated by

$$\Sigma_{(u,v)} = X^T \, W_{(u,v)} \, X \tag{3}$$

where $W_{(u,v)}$ is a diagonal matrix of geographical weights that are generated by the chosen kernel weighting function. The GWPCA at spatial location i can be computed using

$$L \, V \, L^T | (u_i, v_i) = \Sigma_{(u_i, v_i)} \tag{4}$$

where $\sum(u_i, v_i)$ is the GW variance-covariance at that location. The scores matrix at the same location can be found using $T(u_i, v_i) = XL(u_i, v_i)$. On dividing each local eigenvalue by $tr\big(V(u_i, v_i)\big)$, localized versions of the proportion of the total variance (PTV) in the original data accounted for by each component can be found. Thus at each of the 341 sums of this study (i.e. the target locations), a GWPCA provides 23 components, 23 eigenvalues, a component loadings set of size 341 $\times$ 23, and a component scores set of size 341 $\times$ 23.

Bandwidth selection is crucial for the application of any GW model. For GWPCA, bandwidth selection can be guided by a ‘leave-one-out’ residual (LOOR) approach, where scores data are assessed for goodness of fit (GoF) against observed data. The optimal bandwidth is one that corresponds to LOOR data that provides the smallest GoF statistic. This cross-validation procedure and extensive commentaries on choosing bandwidths are provided in Harris et al. (2015). Of note is that the NCR value is decided upon a priori and an optimal bandwidth cannot be found if all m components are retained. Thus the results of this residual-based bandwidth selection procedure are somewhat dependent on a user-specified value of NCR. To counter this dependency, this study proposes two alternative techniques to determine the bandwidth and the NCR value, concurrently. These methodological advances are described and implemented below.

Geographically weighted correlation analysis

A GW correlation analysis (Harris and Brunsdon 2010) on the outputs from the PCA with the raw data is also conducted. Here for variables x and y at spatial location i where the geographical weights w_{ij} again accord to a bi-square function, definitions for a GW standard deviation and a GW correlation coefficient, are respectively

$$s(x_i) = \sqrt{\left. \sum_{j=1}^{n} w_{ij} \left(x_j - m(x_i)\right)^2 \middle/ \sum_{j=1}^{n} w_{ij} \right.} \qquad (5)$$

and

$$\rho(x_i, y_i) = c(x_i, y_i) \big/ \left(s(x_i)s(y_i)\right) \qquad (6)$$

, where a GW mean is

$$m(x_i) = \sum_{j=1}^{n} w_{ij} x_j \Big/ \sum_{j=1}^{n} w_{ij} \qquad (7)$$

and a GW covariance is

$$c(x_i, y_i) = \sum_{j=1}^{n} w_{ij} \left\{ \left(x_j - m(x_i) \right) \left(y_j - m(y_i) \right) \right\} \Big/ \sum_{j=1}^{n} w_{ij} \quad (8)$$

Throughout this study, GWPCA and GW correlations use functions (or adapted functions) from the GWmodel R package (Gollini et al. 2015).

Results

<u>The global PCA</u>

In order to understand any GW model output, it is always important to fit the usual global model for context. In this respect, a PCA was conducted on the 23 temporal variables describing goat populations. Table 1 shows that the first two PCs have eigenvalues greater than unity, and for these two PCs, the cumulative PTV exceeds 90%. This implicitly assumes a uniform temporal trend in goat population across all 341 sums over the 23-year period. The PCA loadings given in Table 2 indicate that the five of the most influential years are 1996-1999 and 2001 for PC1; 1990-1991 and 2010-2012 for PC2.

<u>A GW correlation analysis on the PCA scores and raw data</u>

As the component loadings in Table 2 are the (global) correlation coefficients between the component scores and the raw data, a GW correlation analysis on this data can be used to investigate whether the correlations change across study region. Figure 3 maps the GW correlations between the PCA scores data for PC1 to PC3, and the raw data from the three most influential years. The GW correlations were found using a user-specified bandwidth of 10% (i.e. each local correlation uses the nearest 34 data pairs). As would be expected, spatial coherence for such correlations is highest for PC1, but diminishes through PC2 to PC3. This suggests that the PCA is missing some spatial structure in the data, and as such, an application of GWPCA is worthwhile. Intuitively, this is expected, as the spatio-temporal trend in goat populations is not expected to be uniformly the same across all of Mongolia (as similarly suggested for observations made above, with respect to Figure 2).

<u>GWPCA calibration with dual bandwidth and NCR optimization</u>

As outlined above, in order to calibrate a GWPCA, first the NCR value needs to be user-specified and only then, can an optimal GWPCA bandwidth be found via cross-validation. In previous GWPCA studies, NCR is commonly chosen according to a 80% or 90% threshold of the cumulative PTV from the global PCA. Thus in this study, NCR = 1 or 2 would be appropriate (see Table 1). This bandwidth selection approach is far from ideal, as can be seen in Table 3, where different 'optimal' bandwidths (found by the cross-validation procedure) simply correspond to different choices of NCR (in this case, NCR values from 1 to 10). Furthermore, the results suggest a tendency to a global PCA process for the study data, as eight out of ten bandwidths are taken at 341 suggesting a kernel bandwidth that contains all of the soums data. If this is truly the case (see note 1), then there appears no value in applying GWPCA, and the localized analysis should cease at this juncture.

However, the choice of bandwidth can be investigated more deeply. This is because the results presented in Table 3 are not directly comparable, as given 'optimal' bandwidths correspond to minimized GoF statistics (not shown) where the NCR-specific LOOR data sets have been summarized by their mean. To ensure that the minimized GoF statistics are comparable across different values of NCR, the LOOR data can be summarized instead by their coefficient of variation (CoV) to provide relative (and thus comparable) GoF statistics for each bandwidth and for each NCR value. This leads to a dual optimization approach as shown in Figure 4(i), where the aim to concurrently find the bandwidth and the NCR value that corresponds to minimum GoF (LOOR CoV) value. Again considering only NCR values from 1 to 10, and a clear minimum GoF is reached at 1.296 corresponding to a bandwidth of 247 nearest neighbours and an NCR value of 5. Each individual line in the plot of Fig 4(i) corresponds to a different bandwidth choice, from 5 to 341. This constitutes the first extension to the existing bandwidth selection procedure.

A second alternative is to transfer the usual cumulative PTV approach for NCR selection to a local setting. Globally, a user-specified choice of NCR = 1 or 2 is based on the global cumulative PTV scree plot (e.g. Varmuza and Filzmoser (2009)). This approach can be transferred locally using the local cumulative PTV data from each local PCA from a series of GWPCAs. Local cumulative PTV data were calculated from GWPCAs calibrated with bandwidths ranging from 10 to a maximum of 341 and the resultant local scree plots are depicted in Figure 4(ii). Clearly, the local scree plots suggest that NCR = 2 is the point when some of the local cumulative PTV data exceeds a 90% threshold. Given this, NCR = 2 again appears appropriate for a GWPCA calibration. However, the bandwidth is still required, and unlike the existing approach a bandwidth is identified that has the smallest GoF (LOOR mean) value, but crucially also corresponds to a localized cumulative PTV value exceeding 90% (for all NCR = 2). This indicates a relative tight bandwidth of 198 nearest neighbours.

Thus in summary, there are three possible bandwidths for GWPCA calibration: (a) 341 (via NCR = 1 or 2); (b) 247 (via NCR = 5); and (c) 198 (via NCR = 2). All three should be considered as entirely valid, but where approach (a), the existing approach, strongly suggests a stationary process with respect to a PCA. Given that approach (a) has drawbacks, not only with respect to NCR/bandwidth specification, but also (indirectly) due to current limitations in the GWPCA code (see note 1), it is dropped in favour of the two newly proposed approaches (b and c) which are both viewed as a methodological advance. In the spirit of spatial exploration, which all GW models are eminently designed for, both approaches were investigated further all of the subsequent GWPCA outputs described below are specified with either: (i) a bandwidth of 247 via a NCR value of 5; or (ii) a bandwidth of 198 via a NCR value of 2.

PCA versus GWPCA results

GWPCA is now applied to account for expected spatial heterogeneity in the annual goat population data during 1990-2012 with: (i) a bandwidth of 247 via a NCR value of 5 (call this 'GWPCA-A'); and (ii) a bandwidth of 198 via a NCR value of 2 (call this 'GWPCA-B'). The GWPCA results are compared with those from global PCA, throughout. To compare GWPCA−A, GWPCA-B, and PCA, only the first two components (PC1 and PC2) from each calibration are considered. Observe that once a bandwidth is defined, local components up until any NCR value (in this case NCR = 23) can actually be found and investigated. So in this respect, the NCR values of 2 and 5 from the bandwidth selection procedure do not have to pervade the remainder of the analysis (e.g. Harris et al. 2015).

Scores data

PC1 and PC2 scores from GWPCA-A, GWPCA-B, and the global PCA are mapped in Figure 5. Observe that for GWPCA, a full, n = 341 valued scores data set is available at each location, for each component. Thus, the GWPCA scores data that are mapped are only those that fully correspond to their location. PC1 scores of GWPCA-A and GWPCA-B correlate with those from the global PCA, with r = 0.846

and r = 0.742, respectively. PC2 scores of GWPCA-A and GWPCA-B correlate with those from the global PCA, with correlations of r = 0.943 and r = 0.872, respectively. These moderate to strong correlations simply reflect the relatively large bandwidth sizes used, and such correlations would tend to unity as the bandwidth increases. However these global correlations hide spatial detail, where the study's aim is to see where the local spatial structure in the temporally-changing goat population (via the GWPCA outputs) differs to that found globally (via the PCA outputs). In this respect, the clearest regional differences in both the PC1 and PC2 scores data appear in the north-eastern regions of Mongolia, bordering Russia and also the south-western regions bordering China. Thus the temporal dynamics of goat population change is likely to be clearly different in these regions to that expected nationally.

<u>Percentage PTV data</u>

Globally, the PTV for PC1, and the cumulative PTV for PC1 and PC2 combined, are 84% and 92%, respectively. This suggests a high correlation amongst the goat population data, year on year, throughout the 23-year period. However, the global

PTV values (from PCA) implicitly assume that such relationships are constant across Mongolia - with relatively uniform changes in goat populations everywhere. Mapping the corresponding localized PTV outputs from GWPCA shows where this is the case, and the degree to which it is not (Figure 6).

Focusing on the third row only of Figure 6, regionally the temporal trend in goat population change is actually more uniform than that found globally in central northern regions (coloured dark green), where local PTV data are higher. Conversely, the temporal trend in goat population change is actually less uniform than that found globally in western regions (coloured dark pink), where local PTV data is lower. These changes in regional behaviour broadly confirms that observed for the scores data, above. The PTV maps in the first and second rows of in Figure 6 provide detail of the component contribution to the cumulative PTV maps presented in the third row. Presenting the GWPCA outputs for GWPCA-A and GWPCA-B with their different bandwidths in this way re-affirms the findings, and quantifies how non-stationarities can change at different spatial scales.

<u>Loadings data</u>

In many ways the loadings data from a GWPCA are more difficult to interpret map than the scores and PTV data. In Harris, Brunsdon, and Charlton (2011), three visualizations were proposed, which can only be conducted on a component by component basis: (a) map the 'winning variables' - i.e. those that correspond to largest absolute loading; (b) map the loading sign patterns, e.g. for eight variables, there are 256 possible sign patterns; and (c) map all loadings together using multivariate glyphs, where a spoke's length corresponds to the magnitude of the loading, whilst a spoke's colour corresponds to the sign of the loadings. In this study, the GWPCA loadings data are visualized using the first option. These 'winning year' maps are presented in Figure 7 for PC1 and PC2.

The 'winning year' for PC1 for GWPCA-A and GWPCA-B included 15 and 17 of the 23 years being selected. As so many different years 'win', this is viewed as a confirmation of the generally high correlation amongst the goat population data throughout the 23-year period. Differences between a year providing the highest loading or not, are often extremely small. Thus a 'winning year or variable' map tends to provide little useful information when this happens.

In this instance, greater insight stems from considering the ‘winning year’ maps for the next component (PC2). Now far fewer years are represented (3 to 6 of 23) and the dzud years of 2002 and 2010, strongly dominate in two clear regions; the west and south-west, and the east and north, respectively. This suggests that: (i) the dzud of 2002 and the associated goat population decline was more or less pronounced in the west and south-west than elsewhere; and (ii) the dzud of 2010 and the associated goat population decline was more or less pronounced in the east and north than elsewhere. This strongly indicates that the severity of the dzuds in 2002 and 2010 varied geographically. Visualizing the annual changes in the PCA and GWPCA loadings from PC1 and PC2 for GWPCA-A and GWPCA-B (Figure 8) shows the effects of the 2002 and 2010 dzud years on the loadings, with clear inflection points for both GWPCA fits.

Figure 9 displays the loadings maps for PC2 of GWPCA-A only, for 2001-3, and 2009-11, covering the two dzuds periods. These maps suggest that the 2001-2 dzud and the 2009-10 dzud have different regional and temporal characteristics. The impact of the 2001-2 dzud starts from central/western regions in 2001 and increases in western regions in 2002. The impact of the 2009-10 dzud appears first in western regions in 2009 and then in eastern regions in 2010. This is in contrast to the reporting of dzuds and the devastating damage to livestock populations, which is typically referred to as impacting Mongolia as a whole, and uniformly.

Discussion and conclusions

Understanding the spatio-temporal characteristics of livestock population change is essential for environmental and disaster responses, to sustainably manage grassland environments and to minimize the impact of the dzud in Mongolia. Unfortunately, such analyses are rarely conducted, as they require skilled statistical expertise (Cheng et al. 2014; Shekhar et al. 2015). This study undertook such an analysis for annual goat population data, which are known to have increased over the study period, with abrupt declines following dzud events. The application of a geographically weighted PCA (GWPCA), a spatial version of PCA, to the temporally indexed goat data allowed an understanding of the spatial and temporal variations in goat population change across Mongolia over the 23 year study period.

Mapping GWPCA scores data allowed regional differences to be observed, particularly in the north-eastern regions of Mongolia, bordering Russia and also south-western regions bordering China. Thus the temporal dynamics of goat population change is likely to be different in these regions to that expected nationally. By mapping GWPCA variance proportion data, the temporal trend in goat population change was found to be more uniform, to that found globally, in central northern regions, whilst less uniform (to that found globally) in western regions. Visualizing the ‘winning year’ maps for the GWPCA loadings, suggests that the dzud of 2002 and the associated goat population decline was more or less pronounced in the west and south-west regions and that the dzud of 2010 and the associated goat population decline was more or less pronounced in the east and north regions. This, in turn, suggests that the dzuds of 2002 and 2010 varied geographically in their severity.

It has been reported that 7.7 million livestock died as a result of the 2001-2 dzud and 9.7 million died as a result of the 2009-10 dzud (UNDP and NEMA 2010). This study helps to re-affirm that regionally-specific dzud preparation and response initiatives are required to support different landscape ecological characteristics and management strategies (Fernández-Giménez et al. 2015). This study did not consider change in livestock-type over space and time, and in this respect, future research will seek to explore the full data set of goats, sheep, cattle, camel and horse.

Such an analysis could be achieved via extending GWPCA to a full temporally and geographically weighted PCA form.

This study's application of GWPCA to temporally indexed spatial data is novel and adds to a growing portfolio of GWPCA uses, not only for spatial exploration (Lloyd 2010; Harris, Brunsdon, and Charlton 2011; Harris et al. 2015), but also for spatial anomaly detection (Harris, Brunsdon, et al. 2014; Harris et al. 2015), spatial network re-design (Harris, Clarke, et al. 2014), and spatial classification (Harris et al. 2015; Comber, Harris, and Tsutsumida 2016). Furthermore, this study usefully extended the GWPCA methodology itself to simultaneously optimise the number of components to retain and the kernel weighting bandwidth. This is considered an important advance, and should be adopted in all subsequence GWPCA studies.

Notes

[1] Observe that the current version of the GWmodel R package does not allow adaptive bandwidth values greater than the sample size to be optimally selected. Thus an adaptive bandwidth that is equal to the sample size only directly indicates a stationary spatial process provided a box-car kernel is specified. For any

distance-decay kernel, such as the bi-square, an adaptive bandwidth that is equal to the sample size can only suggest or allude to a stationary spatial process.

Acknowledgements

The authors would like to thank the National Statistics Office of Mongolia for providing livestock population data. We appreciate Dr. Izuru Saizen and Dr. Atsushi Otomo to support this work. This work was supported by Sinfonica Statistical GIS Research Grants, the RIHN (project number D-04), JSPS KAKENHI Grant Number 15K21086, grants for young researchers in GSGES KU, and KU SPIRITS project. For Paul Harris, a UK Biotechnology and Biological Sciences Research Council grant (BBSRC BB/J004308/1).

References

Cheng, T., J. Haworth, B. Anbaroglu, G. Tanaksaranond, and J. Wang. 2014. Spatiotemporal Data Mining. In *Handbook of Regional Science*, eds. M. M. Fischer and P. Nijkamp, 1173–1193. Berlin, Heidelberg: Springer Berlin Heidelberg http://link.springer.com/10.1007/978-3-642-23430-9_68.

Comber, A. J., P. Harris, and N. Tsutsumida. 2016. Improving land cover classification using input variables derived from a geographically weighted principal components analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 119:347–360. http://linkinghub.elsevier.com/retrieve/pii/S0924271616301290.

Demšar, U., P. Harris, C. Brunsdon, A. S. Fotheringham, and S. McLoone. 2013. Principal Component Analysis on Spatial Data: An Overview. *Annals of the Association of American Geographers* 103 (1):106–128. http://www.tandfonline.com/doi/abs/10.1080/00045608.2012.689236.

Felipe-Sotelo, M., L. 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 (38):7421–7436.

Fernandez-Gimenez, M. E. 2006. Land Use and Land Tenure in Mongolia : A Brief History and Current Issues. In *Rangelands of Central Asia: Proceedings of the Conference on Transformations, Issues, and Future Challenges*, eds. D. J. Bedunah, E. D. McArthur, and M. E. Fernandez-Gimenez, 30–36. Salt Lake City: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Fernandez-Gimenez, M. E., B. Batbuyan, and B. Baival. 2012. *Lessons from the Dzud: Adaptation and Resilience in Mongolian Pastoral Social-Ecological Systems*. The world bank. http://www.worldbank.org/en/news/feature/2012/11/06/lessons-from-dzud.

Fernández-Giménez, M. E., B. Batkhishig, and B. Batbuyan. 2012. Cross-boundary and cross-level dynamics increase vulnerability to severe winter disasters (dzud) in Mongolia. *Global Environmental Change* 22 (4):836–851. http://linkinghub.elsevier.com/retrieve/pii/S0959378012000684.

Fernández-Giménez, M. E., B. Batkhishig, B. Batbuyan, and T. Ulambayar. 2015. Lessons from the Dzud: Community-Based Rangeland Management Increases the Adaptive Capacity of Mongolian Herders to Winter Disasters. *World Development* 68:48–65.

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

Geist, H. J., and E. F. Lambin. 2004. Dynamic Causal Patterns of Desertification. *BioScience* 54 (9):817. https://academic.oup.com/bioscience/article-lookup/doi/10.1641/0006-3568(2004)05

4[0817:DCPOD]2.0.CO;2.

Gollini, I., B. Lu, M. Charlton, C. Brunsdon, and P. Harris. 2015. GWmodel : An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software* 63 (17):85–101. http://www.tandfonline.com/doi/abs/10.1080/10095020.2014.917453.

Harris, P., and C. Brunsdon. 2010. Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics. *Computers & Geosciences* 36 (1):54–70. http://dx.doi.org/10.1016/j.cageo.2009.04.012.

Harris, P., C. Brunsdon, and M. Charlton. 2011. Geographically weighted principal components analysis. *International Journal of Geographical Information Science* 25 (10):1717–1736. http://www.tandfonline.com/doi/abs/10.1080/13658816.2011.554838.

Harris, P., C. Brunsdon, M. Charlton, S. Juggins, and A. Clarke. 2014. Multivariate Spatial Outlier Detection Using Robust Geographically Weighted Methods. *Mathematical Geosciences* 46 (1):1–31. http://link.springer.com/10.1007/s11004-013-9491-0.

Harris, P., A. Clarke, S. Juggins, C. Brunsdon, and M. Charlton. 2014. Geographically weighted methods and their use in network re-designs for environmental monitoring. *Stochastic Environmental Research and Risk Assessment* 28 (7):1869–1887. http://link.springer.com/10.1007/s00477-014-0851-1.

———. 2015. Enhancements to a Geographically Weighted Principal Component Analysis in the Context of an Application to an Environmental Data Set. *Geographical Analysis* 47 (2):146–172. http://doi.wiley.com/10.1111/gean.12048.

Huang, B., B. Wu, and M. Barry. 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices eds. M. M. Fischer and P. Nijkamp. *International Journal of Geographical Information Science* 24 (3):383–401. http://link.springer.com/10.1007/978-3-642-23430-9 (last accessed 11 July 2014).

Lasaponara, R. 2006. On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. *Ecological Modelling* 194 (4):429–434. http://linkinghub.elsevier.com/retrieve/pii/S0304380005005454 (last accessed 5 April 2012).

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 (5):389–399. http://dx.doi.org/10.1016/j.compenvurbsys.2010.02.005.

Lu, B., P. Harris, M. Charlton, and C. Brunsdon. 2014. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science* 17 (2):85–101. http://www.tandfonline.com/doi/abs/10.1080/10095020.2014.917453.

Millennium Ecosystem Assessment. 2005. *Ecosystems and Human Well-being: Desertification Synthesis*. Washington, D.C. http://library.wur.nl/WebQuery/clc/1785684.

Obled, C., and J. D. Creutin. 1986. Some Developments in the Use of Empirical Orthogonal Functions for Mapping Meteorological Fields. *Journal of Climate and Applied Meteorology* 25 (9):1189–1204. http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%281986%29025%3C1189%3ASDITUO%3E2.0.CO%3B2.

Research Institute for Humanity and Nature. 2012. *Collapse and restoration of* *ecosystem networks with human activity*. Kyoto.

Saizen, I., A. Maekawa, and N. Yamamura. 2010. Spatial analysis of time-series changes in livestock distribution by detection of local spatial associations in Mongolia. *Applied Geography* 30 (4):639–649. http://linkinghub.elsevier.com/retrieve/pii/S0143622810000032 (last accessed 6 August 2011).

Shekhar, S., Z. Jiang, R. Ali, E. Eftelioglu, X. Tang, V. Gunturi, and X. Zhou. 2015. Spatiotemporal Data Mining: A Computational Perspective. *ISPRS International Journal of Geo-Information* 4 (4):2306–2338. http://www.mdpi.com/2220-9964/4/4/2306/.

Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig. 2007. pcaMethods a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23 (9):1164–1167. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btm069.

Suttie, J. M., S. G. Reynolds, and C. Batello. 2005. *Grasslands of the world* eds. J. . Suttie, S. G. Reynolds, and C. Batello. Rome: Food and Agriculture Organization.

Togtokh, C. 2008. Climate change adaptation strategies for pastoral communities of Mongolia’s central mountainous region. *IHDP Update* :53–58.

Tsutsumida, N., and I. Saizen. 2014. Internal Regional Migration Analysis and Modeling of Population Concentration for Ulaanbaatar, Mongolia. *Papers on Environmental Information Science* 28:25–30.

UNDP and NEMA. 2010. *Dzud national report 2009-2010*.

Varmuza, K., and P. Filzmoser. 2009. *Introduction to multivariate statistical analysis in chemometrics*. Taylor & Francis.

Author professional/contact information

NARUMASA TSUTSUMIDA is an Assistant Professor in the Graduate School of Global Environmental Studies at Kyoto University, Yoshida-honmachi, Sakyo, Kyoto, JP 606-8501. E-mail: naru@kais.kyoto-u.ac.jp. His research interests include the application of geographical information in the field of environmental studies.

PAUL HARRIS is a Senior Research Scientist at Rothamsted Research, North Wyke, Okehampton, Devon, UK EX20 2SB, Email: paul.harris@rothamsted.ac.uk. His research focuses on the development and application of spatial statistics to agricultural, ecological and environmental data.

ALEXIS COMBER holds a chair in Spatial Data Analytics in the School of Geography at University of Leeds, Leeds, UK LS2 9JT, Email: a.comber@leeds.ac.uk. His research develops methods to integrate and analyse high volumes of spatial data to uncover hidden patterns with applications in environmental, socio-economic and public health areas.

Figure captions

Figure 1. Change in livestock populations across Mongolia during 1990–2012.

Figure 2. Correlation matrix of annual goat population data (1990-2012), with the plot size proportional to the correlation.

Figure 3. GW correlation maps between PC1-3 scores of the global PCA and the raw data of the corresponding most influential years (see also Table 2).

Figure 4. GWPCA calibration: (i) GoF (via LOOR CoV) versus NCR values; and (ii) scree plots for local cumulative PTVs versus NCR values. The grey lines have a transparency term added to them. In (i) they represent bandwidths in a range of 5 to 341 and (ii) in a range of 10 to 341. The black line in (i) represents the optimal bandwidth of 247 with NCR = 5, at the minimum GoF. Black line in (ii) represents the 90% threshold of the cumulative PTV.

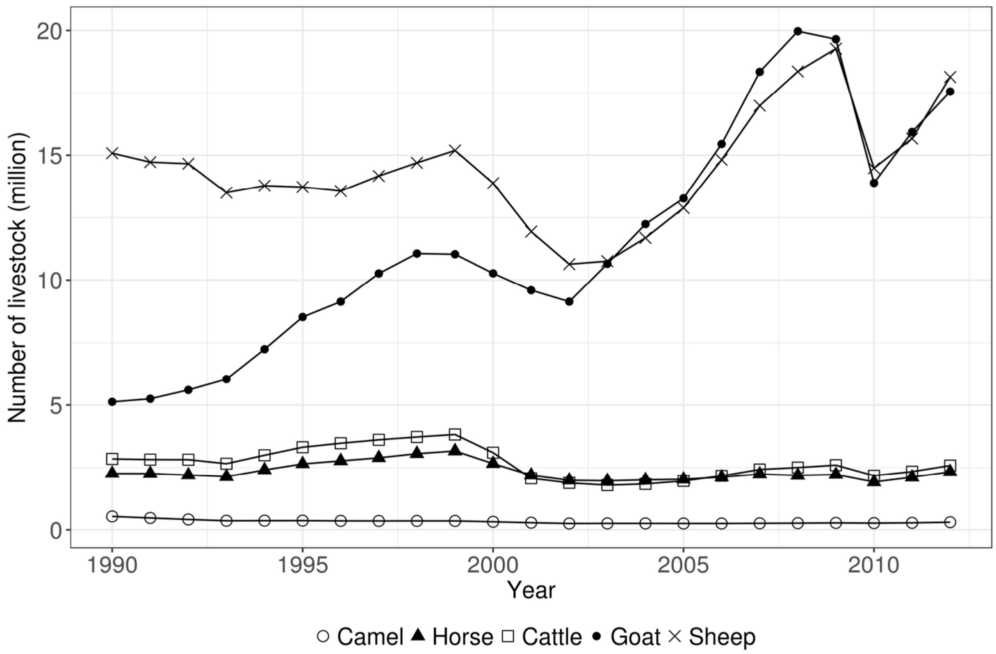
Figure 5. PC1 and PC2 scores maps for GWPCA-A (top row), GWPCA-B (middle row), and the global PCA (bottom row).

Figure 6. GWPCA-A and GWPCA-B PTV maps for PC1 (top row), PC2 (middle row) and PC1/PC2 combined (bottom row).

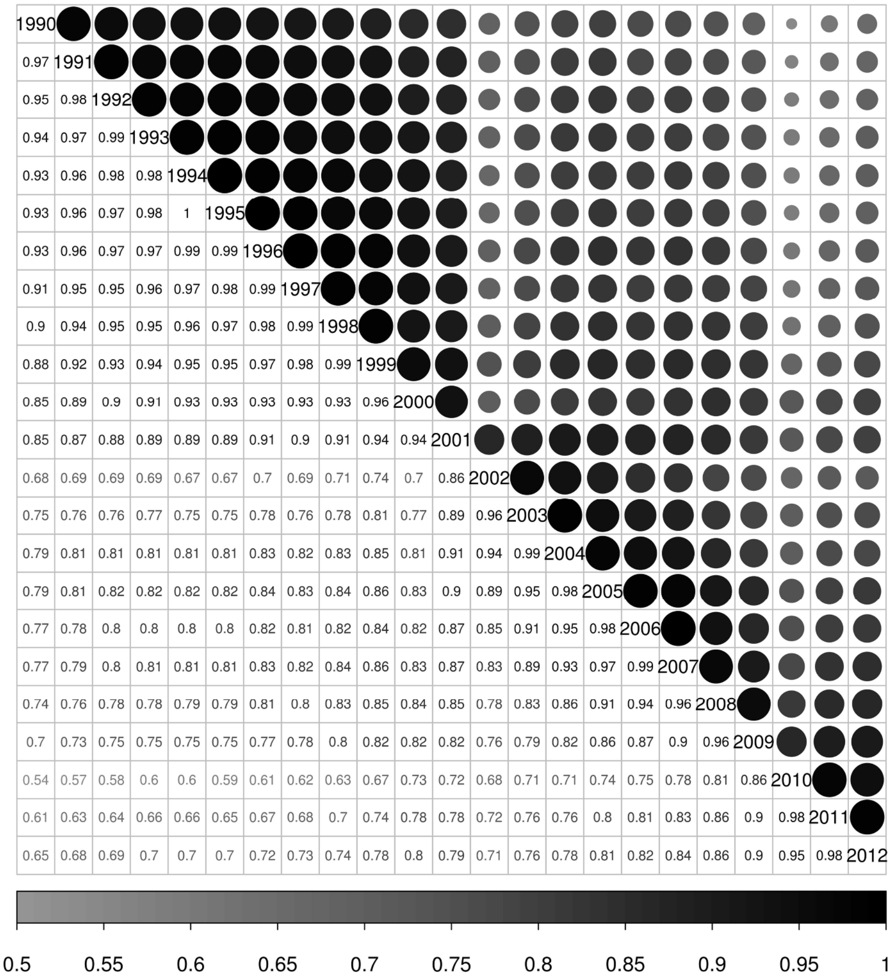
Figure 7. GWPCA-A and GWPCA-B 'winning year' maps (by highest loadings) for PC1 and PC2. Years when dzud occurred are highlighted in grey and black.

Figure 8. GWPCA-A and GWPCA-B loadings for PC1 and PC2, displayed over the 23 study years. The grey lines have a transparency term and represent the loading score at every soum. The black lines represent the loadings from the global PCA. Dark grey rectangles represent dzud periods 2001-2 and 2009-10.

Figure 9. Maps for PC2 loadings from GWPCA-A over dzud periods of 2001-3 (top row) and 2009-11 (bottom row).



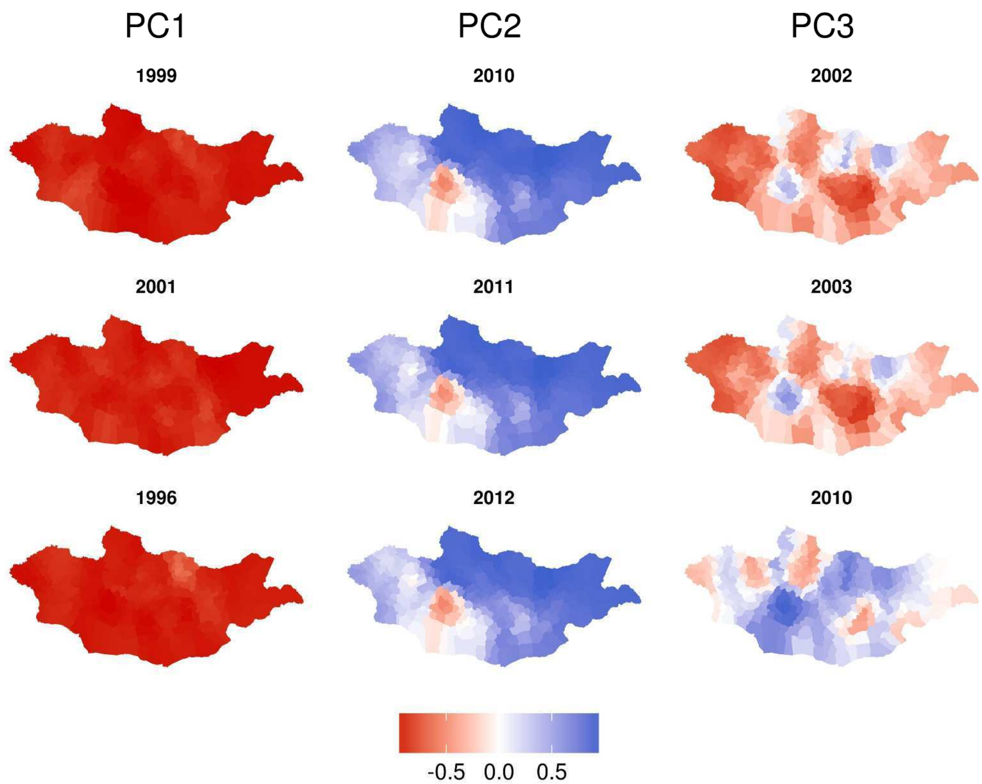
Change in livestock populations across Mongolia during 1990–2012.
Figure 1
152x101mm (300 x 300 DPI)



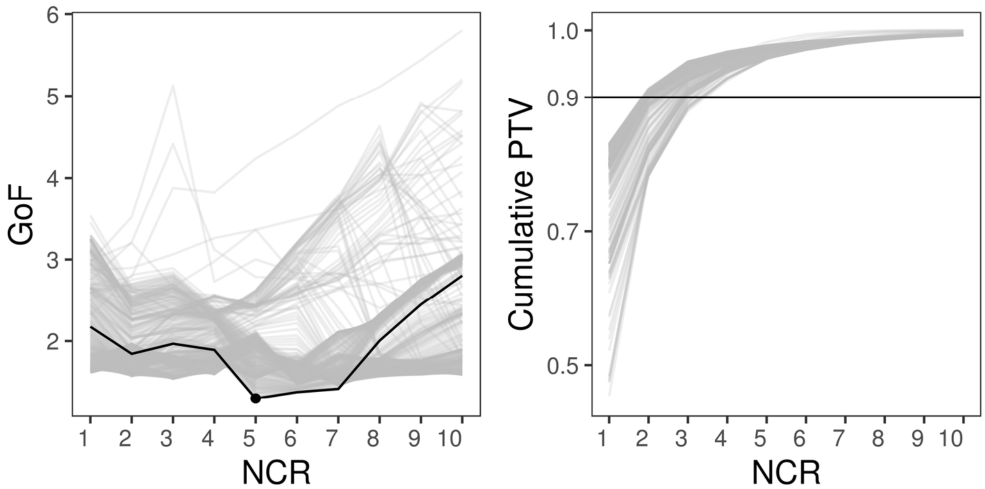
Correlation matrix of annual goat population data (1990-2012), with the plot size proportional to the correlation.

Figure 2

228x228mm (300 x 300 DPI)



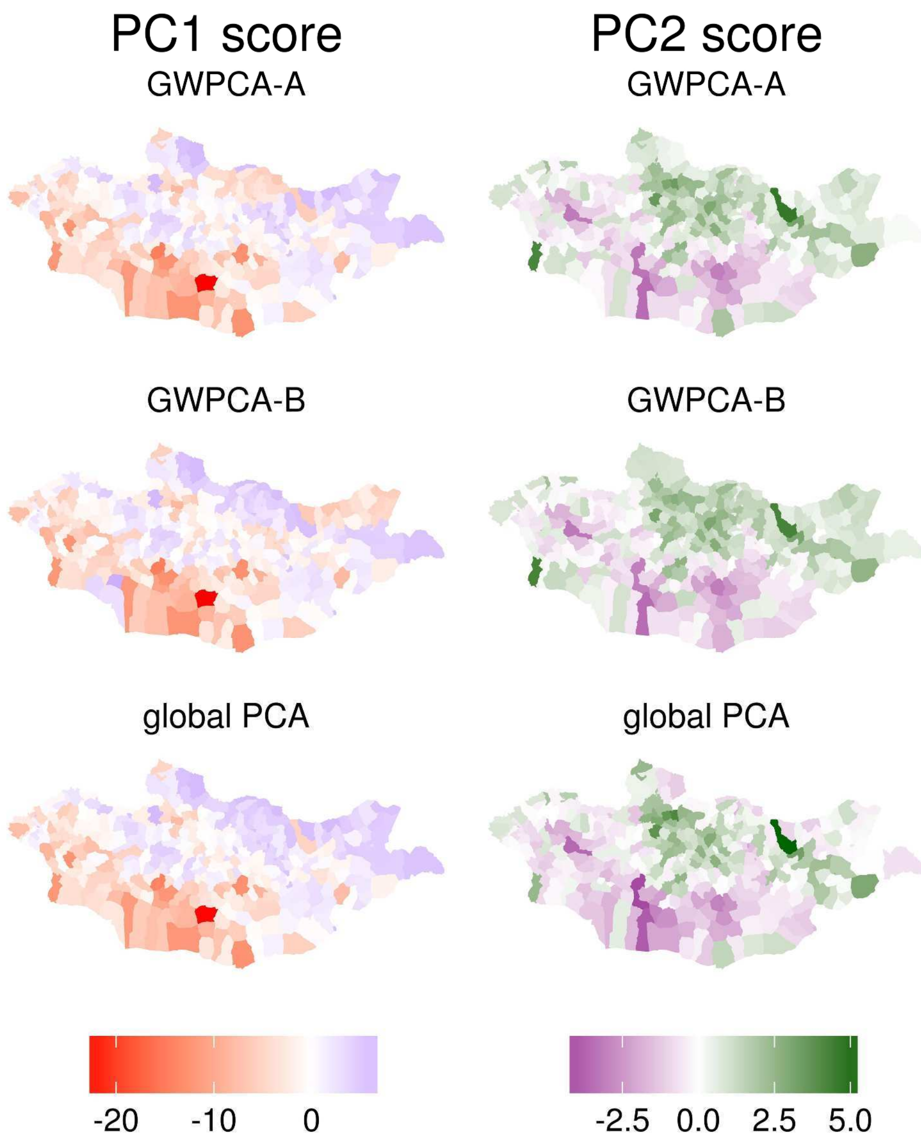
GW correlation maps between PC1-3 scores of the global PCA and the raw data of the corresponding most influential years (see also Table 2).
Figure 3
127x105mm (600 x 600 DPI)



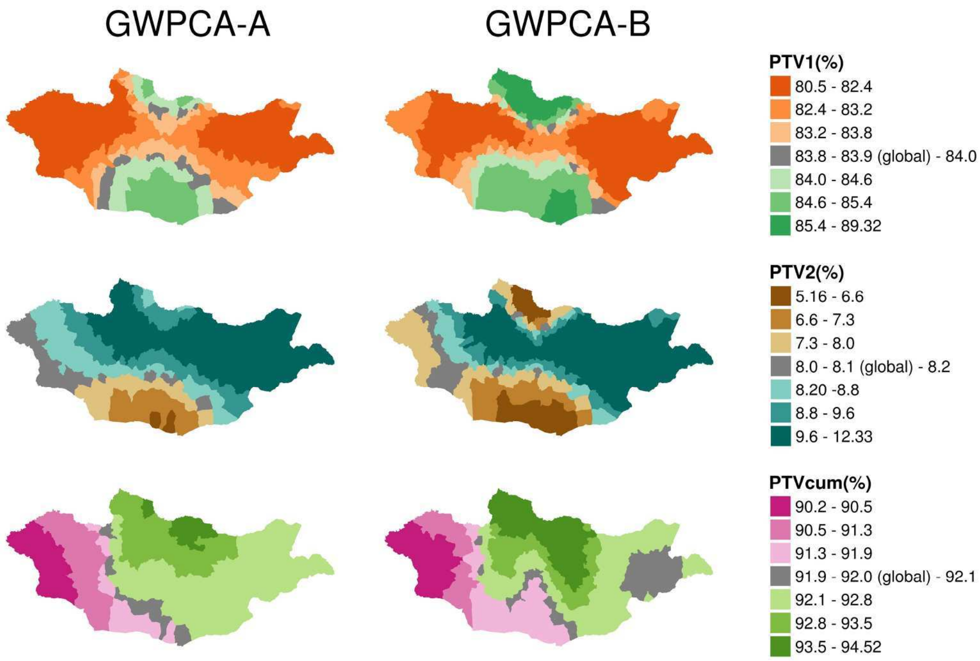
GWPCA calibration: (i) GoF (via LOOR CoV) versus NCR values; and (ii) scree plots for local cumulative PTVs versus NCR values. The grey lines have a transparency term added to them. In (i) they represent bandwidths in a range of 5 to 341 and (ii) in a range of 10 to 341. The black line in (i) represents the optimal bandwidth of 247 with NCR = 5, at the minimum GoF. Black line in (ii) represents the 90% threshold of the cumulative PTV.

Figure 4

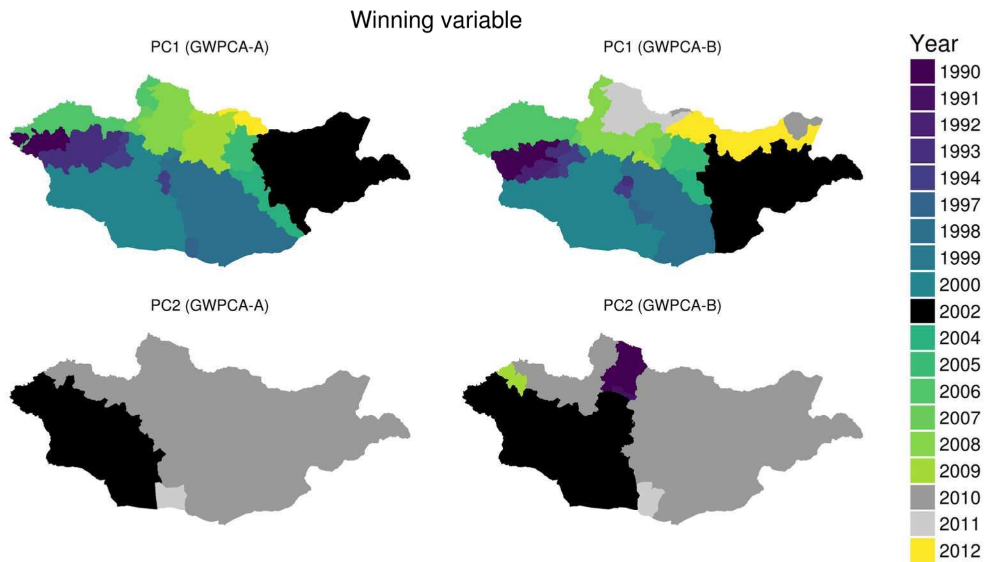
76x38mm (600 x 600 DPI)



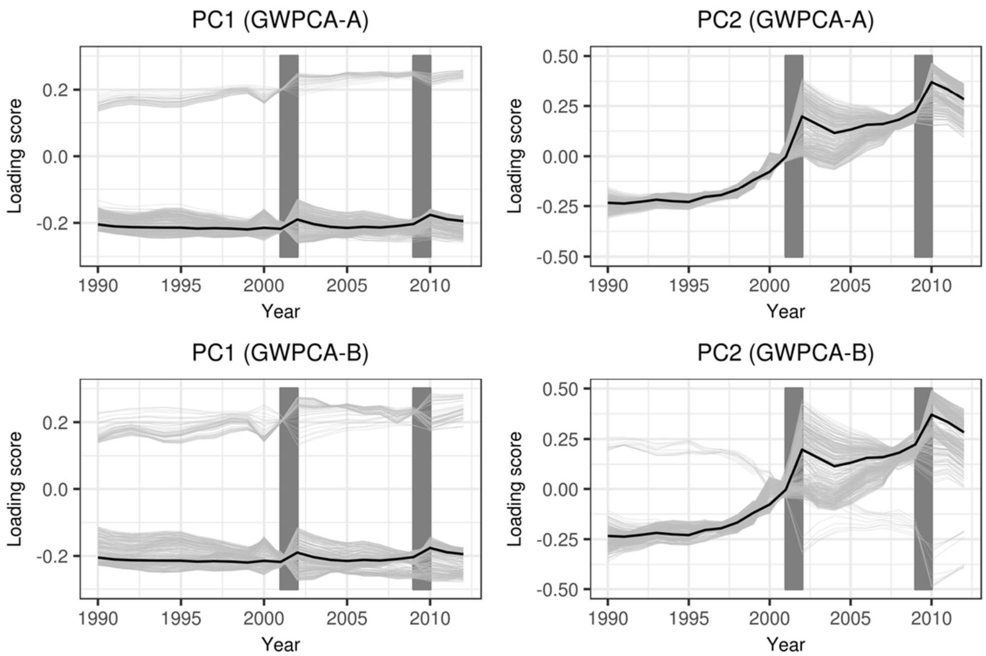
PC1 and PC2 scores maps for GWPCA-A (top row), GWPCA-B (middle row), and the global PCA (bottom row).
Figure 5
127x158mm (600 x 600 DPI)



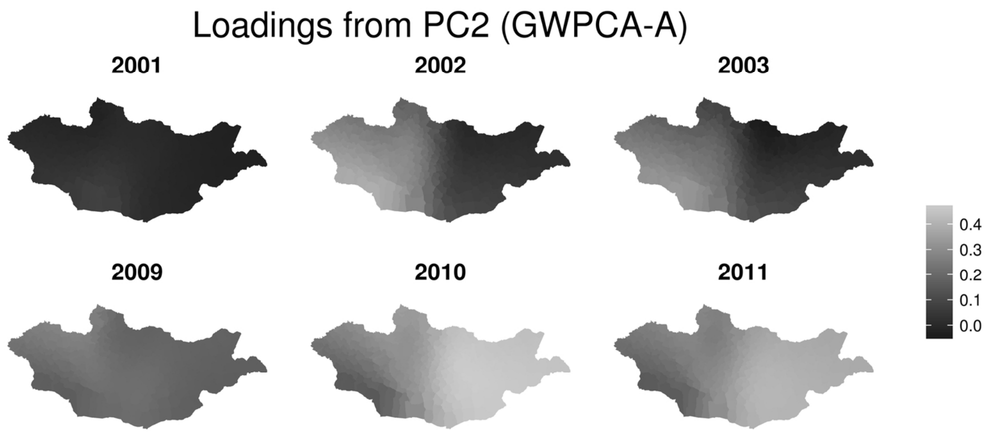
GWPCA-A and GWPCA-B PTV maps for PC1 (top row), PC2 (middle row) and PC1/PC2 combined (bottom row).
Figure 6
101x67mm (600 x 600 DPI)



GWPCA-A and GWPCA-B 'winning year' maps (by highest loadings) for PC1 and PC2. Years when dzud occurred are highlighted in grey and black.
Figure 7
152x101mm (300 x 300 DPI)



GWPCA-A and GWPCA-B loadings for PC1 and PC2, displayed over the 23 study years. The grey lines have a transparency term and represent the loading score at every soum. The black lines represent the loadings from the global PCA. Dark grey rectangles represent dzud periods 2001-2 and 2009-10.
Figure 8
101x67mm (300 x 300 DPI)



Maps for PC2 loadings from GWPCA-A over dzud periods of 2001-3 (top row) and 2009-11 (bottom row).
Figure 9
76x38mm (600 x 600 DPI)

Tables

Table 1. Eigenvalues, PTV, and cumulative PTV for the global PCA. Only the first 5 PCs are shown.

	PC1	PC2	PC3	PC4	PC5
Eigenvalues	4.39	1.36	0.90	0.56	0.46
PTV (%)	83.86	8.09	3.56	1.37	0.91
Cumulative PTV (%)	83.86	91.95	95.51	96.88	97.78

Table 2. Loadings for the first five PCs of the global PCA ordered by contribution.

	PC1	PC2	PC3	PC4	PC5
1st	1999 (-0.220)	2010 (0.371)	2002 (-0.446)	2008 (0.446)	1990 (0.495)
2nd	2001 (-0.218)	2011 (0.334)	2003 (-0.388)	2007 (0.396)	2000 (-0.383)
3rd	1996 (-0.217)	2012 (0.283)	2010 (0.352)	2002 (-0.354)	1999 (-0.346)
4th	1998 (-0.217)	1991 (-0.237)	2004 (-0.341)	2006 (0.351)	2001 (-0.337)
5th	1997 (-0.216)	1990 (-0.234)	2012 (0.371)	2009 (0.269)	1991 (0.327)

Table 3. GWPCA calibration: optimum adaptive bandwidths for different values of NCR.

NCR	1	2	3	4	5	6	7	8	9	10
Bandwidth	341	341	284	277	341	341	341	341	341	341