Article:
Comber, AJ, Harris, P and Tsutsumida, N (2016) Improving land cover classification using input variables derived from a geographically weighted principal components analysis. ISPRS Journal of Photogrammetry and Remote Sensing, 119. pp. 347-360. ISSN: 0924-2716

https://doi.org/10.1016/j.isprsjprs.2016.06.014

Improving land cover classification using input variables derived from a geographically weighted principal components analysis

Alexis J Comber[a*‡], Paul Harris[b] and Narumasa Tsutsumida[c]

[a] Leeds Institute for Data Analytics (LIDA) and School of Geography, University of Leeds, Leeds, LS2 9JT, UK

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

[c] Graduate School of Global Environmental Studies, Kyoto University, Kyoto, 606-8501, Japan

* contact author: ajc36@le.ac.uk

‡ work undertaken at the Centre for Climate and Landscape Research, Department of Geography, University of Leicester, Leicester, LE1 7RH, UK

Abstract

This study demonstrates the use of a geographically weighted principal components analysis (GWPCA) of remote sensing imagery to improve land cover classification accuracy. A principal components analysis (PCA) is commonly applied in remote sensing but generates global, spatially-invariant results. GWPCA is a local adaptation of PCA that locally transforms the image data, and in doing so, can describe spatial change in the structure of the multi-band imagery, thus directly reflecting that many landscape processes are spatially heterogenic. In this research the GWPCA localised loadings of MODIS data are used as textural inputs, along with GWPCA localised ranked scores and the image bands themselves to three supervised classification algorithms. Using a reference data set for land cover to the west of Jakarta, Indonesia the classification procedure was assessed via training and validation data splits of 80/20, repeated 100 times. For each classification algorithm, the inclusion of the GWPCA loadings data was found to significantly improve classification accuracy. Further, but more moderate improvements in accuracy were found by additionally including GWPCA ranked scores as textural inputs, data that provide information on spatial anomalies in the imagery. The critical importance of considering both spatial structure and spatial anomalies of the imagery in the classification is discussed, together with the transferability of the new method to other studies. Research topics for method refinement are also suggested.

Key words: GWmodel, GWPCA, spatial heterogeneity, accuracy

1. Introduction

This paper describes the application of a Geographically Weighted Principal Components Analysis (GWPCA) as a method to improve the reliability of land cover classification from remotely sensed data.

Supervised classification of remote sensing imagery to identify land cover is a clustering process. Training data are collected, typically through field surveys or from higher resolution imagery, and the multivariate image properties of the training data are used to *train* a clustering algorithm. Commonly, this identifies cluster centres for each class, based on the multivariate properties of the training data and the classification proceeds by allocating each image object, typically a pixel, to the cluster to which it is closest in the multivariate image space. Different classification algorithms can vary in the way that they define cluster centres, multivariate distance and in their iteration. Classification algorithms can also differ to whether or not class statistics are calculated (for example, choosing between a logistic regression or support vector machines).

Collinearity occurs when variables exhibit linear relationships and this has been found to affect the reliability of the classification algorithm (Congalton, 1991). PCAs have been used to handle collinearity in remote sensing. The first few components of a PCA frequently capture most of the image data variation and structure by transforming the data into an ordered set of orthogonal components. In remote sensing, PCA approaches have been used to improve classification (e.g. Collins and Woodcock; 1996; Xu et al., 2003; Toutin, 2004; Koutsias et al., 2009), to explore

structure or trends in image data (e.g. Legendre and Legendre, 1998) and to detect anomalies in the outputs (Lasaponara, 2006).

Spatial effects can be important in land cover classifications. They may result in spatial heterogeneity in the relationship between the land cover classes and the imagery (Wang et al., 2005; Propastin, 2012) and the spatial autocorrelation of errors (Congalton, 1988). These arise when the classification algorithm fails to incorporate any spatial effects. To handle such spatial effects some authors have used texture measures constructed from image data as inputs into classifications (Car and Miranda, 1998; Chica-Olmo and Abarca-Hernandez, 2000; Atkinson and Lewis, 2000; Myint, 2003). In these, localised statistics are calculated for one image band at a time (e.g. local variance) using a simple moving window (e.g. a square) of a subjectively specified size.

This study demonstrates the application of a local version of PCA, termed geographically weighted principal components analysis (Fotheringham et al., 2002; Harris et al., 2011). A GWPCA investigates how outputs from a PCA vary spatially. It provides a significant methodological advance on previous approaches. First, all image bands are considered together to provide multivariate localised statistics. Second, a sophisticated distance-decay weighting scheme replaces the moving window approach. This is specified such that it provides a degree of objectivity on the spatial scale at which the local statistic is calculated. In this way, GWPCA is used to create texture variables that account for the spatial heterogeneity in the multi-band image structure. Spatial changes in data dimensionality and multivariate structure can be explored via maps of the GWPCA outputs (Fotheringham et al.,

100 2002; Harris et al., 2011; 2015). GWPCA can also be used to detect multivariate
101 spatial anomalies (Harris et al., 2014b; 2015). This study uses the outputs of a
102 GWPCA applied to 7-band MODIS imagery to classify land cover. In particular,
103 GWPCA loadings for structure and GWPCA scores for anomalies are included as
104 textural inputs, together with the raw image bands themselves as inputs to three
105 standard classification algorithms: latent discriminant analysis, logistic regression
106 and support vector machines. Thus GWPCA outputs provide informative multivariate
107 spatial inputs into the classification process. The study does not seek to directly
108 account for any local dimensionality issues or local collinearity effects in image data,
109 although GWPCA could be used to do so. Rather it aims to capture the local structure
110 in the multi-band image data to improve classification accuracy.
111
112 **2. Background**
113
114 In remote sensing, PCAs are used to transform image data into a new orthogonal set,
115 principal components (PCs), whose observations are called PC scores. Components
116 are ordered by the amount of variance in the original image data they explain and
117 there is always the same number of components as there were image bands. For a
118 PCA to be used for data reduction, it is typically hoped that the first two or first three
119 PCs explain around 80-90% of the original data's variance. Data reduction can then
120 proceed without an undue loss of information, which in turn reduces computational
121 burden of any subsequent analysis. PC loadings are the linear correlation coefficients
122 between the PC scores data and the original data. Thus by investigating the loadings,
123 it is possible to determine which of the original image bands contribute the most to

each PC. As the PCs are uncorrelated they provide a direct way of addressing image band collinearity, commonly found in the visible wavelengths.

PCA has also been used for data reduction to fuse data from multiple sources and platforms (Pohl and Van Genderen, 1998) and to provide greater insight into classification results. For example, Richards (1984) used PCA to monitor brushfire damage and vegetation re-growth in Australia and found that local areas of change were enhanced in some of the lower PCs and Ingebritsen and Lyon (1985) found the first two PCs to be strongly related to soil brightness and vegetation greenness. They have been used in change detection and error analysis and Tewkesbury et al. (2015) note that transformations of multiple image layers provides a convenient method for assessing change within a complex set of time series imagery. Doxani et al. (2011) applied a multivariate alteration detection transformation to identify change objects in VHR imagery. But some research has found that transformed time series data results in the loss of temporal change information (Deng et al. 2008; Tsutsumida et al. 2013).

PCA in remote sensing has been found to be sensitive to the study area being considered through the training or validation samples and the variation in the land cover types that are present (Pohl and Van Genderen, 1998). The implication is that spatial factors affect the relevance and usefulness of the PCA outputs, which only ever reflect the non-spatial properties of the inherently spatial, image data. Some research has sought to address this. Pesaresi and Benediktsson (2001) explored methods for analysing the morphology of panchromatic image data but their approach was not scalable to multivariate data (Soille, 2003).

149

150 A critical issue in remote sensing is the presence and impact of commonly observed
151 *spatial autocorrelation* effects in image data (Spiker and Warner, 2007), which for
152 example results in adjacent pixels being more likely to have similar values
153 (Woodcock et al., 1988) and *spatial heterogeneity* in outputs, such as classification
154 errors (Campbell, 1981). For these reasons, spatially explicit methods have been
155 applied to improve classification accuracy (Congalton, 1988; Steele et al, 1998). The
156 geographically weighted (GW) modelling paradigm provides a suite of models,
157 specifically for spatial heterogeneity effects (Fotheringham et al., 2002; Lu et al.,
158 2014; Gollini et al., 2015), the most commonly used of which is GW regression
159 (Brunsdon et al., 1996). Examples of GW models in remote sensing studies can be
160 found in Atkinson (2004), Wang et al. (2005), Atkinson and Naser (2010), Comber et
161 al. (2012), Johnson et al. (2012) and Propastin (2012). Examples of studies that have
162 sought to account for both spatial autocorrelation and spatial heterogeneity in data
163 include the studies by Car and Miranda (1998) and Chica-Olmo and Abarca-
164 Hernandez (2000). Other work has considered classification accuracy given spatial
165 effects. Foody (2005) modelled local accuracy by interpolating accuracies calculated
166 at regular spaced locations. Riemann et al. (2010) suggested spatial indices to
167 describe classification accuracy and Comber (2013) developed GW approaches to
168 generate maps of user, producer and overall accuracies. Related GW-based
169 approaches are found in Comber et al. (2012) and in Tsutsumida and Comber (2015),
170 where the latter used a PCA to examine the temporal variations in spatial accuracy.
171

172 The PCA method can be adapted to incorporate spatial effects, such as that of
173 autocorrelation or heterogeneity. For the former, Jombart et al. (2008) adapted PCA

using Moran's *I*, whilst for the latter, GWPCA can be used. Spatially-adapted PCA methods have not as yet been applied in a remote sensing context, but these and other methods exist, as reviewed in Demšar et al. (2013). GWPCA is just one of many models based around the GW framework. In this framework, a kernel or moving window is identified and data under the kernel are weighted by their distance to the location being considered under the kernel (i.e. the kernel centre). The *geographically weighted* data are then passed to whatever analysis is being undertaken and the localised model's outputs are mapped to provide a useful investigative tool of spatial heterogeneity. A key challenge in GW modelling is finding the scale at which each localised model should operate, that is choosing the size of the kernel bandwidth. The bandwidth can be user-specified, but preferably guided by some automatic cross-validation routine based on model fit. Similarly, it is not recommended to treat bandwidth optimisation via cross-validation as a purely *black-box* approach (Harris et al. 2014a). A number of GW models have been proposed, including those for summary statistics (Brunsdon et al., 2002), discriminant analysis (Brunsdon et al., 2007), and variograms (Harris et al., 2010).

3. Methods

This section describes the methods used for the application of a GWPCA to the MODIS image data, with the aim of improving land cover classification accuracy. Here the case study data is described, the GWPCA technique is formally presented, the supervised classification algorithms are presented, and finally, the crucial step of GWPCA bandwidth selection is described that accords to the objectives of this study.

198 The Appendix describes the processing times for the computations to provide an
199 overview of the implementation costs: they are not high.
200
201 **3.1. Case Study Area and Data Sets**
202
203 The study area was the Tengarang region to the west of Jakarta in Indonesia (Fig. 1).
204 MODIS surface reflectance from the MOD09A1 product was selected for analysis,
205 dated from 16th March 2012. This product provides a modified version of the
206 ground-level atmospheric scattering or absorption computed from MODIS level 1B
207 product (Vermote et al., 2011). It is an 8-day composite with 7 bands at 464-m
208 spatial resolution. The 7 bands record surface spectral reflectance with wavelengths
209 of 620-670nm, 841-876nm, 459-479nm, 545-565nm, 1230-1250nm, 1628-1652nm
210 and 2105-2155nm. MODIS data can contain noise due to the atmospheric bias,
211 surface anisotropic and sensor problems (Jönsson and Eklundh, 2004) and only data
212 flagged as good or marginal in the MOD09A1 reliability layer were extracted from
213 the original time series data. Band 5 captures short wave infrared reflectance and is
214 sensitive to water vapour. It commonly has a few missing values due to strip noise
215 (Wang et al., 2011) and so an inverse distance weighting interpolation was used to
216 predict (or infill) them. Each MODIS band image consisted of 6200 pixel sites.
217
218 Land cover ground data at 494 randomly selected locations was collected by visual
219 interpretation of the VHR image layers in Google Earth. At each location, the
220 proportions of different land cover types were recorded for an area the size of the
221 MOD09A1 grid cell. Eight land cover types were recorded (Urban, Settlement,
222 Paddyfield, Cultivated, Trees, Grass, Bare and Water) and the land cover with the

largest area in each cell was used to label that cell. This ground data was then associated with its corresponding imagery data.

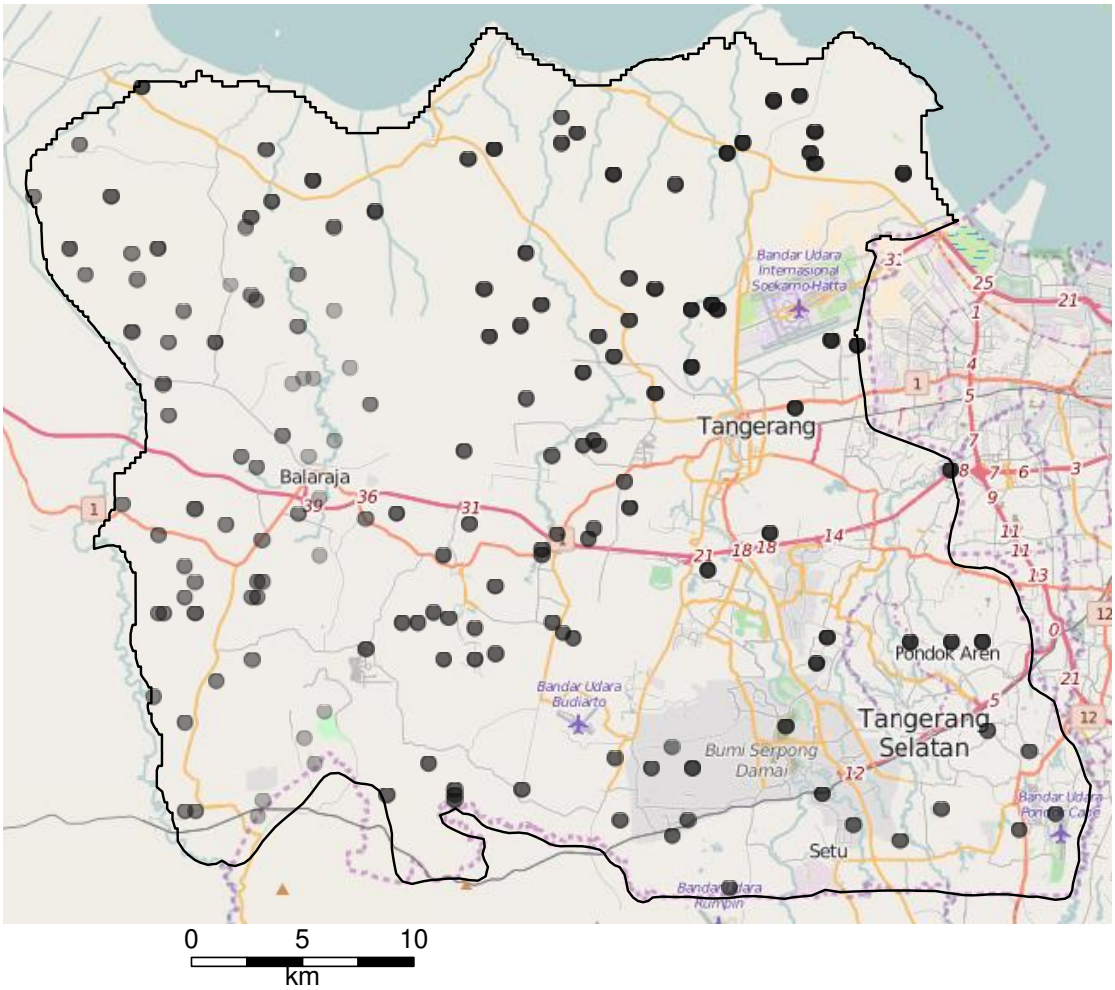


Fig. 1. the study area to the west of Jakarta, Indonesia with the 494 land cover ground data, with a transparency term to show their density, and an OpenStreetMap backdrop.

In order to objectively assess classification accuracy of this study's methods, the combined ground/imagery data were randomly divided into training and validation subsets using a class-stratified 80/20 split. These 80/20 splits were repeated 100 times and the classification procedures applied to the 100 different splits. The distribution of land cover classes for the described training/validation split is given in

Table 1. Observe that the training data set size is relatively small, and as such, will provide a further challenge to this study's methods.

Table 1. Class-stratified training/validation split for land cover ground data.

	Urban	Settlement	Paddyfield	Cultivated	Trees	Grass	Bare	Water
Training	26	96	190	22	11	32	15	4
Validation	6	24	47	5	3	8	4	1

3.2. GWPCA

For GWPCA, a localised PCA is computed at target locations**,** allowing a local identification of any change in the structure of a multivariate data set. Formally, if spatial location i has coordinates (u, v), then GWPCA involves a vector of observed variables $\mathbf{x}_i$ being conceptualised as having a certain dependence on its location i, where $\boldsymbol{\mu}(u_i, v_i)$ and $\boldsymbol{\Sigma}(u_i, v_i)$ are the local mean vector and the local variance-covariance matrix, respectively. The local variance-covariance matrix is

$$\boldsymbol{\Sigma}(u_i, v_i) = \mathbf{X}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\mathbf{X}, \tag{1}$$

where $\mathbf{X}$ is the data matrix (with n rows for the observations, and m columns for the variables), and $\mathbf{W}(u_i, v_i)$ is a diagonal matrix of geographic weights. For this study these were generated using a bi-square kernel function

$$w_{ij} = \left(1 - \left(d_{ij}/r\right)^2\right)^2 \text{ if } d_{ij} \le r \quad w_{ij} = 0 \text{ otherwise,} \tag{2}$$

where the bandwidth is the geographic distance r, and d_{ij} is the distance between spatial locations of the i^{th} and j^{th} rows in $\mathbf{X}$. As with any GW model, other kernel shapes are possible (Gollini et al., 2015). To find the local PCs at location (u_i, v_i), the decomposition of the local variance-covariance matrix provides the local eigenvalues and local eigenvectors (or loading vectors) with

$$\mathbf{L}(u_i, v_i)\mathbf{V}(u_i, v_i)\mathbf{L}(u_i, v_i)^{\mathrm{T}} = \mathbf{\Sigma}(u_i, v_i), \tag{3}$$

where $\mathbf{L}(u_i, v_i)$ is a matrix of local eigenvectors, and $\mathbf{V}(u_i, v_i)$ is a diagonal matrix of local eigenvalues. A matrix of local component scores $\mathbf{T}(u_i, v_i)$ can be found using

$$\mathbf{T}(u_i, v_i) = \mathbf{X}\mathbf{L}(u_i, v_i), \tag{4}$$

where the product of the i^{th} row of the data matrix with the local eigenvectors for the i^{th} location provides the i^{th} row of local component scores. If each local eigenvalue is divided by $\mathrm{tr}(\mathbf{V}(u_i, v_i))$, then localised versions of the proportion of the total variance in the original data accounted for by each component are found (see section 4.1).

Thus at each observed location for a GWPCA with m variables, there are m components, m eigenvalues, m sets of component loadings (each of size $m \times m$), and m sets of component scores (each of size $n \times m$). Eigenvalues and their associated eigenvectors at unobserved locations can be obtained, although as no data exist for these locations component scores cannot be obtained.

278

279 A GWPCA was used to generate spatially-varying PCAs for the image data. A GWPCA
280 loadings data set for a given image band, for a given PC, reflects a spatially-
281 distributed set of correlations between the observations of the original band and the
282 GWPCA scores for the chosen PC. GWPCA loadings provide a local summary of each
283 band's local variance together with the local covariances, and because of this they
284 succinctly encapsulate the multivariate spatial structure in the image data. This is the
285 prime reason why they are considered worthy as input variables to improve land
286 cover classification accuracy. For the case study data (and for a given GWPCA
287 bandwidth), $7 \times 7 = 49$ GWPCA loadings data sets are generated, together with
288 $6200 \times 7 = 43400$ GWPCA scores data sets. Thus a considerable amount of data is
289 generated.

290

291 Both PCA and GWPCA results are presented, where for the PCA the image bands
292 were standardised to specify the covariance matrix. The same globally standardised
293 data were also used in the GWPCA, which is similarly specified with (localised)
294 covariance matrices. As with any PCA-based study there are consequences of these
295 data pre-processing decisions and different results may occur (e.g. Eklundh and
296 Singh 1993). Furthermore, for GWPCA, data that are globally standardised does not
297 guarantee that the data will retain their associated properties at the scale of each
298 localised PCA. A detailed presentation on the consequences of these data pre-
299 processing decisions when applying (PCA and) GWPCA, together with a list of
300 pragmatic data checks, is given in Harris et al. (2015).

301

302 **3.3. Supervised Classification**

303

304 In remote sensing, supervised classification proceeds by examining the
305 characteristics of the training data (image and ground data) to be used in the
306 classification and allocates image objects to classes based on their characteristics at
307 the validation sites. In this study, the image input data was supplemented with the
308 GWPCA loadings and then with the GWPCA scores of the image data itself.

309

310 Three classification algorithms were applied: (a) a latent discriminant analysis (LDA)
311 implementation of maximum likelihood, (b) a logistic regression (LR), and (c) support
312 vector machines (SVM). These were implemented using the following functions and
313 associated R packages, respectively: *lda* in MASS (Venables and Ripley, 2002),
314 *multinom* in nnet (Ripley, 2013) and *svm* in e1071 (Meyer et al., 2012). In all cases,
315 the default arguments for the parameterisation of the classifiers were retained. For
316 details, please refer to the R package manuals.

317

318 Classification algorithms were chosen according to their common usage and the fact
319 that each classifier could be reliably run without additional manipulation or input
320 parameters. Furthermore, the LDA and LR classifiers (which are broadly similar)
321 provide a useful contrast to SVM which takes a quite different (machine learning)
322 approach to classification. This rather naïve selection of algorithms provides some
323 objectivity to this study, as it provides a focus to the performance of GWPCA-derived
324 input variables, not the classification algorithms themselves. Future work could
325 expend the choice of algorithms and more accurately assess whether a given
326 algorithm is particularly suited to GWPCA-derived input variables.

327

3.4. Bandwidth Selection for GWPCA

Bandwidth choice is of great importance to any GW approach. Small bandwidths result in greater spatial variation in the local outputs and the results of using large bandwidths get increasingly close to the global metric. Bandwidths can be found in an adaptive form, where the number of nearest neighbours is fixed, or in a fixed form, where the distance is fixed. In this study, only adaptive bandwidths were specified. For a standard implementation of GWPCA, an automatic bandwidth can be found using a cross-validation procedure as detailed in Harris et al. (2011; 2015). This procedure optimally selects the bandwidth according to a minimised fit between the raw data and the scores data.

The aim was to use GWPCA outputs as inputs to improve land cover classification accuracy. As such, it made sense to find a GWPCA calibration (i.e. its bandwidth) whose outputs provided the most accurate classification. Only the GWPCA loadings needed to be considered in this exercise as the GWPCA scores data should be found from a small, user-specified bandwidth reflecting their use for anomaly detection. The bandwidth selection procedure used in this study is described as follows:

i. GWPCA loadings data were generated at all 6200 pixel sites of the full image using GWPCAs calibrated with adaptive bandwidths of 1%, 5%, 10%, 15%, and continuing in increments of 5%, to a maximum of 100%. Thus for a bandwidth of 1%, localised PCAs were found using only their nearest 62 neighbours. For a bandwidth of 5%, localised PCAs were found using their nearest 310 neighbours, and so on. This results in 21 instances of GWPCA loadings data sets.

ii. The whole 80/20 training/validation classification assessment (i.e. now at the 494 ground data sites) was repeatedly re-run using the same raw image data, but for each run a different set of GWPCA loadings data was used from step (i). Here it soon became apparent that GWPCA loadings data via a 20% bandwidth would provide the most accurate classification results (at least on average for each classifier over the 100 runs). Thus for clarity, the accuracies in Table 3 were found 21 times corresponding to the most accurate results.

Observe that step (ii) of this procedure is sub-optimal in that a more accurate set of results would be possible if an optimal bandwidth was retained for: (a) each individual training/validation data split and (b) each classifier (i.e. LDA, LR and SVM). However, such level of detail would distract from this paper's narrative. It was also considered useful to have a broad understanding of the spatial scale at which the image-derived GWPCA loadings were best able to discriminate between land cover classes. A single bandwidth allows this, where a 20% bandwidth uses the nearest 1240 neighbouring pixels. Thus in summary, a 20% bandwidth was user-specified but was strongly guided by the given validation exercise in step (ii) above.

Also observe that the bandwidth selection procedure is potentially compromised in step (i) in that any given set of ground data validation sites (always some class-stratified random allocation of 98 sites from 494 sites) is always included in the bandwidth selection procedure. That is, each set of GWPCA loadings data was in part derived from image information at the 98 validation sites, where the extent of *contamination* at any one of 6200 pixels accorded to its proximity to a validation site.

377 The question then arises - is this a serious oversight and if so, should all validation
378 data sets be entirely unseen until the final accuracy assessment?
379
380 Although, it would have been possible to remove such validation sites from step (i),
381 and still provide GWPCA loadings data at these now unobserved sites in step (ii) (see
382 the GWPCA algorithm in section 3.2) – thus negating this issue altogether, a revision
383 was not undertaken for the following three reasons:
384
385 a. It would have entailed that in step (i), the GWPCA algorithm would have had to
386 run 21x100 =2100 times to reflect the 21 bandwidth choices together with the
387 100 training/validation data splits.
388 b. It was likely that each set of GWPCA loadings data would change little if the image
389 data at the 98 validation sites (1.6% of the image) were included or not. In turn,
390 the final selection of a 20% bandwidth would still be likely.
391 c. The chosen 20% bandwidth was itself a (deliberately) sub-optimal selection.
392
393 Thus in the interest of parsimony and pragmatism, such a revision was not followed.
394 All further results of this study were considered similarly unaffected by this decision.
395
396 Furthermore, this issue is only concerned with the creation of variables for input into
397 a classification algorithm. It is not concerned about the testing of the classification
398 itself, as is usually the case in a training/validation exercise – and here, in step (ii),
399 the validation data still remained unseen in this sense. A final point worth noting is
400 that there would be no advantage to only focus on the 494 ground data sites (i.e.
401 ignore the full 6200 image data altogether) for bandwidth selection, say to save

computationally. A bandwidth found using this relatively sparse data (see Fig. 1b) would not directly transfer to that which is required for the full image data.

4. Results

4.1. PCA

For any GW model application, it is informative to consider its global counterpart for reference. The PCA results are shown in Table 2 and indicate that a subsequent analysis could justifiably proceed retaining only the first (PC1) and second (PC2) components as both have eigenvalues that are greater than 1 and together they account for 87.4% of the total variance. This level of explained variance amongst only the first two PCs reflects strong levels of collinearity amongst the MODIS bands, which is not unexpected with this type of data. Interrogation of a simple correlation matrix confirms this, with strong correlations ($p > 0.85$) between Bands 1 and 3, Bands 1 and 4, Bands 2 and 5, Bands 3 and 4, Bands 5 and 6, Bands 6 and 7. The PCA loadings indicate that Band 6 contributes most to PC1 and that Bands 2 and 5 equally contribute the most to PC2.

Table 2. PCA outputs from the 7-band MODIS data.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalues	4.082	2.036	0.695	0.090	0.046	0.036	0.015
Variance %	58.3	29.1	9.9	1.3	0.7	0.5	0.2
Cumulative variance %	58.3	87.4	97.3	98.6	99.3	99.8	100
Band 1 loadings	0.388	0.415	0.030	0.344	0.493	-0.561	0.001
Band 2 loadings	0.284	-0.491	0.484	-0.024	-0.380	-0.490	-0.241
Band 3 loadings	0.390	0.389	0.126	-0.820	-0.059	0.010	0.072
Band 4 loadings	0.392	0.334	0.406	0.428	-0.296	0.542	-0.079
Band 5 loadings	0.340	-0.491	0.119	-0.111	0.659	0.389	-0.178
Band 6 loadings	0.435	-0.286	-0.269	0.111	-0.160	0.004	0.787
Band 7 loadings	0.398	-0.032	-0.705	0.038	-0.249	0.017	-0.529

422

423 **4.2. GWPCA**

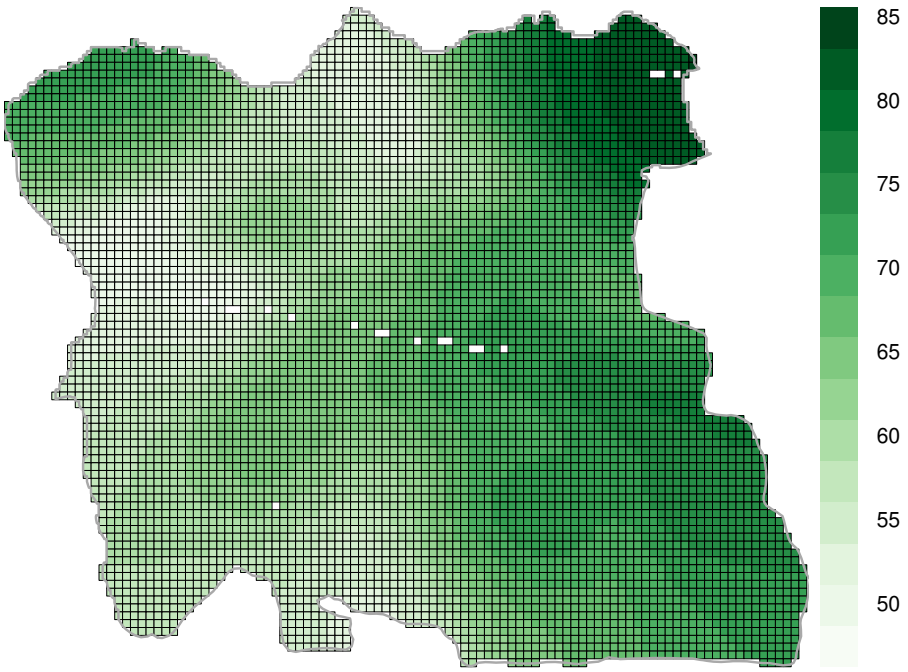
424

425 A GWPCA was then applied to the same data, generating local eigenvalues, local
426 variance proportions, local cumulative variance proportions, local loadings data sets
427 and local scores data sets, across the 6200 image sites. The GWPCA can assess how
428 the dimensionality in the imagery can vary across the study region via the local
429 variances and how the multivariate structure of the imagery can vary via the local
430 loadings data sets. Fig. 2a shows the spatial distribution of the variance proportions
431 accounted for by PC1 and how they vary geographically from the global value of
432 58.3%. Fig. 2b shows the distribution of cumulative variance proportions for PC1
433 and PC2 combined, which was 87.4% globally. It is evident that PC1 explains much
434 more of the variance in the north eastern corner of the study region (Fig. 2a), whilst
435 together PC1 and PC2 explain more of the cumulative variance in the northern and
436 eastern areas of the study region (Fig. 2b). These areas are also most likely to exhibit
437 the strongest levels of (local) collinearity amongst the image bands. Image
438 dimensionality clearly varies across the study region, where for all areas the
439 retention of the first two PC's will at least account for 78% of the total variance.

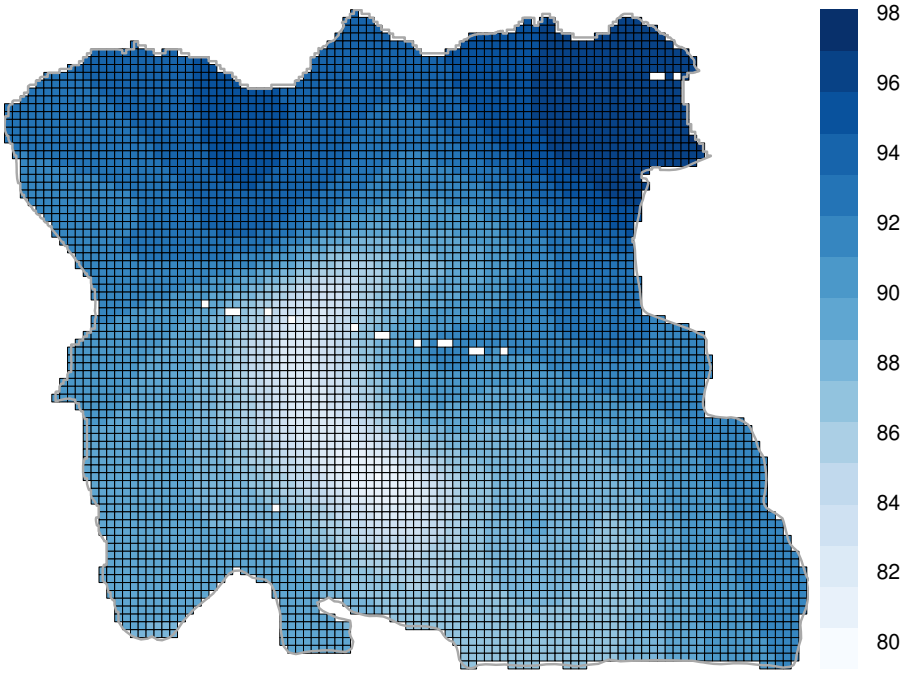
440

441 Investigating and visualising the loadings data from GWPCA is a challenge, and in this
442 respect, various visualisation tools can be found in Harris et al. (2011; 2015). The
443 difficulty lies in the fact that at every local PCA location, the loadings data for each
444 band and PC needs to be somehow viewed and related to each other. For this study,
445 a simple visualisation is adopted where the image bands with the largest (absolute)
446 loadings for the first two PCs only are mapped (Fig. 3). It is clear that different bands

447 dominate PC1 within the study region and that Band 6, which dominates globally,
448 only dominates in a north western and a north eastern area (Fig. 3a). Locally, it now
449 appears that Bands 1 and 4 contribute more to PC1, than Band 6 does. Similarly,
450 Bands 2 and 5 that contributed the most to PC2, do not dominate PC2 throughout
451 the study region (Fig. 3b). Although in this instance, Band 2 displays a degree of
452 homogeneity - as it provides the highest absolute loading for over 50% of the region.
453
454 Although this has intentionally been only a brief demonstration of GWPCA, it has
455 highlighted that both dimensionality and structure in the image data can vary across
456 the study region. In particular, as image-band structure (via the loadings) clearly
457 exhibits spatial variation then this information has the potential to act as a useful
458 discriminator of land cover, which is now assessed.

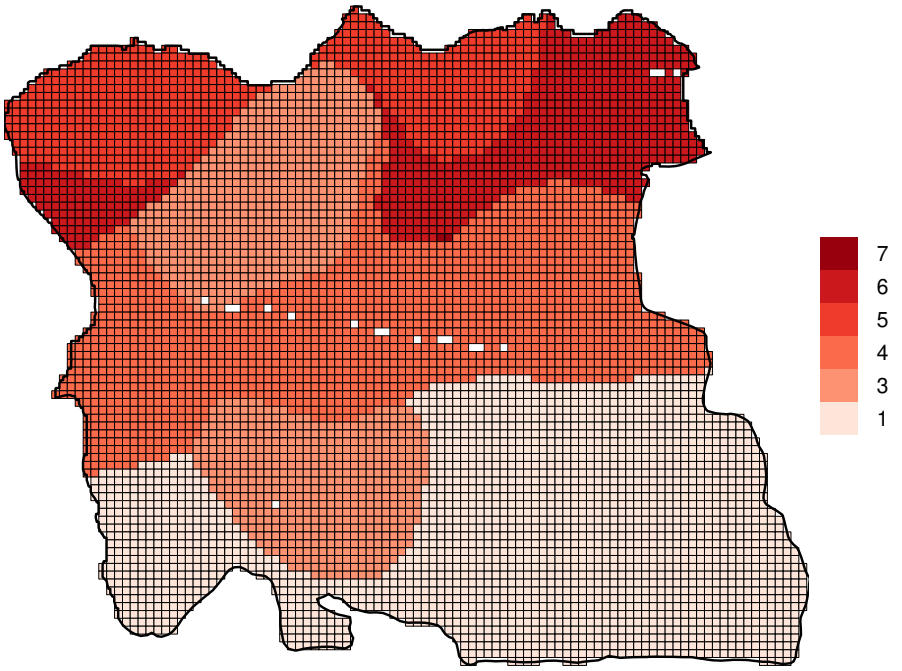


a)

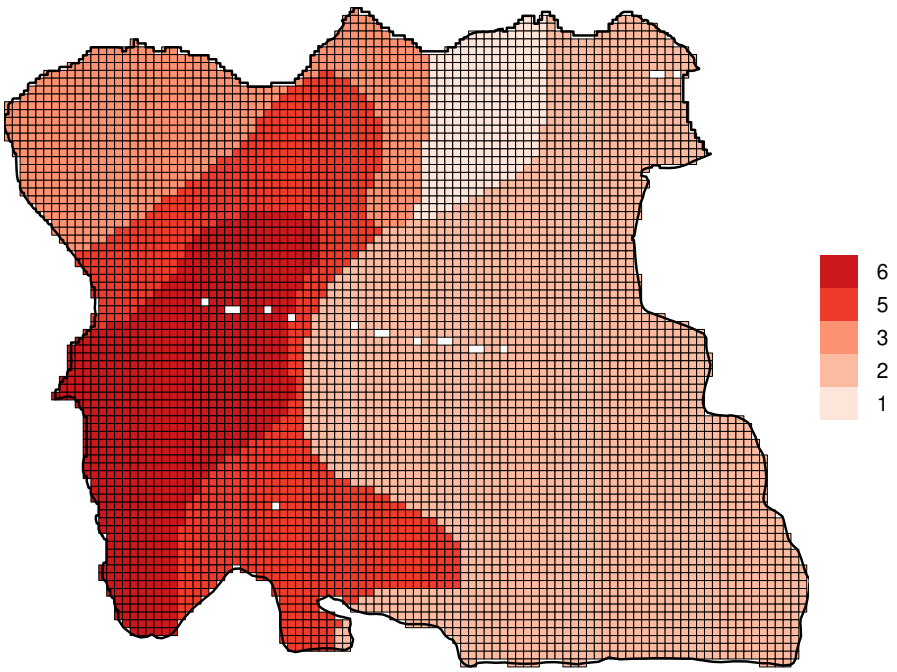


b)

459 **Fig. 2.** Spatial distribution of the local variances proportions explained by a) PC1 and
460 b) PC1 and PC2 combined, from a GWPCA. Global variance proportion for PC1 was
461 58.3%. Global cumulative variance proportion for PC1 and PC2 combined was 87.4%.



a)



b)

462 **Fig. 3.** Spatial distribution of the image bands with largest (absolute) local loadings
463 for a) PC1 and b) PC2, from a GWPCA. Corresponding global PCA results were
464 respectively Band 6, and Bands 2 and 5, jointly.
465

466 **4.3. Land Cover Classification**

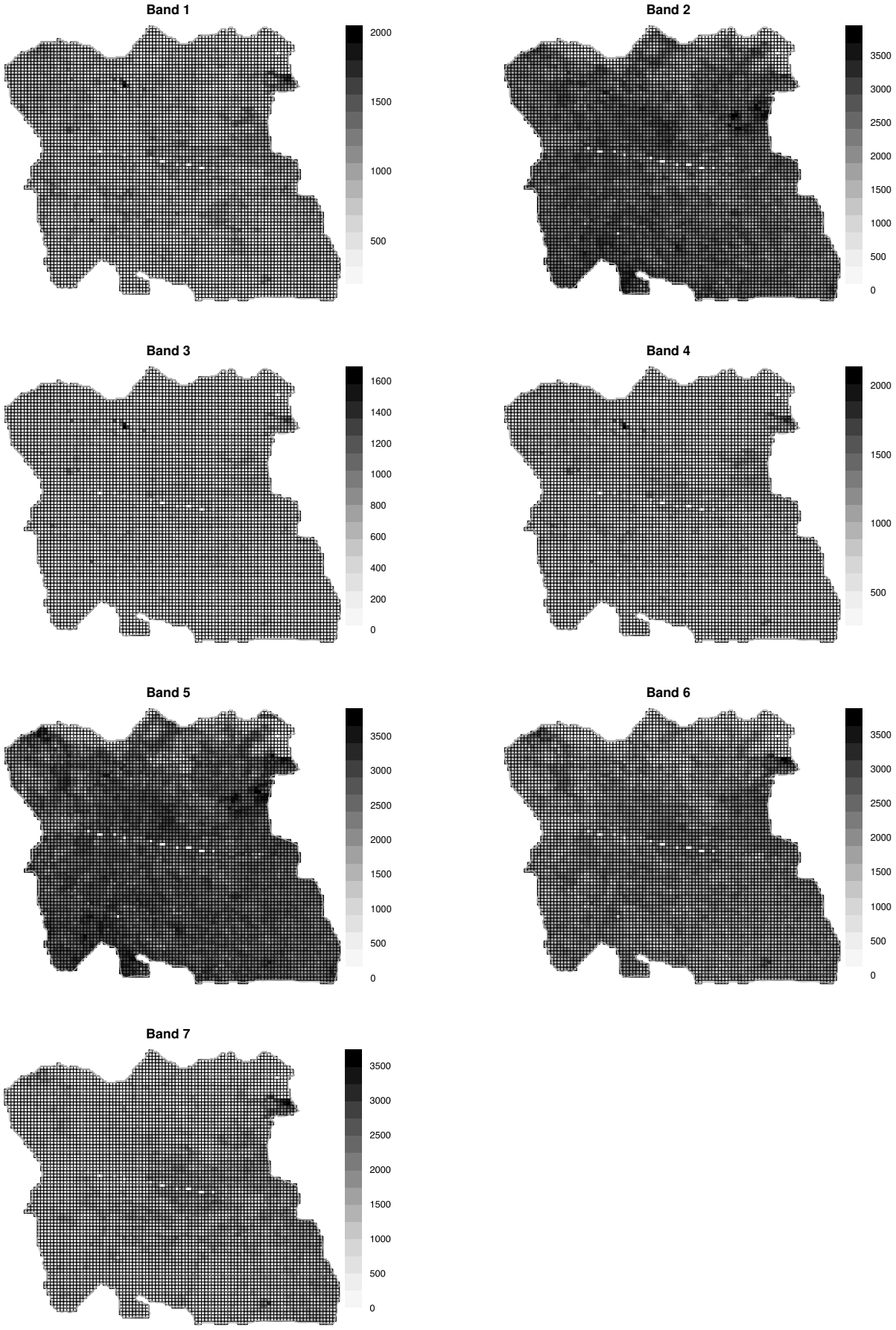
467

468 A series of supervised classifications were undertaken with four groups of input
469 variables: (1) the seven image bands only (2) the image-derived GWPCA loadings for
470 PC1 and PC2 only, (3) the image data plus the GWPCA loadings, and (4) the image
471 data plus the GWPCA loadings plus the GWPCA ranked scores. As the global PCA
472 indicated that the first two PCs were the most important, then this was assumed to
473 be true for the GWPCA with respect to the loadings. To provide some spatial context
474 to the different input data and how they vary across the study region, Figures 4, 5
475 and 6 shows in their full form the MODIS imagery and the GWPCA image band
476 loadings for PC1 and for PC2.

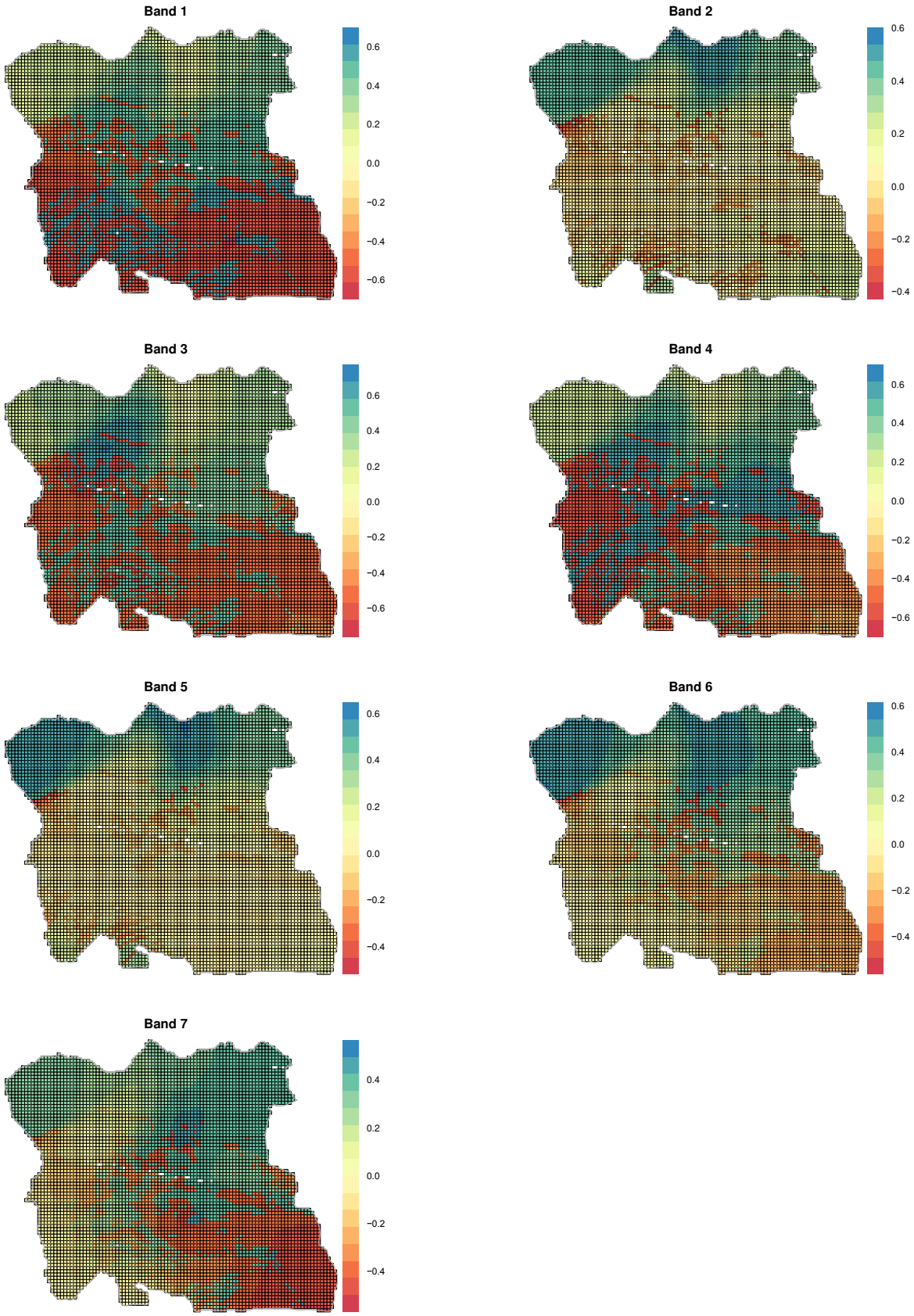
477

For input variable group 4, the GWPCA ranked scores data capture local multi-band outlier information. Harris et al. (2014b) found that the most promising outlier detection method resulted from observations with extreme local scores values from either the first PC or from the last PC. Thus two additional texture input variables were constructed to reflect the ranking of the local scores data for each of PC1 and for PC7, where the lower the ranking, the more likely that the MODIS pixel is locally anomalous in a multi-band (or multivariate) sense. These extra input variables (simply termed GWPCA ranked scores) may help classify a land cover that is somewhat obscure, or occurs in an unexpected location/setting. In this instance, the GWPCA run was calibrated with a much smaller (user-specified) bandwidth of 2.5%, as GWPCA was now being used to detect anomalies.

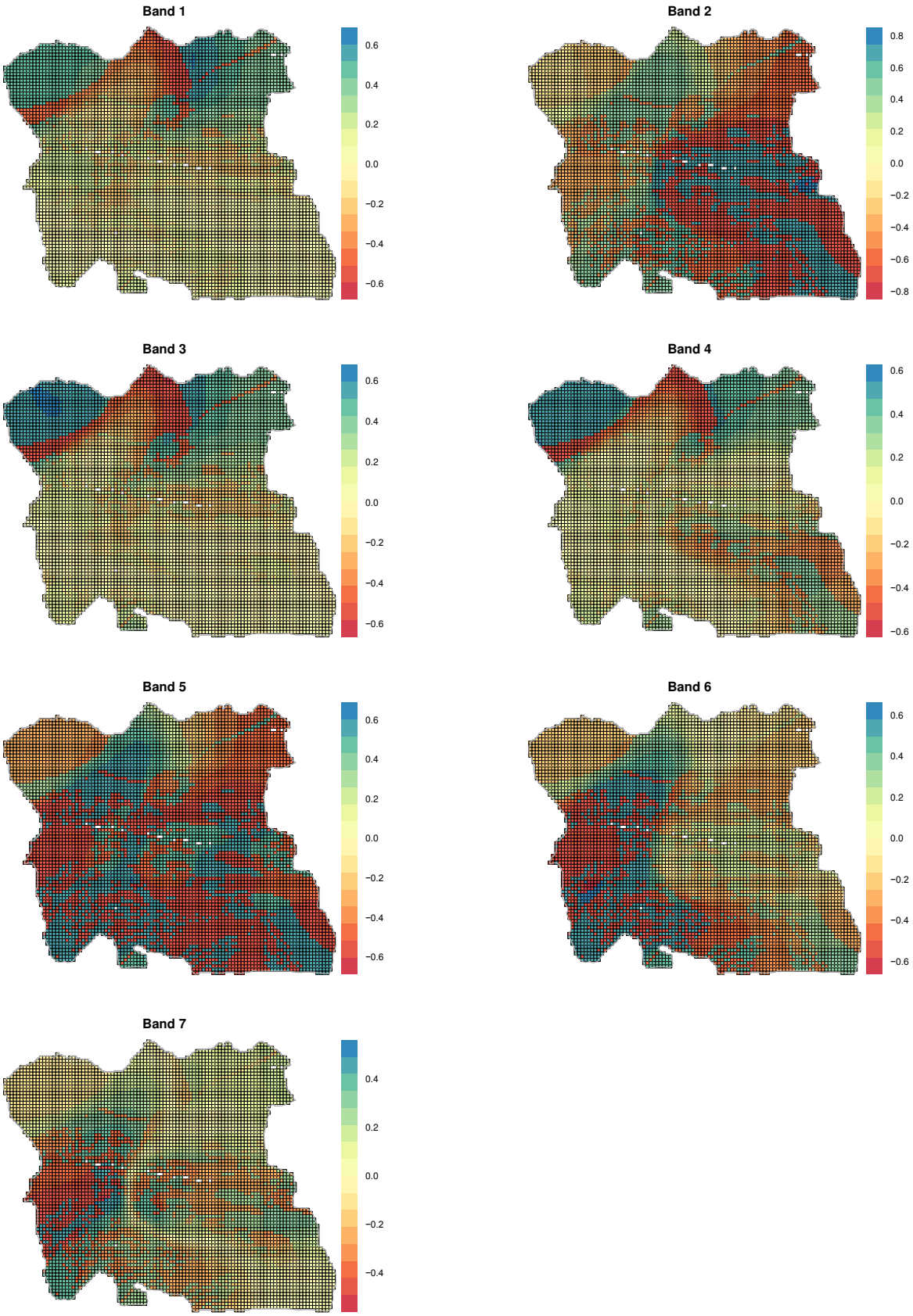
491



492 **Fig. 4.** Inputs into land cover classifications: the MODIS imagery



493 **Fig. 5.** Inputs into land cover classifications: the GWPCA band loadings for PC1
494



495 **Fig. 6.** Inputs into land cover classifications: the GWPCA band loadings for PC2
496

Following the procedures described in section 3, LDA, LR and SVM classifiers were run 100 times across 494 ground data sites using class-stratified 80/20 training/validation data splits (see Table 1) with the four input variable groups described. For each run, the overall accuracy percentage for each classifier was determined from the diagonal of a standard correspondence matrix, comparing the class of the validation data with the predicted class. The resultant mean overall accuracies for 100 runs are presented in Table 3, indicating that classification accuracy is broadly similar for each of the three classifiers when just the image data are used, and also when just the GWPCA loadings are used. However when the image data are combined with the GWPCA loadings, the accuracies increase markedly. This suggests that including variables that describe the spatial multivariate structure of the imagery improves classification predictive strength. There are only slight improvements in accuracy when the GWPCA ranked scores data were included as inputs. However, these marginal improvements were entirely expected given that the focus was on anomalies and by definition, they should be fairly rare. On average over the 100 runs, LR is consistently the most accurate classifier, in this instance.

Table 3. The mean overall accuracy percentages for four different input variable groups to a set of three different classification algorithms. Corresponding standard errors of the means (SEMs) are in brackets. The number of input variables per group were 7, 14, 21 and 23, respectively.

	Image	GWPCA loadings	Image + GWPCA loadings	Image + GWPCA loadings + GWPCA ranked scores
LDA	63.6 (0.39)	61.9 (0.36)	68.7 (0.38)	70.4 (0.38)
LR	66.0 (0.37)	65.5 (0.40)	75.5 (0.40)	77.4 (0.38)
SVM	65.3 (0.25)	64.3 (0.25)	69.4 (0.30)	74.9 (0.34)

To provide a fuller understanding of the results of Table 3, Table 4 describes the results per land cover class. It show the improvement from using the image data only (input variable group 1) to using the image data plus the GWPCA loadings plus the GWPCA ranked scores (input variable group 4). These results need to be viewed in context of the training/validation data splits given in Table 1, where for land cover classes that are poorly represented (or rare), the classification improvement is often quite marked, whereas for land cover classes that are relatively well represented (Settlement and Paddyfield), classification accuracy is sometimes marginally reduced. These results provide clear value in the GWPCA-based methodology to accurately classify land cover across the full spectrum of possible classes, and in doing so, goes someway in justifying the extra complexity that the new methodology introduces into the classification procedure.

Table 4. Changes in mean overall accuracy for each land cover class, comparing the 'Image' input group to the 'Image+GWPCA loadings+GWPCA ranked scores' input group.

	Urban	Settlement	Paddyfield	Cultivated	Trees	Grass	Bare	Water
LDA	38.2 to 68.0	71.2 to 68.1	86.2 to 79.3	8.2 to 75.4	20.0 to 20.0	18.6 to 54.5	1.2 to 35.5	0 to 68
LR	42.5 to 63.5	74.0 to 78.0	88.9 to 84.8	6.6 to 80.2	19.3 to 20.7	16.6 to 59.9	1.5 to 62.5	0 to 99
SVM	1.8 to 45.5	83.4 to 82.9	92.5 to 92.8	0 to 48.6	1.3 to 13.3	7.4 to 45.5	0 to 22.8	0 to 0

Fig.7 shows the full distributions of the accuracy results as summarised in Table 3, from all 100 of the class-stratified 80/20 training/validation data splits. The boxplots of the accuracy distributions clearly shows that the use of GWPCA-derived inputs improves classification accuracy. The greatest improvements were found with the LR classifier, where for some training/validation data splits, land cover classification accuracy exceeds 80%. The boxplots also confirm that the SVM classifier consistently has lower variance in the results than the LDA or LR classifiers. Paired *t*-tests were

used to test for significance differences in the means of selected accuracy distributions and are summarised in Table 5. All of the key differences that have been reported are highly significant.

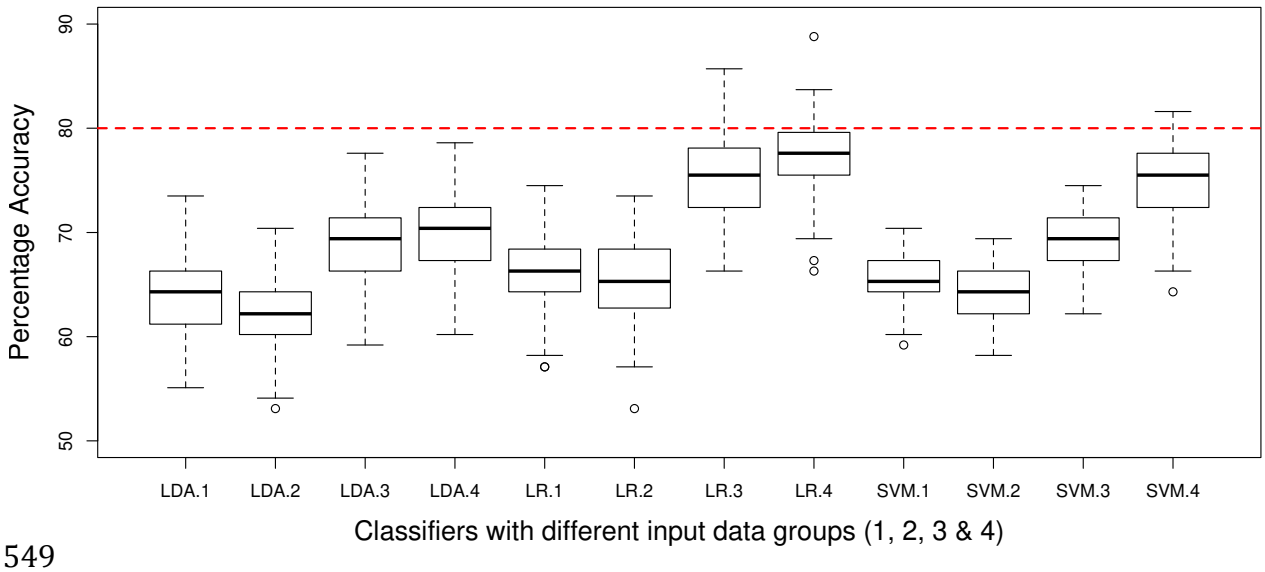


Fig. 7. Distributions of classification accuracy data from 100 runs of the training/validation data splits. Classifiers (LDA, LR, SVM) denoted with input groups (1-Image, 2-GWPCA loadings, 3-Image+GWPCA loadings, 4-Image+GWPCA loadings+GWPCA ranked scores). Given with an 80% accuracy line for context.

Table 5. Paired *t*-test results for differences in mean overall accuracies, with input groups as follows: 1-Image, 2-GWPCA loadings, 3-Image+GWPCA loadings, 4-Image+GWPCA loadings+GWPCA ranked scores.

	Groups 1 vs. 2	**Groups 1 vs. 3**	**Groups 1 vs. 4**	**Groups 3 vs. 4**
LDA	$p < 0.0019$	$p < 0.0000$	$p < 0.0000$	$p < 0.0018$
LR	$p < 0.2945$	$p < 0.0000$	$p < 0.0000$	$p < 0.0005$
SVM	$p < 0.0083$	$p < 0.0000$	$p < 0.0000$	$p < 0.0000$

5. Discussion

The results of Table 3 indicate that use of the MODIS image data alone (Fig. 4) or the GWPCA loadings alone (Fig. 5 and 6) result in similar land cover classification accuracies. However, accuracy improves when the image data is supplemented with

565 the GWPCA loadings as texture variables. This improvement is because the local
566 loadings arising from GWPCA capture important spatial heterogenic effects in the
567 multi-band structure of the image data via variances and covariances. This
568 information reflects spatially-distributed sets of correlations between the
569 observations of the original bands and the GWPCA scores and describes the local
570 contribution of each band to each PC. The results imply that if a loadings structure
571 were to be associated with each land cover class it would not be fixed, but instead
572 would vary geographically placing land cover in context with its locality. Further, but
573 more marginal improvements in accuracy were found when the GWPCA ranked
574 scores were included as inputs. Here only slight improvements were expected given
575 that these inputs should only help classify a land cover that is somewhat obscure, or
576 occurs in an unexpected location/setting. The results of Tables 4 and 5 and Fig. 7
577 were given to provide clarity and detail to those summarised in Table 3, providing
578 added value to the GWPCA-based land cover classification methodology.
579
580 In this study the classifications were undertaken from a standpoint of naivety. It is
581 well known that collinearity might be expected between certain image bands and
582 this was the case here. Furthermore, many of the 14 different GWPCA loadings
583 datasets are themselves highly collinear. Evidence of this can be seen in Fig. 5 and 6
584 and is not surprising given the collinearity of the image bands. Collinearity can result
585 in a loss of model precision and a loss of power in a classification model's parameter
586 estimates. There are a number of ways to reduce such global collinearities. An initial
587 step of image band selection could be applied to help identify specific types of
588 features, for example, red and infra-red bands to support biomass analyses. Also, the
589 image bands and GWPCA loadings data that are highly collinear could be removed,

590 the first two global PCs of the image data could be used with the GWPCA loadings, or
591 a classification technique could be specifically designed to accommodate collinear
592 variables such as a penalised or shrinkage approach (see Tibshirani, 1996). Removal
593 of collinear image bands and GWPCA loadings data and using only the first two
594 global PCs of the image data were experimented with, but found to make little
595 difference to the classification results. This was not surprising given that collinearity
596 tends to have more of an effect on model inference rather than the model's ability
597 to provide accurate predictions.
598
599 Of note however, is that of the three classifiers, SVM is itself a penalised or shrinkage
600 method in that it has a regularisation term. This may in part explain why the SVM
601 classifier consistently performed better than the LDA or LR classifiers with respect to
602 its spread of results (see Table 3 and Fig. 7). This is because the standard error of the
603 means (SEMs) are consistently smaller than that found with LDA and LR. However, as
604 only the default arguments for the optimisation of its parameters were used, the
605 SVM classifier would still require some user-input to check that the regularisation
606 term and any kernel parameters were correctly calibrated.
607
608 In any land cover classification study the results are only ever specific to the
609 particular properties of data and the region it covers. As a consequence, it can be
610 difficult to infer the transferability any methodology to other studies, especially those
611 with very different spatial and attribute properties. In this work a case study was
612 chosen in which the spatial resolution of the imagery was coarse relative to the study
613 area as a means to demonstrate this paper’s methodology. There is no reason why
614 the methodology would not perform similarly well if applied to data at a much finer

615 spatial resolution or over a much larger area. Thus issues of scale should not unduly
616 compromise this new classification method. What is important to local techniques is
617 that sufficient information is available so that important spatial heterogeneities can
618 be reliably captured. Future work will apply the methodology within a simulation
619 experiment to data generated with known and user-controlled properties.
620
621 Unlike many moving window or partitioned-based methods, a GW approach makes
622 much better use of available information, as it is still possible to use all the data
623 whilst still modelling local effects. For example, a 100% bandwidth of this study was
624 still able to provide localised PCAs because a distance-decay (bi-square) kernel was
625 specified. Similarly, much attention in GW modelling is placed in finding optimal
626 kernel bandwidths so that the scale at which each localised model operates is
627 appropriately determined. Future work will compare similarly localised classification
628 methods to the one demonstrated here, providing further context to the GWPCA-
629 based method.
630
631 In this study GWPCA was used to improve land cover classification accuracy.
632 However, GWPCA may also provide local solutions to other modelling issues of
633 interest to the remote sensing community. GWPCA could be applied for an
634 optimisation of the image data collection, or the detection of image band outliers
635 (e.g. Harris et al., 2014a; b). It could also be applied as a local data reduction
636 technique to hyperspectral imagery or in a data fusion exercise with inputs operating
637 at varying spatial scales. This study adds to a growing body of work that has used
638 GWPCA to provide a greater understanding on how dimensionality and structure in
639 multivariate data can vary spatially including research in the social sciences

(Fotheringham et al., 2002; Lloyd, 2010, Harris et al., 2011) and the environmental sciences (Kumar et al., 2012; Harris et al., 2015).

The results from this study strongly suggest that land cover classes have different image properties in different portions of the image scene. However the GWPCA-based methodology is such that this relationship heterogeneity is only indirectly accounted for since ultimately, only non-spatial classifiers are used. Thus a GW logistic regression or GW discriminant analysis could be applied with land cover as the response variable and the image bands as the predictors. This would enable a direct and spatially-informative investigation of such heterogeneous relationships. However as the image bands are not only highly collinear globally, but also locally, then these GW models may be somewhat compromised (e.g. Páez et al., 2011), and if so, one way to address this would be to replace the raw image data with their (locally orthogonal) GWPCA scores data. These observations present an interesting philosophical challenge in remote sensing as it implies that the definition of any given land cover class as a position in multivariate feature / image space, actually varies geographically. It also has potentially interesting applications and implications for socio-economic classifications such as *land use* which are hard to detect directly from remote sensing imagery alone.

6. Conclusions

This research has found that the use of spatial measures of imagery structure and anomalies in the form of GWPCA loadings and GWPCA ranked scores can provide significant improvements in land cover classification accuracy. Such improvements

665 have the potential to provide immediate benefits and thereby may offer greater
666 information value than many technology-led developments (new sensors, finer scales
667 etc.). In this way, remote sensing as a discipline could benefit from a greater focus on
668 the many existing as well emergent techniques arising from spatial science.
669
670 **Acknowledgements**
671 The authors would like to thank Prof Heiko Balzter and the *Centre for Landscape and*
672 *Climate Research* at the University of Leicester, and the JSPS program *International*
673 *network-hub for future earth: research for global sustainability* for supporting this
674 research. Research is also supported by the Biotechnology and Biological Sciences
675 Research Council of the UK (BBSRC BB/J004308/1). All of the statistical analysis and
676 mapping were implemented in R version 3.2.0, the open source statistical software
677 (http://cran.r-project.org). The code and data used in this analysis will be provided
678 to interested researchers on request. The authors would like to thank the
679 anonymous reviewers whose comments helped significantly improve this article.
680
681 **References**
682 Abdel-Aziz, Y. I., & Karara, H. M. (2015). Direct Linear Transformation from
683 Comparator Coordinates into Object Space Coordinates in Close-Range
684 Photogrammetry. *Photogrammetric Engineering & Remote Sensing*, *81*(2), 103-
685 107.
686 Atkinson, P.M., & Lewis, P. (2000) Geostatistical classification for remote sensing: an
687 introduction. Computers & Geosciences 26, 361-371

688 Atkinson, P.M. (2004) Spatially weighted supervised classification for remote
689 sensing. *International Journal of Applied Earth Observation and*
690 *Geoinformation* 5(4), 277-291.
691 Atkinson, P.M., & Naser. D.K. (2010) A Geostatistically Weighted k‐NN Classifier for
692 Remotely Sensed Imagery. *Geographical Analysis* 42(2), 204-225.
693 Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and
694 data representation. *Neural Computation*, *15*(6), 1373-1396.
695 Brunsdon, C., Fotheringham, A.S. & Charlton, M.E.(2002) Geographically weighted
696 summary statistics - a framework for localised exploratory data analysis.
697 *Computers, Environment and Urban Systems*, 26, 501-524.
698 Brunsdon, C., Fotheringham, A.S. & Charlton, M. (2007), Geographically Weighted
699 Discriminant Analysis, *Geographical Analysis*, 39, 376-396.
700 Brunsdon, C., Fotheringham, A.S. & Charlton, M. (1996), Geographically Weighted
701 Regression: A Method for Exploring Spatial Nonstationarity. *Geographical*
702 *Analysis*, 28(4), 281-298.
703 Campbell, J., (1981). Spatial correlation effects upon accuracy of supervised
704 classification of land cover. *Photogrammetric Engineering of Remote Sensing*,
705 47, pp. 355–364.
706 Car, J.R., & Miranda, F.P. (1998) The semivariogram in comparison to the co-
707 occurrence matrix for classification of image texture. IEEE Transactions on
708 Geoscience and Remote Sensing 36(6), 1945-1952
709 Chica-Olmo, M. & Abarca-Hernandez, F. (2000). Computing geostatistical image
710 texture for remotely sensed data classification. *Computers & Geosciences*,
711 *26*(4), 373-383.

712 Collins, J. B. & Woodcock, C. E. (1996). An assessment of several linear change
713 detection techniques for mapping forest mortality using multitemporal Landsat
714 TM data. *Remote Sensing of Environment*, *56*(1), 66-77.
715 Comber A.J., (2013). Geographically weighted methods for estimating local surfaces
716 of overall, user and producer accuracies. *Remote Sensing Letters*, 4(4): 373-
717 380.
718 Comber, A., Fisher, P.F., Brunsdon, C. & Khmag, A. (2012). Spatial analysis of remote
719 sensing image classification accuracy. *Remote Sensing of Environment*, 127:
720 237–246
721 Congalton, R. G., (1988). Using spatial auto-correlation analysis to explore the errors
722 in maps generated from remotely sensed data. *Photogrammetric Engineering*
723 *and Remote Sensing*, 54, 587-592.
724 Congalton, R. G. (1991). Remote Sensing and Geographic Information System Data
725 Integration: Error Sources and. *Photogrammetric Engineering & Remote*
726 *Sensing*, *57*(6), 677-687.
727 Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A.S., & McLoone, S. (2013).
728 Principal components analysis on spatial data: an overview. *Annals of the*
729 *Association of American Geographers* 103, 106–128.
730 Deng, J.S. Wang, K. Deng, Y.H. & Qi, G.J. (2008). PCA-based land-use change
731 detection and analysis using multitemporal and multisensor satellite data.
732 *International Journal of Remote Sensing*, 29, 4823–4838.
733 Doxani, G., Karantzalos, K., & Tsakiri-Strati, M. (2011). Monitoring urban changes
734 based on scale-space filtering and object-oriented classification. *International*
735 *Journal of Applied Earth Observation and Geoinformation*, 15, 38–48.

736 Eklundh, L., & Singh, A. (1993). A comparative analysis of standardised and
737 unstandardised principal components analysis in remote sensing. *International*
738 *Journal of Remote Sensing*, 14(7), 1359-1370.
739 Filzmoser, P., R. Garrett, & Reimann, C. (2005). Multivariate Outlier Detection in
740 Exploration Geo-chemistry. *Computers & Geosciences*, 31, 579–87.
741 Filzmoser, P., R. Maronna, & Werner, M. (2008). Outlier Identification in High
742 Dimensions. *Computational Statistics and Data Analysis* 52, 1694–711.
743 Foody, G.M. (2005) Local characterization of thematic classification accuracy
744 through spatially constrained confusion matrices. *International Journal of*
745 *Remote Sensing*, 26, 1217-1228.
746 Fotheringham, A.S., Brunsdon, C., Charlton, M. (2002), *Geographically Weighted*
747 *Regression – the analysis of spatially varying relationships*. Wiley, Chichester
748 Getis, A., Ord, J.K. (1992) The Analysis of Spatial Association by Use of Distance
749 Statistics. *Geographical Analysis* 24, 189-206.
750 Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: an R
751 package for exploring spatial heterogeneity using geographically weighted
752 models. *Journal of Statistical Software,* 63.
753 Hansen, M., Dubayah, R., & DeFries, R. (1996). Classification trees: an alternative to
754 traditional land cover classifiers. *International Journal of Remote Sensing*,
755 *17*(5), 1075-1081.
756 Harris, P., Charlton, M., Fotheringham, A.S. (2010) Moving window kriging with
757 geographically weighted variograms. Stochastic Environmental Research and
758 Risk Assessment 24, 1193-1209

759 Harris P., Brunsdon C., Charlton M. (2011) Geographically weighted principal
760 components analysis. *International Journal of Geographical Information*
761 *Science*, 25:1717-1736.
762 Harris P., Clarke A., Juggins S., Brunsdon C., Charlton M. (2014a) Geographically
763 weighted methods and their use in network re-designs for environmental
764 monitoring. *Stochastic Environmental Research and Risk Assessment*, 28: 1869-
765 1887.
766 Harris P., Brunsdon C., Charlton M., Juggins S., Clarke A. (2014b) Multivariate spatial
767 outlier detection using robust geographically weighted methods. *Mathematical*
768 *Geosciences*, 46(1) 1-31.
769 Harris P., Clarke A., Juggins S., Brunsdon C., Charlton M. (2015) Enhancements to a
770 geographically weighted principal components analysis in the context of an
771 application to an environmental data set. *Geographical Analysis*, 47: 146-172.
772 Harsanyi, J. C., & Chang, C. I. (1994). Hyperspectral image classification and
773 dimensionality reduction: an orthogonal subspace projection approach.
774 *Geoscience and Remote Sensing, IEEE Transactions on*, *32*(4), 779-785.
775 Hubert, M., P. J. Rousseeuw, & K. Vanden Branden, (2005). ROBPCA: A New
776 Approach to Robust Principal Component Analysis. *Technometrics*, 47, 64–79.
777 Ingebritsen, S. E., & Lyon, R. J. P. (1985). Principal components analysis of
778 multitemporal image pairs. *International Journal of Remote Sensing*, *6*(5), 687-
779 696.
780 Johnson, B., Ryutaro, T, Zhixiao, Xie. (2012) Using geographically weighted variables
781 for image classification. *Remote Sensing Letters* 3(6), 491-499.
782 Jönsson, P., & Eklundh, L. (2004). TIMESAT—a program for analyzing time-series of
783 satellite sensor data. *Computers & Geosciences*, *30*(8), 833–845.

784 Koutsias, N., Mallinis, G., & Karteris, M. (2009). A forward/backward principal
785 component analysis of Landsat-7 ETM+ data to enhance the spectral signal of
786 burnt surfaces. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(1),
787 37-46.
788 Kumar, S., R. Lal, & Lloyd, C. (2012). Assessing spatial variability in soil characteristics
789 with geographically weighted principal components analysis. *Computational*
790 *Geosciences*, 16, 827–835.
791 Legendre, P., & Legendre, L. (1998). *Numerical Ecology,* 2nd ed. Amsterdam: Elsevier.
792 Lloyd, C. D. (2010). Analysing population characteristics using geographically
793 weighted principal components analysis: a case study of Northern Ireland in
794 2001. *Computers, Environment and Urban Systems*, 34(5), 389-399.
795 Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2012). e1071:
796 Misc Functions of the Department of Statistics (e1071), TU Wien, 2012. *R*
797 *package version*, 1-6.
798 Myint, S. W. (2003). Fractal approaches in texture analysis and classification of
799 remotely sensed data: Comparisons with spatial autocorrelation techniques
800 and simple descriptive statistics. *International Journal of Remote Sensing*,
801 *24*(9), 1925-1947.
802 Páez, A.**,** Farber, S., Wheeler, D. (2011) A simulation-based study of geographically
803 weighted regression as a method for investigating spatially varying
804 relationships, Environment and Planning A 43(12), 2992-3010.
805 Pesaresi M. & J. A. Benediktsson, (2001). A new approach for the morphological
806 segmentation of high-resolution satellite imagery. *IEEE Transactions on*
807 *Geoscience and Remote Sensing*, 39, 309–320.

808 Pohl C. & Van Genderen J.L., (1998). Multisensor image fusion in remote sensing:
809 Concepts, methods and applications. *International Journal of Remote Sensing*,
810 19(5), 823-854.
811 Propastin, P. (2012). Modifying geographically weighted regression for estimating
812 aboveground biomass in tropical rainforests by multispectral remote sensing
813 data. *International Journal of Applied Earth Observation and Geoinformation*,
814 18, 82-90.
815 Pugh, S.A. & R.G. Congalton (2001). Applying spatial autocorrelation analysis to
816 evaluate error in New England forest-cover-type maps derived from Landsat
817 Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*
818 67(5), 613-620.
819 Richards, J.A. (1984). Thematic mapping from multitemporal image data using the
820 principal components transformation. *Remote Sensing of Environment*, 16:35-
821 46.
822 Riemann, R., Wilson, B. T., Lister, A., & Parks, S. (2010). An effective assessment
823 protocol for continuous geospatial datasets of forest characteristics using USFS
824 Forest Inventory and Analysis (FIA) data. *Remote Sensing of Environment*, 114,
825 2337-2352.
826 Ripley, B. (2013). Feed-forward neural networks and multinomial log-linear
827 models,"nnet" package, version 7.3. 6. *URL http://www. stats. ox. ac.*
828 *uk/pub/MASS4*.
829 Soille P (2003). *Morphological Image Analysis: Principles and Applications*, Springer,
830 New York, NY, USA, 2nd edition.

831 Spiker, J. S., & Warner, T. A. (2007). Scale and spatial autocorrelation from a remote
832 sensing perspective. In *Geo-Spatial Technologies in Urban Environments* (pp.
833 197-213). Springer Berlin Heidelberg.
834 Steele, B. M., Winne, J. C., & Redmond, R. L. (1998). Estimation and mapping of
835 misclassification probabilities for thematic land cover maps. *Remote Sensing of*
836 *Environment*, 66(2), 192–202.
837 Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015). A critical
838 synthesis of remotely sensed optical image change detection techniques.
839 *Remote Sensing of Environment*, 160: 1-14.
840 Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the*
841 *Royal Statistical Society B*, 58, 267-288
842 Toutin, T. (2004). Review article: Geometric processing of remote sensing images:
843 models, algorithms and methods. *International Journal of Remote Sensing*,
844 *25*(10), 1893-1924.
845 Tsutsumida, N., Saizen, I., Matsuoka, M., & Ishii, R. (2013). Land cover change
846 detection in Ulaanbaatar using the breaks for additive seasonal and trend
847 method. *Land,* 2(4), 534-549.
848 Tsutsumida N. & Comber A.J. (2015). Measures of spatio-temporal accuracy for time
849 series land cover data. *International Journal of Applied Earth Observation and*
850 *Geoinformation*, 41: 46-55.
851 Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth
852 Edition. Springer, New York.
853 Vermote, E. F., S. Y. Kotchenova, & J. P. Ray. (2011) *MODIS surface reflectance user's*
854 *guide, MODIS Land Surface Reflectance Science Computing Facility*, version 1.3.

855 Lasaponara, R. (2006). On the use of principal component analysis (PCA) for
856 evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI
857 temporal series. *Ecological Modelling*, *194*(4), 429-434.
858 Wang, Q., Ni, J., & Tenhunen, J. (2005). Application of a geographically-weighted
859 regression analysis to estimate net primary production of Chinese forest
860 ecosystems. *Global Ecology and Biogeography*, *14*(4), 379-393.
861 Wang, R., Zeng, C., Li, P., & Shen, H. (2011). Terra MODIS band 5 Stripe noise
862 detection and correction using MAP-based algorithm. In *2011 International*
863 *Conference on Remote Sensing, Environment and Transportation Engineering*
864 *(RSETE)*, (pp. 8612-8615). IEEE.
865 Woodcock, C. E., Strahler, A. H., & Jupp, D. L. (1988). The use of variograms in
866 remote sensing: I. Scene models and simulated images. *Remote Sensing of*
867 *Environment*, 25(3), 323-348.
868 Xu, B., Gong, P., & Pu, R. (2003). Crown closure estimation of oak savannah in a dry
869 season with Landsat TM imagery: comparison of various indices through
870 correlation analysis. *International Journal of Remote Sensing*, 24(9), 1811-1822.
871

872 **Appendix: Computation times**

873 The main computational cost associated with GWPCA-derived input variables in the
874 classification is in selecting the GWPCA bandwidth. For this study, we investigated 21
875 bandwidths, as detailed in Section 3.4. On a medium specs laptop (Intel® core ™ i7-
876 4600U CPU @ 2.10Ghz to 2.70GHz with 16.0 GB using a 64-bit OS), bandwidth
877 selection took 54 minutes and 47 seconds. Each of the 21 runs took from 142 to 175
878 seconds to complete, where longer runs tended to be for the larger bandwidths. The
879 run with the chosen bandwidth of 20% took 149 seconds to output *all* of this study's

880 results. Breaking this particular run down further, the classification comparisons
881 took: (i) 27 seconds using only input variable group 1 (image data), and (ii) 95
882 seconds, using only input variable group 4 (image plus all GWPCA-derived input
883 variables), where 60 of the 95 seconds were used to calculate the GWPCA-derived
884 input variables. Thus, for this particular case study, additional computational
885 complexity is not an issue. To transfer the GWPCA-based classification methodology
886 to larger data sets, immediate computational costs could be reduced by not
887 conducting all of this study's comparisons (i.e. choose only one classifier and only the
888 fourth input variable group). Furthermore, work is at an advanced stage in
889 developing more efficient GWPCA code, incorporating both mathematical and
890 hardware solutions. Judged short-cuts in the key step of bandwidth selection could
891 also be employed, for example using thinned but spatially-representative image
892 data.
893