

Rothamsted Repository Download

A - Papers appearing in refereed journals

Tsutsumida, N, Harris, P. and Comber, A. 2017. The application of a geographically weighted principal components analysis for exploring twenty-three years of goat population change across Mongolia. *Annals of the American Association of Geographers*. 107 (5), pp. 1060-1074.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.1080/24694452.2017.1309968>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/84710>.

© 31 May 2017, Rothamsted Research. Licensed under the Creative Commons CC BY.

1

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

4

5 **Author names/Affiliations:**

6 Narumasa Tsutsumida¹, Paul Harris², Alexis Comber³

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

9 ² Rothamsted Research, North Wyke, Okehampton, Devon, UK EX20 2SB

10 ³ School of Geography, University of Leeds, Leeds, UK LS2 9JT

11 **Abstract**

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

26

27

28 **Keywords**

29 Spatio-temporal; GWmodel; Livestock; Grasslands; Sustainability

30

31 **Introduction**

32 It is important to evaluate the impacts of disasters to improve and support
33 agricultural planning. In Mongolia, deep snow, severe cold and associated
34 conditions, called *dzud*, occur repeatedly and make forage unavailable or
35 inaccessible to livestock. This results in high livestock mortality (Fernandez-
36 Gimenez, Batbuyan, and Baival 2012; Fernández-Giménez et al. 2015) and huge
37 economic losses, as livestock in Mongolia represents 16% of national GDP (UNDP
38 and NEMA 2010). Traditional nomadic pastoralism is one of the most sustainable
39 ways of life on grasslands and sparsely vegetated lands, as are commonly found in
40 Mongolia (Millennium Ecosystem Assessment 2005; Research Institute for Humanity
41 and Nature 2012). Vegetation availability depends on the impacts of livestock
42 grazing which has been well managed by nomadic herders over thousands of years
43 (Research Institute for Humanity and Nature 2012), and is not suited to intensive
44 livestock and crop production. In particular, excessive livestock populations, whether
45 managed commercially or traditionally, endangers sustainability (Geist and Lambin
46 2004; Suttie, Reynolds, and Batello 2005). Recent changes to the Mongolian
47 livestock industry, which has become swamped with non-professional herders due to

48 de-regularisation, has made the grasslands vulnerable to environmental change and
49 to extreme weather events. Thus there is a clear need to understand the spatio-
50 temporal trends in Mongolia's livestock populations, accounting for the impacts of the
51 dzuds.

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

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

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

81

82 **Background**

83 Livestock populations in Mongolia

84 Nomadic pastoralism has provided a sustainable way of life for thousands of
85 years in Mongolia (Research Institute for Humanity and Nature 2012). Although
86 Mongolian grasslands have been well-managed, there are concerns about the
87 impacts of increases in livestock populations. The lives of nomadic pastoralists have
88 been strongly influenced by political changes, especially the move from a planned
89 economy to a free-market economy in 1992 (Fernandez-Gimenez 2006). Prior to
90 this, livestock production was managed centrally and nomadic herders raised state-
91 owned livestock, restricting excessive livestock production. The government
92 encouraged herders to organize their collectives locally, and gave salaried
93 (professional) herders the responsibility of breeding livestock. Collectives were self-
94 regulated in their land use and their seasonal long-distance travel, resulting in good
95 pasture maintenance with advance preparedness for keeping livestock secure from
96 extreme events (Fernandez-Gimenez 2006). Since the transition to a free-market
97 economy, pastures have been managed by individual herders, leading to serious

98 sustainability and land management issues, as herders are now focussed on profit
99 and their number has more than doubled (Togtokh 2008) – all of which makes the
100 livestock industry more vulnerable. Five main livestock types are found in Mongolia
101 (sheep, goat, horse, cattle and camel), and the country-wide goat population has
102 rapidly increased since the government policy change in 1992 (Figure 1). This
103 increase is primarily due to the strong demand for goat cashmere (Saizen,
104 Maekawa, and Yamamura 2010), but unfortunately, the rate of increase threatens
105 livestock sustainability and the nomadic lives of herders.

106 Livestock losses occur during periods of the dzud as a result of deep snow and
107 severe cold (Fernandez-Gimenez, Batbuyan, and Baival 2012; Tsutsumida and
108 Saizen 2014). Additional pressure is also placed on herders as the dzud directly
109 results in reduced opportunities for grazing in the summer that follows, as a result of
110 droughts. Effects of this combination of winter dzuds and summer droughts can be
111 seen in Figure 1 for the years 2001-2 and 2009-10, where declines in the sheep and
112 goat populations are clearly evident. As a result of the 2009-10 dzud, approximately
113 20% of the country's livestock population were killed, affecting 28% of Mongolia's
114 human population (Fernandez-Gimenez, Batbuyan, and Baival 2012; Fernández-

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

118 Little attention has been paid into the geographical dynamics of the Mongolian
119 livestock population, over this 23-year period of change. Although some research
120 has been conducted, notably by Saizen, Maekawa, and Yamamura (2010) who
121 found areas of goat population increase to be independent of land cover. Saizen,
122 Maekawa, and Yamamura (2010) also noted that in more severe conditions, goat
123 herders were not restricted to the grazing pastures close to Ulaanbaatar, as goats
124 are more resilient to severe conditions, and the fact that a key goat product,
125 cashmere, is relatively portable. Liu et al. (2013) investigated the relationship
126 between goat population density and various climatic factors and suggested that the
127 marked increase in goat population density was a key non-climatic factor affecting
128 grassland degradation. Hilker et al. (2014) observed that livestock population
129 increases, associated with vegetation greenness, were primarily in the western part
130 of Mongolia. Thus previous research has tended to focus on environmental issues
131 and not the vulnerability of the livestock populations due to dzuds, even though they

132 are relatively common. This study seeks to address this oversight by investigating
133 the spatio-temporal pattern of goat population change in relation to the varying
134 impacts of dzuds, via a GWPCA approach.

135

136 PCA and geographically weighted PCA

137 PCA is standard information reduction technique, commonly employed in many
138 areas of data analysis. It transforms a set of m correlated variables into a new set of
139 m uncorrelated variables called components. The components are linear
140 combinations of the original variables and can allow for a better understanding of
141 differing sources of variation and key trends in data. Its use as a dimension reduction
142 technique is viable if the first few components account for most (say, 80 to 90%) of
143 the variation in the original data. Component scores and component loadings data
144 are produced, where the latter display how much each of the original variables
145 attribute to the dimensional variance of the overall data. For details, see Jolliffe
146 (2002).

147 There are a number of ways that a PCA can be usefully applied to multivariate
148 spatio-temporal data sets, such as the livestock data sets for this study (when all five

149 livestock types are considered). Demšar et al. (2013) provide a review in this
150 respect, where the many dimensional groups can be treated in a variety of ways.
151 This study applies a PCA to the goat population data, collected over a 23-year time
152 period. Thus the application of PCA is to a set of 23 time-stamped geographic
153 variables, where each variable measures goat population for a different year. This
154 means that the PCA only accounts for the temporal correlations in the data.

155 PCAs have been used to identify spatio-temporal data characteristics in many
156 scientific fields (e.g. Felipe-Sotelo et al. 2006; Lasaponara 2006; and see Demsar et
157 al. 2013 therein). For example, Lasaponara (2006) applied PCA for the evaluation of
158 vegetation anomalies from multi-temporal remote sensing data; and found that the
159 first principal component (PC1) related to a general vegetation distribution pattern,
160 while the second (PC2) indicated a decreasing trend of vegetation amount. In the
161 atmospheric sciences, PCAs are commonly applied to spatio-temporal (univariate)
162 data, and is referred to as an empirical orthogonal function (EOF) analysis (e.g.
163 Obled and Creutin (1986)). However for EOFs, the time series data is sufficiently
164 long enough to consider PCA in Q-mode (rather than the usual R-mode), thus spatial
165 correlations are captured as the data matrix is transposed. If the livestock population

166 data of this study was considered temporally long enough (i.e. collected over 100
167 years, say), then such an application of Q-mode PCA could also have been
168 considered. Instead, an R-mode PCA is applied and thus only temporal correlations
169 in the goat data are captured. Note that applications of PCA to spatio-temporal data
170 entails that Q-mode PCA is often referred to as S-mode PCA, where "S" denotes
171 spatial, and R-mode PCA is often referred to as T-mode PCA, where "T" denotes
172 temporal. The idea being that Q-mode and R-mode PCAs are reserved for attribute
173 sub-space applications with no spatio-temporal context.

174 However, a standard (R-mode) PCA application to this study's goat data does
175 not account for any spatial effects, because it only ensures a non-spatial linear
176 transform (Demšar et al. 2013). In order to deal with such a naïve application, but
177 from a spatial effects point of view only, GWPCA can be used. This adaptation of
178 PCA provides a better description of any spatial phenomenon in the structure of the
179 data. It uses a moving window weighting technique and constructs a localized PCA
180 at all target locations (e.g. a grid, such as the application by Comber, Harris, and
181 Tsutsumida (2016)). It is important to note, that although spatio-temporal correlations
182 in the goat population data are captured via GWPCA, only spatial dependencies in

183 the data are fully captured. Temporal dependencies such as those between
184 neighbouring years, are not fully captured nor are true spatio-temporal
185 dependencies. That requires a further extension to GWPCA to a full spatio-temporal
186 approach, similar that proposed for GW regression by Huang, Wu, and Barry (2010).
187 Thus in this study, both PCA and GWPCA are applied in order to provide a better
188 understanding of the dynamics of the Mongolian goat population data, at a soum-
189 level scale, across the period 1990–2012.

190

191 **Study data**

192 Annual livestock population data were obtained from the National Statistical
193 Office (NSO) of Mongolia for the period 1990–2012. Populations were summarized
194 per soum, an administrative sub-division area. Since local governments collect taxes
195 from herders according to herd size, the data are assumed to reflect livestock
196 numbers reasonably well (Saizen, Maekawa, and Yamamura 2010). Administrative
197 boundaries slightly changed during the 23-year study period. To cater for this, the
198 data were merged accounting for all 341 soums, using the most recent boundaries.
199 Thus all data are taken into account when a soum changed or was incorporated into

200 a neighbour. Missing data that arose because of these changes, were infilled using a
201 probabilistic PCA method provided in the pcaMethod R package (Stacklies et al.
202 2007). This infilling was fairly minor and was not considered an issue for subsequent
203 analyses.

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

211

212 **Methods**

213 Principal components analysis (PCA)

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

$$216 \quad L V L^T = S \quad (1)$$

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

$$224 \quad T = XL \quad (2)$$

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

227

228 Geographically weighted principal components analysis (GWPCA)

229 A GWPCA utilises a kernel weighting approach where localised PCs are found
230 at target locations. At a target location, neighbouring observations are weighted by a
231 distance-decay weighting function, and then a standard PCA is locally applied to its
232 own specific weighted data subset. The size of the window over which this localised
233 PCA might apply is controlled by the kernel's bandwidth. Small bandwidths lead to

234 more rapid spatial variation in the results whereas large bandwidths yield results
 235 increasingly close to the global PCA solution. This study identifies an adaptive
 236 bandwidth corresponding to a bi-square kernel, a discontinuous function that
 237 generates distance-decaying weights data points within the set bandwidth.
 238 Observations outside of the bandwidth's range receive weights of zero, and hence
 239 the discontinuity. For details, see Gollini et al. (2015).

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

$$245 \quad \Sigma_{(u,v)} = X^T W_{(u,v)} X \quad (3)$$

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

$$249 \quad L V L^T |(u_i, v_i) = \Sigma_{(u_i, v_i)} \quad (4)$$

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

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

267 to determine the bandwidth and the NCR value, concurrently. These methodological
268 advances are described and implemented below.

269

270 Geographically weighted correlation analysis

271 A GW correlation analysis (Harris and Brunson 2010) on the outputs from the

272 PCA with the raw data is also conducted. Here for variables x and y at spatial

273 location i where the geographical weights w_{ij} again accord to a bi-square function,

274 definitions for a GW standard deviation and a GW correlation coefficient, are

275 respectively

$$276 \quad s(x_i) = \sqrt{\frac{\sum_{j=1}^n w_{ij} (x_j - m(x_i))^2}{\sum_{j=1}^n w_{ij}}} \quad (5)$$

277 and

$$278 \quad \rho(x_i, y_i) = \frac{c(x_i, y_i)}{(s(x_i)s(y_i))} \quad (6)$$

279 , where a GW mean is

$$280 \quad m(x_i) = \frac{\sum_{j=1}^n w_{ij} x_j}{\sum_{j=1}^n w_{ij}} \quad (7)$$

281 and a GW covariance is

282
$$c(x_i, y_i) = \frac{\sum_{j=1}^n w_{ij} \{(x_j - m(x_i))(y_j - m(y_i))\}}{\sum_{j=1}^n w_{ij}} \quad (8)$$

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

285

286 **Results**

287 The global PCA

288 In order to understand any GW model output, it is always important to fit the
289 usual global model for context. In this respect, a PCA was conducted on the 23
290 temporal variables describing goat populations. Table 1 shows that the first two PCs
291 have eigenvalues greater than unity, and for these two PCs, the cumulative PTV
292 exceeds 90%. This implicitly assumes a uniform temporal trend in goat population
293 across all 341 sums over the 23-year period. The PCA loadings given in Table 2
294 indicate that the five of the most influential years are 1996-1999 and 2001 for PC1;
295 1990-1991 and 2010-2012 for PC2.

296

297 A GW correlation analysis on the PCA scores and raw data

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

311

312 GWPCA calibration with dual bandwidth and NCR optimization

313 As outlined above, in order to calibrate a GWPCA, first the NCR value needs to
314 be user-specified and only then, can an optimal GWPCA bandwidth be found via

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

325 However, the choice of bandwidth can be investigated more deeply. This is
326 because the results presented in Table 3 are not directly comparable, as given
327 ‘optimal’ bandwidths correspond to minimized GoF statistics (not shown) where the
328 NCR-specific LOOR data sets have been summarized by their mean. To ensure that
329 the minimized GoF statistics are comparable across different values of NCR, the
330 LOOR data can be summarized instead by their coefficient of variation (CoV) to
331 provide relative (and thus comparable) GoF statistics for each bandwidth and for

332 each NCR value. This leads to a dual optimization approach as shown in Figure 4(i),
333 where the aim to concurrently find the bandwidth and the NCR value that
334 corresponds to minimum GoF (LOOR CoV) value. Again considering only NCR
335 values from 1 to 10, and a clear minimum GoF is reached at 1.296 corresponding to
336 a bandwidth of 247 nearest neighbours and an NCR value of 5. Each individual line
337 in the plot of Fig 4(i) corresponds to a different bandwidth choice, from 5 to 341. This
338 constitutes the first extension to the existing bandwidth selection procedure.

339 A second alternative is to transfer the usual cumulative PTV approach for NCR
340 selection to a local setting. Globally, a user-specified choice of NCR = 1 or 2 is
341 based on the global cumulative PTV scree plot (e.g. Varmuza and Filzmoser (2009)).
342 This approach can be transferred locally using the local cumulative PTV data from
343 each local PCA from a series of GWPCAs. Local cumulative PTV data were
344 calculated from GWPCAs calibrated with bandwidths ranging from 10 to a maximum
345 of 341 and the resultant local scree plots are depicted in Figure 4(ii). Clearly, the
346 local scree plots suggest that NCR = 2 is the point when some of the local
347 cumulative PTV data exceeds a 90% threshold. Given this, NCR = 2 again appears
348 appropriate for a GWPCA calibration. However, the bandwidth is still required, and

349 unlike the existing approach a bandwidth is identified that has the smallest GoF
350 (LOOR mean) value, but crucially also corresponds to a localized cumulative PTV
351 value exceeding 90% (for all NCR = 2). This indicates a relative tight bandwidth of
352 198 nearest neighbours.

353 Thus in summary, there are three possible bandwidths for GWPCA calibration:

354 (a) 341 (via NCR = 1 or 2); (b) 247 (via NCR = 5); and (c) 198 (via NCR = 2). All
355 three should be considered as entirely valid, but where approach (a), the existing
356 approach, strongly suggests a stationary process with respect to a PCA. Given that
357 approach (a) has drawbacks, not only with respect to NCR/bandwidth specification,
358 but also (indirectly) due to current limitations in the GWPCA code (see note 1), it is
359 dropped in favour of the two newly proposed approaches (b and c) which are both
360 viewed as a methodological advance. In the spirit of spatial exploration, which all
361 GW models are eminently designed for, both approaches were investigated further
362 all of the subsequent GWPCA outputs described below are specified with either: (i) a
363 bandwidth of 247 via a NCR value of 5; or (ii) a bandwidth of 198 via a NCR value of
364 2.

365

366 PCA versus GWPCA results

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

377

378 Scores data

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

383 GWPCA-A and GWPCA-B correlate with those from the global PCA, with $r = 0.846$
384 and $r = 0.742$, respectively. PC2 scores of GWPCA-A and GWPCA-B correlate with
385 those from the global PCA, with correlations of $r = 0.943$ and $r = 0.872$,
386 respectively. These moderate to strong correlations simply reflect the relatively large
387 bandwidth sizes used, and such correlations would tend to unity as the bandwidth
388 increases. However these global correlations hide spatial detail, where the study's
389 aim is to see where the local spatial structure in the temporally-changing goat
390 population (via the GWPCA outputs) differs to that found globally (via the PCA
391 outputs). In this respect, the clearest regional differences in both the PC1 and PC2
392 scores data appear in the north-eastern regions of Mongolia, bordering Russia and
393 also the south-western regions bordering China. Thus the temporal dynamics of goat
394 population change is likely to be clearly different in these regions to that expected
395 nationally.

396

397 Percentage PTV data

398 Globally, the PTV for PC1, and the cumulative PTV for PC1 and PC2
399 combined, are 84% and 92%, respectively. This suggests a high correlation amongst

400 the goat population data, year on year, throughout the 23-year period. However, the
401 global PTV values (from PCA) implicitly assume that such relationships are constant
402 across Mongolia - with relatively uniform changes in goat populations everywhere.
403 Mapping the corresponding localized PTV outputs from GWPCA shows where this is
404 the case, and the degree to which it is not (Figure 6).

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

416

417 Loadings data

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

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

434 In this instance, greater insight stems from considering the ‘winning year’ maps
435 for the next component (PC2). Now far fewer years are represented (3 to 6 of 23)
436 and the dzud years of 2002 and 2010, strongly dominate in two clear regions; the
437 west and south-west, and the east and north, respectively. This suggests that: (i) the
438 dzud of 2002 and the associated goat population decline was more or less
439 pronounced in the west and south-west than elsewhere; and (ii) the dzud of 2010
440 and the associated goat population decline was more or less pronounced in the east
441 and north than elsewhere. This strongly indicates that the severity of the dzuds in
442 2002 and 2010 varied geographically. Visualizing the annual changes in the PCA
443 and GWPCA loadings from PC1 and PC2 for GWPCA-A and GWPCA-B (Figure 8)
444 shows the effects of the 2002 and 2010 dzud years on the loadings, with clear
445 inflection points for both GWPCA fits.

446 Figure 9 displays the loadings maps for PC2 of GWPCA-A only, for 2001-3,
447 and 2009-11, covering the two dzuds periods. These maps suggest that the 2001-2
448 dzud and the 2009-10 dzud have different regional and temporal characteristics. The
449 impact of the 2001-2 dzud starts from central/western regions in 2001 and increases
450 in western regions in 2002. The impact of the 2009-10 dzud appears first in western

451 regions in 2009 and then in eastern regions in 2010. This is in contrast to the
452 reporting of dzuds and the devastating damage to livestock populations, which is
453 typically referred to as impacting Mongolia as a whole, and uniformly.

454

455 **Discussion and conclusions**

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

466 Mapping GWPCA scores data allowed regional differences to be observed,
467 particularly in the north-eastern regions of Mongolia, bordering Russia and also

468 south-western regions bordering China. Thus the temporal dynamics of goat
469 population change is likely to be different in these regions to that expected nationally.
470 By mapping GWPCA variance proportion data, the temporal trend in goat population
471 change was found to be more uniform, to that found globally, in central northern
472 regions, whilst less uniform (to that found globally) in western regions. Visualizing the
473 'winning year' maps for the GWPCA loadings, suggests that the dzud of 2002 and
474 the associated goat population decline was more or less pronounced in the west and
475 south-west regions and that the dzud of 2010 and the associated goat population
476 decline was more or less pronounced in the east and north regions. This, in turn,
477 suggests that the dzuds of 2002 and 2010 varied geographically in their severity.

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

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

487 This study's application of GWPCA to temporally indexed spatial data is novel
488 and adds to a growing portfolio of GWPCA uses, not only for spatial exploration
489 (Lloyd 2010; Harris, Brunsdon, and Charlton 2011; Harris et al. 2015), but also for
490 spatial anomaly detection (Harris, Brunsdon, et al. 2014; Harris et al. 2015), spatial
491 network re-design (Harris, Clarke, et al. 2014), and spatial classification (Harris et al.
492 2015; Comber, Harris, and Tsutsumida 2016). Furthermore, this study usefully
493 extended the GWPCA methodology itself to simultaneously optimise the number of
494 components to retain and the kernel weighting bandwidth. This is considered an
495 important advance, and should be adopted in all subsequent GWPCA studies.

496

497 **Notes**

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

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

504

505 **Acknowledgements**

506 The authors would like to thank the National Statistics Office of Mongolia for
507 providing livestock population data. We appreciate Dr. Izuru Saizen and Dr. Atsushi
508 Otomo to support this work. This work was supported by Sinfonica Statistical GIS
509 Research Grants, the RIHN (project number D-04), JSPS KAKENHI Grant Number
510 15K21086, grants for young researchers in GSGES KU, and KU SPIRITS project.
511 For Paul Harris, a UK Biotechnology and Biological Sciences Research Council
512 grant (BBSRC BB/J004308/1).

513

514 **References**

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

519 Comber, A. J., P. Harris, and N. Tsutsumida. 2016. Improving land cover
520 classification using input variables derived from a geographically weighted principal
521 components analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*
522 119:347–360. <http://linkinghub.elsevier.com/retrieve/pii/S0924271616301290>.

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

527 Felipe-Sotelo, M., L. Gustems, I. Hernández, M. Terrado, and R. Tauler. 2006.
528 Investigation of geographical and temporal distribution of tropospheric ozone in
529 Catalonia (North-East Spain) during the period 2000-2004 using multivariate data
530 analysis methods. *Atmospheric Environment* 40 (38):7421–7436.

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

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

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

544 Fernández-Giménez, M. E., B. Batkhishig, B. Batbuyan, and T. Ulambayar. 2015.
545 Lessons from the Dzud: Community-Based Rangeland Management Increases the
546 Adaptive Capacity of Mongolian Herders to Winter Disasters. *World Development*
547 68:48–65.

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

550 Geist, H. J., and E. F. Lambin. 2004. Dynamic Causal Patterns of Desertification.
551 *BioScience* 54 (9):817. [https://academic.oup.com/bioscience/article-](https://academic.oup.com/bioscience/article-lookup/doi/10.1641/0006-3568(2004)054[0817:DCPOD]2.0.CO;2)
552 [lookup/doi/10.1641/0006-3568\(2004\)054\[0817:DCPOD\]2.0.CO;2.](https://academic.oup.com/bioscience/article-lookup/doi/10.1641/0006-3568(2004)054[0817:DCPOD]2.0.CO;2)

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

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

561 Harris, P., C. Brunsdon, and M. Charlton. 2011. Geographically weighted principal
562 components analysis. *International Journal of Geographical Information Science* 25
563 (10):1717–1736.
564 <http://www.tandfonline.com/doi/abs/10.1080/13658816.2011.554838>.

565 Harris, P., C. Brunsdon, M. Charlton, S. Juggins, and A. Clarke. 2014. Multivariate
566 Spatial Outlier Detection Using Robust Geographically Weighted Methods.
567 *Mathematical Geosciences* 46 (1):1–31. [http://link.springer.com/10.1007/s11004-](http://link.springer.com/10.1007/s11004-013-9491-0)
568 [013-9491-0](http://link.springer.com/10.1007/s11004-013-9491-0).

569 Harris, P., A. Clarke, S. Juggins, C. Brunsdon, and M. Charlton. 2014.

570 Geographically weighted methods and their use in network re-designs for
571 environmental monitoring. *Stochastic Environmental Research and Risk Assessment*
572 28 (7):1869–1887. <http://link.springer.com/10.1007/s00477-014-0851-1>.
573 — — —. 2015. Enhancements to a Geographically Weighted Principal Component
574 Analysis in the Context of an Application to an Environmental Data Set.
575 *Geographical Analysis* 47 (2):146–172. <http://doi.wiley.com/10.1111/gean.12048>.
576 Huang, B., B. Wu, and M. Barry. 2010. Geographically and temporally weighted
577 regression for modeling spatio-temporal variation in house prices eds. M. M. Fischer
578 and P. Nijkamp. *International Journal of Geographical Information Science* 24
579 (3):383–401. <http://link.springer.com/10.1007/978-3-642-23430-9> (last accessed 11
580 July 2014).
581 Lasaponara, R. 2006. On the use of principal component analysis (PCA) for
582 evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI
583 temporal series. *Ecological Modelling* 194 (4):429–434.
584 <http://linkinghub.elsevier.com/retrieve/pii/S0304380005005454> (last accessed 5 April
585 2012).
586 Lloyd, C. D. 2010. Analysing population characteristics using geographically

587 weighted principal components analysis: A case study of Northern Ireland in 2001.
588 *Computers, Environment and Urban Systems* 34 (5):389–399.
589 <http://dx.doi.org/10.1016/j.compenvurbsys.2010.02.005>.

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

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

597 Obled, C., and J. D. Creutin. 1986. Some Developments in the Use of Empirical
598 Orthogonal Functions for Mapping Meteorological Fields. *Journal of Climate and
599 Applied Meteorology* 25 (9):1189–1204.
600 [http://journals.ametsoc.org/doi/abs/10.1175/1520-
601 0450%281986%29025%3C1189%3ASDITUO%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0450%281986%29025%3C1189%3ASDITUO%3E2.0.CO%3B2).

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

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

609 Shekhar, S., Z. Jiang, R. Ali, E. Eftelioglu, X. Tang, V. Gunturi, and X. Zhou. 2015.
610 Spatiotemporal Data Mining: A Computational Perspective. *ISPRS International*
611 *Journal of Geo-Information* 4 (4):2306–2338. [http://www.mdpi.com/2220-
612 9964/4/4/2306/](http://www.mdpi.com/2220-9964/4/4/2306/).

613 Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig. 2007. pcaMethods
614 a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*
615 23 (9):1164–1167. [https://academic.oup.com/bioinformatics/article-
616 lookup/doi/10.1093/bioinformatics/btm069](https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btm069).

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

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

- 621 Tsutsumida, N., and I. Saizen. 2014. Internal Regional Migration Analysis and
622 Modeling of Population Concentration for Ulaanbaatar, Mongolia. *Papers on*
623 *Environmental Information Science* 28:25–30.
- 624 UNDP and NEMA. 2010. *Dzud national report 2009-2010*.
- 625 Varmuza, K., and P. Filzmoser. 2009. *Introduction to multivariate statistical analysis*
626 *in chemometrics*. Taylor & Francis.
- 627
- 628

629 **Author professional/contact information**

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

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

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

643

644

645

646

647 **Figure captions**

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

649

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

652

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

655

656 Figure 4. GWPCA calibration: (i) GoF (via LOOR CoV) versus NCR values; and (ii)

657 scree plots for local cumulative PTVs versus NCR values. The grey lines have a

658 transparency term added to them. In (i) they represent bandwidths in a range of 5 to

659 341 and (ii) in a range of 10 to 341. The black line in (i) represents the optimal

660 bandwidth of 247 with NCR = 5, at the minimum GoF. Black line in (ii) represents the

661 90% threshold of the cumulative PTV.

662

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

665

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

668

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

671

672 Figure 8. GWPCA-A and GWPCA-B loadings for PC1 and PC2, displayed over the
673 23 study years. The grey lines have a transparency term and represent the loading
674 score at every soum. The black lines represent the loadings from the global PCA.

675 Dark grey rectangles represent dzud periods 2001-2 and 2009-10.

676

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

679

