Landscape-scale Remote Sensing and Classification of Pond and Lake Habitat Diversity in a Tropical City

Hoong Chen Teo (hcteo@u.nus.edu)
National University of Singapore https://orcid.org/0000-0003-0127-978X

Matthew J. Hill
University of Huddersfield

Alex M. Lechner
University of Lincoln

Fang Yenn Teo
University of Nottingham - Malaysia Campus

Christopher N. Gibbins
University of Nottingham - Malaysia Campus

Research Article

Keywords: Freshwater conservation, land cover mapping, lentic, pond network, urban, wetland

Posted Date: May 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-468321/v1

Landscape-scale remote sensing and classification of pond and lake habitat diversity in a tropical city

Hoong Chen Teo[1,2,3], Matthew J. Hill[4], Alex M. Lechner[1,5], Fang Yenn Teo[6], and Christopher N. Gibbins[1]

Journal: Wetlands

[1] School of Environmental and Geographical Sciences, University of Nottingham Malaysia, Jalan Broga, 43500, Semenyih, Selangor, Malaysia. Email: Christopher.Gibbins@nottingham.edu.my

[2] Centre for Nature-based Climate Solutions, National University of Singapore, 14 Science Drive 4, 117543, Singapore. Email: hcteo@u.nus.edu

[3] Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117543, Singapore.

[4] School of Applied Sciences, University of Huddersfield, Huddersfield, HD1 3DH, UK. Email: m.hill@hud.ac.uk

[5] Lincoln Centre for Water and Planetary Health, School of Geography, University of Lincoln, Lincoln, LN6 7TS, UK. Email: alechner@lincoln.ac.uk

[6] Faculty of Science and Engineering, University of Nottingham Malaysia, 43500, Semenyih, Selangor, Malaysia. Email: fangyenn.teo@nottingham.edu.my

Author for correspondence

Hoong Chen TEO

Centre for Nature-based Climate Solutions

Department of Biological Sciences

National University of Singapore

14 Science Drive 4, Singapore 117543

E-mail: hcteo@u.nus.edu

ORCID IDs

Hoong Chen Teo (0000-0003-0127-978X), Matthew J. Hill (0000-0001-8008-2197), Alex M. Lechner (0000-0003-2050-9480), Fang Yenn Teo (0000-0002-5529-1381)

Abstract

Ponds and lakes are common freshwater habitats in urban landscapes, and often have a high biodiversity and conservation value. The importance of landscape-scale conservation of pond networks has recently been recognised, yet the categorisation and classification of pond network spatial structures is missing. Developing spatial methods and tools to characterise and understand pond networks is a critical first step to accurately conserve pond habitats and inhabiting species. This paper presents an inventory of ponds and lakes in Greater Kuala Lumpur, Malaysia, characterising their distribution, abundance and type. Remote sensing was first employed to map and characterise these habitats, followed by multivariate cluster analysis to classify and develop a typology of the ponds identified. Physicochemical data was collected from a sample (n=60) of ponds to compare with the remotely sensed pond classification. Results demonstrated that multi-source remote sensing can be highly accurate and effective in inventorying ponds and lakes of varying sizes. A total of 1013 ponds and lakes were identified within the Greater Kuala Lumpur region and were found to be highly environmentally heterogeneous. Typology clusters were driven by land cover rather than local physicochemical variables demonstrating that specific remotely-sensed variables may be sufficient proxies for certain chemical variables. Landscape-scale conservation and management of pond networks should utilise remote sensing tools, to establish pond network structure, and to maintain wide environmental heterogeneity among pond habitats. Incorporating remote sensing tools into pond conservation will ensure that effective pond conservation can be achieved and biodiversity protection can be maximised.

Keywords: Freshwater conservation, land cover mapping, lentic, pond network, urban, wetland

1. Introduction

Urbanisation typically poses a severe threat to natural ecosystems by driving their removal, degradation and biotic homogenisation (McKinney and Lockwood, 2001; Seto et al., 2013; Zheng et al., 2021), resulting in the loss of biodiversity within urban freshwater and terrestrial habitats (McKinney, 2006). However, while natural habitats are being lost or fragmented as a result of urban development, novel urban ecosystems are being created from new landscape features (e.g., stormwater retention ponds and ornamental ponds) or from significantly modified natural habitats (Jonsson et al., 2011; Kowarik, 2011; Southwood, 1988). Historically, the biodiversity of urban ecosystems such as ponds and lakes has been largely ignored, but recent research has demonstrated that these habitats can be an important source of ecosystem services and biodiversity in urban landscapes (Elmqvist et al., 2015; Hill et al., 2021).

While urbanisation has caused a decline in biodiversity in terrestrial and lotic habitats (Knop, 2016; Roy et al., 2003), the negative effects of urbanisation on pond communities are less evident, with biodiversity among urban ponds habitats reported to be similar to non-urban ponds (Hassall and Anderson, 2015; Hill et al., 2017a). Individually, ponds can support highly variable biodiversity, but when considered together at a landscape-scale (i.e. pond networks), ponds have been demonstrated to support a higher biodiversity than other freshwater habitats, such as rivers, lakes and ditches (Davies et al., 2008). The prevalence of urban ponds and their wide environmental heterogeneity indicates that complex pond networks may exist in urban areas, and can encourage high biodiversity at a landscape-scale (Hassall, 2014; Hill et al., 2018).

Most studies on urban ponds are limited in geographic extent, based in the manual identification of study sites (typically no more than a few dozen) from remote sensing imagery (De Marco et al., 2014; Hill et al., 2015; Jeffries, 2008). However, ponds contribute most to freshwater biodiversity at the landscape-scale, as many ecological processes (such as dispersal) operate at larger scales and therefore there is a need for their conservation, management and research to be undertaken at landscape-scales

(Hill et al., 2018; Jeffries, 2008). This paper aims to present a new method for the landscape-scale characterisation of pond habitats in order to provide a foundation for landscape-scale pond conservation in large cities, using Greater Kuala Lumpur (GKL), Malaysia as a case study. GKL is especially interesting; like most of Southeast Asia, it is found within one of the global biodiversity hotspots (Lechner et al., 2020; Woodruff, 2010). Assessments are critical in cities in the Global South, because the rapid pace of development is driving pond loss, eroding diversity and threatening ecosystem service provision to degrees not currently seen in urban areas in the Global North. Furthermore, research on urban ponds in countries in the Global South is underrepresented in the academic literature (Oertli and Parris, 2019) and urban ponds may be overlooked by decision makers as these countries are often data poor (Farrell, 2017; Lourdes et al., 2021; Teo et al., 2020a).

With urban land cover expected to increase in the future (Seto et al., 2012), the recognition of urban landscapes, particularly urban ponds and lakes, for biodiversity conservation is growing. This study identifies and maps ponds at the landscape-scale to characterise their abundance, distribution and type. Understanding the distribution, abundance and ecology of different pond and lake types present in a pond network can guide conservation prioritisation (Arponen et al., 2008; Trakhtenbrot and Kadmon, 2005). At the scale of a large city, satellite remote sensing approaches may provide the most effective method to acquire spatial and spectral data to produce a comprehensive inventory (Lopez et al., 2013). Open access multispectral satellite imagery such as Landsat and Sentinel have sufficient radiometric resolution to identify ponds and lakes by spectrally separating water from other land covers; identification this way is typically straightforward and can by highly accurate (Verpoorter et al., 2014). High temporal resolution is only necessary if ephemeral ponds are prevalent, reflecting their hydrological variability. In this paper, ponds are defined as standing waterbodies between 1 m^2 to 2 ha (Biggs et al., 2005), with any water body larger than 2 ha considered to be a lake. Spatial resolution is a key challenge in freshwater pond habitat identification, as larger pixels will increase the minimum mappable unit (MMU) at which an object can be reliably identified. The MMU is theoretically 4 pixels but in practice can be 11-12 pixels (Lechner et al., 2009). At pixels 30m or larger, small or medium-

sized ponds will be excluded, which potentially leads to underestimation of pond abundance and environmental heterogeneity. Thus, the priority for the remote sensing methods applied in this paper is to increase spatial resolution to detect smaller ponds, by applying pansharpening and using multi-sensor data.

Classifications and typologies are frequently used in ecology, such as to characterise ecological communities, through cluster analysis of continuous abiotic and biotic multivariate data to identify ecological mapping units which share similar characteristics. Clusters identified from abiotic and biotic data are likely to share similar ecological processes and controls, and similar biodiversity (Burress et al., 2016; James and McCulloch, 1990), facilitating targeted management. Habitat types/ecological mapping units can be classified in several ways, based on their origin/purpose, and biological or environmental variables (de Cáceres and Wiser, 2012; Lechner et al., 2016). In GKL, while some pond and lake (hereby collectively referred to as 'wetlands') habitats have a clear origin or purpose (such as golf course ponds or reservoirs), others may have unclear or overlapping origins. Typing using origin or biological data is likely to be unfeasible for large areas or wetland numbers, however typologies based on environmental variables acquired from remote sensing data offers significant potential for the rapid classification of large numbers of wetlands. Remote sensing can provide multivariate data on wetland attributes, surrounding land cover, and spectral information, and so provides a robust basis for cluster analysis. Although physical variables can be generated from remote sensing data, and inferences can be made about certain chemical variables (such as turbidity), most chemical variables can only be analysed from water samples acquired from fieldwork. Physical and chemical variables provide a direct indicator of wetland water quality, and their environmental heterogeneity has been demonstrated to be important drivers of patterns in their biodiversity (Bagella et al., 2010; Gioria et al., 2010; Sun et al., 2018). Thus, comparing physicochemical data collected from a subset of the wetlands across GKL, with remote sensing data may provide important insights and determine the suitability of remote sensing data to accurately classify pond types environmentally and capture environmental heterogeneity.

This study characterises the distribution, abundance and types of ponds and lakes (wetlands) present across a rapidly expanding tropical city (Greater Kuala Lumpur: GKL). Specifically, this study (1) uses remote sensing to identify, map and physically characterise pond and lake habitats at the scale of the whole city, (2) applies a cluster analysis to develop a typology of these habitats and provide the basis for assessing spatial patterns in wetland types that might reflect major underlying controls, related to natural or man-made factors and (3) collect physicochemical data for a sample of wetlands (n=60) spanning all clusters/types to assess whether remotely-sensed physical characteristics can predict chemical characteristics.

2. Methods

2.1. Study area and overall approach

The study region of GKL covers and area 2,950 km^2 and has a tropical climate with a mean annual temperature of 28°C and a mean annual precipitation of 2500-3000 mm (Meterologi Malaysia, 2020). Due to high and regular rainfall, most ponds and lakes in the city are perennial. Although lentic habitats exist naturally in the tropics (Low et al., 2016; Sharip and Jusoh, 2010), the vast majority of lentic habitats in GKL are anthropogenic in origin. These include tin mining pools and granite quarries remaining from historical mining activities (Arumugam, 1994), and ponds and lakes created in more recent decades for purposes such as water storage, stormwater retention, aesthetic, recreation and biodiversity (Caddis et al., 2012).

The approach was to obtain data at two spatial scales to map and characterise ponds. The first step involved remote sensing which, via image classification and ordination analyses, was used to identify and map ponds across the whole of GKL. The second step was to produce a typology of all the ponds identified, using features extracted from the remote sensing data. Thirdly, ponds that were too small to be detected from satellite data were identified, by manually checking high resolution maps. Together steps one and three provided an estimate of the total number of ponds across the study area. The fourth

step involved collecting field data for a sample of ponds from each of the types identified in step 2. For these sample ponds, satellite and field data were used together to help understand how individual ponds and the pond types differed in their environmental condition. These steps are outlined in detail below.

2.2. Remote sensing mapping of ponds

Google Earth Engine (GEE) was used to generate Sentinel-2 visual and infrared (IR) bands for the study area for 2018. The images were cloud masked and 20m bands (B5, 6, 7, 8A, 11, 12) pansharpened to 10m, using the average of the 10m bands (B2, 3, 4, 8) as a simulated panchromatic band (Kaplan, 2018). However, the only existing pansharpening method available in GEE, the Hue-Saturation-Value (HSV) spectral transformer, can only pansharpen three bands at a time and is unsuitable for pansharpening multiple bands. To overcome this, a novel implementation of a high pass filter (HPF) resolution merge technique (Gangkofner et al., 2007) was created in GEE (see code in Supplementary Material A). Next, to improve classification performance, the Normalised Difference Vegetation Index (NDVI, $\frac{B8a - B4}{B8a + B4}$), Normalised Difference Water Index (NDWI, $\frac{B8a - B11}{B8a + 11}$), Normalised Difference Built-up Index (NDBI, $\frac{B11 - B8}{B11 + B8}$) and Bare Soil Index (BSI, $\frac{(B11 + B4) - (B8 + B2)}{(B11 + B4) + (B8 + B2)}$) were generated and stacked together with the IR bands. Finally, high-resolution Airbus Pleiades imagery for the study area was downloaded from the online mapping service, Yandex Maps, available at 2.4m, were then imported into GEE to further pansharpen the Sentinel-2 image to 2.4m resolution using the method described above.

Training areas were selected with the aid of higher resolution (50 cm) World View 3 reference data from Google Earth, Street View and personal expertise. In GEE, a random forest classifier was trained and used to classify land cover (Table 1). A roads mask was added using data from OpenStreetMap. Standing water bodies were manually identified by eliminating rivers, then manually edited to reflect the correct spatial extent. Aquaculture ponds were excluded from the analysis as they are highly managed and manipulated (e.g. water level drawdowns, chemical treatments related to fish husbandry),

and are stocked with a high densities of fish. Finally, a raster land cover map and a vector ponds shapefile were produced.

An accuracy assessment was conducted for the GKL land cover map using 50 random points per class. Furthermore, to provide an estimate for the number of smaller ponds either too small to be reliably detected or completely undetected by remote sensing classification, 50cm resolution World View 3 images in Google Earth were used to manually count the number of undetected ponds over a sample of areas across GKL. This sample of areas consisted of 24 evenly spaced points across GKL each buffered by a 2 km radius.

2.3. Pond metric selection

Using the land cover map and ponds areal extents, various environmental and spatial metrics were extracted for each pond. Eight explanatory variables, representing five environmental characteristics of ponds, which could be mapped with remote sensing were assessed: dimension, shape, proximity, surrounding land cover and sediment (Table 2). A correlation matrix (Fig. 1 and Table 3) was used to investigate relationships between metrics. Area and perimeter were highly correlated (Pearson's r = 0.96, $p < 0.01$) and as a result only area was selected for further analysis. Perimeter-Area Ratio (PAR), shape index and fractal dimension are all measures of shape complexity, so using only one of these helped reduce metric redundancy (Cushman et al., 2008). Since PAR and fractal dimension proved highly correlated (Pearson's $r = 0.79$, $p < 0.01$), and the Shape Index is merely a normalised PAR, to avoid the size-dependency problem of PAR (Forman and Godron, 1986), the Shape Index was selected to represent pond shape complexity. A measure of pond proximity is typically recorded as the number of water bodies within 500 m for smaller-scale studies (Waterkeyn *et al.*, 2008; Hill, Heino *et al.*, 2017). However, in the present study the mean distance to the nearest three ponds was used as the analysis is at a larger scale and many water bodies are located further than 500 m from their neighbours. To represent the surrounding land cover, percentage vegetation, soil, water and impervious land cover

values within 100m, 250m, 500m and 1km of each pond were generated and tested for correlation. Of these, the characteristics at the 100m and 1km buffer distances proved least correlated (Table 3) and so these two were used for further analyses; these are useful as representing characteristics and possible anthropogenic influences over more at proximate and larger areas. The variables were reduced using a PCA and are represented by component 1. The final list of metrics used for clustering is presented in Table 4. All statistical analyses were conducted with R 3.5.3 (R Core Team, 2019).

2.4. Clustering of ponds

Pond clustering (n=1013) was based on the final set of remotely-sensed metrics and was performed using three methods – hierarchical clustering, k-means and partitioning around medoids (PAM). Three clustering validation measures (connectivity, Dunn index, silhouette width) were generated for cluster numbers from 5 to 10 to identify the clustering method and number of clusters that would best differentiate between ponds and generate distinct clusters. Since connectivity should be minimised, and both Dunn index and silhouette width should be maximised, connectivity was multiplied by -1, and each clustering validation measure was normalised (Table 5). PAM with 6 clusters delivered the highest values and was used for the main clustering. Using a Bray-Curtis dissimilarity matrix, an analysis of similarities (ANOSIM) was conducted to determine whether similarity within clusters is greater than similarity between clusters. In addition, similarity percentages (SIMPER) analysis was undertaken to determine the most important contributors to differences between clusters. ANOSIM and SIMPER analyses used the vegan (Oksanen et al., 2019) and clValid (Brock et al., 2020) packages in R.

2.5. Field data

In total, 60 ponds (10 from each cluster) were sampled between 18 Dec 2019 and 22 Jan 2020. The choice of which pond to sample was governed by the need to include ponds from across the whole of GKL, so that a wide range of environmental conditions across different areascould be incorporated. Water samples were taken from three different locations in each pond, at around 5m from the shore.

From the water samples, turbidity (NTU) was recorded using a turbidity meter (Hach 2100Q), while conductivity (µS/cm), pH and dissolved oxygen (mg/L) were recorded using a multiparameter probe (YSI Pro Plus). Average values for turbidity, conductivity, pH and dissolved oxygen were calculated from the three samples. The following variables were visually estimated for each pond: percentage of pond margin shaded (PMS); percentage surface covered by emergent, floating or submerged macrophytes (EM, FM, SM) and; percentage of bank made from bare earth, wood, concrete or stone material. Prior to ordination, fieldwork variables for macrophyte coverage (EM, SM, FM) and bank material (earth, concrete, wood, stone) were reduced to the 1st principal component (PC) respectively, similar to the land cover variables as described in Section 2.3. Pearson's correlation was used to compare variables for redundancy, with one from each pair of correlated variables removed.

2.6. Comparison and clustering of remote sensing and fieldwork variables

Ordinations based on (i) data extracted from satellite images and (ii) data collected from field work were compared using co-inertia analysis (Co-IA). Co-IA compares simultaneous features in two sets of multivariate ordinated data and determines to what extent the two datasets as a whole are co-occurring (Dray et al., 2003).

Cluster analysis was conducted on the sampled ponds (n=60). Clustering validation measures were computed as described in Section 2.4, but using only 3-6 clusters due to the smaller number of ponds. Three cluster scenarios with different variable combinations were run. Firstly, for remote sensing variables and fieldwork variables of sampled ponds, a 4-cluster k-means model performed best (Table 6a). Secondly, for remote sensing variables of sampled ponds only, a 3-cluster hierarchical model performed best (Table 6b). Thirdly, for fieldwork variables of sampled ponds only, a 4-cluster k-means model performed best (Table 6c). Bray-Curtis dissimilarity matrices, ANOSIM and SIMPER analyses were conducted for each clustering scenario, following the procedure outlined in Section 2.4.

3. Results

3.1. Pond and lake numbers

The land cover map of GKL (2,950 km^2) derived from remote sensing achieved an overall classification accuracy of 82.0% (Fig. 2; see Table B1 in Supplementary Material B for confusion matrix). In total, 1271 standing waterbodies were identified; of these, 258 aquaculture ponds were excluded, leaving 1,013 ponds and lakes (Fig. 3a). The smallest pond was 0.0819 ha (819 m^2) and the largest lake was 418 ha (4.18 km^2). Based on the number of smaller undetected ponds per km^2 within the 24 sampled areas, it is estimated that another 440 ± 243 (90% CI) smaller undetected ponds may exist over the entire GKL (Fig. 3b). These values suggest that the satellite-based analysis detected 74.3% (90% CI: 65.0%-86.6%) of all ponds likely to exist in GKL. No clear spatial pattern in smaller undetected ponds could be discerned.

3.2. Classification of ponds

A 6-cluster PAM model was found to perform best for clustering all 1013 ponds using the remote sensing physical variables. Overall, similarity within clusters was greater than similarity between clusters (ANOSIM R=0.64, p=0.001), indicating that the clusters are distinct. The clusters were mostly separated based on land cover, as supported by the PCA ordination biplot (Fig. 4) and SIMPER analysis (Table 7).

Characteristics of each cluster (based on a sample of 60 ponds, with both remote sensing and field data), are described in Tables 8 and 9. Detailed descriptions of each pond cluster are presented in the Supplementary Material (Supplementary Material C). Although all variables show considerable overlap between clusters, one-way ANOVAs showed significant differences between clusters for all remote sensing variables, as well as PMS, temperature and DO (Table 8). Overall, clusters 1-3 represent ponds surrounded by vegetation while clusters 4-6 represent those surrounded impervious material. Red band values are lower in ponds from clusters 1-3 (M=0.082, SD=0.027) than clusters 4-6 (M=0.096,

286 SD=0.033) (t-test: t(1011)=-7.49, p<0.001). Turbidity is also lower in ponds from clusters 1-3 (M=27.0,
287 SD=5.2) than clusters 4-6 (M=44.1, SD=6.6) (t-test: t(29)=-2.22, p=0.034). All clusters have a large
288 range of sizes and all contain smaller ponds; on the upper end of the range of sizes, clusters 1-2 include
289 sizes up to the largest lakes, clusters 3-4 include sizes up to small lakes, cluster 5 includes sizes up to
290 large lakes, and cluster 6 includes sizes up to medium lakes.

291

292 The spatial distribution and abundance of each cluster is shown in Fig. 5 and Table 10. Peri-urban and
293 suburban areas of Selayang, Shah Alam and Sepang have a greater diversity of pond types than the
294 urban core of the city (Kuala Lumpur, Petaling, Subang), natural forested areas (northern Kajang), rural
295 areas (northern Klang) or the industrial and port areas (southern Klang). Cluster 1 and 2 are mostly
296 absent from the urban core and largely recorded near peri-urban and suburban areas. Cluster 3 can be
297 found in the urban core, suburban and peri-urban areas. Cluster 4 is predominantly found in the peri-
298 urban areas. Clusters 5 and 6 are mostly found in the urban core and suburban areas. Clusters 3, 5 and
299 6 are found in all districts.

300

301 *3.3. Comparison of remote sensing and fieldwork data*

302 Using the same clusters assigned to each pond from the 6-cluster PAM model for remote sensing
303 variables, similarity within clusters for the fieldwork variables were not found to be greater than
304 similarity between clusters (ANOSIM R=0.019, p=0.25). Co-inertia analysis between the PCA
305 ordinations of remote sensing variables and fieldwork variables resulted in an RV coefficient of 0.22
306 (Fig. 6), indicating that the fieldwork variables did not cluster into as similar groups as the remote-
307 sensing variables. However, the remote sensing-derived red band values are positively correlated with
308 turbidity and appear to be an acceptable proxy (Table 11). pH is positively correlated with area and
309 shape index, and negatively correlated with the red band value. DO is negatively correlated with mean
310 distance to the 3 nearest ponds.

311

3.4. Ordination and clustering of all variables for sampled ponds

Using the same clusters assigned to each pond from the 6-cluster PAM model for remote sensing variables, similarity within clusters for all (remote sensing and fieldwork) variables were not found to be greater than similarity between clusters (ANOSIM R=-0.033, p=0.84). However, the PCA biplot suggests that although clusters overlap significantly, clusters 1-3 are generally separated from clusters 4-6, with this separation occurring along land cover, median(RED), turbidity and the percentage of pond margin shaded (Fig. 7a). A new clustering of all variables (remotely sensed and fieldwork) for sampled ponds (n=60) using the best model, a 4-cluster k-means model (Fig. 7b), generated significantly distinct clusters (ANOSIM R=0.65, p=0.001). A SIMPER analysis suggested that the k-means clusters are mostly separated by turbidity and conductivity (Table 12). All variables were subjected to an ANOVA to determine whether clusters were significantly different; clusters only displayed statistically significant differences for median(RED), turbidity and conductivity (Table 13).

Based on the above results, a new clustering of remote sensing variables for sampled ponds (n=60) using the best model, a 3-cluster hierarchical model (Fig. 8a), generated clusters more similar within than between clusters (ANOSIM R=0.86, p=0.001). A SIMPER analysis suggested that the clusters are now mostly separated by land cover (Table 14), which is comparable to the separation of clusters of remote sensing variables for all ponds (Table 9). Clustering of fieldwork variables for sampled ponds (n=60) using the best model, a 4-cluster k-means model (Fig. 8b), generated clusters more similar within than between clusters (ANOSIM R=0.66, p=0.001). A SIMPER analysis suggested that the clusters are now mostly separated by turbidity and conductivity (Table 15), just as the clusters of all variables for sampled ponds (Table 9).

4. Discussion

Surveying and assessing natural habitats in urban landscapes is challenging as they tend to be fragmented and small in size (Holgerson and Raymond, 2016). Landscape-scale urban pond conservation is increasingly recognised as important, but this requires methods for spatial data acquisition and analysis suited to this scale. To date, almost no attention has been given to how best to address the challenge of surveying urban wetlands. Furthermore, our methods can address data and research gaps in Global South countries (Oertli and Parris, 2019), which are often data poor, with small urban ponds not commonly captured in existing land cover datasets and/or those datasets may be out of date due to the rapid pace of urbanisation in the global south (Farrell, 2017; Lourdes et al., 2021). While in the Global North, medium-large urban ponds may already be captured in existing topographical maps which are typically at produced at much higher spatial resolutions. The UK Ordnance Survey topographical maps are available at 1:1250 scale, which is the equivalent of 12.5 m spatial resolution (UK Ordnance Survey, 2021).

Globally, pond numbers have been estimated to be between 547 million – 3.19 billion globally, demonstrating the uncertainty surrounding pond numbers at national and international scales (Holgerson and Raymond, 2016). Limited attempts have also been made to quantify pond numbers at regional or national scales (Al Sayah et al., 2019; Lacaux et al., 2007), with similar high levels of uncertainty, reflecting the spatial resolution of remote sensing and identification of the smallest ponds. However, this study has demonstrated how a multi-source remote sensing and classification approach to inventorying urban ponds and lakes at the scale of a large metropolitan area can be highly accurate and conducted with open-access data. This study successfully captured most ponds (>74% with a spatial resolution of 2.4m) likely to exist in GKL, except some of the smallest ones and demonstrated the utility of a multi-source data fusion approach in enabling a higher spatial resolution and accuracy to comprehensively inventorise lentic features of varying sizes. The creation of pond inventories is critical to understanding their spatial and landscape dynamics, and facilitates the targeted sampling of a manageable number of these features.

This study also demonstrated how multi-source data can facilitate the use of spectral information from satellite imagery to characterise the environmental conditions of pond habitats. Pond networks have been widely documented to have high environmental heterogeneity at the landscape-scale, albeit this is based on research typically at small geographical scales (Scheffer *et al.*, 2006; Thornhill et al. 2017). The wide environmental heterogeneity among pond networks has been identified as the key driver of high biodiversity recorded among these habitats (Davies et al., 2008; Williams et al., 2004). As a result, there have been calls for pond conservation to focus on networks of 'pondscapes' rather than individual ponds, to maximise environmental heterogeneity and biodiversity (Hassall et al., 2012; Hill et al., 2016). The inventory and analysis of 1013 ponds and lakes across a large metropolitan area provides robust evidence of the environmental heterogeneity of urban lentic habitats and may facilitate the identification of pond networks that maximise environmental heterogeneity at larger spatial scales.

Landscape-scale pond management also requires analytical methods for large datasets. Cluster analysis spatial data can provide a classification of pond types, which can yield insights into pond environmental characteristics at landscape-scale and guide pond management. Remote sensing variables provide clusters separated primarily by land cover rather than patch metrics, suggesting that land cover is a stronger driver of clustering than patch metrics. As a whole, physicochemical variables derived from fieldwork do not group into the same clusters as remote sensing variables, but specific chemical variables appear to be moderately correlated with specific remote sensing variables, which may be used as a proxy. Turbidity and conductivity are the most dominant factors driving clustering when fieldwork variables are clustered alone, and when remote sensing variables are grouped together with fieldwork variables. Conductivity and turbidity have been shown to be primary drivers of freshwater diversity and composition (Pérez-Quintero, 2011; Svitok et al., 2016), with species often demonstrating specific conductivity preferences (Boets et al., 2013; Coviaga et al., 2018). By classifying the variability in these variables at larger-scales practitioners can identify pond areas where gradients of important environmental conditions can be maintained. Also, the identification of a pond typology through cluster

analysis and the subsequent mapping of those clusters can serve as the basis for landscape-scale conservation, creation and management aiming to maintain ponds of variable environmental conditions across the landscape. For example, the data can be interrogated to identify pond clusters where the surrounding land cover have the highest levels of anthropogenic disturbances to target for rehabilitation. Alternatively, pond clusters that represent areas surrounded by natural land cover could be targeted for restoration. Summarising complex environmental pond data through the statistical analyses represented in this paper is critical step to support decision makers and avoid manually and subjectively assessing 1013 ponds.

This study has also shown that urban structure can be an important driver of pond environmental heterogeneity, since peri-urban and suburban areas supported the greatest number of pond types, compared to urban, rural, or forested landscapes. Pond diversity is crucial to the overall health of pond ecosystems at landscape-scale, providing a wide range of environmental niches for taxa to exploit, facilitating dispersal and colonisation and acting as buffers against stochastic events affecting individual ponds (Scheffer et al., 2006). Results from this study suggest that cities with larger peri-urban and suburban areas can support more diverse pond types, so pond management needs to consider the opportunities and constraints arising from urban structure and development dynamics. In GKL, one cluster (3) with more natural influence can be found in all districts and across the urban core, suburban and peri-urban areas. This suggests that there are blue-green spaces across all areas of GKL that can provide natural spaces and corridors for dispersal, although this is heavily dependent on species and obstacles (e.g., high rise developments) present.

The predominant pond management paradigm in developed cities focuses strongly on preventing the further loss of ponds, with urbanisation viewed as driving the disappearance of these habitats (Hassall, 2014; Jeanmougin et al., 2014; Wood et al., 2003). This is less applicable to cities like GKL, where ponds are mostly man-made and abundant. This difference reflects local societal needs and the provision of ecosystem services (e.g., flood alleviation strategies) that have spurred their creation,

especially in rapidly growing peri-urban and suburban areas. Novel habitats, such as urban ponds, have been shown to make an important contribution to urban biodiversity due their high spatiotemporal environmental heterogeneity (Kowarik, 2011; McKinney and Lockwood, 2001; Pyšek, 1993).. However, urban habitats are no substitute for natural habitats, since natural habitats support more native species, and typically record higher biodiversity especially in non-plant taxonomic groups (Hansen et al., 2005). Where urbanisation is unavoidable, man-made novel urban ecosystems can still maximise biodiversity and ecosystem services benefits to urban areas, and form part of nature-based solutions for sustainable cities (Lechner et al., 2020; Teo et al., 2021, 2020b). However, in the context of tropical cities, urban ponds have perhaps a more important role than other parts of the world, as tropical countries tend to also be high in biodiversity and therefore these ponds may make important contributions to freshwater biodiversity conservation. Furthermore, urban ponds provide a vital role in ecosystem service provision (hence the sheer number recorded in our GKL case study), mitigating the impacts of extremes in climate such as high and intense rainfall and urban heat island, as well as providing cultural ecosystem services such as semi-natural space for exercise and relaxation in cities, in landscapes where large areas blue and green spaces have being rapidly lost to development (Drillet et al., 2020; Nath et al., 2018).

5. Conclusion

This study has highlighted the application of multi-source remote sensing for the accurate mapping and typing of lentic habitats across an urban landscape. It is one of very few studies of its kind, and has two important outcomes. The first is that the approach provides a robust way of identifying pond network spatial structure, and hence provides the evidence-base needed for the maintenance and management of environmental heterogeneity at large spatial scales. The second it that the methods used in this study provides an accurate inventory of urban lentic habitats. The satellite and follow-up map-based analysis suggest that approximately 1700 wetlands exist across GKL, along with another 258 aquaculture ponds. By quantifying the number and distribution of ponds and lakes in major urban cities, an accurate assessment of the freshwater resource can be determined, enabling practitioners to ensure that

ecosystem services (e.g., wellbeing and stormwater collection) can continue to be provided by these freshwater habitats to urban communities, urban biodiversity can be maximised (through targeted conservation and management), and areas can be identified where the creation of urban freshwater habitats are needed to support communities and biodiversity.

Declarations

Funding

This work was supported by the Malaysian Ministry of Higher Education Fundamental Research Grant Scheme (FRGS) FRGS/1/2019/WAB13/UNIM/01/1.

Conflicts of interest

No conflicts of interest are declared.

Availability of data and material

Data and material are available in the supplementary. Additional data available upon request from the authors.

Code availability

Code is available in the supplementary.

Author's contributions

HCT, MJH, AML, FYT and CNG conceptualised the study, designed the methods, reviewed and edited the manuscript. HCT performed the data collection and wrote the initial manuscript. CNG, AML and MJH supervised the study. All authors read and approved the final manuscript.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

References

Al Sayah, M.J., Nedjai, R., Kaffas, K., Abdallah, C., Khouri, M., 2019. Assessing the Impact of Man–Made Ponds on Soil Erosion and Sediment Transport in Limnological Basins. Water 11, 2526. https://doi.org/10.3390/w11122526

Arponen, A., Moilanen, A., Ferrier, S., 2008. A successful community-level strategy for conservation prioritization. J. Appl. Ecol. https://doi.org/10.1111/j.1365-2664.2008.01513.x

Arumugam, P.T., 1994. Present utilization and recommendations for management of tin-mine lakes in Malaysia. SIL Commun. 1953-1996 24, 265–272. https://doi.org/10.1080/05384680.1994.11904047

Bagella, S., Gascón, S., Caria, M.C., Sala, J., Mariani, M.A., Boix, D., 2010. Identifying key environmental factors related to plant and crustacean assemblages in Mediterranean temporary ponds. Biodivers. Conserv. 19, 1749–1768. https://doi.org/10.1007/s10531-010-9801-5

Biggs, J., Williams, P., Whitfield, M., Nicolet, P., Weatherby, A., 2005. 15 Years of pond assessment in Britain: Results and lessons learned from the work of Pond Conservation, in: Aquatic Conservation: Marine and Freshwater Ecosystems. pp. 693–714. https://doi.org/10.1002/aqc.745

Boets, P., Lock, K., Goethals, P.L.M., 2013. Modelling habitat preference, abundance and species richness of alien macrocrustaceans in surface waters in Flanders (Belgium) using decision trees. Ecol. Inform. 17, 73–81. https://doi.org/10.1016/j.ecoinf.2012.06.001

Brock, G., Pihur, V., Datta, Susmita, Datta, Somnath, 2020. clValid: Validation of Clustering Results. R package version 0.6-7.

Burress, E.D., Holcomb, J.M., Armbruster, J.W., 2016. Ecological clustering within a diverse minnow assemblage according to morphological, dietary and isotopic data. Freshw. Biol. https://doi.org/10.1111/fwb.12710

Caddis, B., Nielsen, C., Hong, W., Anun Tahir, P., Teo, F.Y., 2012. Guidelines for floodplain development – a Malaysian case study. Int. J. River Basin Manag. 10, 161–170. https://doi.org/10.1080/15715124.2012.688750

Coviaga, C., Cusminsky, G., Pérez, P., 2018. Ecology of freshwater ostracods from Northern Patagonia and their potential application in paleo-environmental reconstructions. Hydrobiologia 816, 3–20. https://doi.org/10.1007/s10750-017-3127-1

Cushman, S.A., McGarigal, K., Neel, M.C., 2008. Parsimony in landscape metrics: Strength, universality, and consistency. Ecol. Indic. 8, 691–703. https://doi.org/10.1016/J.ECOLIND.2007.12.002

Davies, B., Biggs, J., Williams, P., Whitfield, M., Nicolet, P., Sear, D., Bray, S., Maund, S., 2008. Comparative biodiversity of aquatic habitats in the European agricultural landscape. Agric. Ecosyst. Environ. https://doi.org/10.1016/j.agee.2007.10.006

de Cáceres, M., Wiser, S.K., 2012. Towards consistency in vegetation classification. J. Veg. Sci. 23, 387–393. https://doi.org/10.1111/j.1654-1103.2011.01354.x

De Marco, P., Nogueira, D.S., Correa, C.C., Vieira, T.B., Silva, K.D., Pinto, N.S., Bichsel, D., Hirota, A.S.V., Vieira, R.R.S., Carneiro, F.M., de Oliveira, A.A.B., Carvalho, P., Bastos, R.P., Ilg, C., Oertli, B., 2014. Patterns in the organization of Cerrado pond biodiversity in Brazilian pasture landscapes. Hydrobiologia 723, 87–101. https://doi.org/10.1007/s10750-013-1695-2

Dray, S., Chessel, D., Thioulouse, J., 2003. Co-inertia analysis and the linking of ecological data tables. Ecology. https://doi.org/10.1890/03-0178

Drillet, Z., Fung, T., Leong, R., Sachidhanandam, U., Edwards, P., Richards, D., 2020. Urban Vegetation Types are Not Perceived Equally in Providing Ecosystem Services and Disservices. Sustainability 12, 2076. https://doi.org/10.3390/su12052076

Elmqvist, T., Setälä, H., Handel, S.N., van der Ploeg, S., Aronson, J., Blignaut, J.N., Gómez-Baggethun, E., Nowak, D.J., Kronenberg, J., de Groot, R., 2015. Benefits of restoring ecosystem services in urban areas. Curr. Opin. Environ. Sustain. 14, 101–108. https://doi.org/10.1016/j.cosust.2015.05.001

Farrell, K., 2017. The Rapid Urban Growth Triad: A New Conceptual Framework for Examining the Urban Transition in Developing Countries. Sustainability 9, 1407. https://doi.org/10.3390/su9081407

Forman, R.T.T., Godron, M., 1986. Landscape Ecology, 1st ed. Wiley, Chichester.

Gangkofner, U.G., Pradhan, P.S., Holcomb, D.W., 2007. Optimizing the High-Pass Filter Addition Technique for Image Fusion. Photogramm. Eng. Remote Sens. 73, 1107–1118. https://doi.org/10.14358/PERS.73.9.1107

Gioria, M., Bacaro, G., Feehan, J., 2010. Identifying the drivers of pond biodiversity: The agony of model selection. Community Ecol. 11, 179–186. https://doi.org/10.1556/ComEc.11.2010.2.6

Hansen, A.J., Knight, R.L., Marzluff, J.M., Powell, S., Brown, K., Gude, P.H., Jones, K., 2005. Effects of exurban development on biodiversity: Patterns, mechanisms, and research needs. Ecol. Appl. https://doi.org/10.1890/05-5221

Hassall, C., 2014. The ecology and biodiversity of urban ponds. Wiley Interdiscip. Rev. Water. https://doi.org/10.1002/wat2.1014

Hassall, C., Anderson, S., 2015. Stormwater ponds can contain comparable biodiversity to unmanaged wetlands in urban areas. Hydrobiologia 745, 137–149. https://doi.org/10.1007/s10750-014-2100-5

Hassall, C., Hollinshead, J., Hull, A., 2012. Temporal dynamics of aquatic communities and implications for pond conservation. Biodivers. Conserv. https://doi.org/10.1007/s10531-011-0223-9

Hill, M.J., Biggs, J., Thornhill, I., Briers, R.A., Gledhill, D.G., White, J.C., Wood, P.J., Hassall, C., 2017a. Urban ponds as an aquatic biodiversity resource in modified landscapes. Glob. Chang. Biol. 23, 986–999. https://doi.org/10.1111/gcb.13401

Hill, M.J., Hassall, C., Oertli, B., Fahrig, L., Robson, B.J., Biggs, J., Samways, M.J., Usio, N., Takamura, N., Krishnaswamy, J., Wood, P.J., 2018. New policy directions for global pond conservation. Conserv. Lett. 11, 1–8. https://doi.org/10.1111/conl.12447

Hill, M.J., Heino, J., Thornhill, I., Ryves, D.B., Wood, P.J., 2017b. Effects of dispersal mode on the environmental and spatial correlates of nestedness and species turnover in pond communities. Oikos 126, 1575–1585. https://doi.org/10.1111/oik.04266

Hill, M.J., Mathers, K.L., Wood, P.J., 2015. The aquatic macroinvertebrate biodiversity of urban ponds in a medium-sized European town (Loughborough, UK). Hydrobiologia 760, 225–238. https://doi.org/10.1007/s10750-015-2328-8

Hill, M.J., Ryves, D.B., White, J.C., Wood, P.J., 2016. Macroinvertebrate diversity in urban and rural ponds: Implications for freshwater biodiversity conservation. Biol. Conserv. 201, 50–59. https://doi.org/10.1016/j.biocon.2016.06.027

Hill, M.J., White, J.C., Biggs, J., Briers, R.A., Gledhill, D., Ledger, M.E., Thornhill, I., Wood, P.J., Hassall, C., 2021. Local contributions to beta diversity in urban pond networks: Implications for biodiversity conservation and management. Divers. Distrib. 00, ddi.13239. https://doi.org/10.1111/ddi.13239

Holgerson, M.A., Raymond, P.A., 2016. Large contribution to inland water CO2 and CH4 emissions from very small ponds. Nat. Geosci. 9, 222–226. https://doi.org/10.1038/ngeo2654

James, F.C., McCulloch, C.E., 1990. Multivariate Analysis in Ecology and Systematics: Panacea or Pandora's Box? Annu. Rev. Ecol. Syst. 21, 129–166. https://doi.org/10.1146/annurev.es.21.110190.001021

Jeanmougin, M., Leprieur, F., Loïs, G., Clergeau, P., 2014. Fine-scale urbanization affects Odonata species diversity in ponds of a megacity (Paris, France). Acta Oecologica. https://doi.org/10.1016/j.actao.2014.05.008

Jeffries, M., 2008. The spatial and temporal heterogeneity of macrophyte communities in thirty small, temporary ponds over a period of ten years. Ecography (Cop.). 31, 765–775. https://doi.org/10.1111/j.0906-7590.2008.05487.x

Jonsson, B., Jonsson, N., Jonsson, B., Jonsson, N., 2011. Habitats as Template for Life Histories, in: Ecology of Atlantic Salmon and Brown Trout. https://doi.org/10.1007/978-94-007-1189-1_1

Kaplan, G., 2018. Sentinel-2 Pan Sharpening—Comparative Analysis. Proceedings 2, 345. https://doi.org/10.3390/ecrs-2-05158

Knop, E., 2016. Biotic homogenization of three insect groups due to urbanization. Glob. Chang. Biol. 22, 228–236. https://doi.org/10.1111/gcb.13091

Kowarik, I., 2011. Novel urban ecosystems, biodiversity, and conservation. Environ. Pollut. https://doi.org/10.1016/j.envpol.2011.02.022

Lacaux, J.P., Tourre, Y.M., Vignolles, C., Ndione, J.A., Lafaye, M., 2007. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. Remote Sens. Environ. 106, 66–74. https://doi.org/10.1016/j.rse.2006.07.012

Lechner, A.M., Gomes, R.L., Rodrigues, L., Ashfold, M.J., Selvam, S.B., Wong, E.P., Raymond, C.M., Zieritz, A., Sing, K.W., Moug, P., Billa, L., Sagala, S., Cheshmehzangi, A., Lourdes, K., Azhar, B., Sanusi, R., Ives, C.D., Tang, Y.-T., Tan, D.T., Chan, F.K.S., Nath, T.K., Sabarudin, N.A.B., Metcalfe, S.E., Gulsrud, N.M., Schuerch, M., Campos-Arceiz, A., Macklin, M.G., Gibbins, C., 2020. Challenges and considerations of applying nature-based solutions in low- and middle-income countries in Southeast and East Asia. Blue-Green Syst. https://doi.org/10.2166/bgs.2020.014

Lechner, A.M., McCaffrey, N., McKenna, P., Venables, W.N., Hunter, J.T., 2016. Ecoregionalization classification of wetlands based on a cluster analysis of environmental data. Appl. Veg. Sci. https://doi.org/10.1111/avsc.12248

Lechner, A.M., Stein, A., Jones, S.D., Ferwerda, J.G., 2009. Remote sensing of small and linear features: Quantifying the effects of patch size and length, grid position and detectability on land cover mapping. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2009.06.002

Lopez, R.D., Lyon, J.Z., Lyon, L.K., Lopez, D.K., 2013. Wetland landscape characterization: Practical tools, methods and approaches for landscape ecology, 2nd ed. Boca Raton, FL.

Lourdes, K.T., Gibbins, C.N., Hamel, P., Sanusi, R., Azhar, B., Lechner, A.M., 2021. A Review of Urban Ecosystem Services Research in Southeast Asia. Land 10, 40. https://doi.org/10.3390/land10010040

Low, K.H., Koki, I.B., Juahir, H., Azid, A., Behkami, S., Ikram, R., Mohammed, H.A., Zain, S.M.,

602 2016. Evaluation of water quality variation in lakes, rivers, and ex-mining ponds in Malaysia
603 (review). Desalin. Water Treat. 57, 28215–28239.
604 https://doi.org/10.1080/19443994.2016.1185382

605 McKinney, M.L., 2006. Urbanization as a major cause of biotic homogenization. Biol. Conserv.
606 https://doi.org/10.1016/j.biocon.2005.09.005

607 McKinney, M.L., Lockwood, J.L., 2001. Biotic homogenization: a sequential and selective process,
608 in: McKinney, M.L., Lockwood, J.L. (Eds.), Biotic Homogenization. Plenum Publishers, New
609 York, pp. 1–17.

610 Meterologi Malaysia, 2020. Climate of Malaysia [WWW Document]. URL
611 http://www.met.gov.my/pendidikan/iklim/iklimmalaysia (accessed 4.6.20).

612 Nath, T.K., Zhe Han, S.S., Lechner, A.M., 2018. Urban green space and well-being in Kuala Lumpur,
613 Malaysia. Urban For. Urban Green. 36, 34–41. https://doi.org/10.1016/j.ufug.2018.09.013

614 Oertli, B., Parris, K.M., 2019. Review: Toward management of urban ponds for freshwater
615 biodiversity. Ecosphere 10, e02810. https://doi.org/10.1002/ecs2.2810

616 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R.,
617 O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019.
618 vegan: Community Ecology Package. R package version 2.5-6.

619 Pérez-Quintero, J.C., 2011. Freshwater mollusc biodiversity and conservation in two stressed
620 Mediterranean basins. Limnologica 41, 201–212. https://doi.org/10.1016/j.limno.2010.09.007

621 Pyšek, P., 1993. Factors affecting the diversity of flora and vegetation in central European
622 settlements. Vegetatio 106, 89–100. https://doi.org/10.1007/BF00044860

623 R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for
624 Statistical Computing, Vienna, Austria.

625 Roy, A.H., Rosemond, A.D., Leigh, D.S., Paul, M.J., Wallace, J.B., 2003. Habitat-specific responses
626 of stream insects to land cover disturbance: Biological consequences and monitoring

implications. J. North Am. Benthol. Soc. 22, 292–307. https://doi.org/10.2307/1467999

Scheffer, M., Van Geest, G.J., Zimmer, K., Jeppesen, E., Søndergaard, M., Butler, M.G., Hanson, M.A., Declerck, S., De Meester, L., 2006. Small habitat size and isolation can promote species richness: Second-order effects on biodiversity in shallow lakes and ponds. Oikos. https://doi.org/10.1111/j.0030-1299.2006.14145.x

Seto, K.C., Güneralp, B., Hutyra, L.R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proc. Natl. Acad. Sci. U. S. A. 109, 16083–16088. https://doi.org/10.1073/pnas.1211658109

Seto, K.C., Parnell, S., Elmqvist, T., 2013. A Global Outlook on Urbanization, in: Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities. Springer Netherlands, Dordrecht, pp. 1–12. https://doi.org/10.1007/978-94-007-7088-1_1

Sharip, Z., Jusoh, J., 2010. Integrated lake basin management and its importance for Lake Chini and other lakes in Malaysia. Lakes Reserv. Res. Manag. 15, 41–51. https://doi.org/10.1111/j.1440-1770.2010.00421.x

Southwood, T.R.E., 1988. Tactics, Strategies and Templets. Oikos. https://doi.org/10.2307/3565974

Sun, Z., Brittain, J.E., Sokolova, E., Thygesen, H., Saltveit, S.J., Rauch, S., Meland, S., 2018. Aquatic biodiversity in sedimentation ponds receiving road runoff – What are the key drivers? Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2017.06.080

Svitok, M., Hrivnák, R., Kochjarová, J., Oťaheľová, H., Paľove-Balang, P., 2016. Environmental thresholds and predictors of macrophyte species richness in aquatic habitats in central Europe. Folia Geobot. 51, 227–238. https://doi.org/10.1007/s12224-015-9211-2

Teo, H.C., Campos-Arceiz, A., Li, B. V., Wu, M., Lechner, A.M., 2020a. Building a green Belt and Road: A systematic review and comparative assessment of the Chinese and English-language literature. PLoS One 15, e0239009. https://doi.org/10.1371/journal.pone.0239009

Teo, H.C., Lechner, A.M., Sagala, S., Campos-Arceiz, A., 2020b. Environmental Impacts of Planned

652 Capitals and Lessons for Indonesia's New Capital. Land 9, 438.
653 https://doi.org/10.3390/land9110438

654 Teo, H.C., Zeng, Y., Sarira, T.V., Fung, T.K., Zheng, Q., Song, X.P., Chong, K.Y., Koh, L.P., 2021.
655 Global urban reforestation can be an important natural climate solution. Environ. Res. Lett. 16,
656 034059. https://doi.org/10.1088/1748-9326/abe783

657 Trakhtenbrot, A., Kadmon, R., 2005. Environmental cluster analysis as a tool for selecting
658 complementary networks of conservation sites. Ecol. Appl. https://doi.org/10.1890/04-0077

659 UK Ordnance Survey, 2021. OS MasterMap Topography Layer Support [WWW Document]. URL
660 https://www.ordnancesurvey.co.uk/business-government/tools-support/mastermap-topography-
661 support (accessed 4.8.21).

662 Verpoorter, C., Kutser, T., Seekell, D.A., Tranvik, L.J., 2014. A global inventory of lakes based on
663 high-resolution satellite imagery. Geophys. Res. Lett. https://doi.org/10.1002/2014GL060641

664 Waterkeyn, A., Grillas, P., Vanschoenwinkel, B., Brendonck, L., 2008. Invertebrate community
665 patterns in Mediterranean temporary wetlands along hydroperiod and salinity gradients. Freshw.
666 Biol. https://doi.org/10.1111/j.1365-2427.2008.02005.x

667 Williams, P., Whitfield, M., Biggs, J., Bray, S., Fox, G., Nicolet, P., Sear, D., 2004. Comparative
668 biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern
669 England. Biol. Conserv. https://doi.org/10.1016/S0006-3207(03)00153-8

670 Wood, P.J., Greenwood, M.T., Agnew, M.D., 2003. Pond biodiversity and habitat loss in the UK.
671 Area 35, 206–216. https://doi.org/10.1111/1475-4762.00249

672 Woodruff, D.S., 2010. Biogeography and conservation in Southeast Asia: How 2.7 million years of
673 repeated environmental fluctuations affect today's patterns and the future of the remaining
674 refugial-phase biodiversity. Biodivers. Conserv. 19, 919–941. https://doi.org/10.1007/s10531-
675 010-9783-3

676 Zheng, Q., Teo, H.C., Koh, L.P., 2021. Artificial Light at Night Advances Spring Phenology in the

United States. Remote Sens. https://doi.org/10.3390/rs13030399

680 **Tables**

681 **Table 1**. Kuala Lumpur land cover map classification scheme

Class	Sub-class	Description
Bare soil	Bare soil	Open areas covered by bare soil or sand, e.g. cleared areas/construction sites
Water	Water	Water bodies, e.g. sea, lakes, ponds, swamps
Impervious	Asphalt	Urban ground surfaces covered in asphalt, e.g. roads, car parks
	Built-up	Other impervious built-up surfaces, e.g. concrete/metal/tile roofs
Vegetation	Grass	Surfaces covered by grass, e.g. fields, parks, golf courses, rural lots
	Crop	Cultivated crops, e.g. rice and fruit farms
	Trees	Patches of trees in forest or urban areas
	Oil palm	Cultivated oil palm plantations
	Mangroves	Mangrove vegetation growing in saline/brackish water

682

683

684 **Table 2**. Pond metrics and methodology.

Category	Purpose	Metric	ArcGIS tool
Pond dimensions	Measure of size	Area Perimeter	Calculate Geometry tool
Pond shape	Measure of shape complexity	Perimeter-Area Ratio (PAR) Shape Index Fractal dimension	Vector-based Landscape Analysis Tools (V-LATE) extension
Pond proximity	Proxy for landscape connectivity	Mean distance of nearest 3 ponds	Near tool
Land cover	Proxy for anthropogenic influences	% vegetation, soil, water, impervious within 100 m, 250 m, 500 m & 1 km, reduced to 1st principal component (PC)	Tabulate Area 2 tool
Red spectrum	Proxy for sediment content	Median Sentinel-2 red band value	Zonal Statistics tool

685

686

687 **Table 3**. Summary of Pearson's correlations between the 1st principal component (PC1) of land cover

688 (% vegetation, soil, water, impervious) at 100 m, 250 m, 500 m and 1 km buffers.

(Pearson's *r*, *p*-value)	PCA100m	PCA250m	PCA500m
PCA250m	0.67, <0.01		
PCA500m	0.53, <0.01	0.92, <0.01	
PCA1km	0,43, <0.01	0.78, <0.01	0.92, <0.01

689

690

691 **Table 4**. Selected metrics for clustering.

Category	Purpose	Metric
Pond dimensions	Measure of size	Area
Patch metrics	Measure of shape complexity	Shape Index
Pond proximity	Proxy for landscape connectivity	Mean distance of nearest 3 ponds
Land cover	Proxy for anthropogenic influences	PC1 of land cover within 100m buffer PC1 of land cover within 1km buffer
Red spectrum	Proxy for sediment content	Median Sentinel-2 red band value

692

693

694 **Table 5**. Average of normalised clustering validation measures for hierarchical, k-means and PAM

695 clustering methods on a clustering of remotely-sensed variables for all ponds (*n*=1013).

Cluster nos.	5	6	7	8	9	10
Hierarchical	-0.071	-0.295	-0.416	-0.489	-0.664	-0.595
k-means	-0.151	-0.149	-0.065	-0.128	-0.68	-0.556
PAM	0.145	**0.266**	-0.2	0	-0.746	-0.891

696

697

Table 6. Average of normalised clustering validation measures for hierarchical, k-means and PAM clustering methods on (a) a clustering of remotely-sensed variables and fieldwork variables of sampled ponds (n=60), (b) a clustering of remotely-sensed variables of sampled ponds and (c) a clustering of fieldwork variables of sampled ponds.

(a)

Cluster nos.	3	4	5	6
Hierarchical	0.387	0.214	-0.175	-0.286
k-means	0.572	**1.17**	-0.792	-0.555
PAM	-0.330	0.166	-0.668	-0.129

(b)

Cluster nos.	3	4	5	6
Hierarchical	**0.741**	-0.063	-0.255	-0.182
k-means	-0.270	-0.027	-0.041	-0.046
PAM	0.664	-0.625	-0.172	-0.569

(c)

Cluster nos.	3	4	5	6
Hierarchical	0.388	0.215	-0.176	-0.287
k-means	0.572	**1.18**	-0.788	-0.556
PAM	-0.322	-0.183	-0.672	-0.134

708 **Table 7**. SIMPER analysis on cluster pairs for 6-cluster PAM model, showing cumulative contributions

709 of most influential variable.

Cluster pairs	Cumulative contributions of most influential variable		
1 vs 2	pca100 (0.33)	pca1k(0.65)	area.log (0.85)
1 vs 3	pca1k (0.44)	area.log (0.67)	pca100 (0.82)
1 vs 4	pca100 (0.54)	pca1k (0.71)	
1 vs 5	pca100 (0.40)	pca1k (0.77)	
1 vs 6	pca1k (0.47)	pca100 (0.71)	
2 vs 3	pca100 (0.49)	area.log (0.68)	pca1k (0.81)
2 vs 4	pca100 (0.42)	pca1k (0.61)	area.log (0.78)
2 vs 5	pca100 (0.37)	pca1k (0.62)	area.log (0.83)
2 vs 6	pca1k (0.38)	area.log (0.61)	pca100 (0.78)
3 vs 4	pca100 (0.62)	mean(NEAR3).log (0.75)	
3 vs 5	pca100 (0.57)	pca1k (0.77)	
3 vs 6	pca100 (0.41)	pca1k (0.71)	
4 vs 5	pca1k (0.38)	pca100 (0.63)	mean(NEAR3).log (0.80)
4 vs 6	pca100 (0.38)	pca1k (0.72)	
5 vs 6	pca100 (0.49)	area.log (0.77)	

710

Table 8. Mean ± SD of pond variables for each cluster. For remote sensing variables n = 1,013, and for fieldwork variables n = 60. SI = Shape Index, Mean(NEAR3) = mean distance to the nearest 3 ponds, Median(RED) = median red band value, VEG = vegetation land cover, IMP = impervious land cover, PMS = Pond Margin Shaded, EM = Emergent Macrophytes, SM = Submerged Macrophytes, FM = Floating Macrophytes, EARTH = earthen bank material, CONCR = concrete bank material, WOOD = wooden bank material, STONE = stone bank material, TEMP = water temperature, TURB = turbidity, COND = conductivity, and DO = Dissolved Oxygen.

		Variable	Mean ± SD of ponds in cluster [min, max]						
			1	**2**	**3**	**4**	**5**	**6**	One-way ANOVA
Remote sensing	Physical	Area (ha)	6.41 ± 30 [0.098, 304]	5.77 ± 35 [0.12, 418]	1.38 ± 3.0 [0.085, 25]	1.89 ± 3.9 [0.082, 25]	3.70 ± 16 [0.14, 186]	3.99 ± 10 [0.16, 72]	$F_{(5,1007)} = 7.6$, $p < 0.01$
		SI	1.9 ± 0.7 [1.1, 7.3]	2.0 ± 0.8 [1.1, 9.1]	1.79 ± 0.5 [1.1, 4.1]	1.7 ± 0.5 [1.1, 4]	1.8 ± 0.5 [1.1, 5.1]	1.8 ± 0.5 [1.1, 5.4]	$F_{(5,1007)} = 4.9$, $p < 0.01$
		mean(NEAR3) (m)	743 ± 1068 [30, 9617]	595 ± 524 [19, 3542]	633 ± 566 [62, 2794]	732 ± 1018 [33, 8143]	791 ± 564 [43, 3466]	784 ± 541 [68, 2895]	$F_{(5,1007)} = 8.2$, $p < 0.01$
		median(RED)	0.080 ± 0.029 [0.048, 0.213]	0.090 ± 0.028 [0.056, 0.196]	0.080 ± 0.023 [0.055, 0.177]	0.107 ± 0.041 [0.056, 0.248]	0.101 ± 0.032 [0.057, 0.200]	0.082 ± 0.020 [0.054, 0.140]	$F_{(5,1007)} = 25$, $p < 0.01$
		% VEG in 100m	78 ± 13 [39, 99]	47 ± 11 [16, 73]	75 ± 9 [46, 97]	20 ± 14 [0, 54]	17 ± 10 [0, 48]	47 ± 12 [24, 80]	$F_{(5,1007)} = 910$, $p < 0.01$
		% IMP in 100m	10 ± 9 [0, 42]	34 ± 15 [4, 73]	15 ± 9 [0, 37]	43 ± 20 [4, 87]	62 ± 19 [1, 97]	41 ± 14 [6, 74]	$F_{(5,1007)} = 335$, $p < 0.01$
		% VEG in 1km	71 ± 12 [27, 98]	48 ± 11 [18, 80]	42 ± 11 [12, 61]	53 ± 14 [23, 88]	21 ± 9 [6, 42]	25 ± 8 [5, 41]	$F_{(5,1007)} = 607$, $p < 0.01$
		% IMP in 1km	17 ± 9 [0, 36]	37 ± 10 [13, 52]	47 ± 10 [28, 78]	27 ± 11 [4, 50]	65 ± 12 [38, 90]	66 ± 9 [50, 93]	$F_{(5,1007)} = 725$, $p < 0.01$

Fieldwork		PMS (%)	50 ± 25 [10, 80]	32 ± 21 [10, 70]	36 ± 24 [0, 80]	16 ± 15 [0, 40]	17 ± 14 [0, 50]	11 ± 8 [0, 30]	$F_{(5,54)} = 6.2$, $p < 0.01$
		EM (%)	18 ± 14 [2, 50]	24 ± 23 [0, 80]	25 ± 27 [0, 70]	23 ± 28 [0, 95]	12 ± 11 [0, 40]	17 ± 22 [0, 60]	$F_{(5,54)} = 0.54$, $p = 0.75$
		SM (%)	1 ± 3 [0, 10]	0 [0, 0]	4 ± 7 [0, 20]	0 [0, 0]	1 ± 3 [0, 10]	0 [0, 0]	$F_{(5,54)} = 1.7$, $p = 0.15$
		FM (%)	5 ± 12 [0, 40]	2 ± 5 [0, 15]	2 ± 3 [0, 10]	0 [0, 0]	5 ± 8 [0, 20]	0.2 ± 0.6 [0, 2]	$F_{(5,54)} = 1.1$, $p = 0.40$
		EARTH (%)	92 ± 9 [70, 100]	88 ± 25 [20, 100]	87 ± 22 [30, 100]	63 ± 38 [0, 100]	60 ± 44 [0, 100]	81 ± 30 [0, 100]	$F_{(5,54)} = 2.1$, $p = 0.08$
		CONCR (%)	3 ± 3 [0, 5]	12 ± 21 [0, 70]	6 ± 13 [0, 40]	23 ± 41 [0, 100]	27 ± 41 [0, 100]	14 ± 30 [0, 100]	$F_{(5,54)} = 1.1$, $p = 0.36$
		WOOD (%)	1 ± 3 [0, 10]	0 [0, 0]	0 [0, 0]	0 [0, 0]	0 [0, 0]	2 ± 5 [0, 15]	$F_{(5,54)} = 0.8$, $p = 0.54$
		STONE (%)	5 ± 10 [0, 30]	1 ± 3 [0, 10]	8 ± 12 [0, 30]	23 ± 39 [0, 95]	14 ± 32 [0, 100]	4 ± 6 [0, 15]	$F_{(5,54)} = 1.4$, $p = 0.23$
	Chemical	TEMP (ºC)	32.0 ± 2.5 [28.4, 36.4]	31.1 ± 2.1 [28.5, 34.2]	30.6 ± 0.9 [28.9, 31.5]	32.3 ± 1.6 [29.6, 35.7]	30.0 ± 0.9 [28.6, 31.7]	30.4 ± 1.3 [28.9, 32.7]	$F_{(5,54)} = 3.1$, $p = 0.017$
		TURB (NTU)	25.1 ± 19.1 [4.2, 61.1]	29.2 ± 27.5 [6.6, 94.7]	26.6 ± 25.7 [6.3, 83.5]	40.7 ± 33.3 [9.9, 116.6]	40.6 ± 28.9 [12.1, 107.1]	51.2 ± 63.7 [11.3, 224.6]	$F_{(5,54)} = 0.8$, $p = 0.55$
		COND (µS/cm)	78.2 ± 50.5 [20.8, 179.8]	94.7 ± 55.5 [50.3, 194.7]	89.1 ± 46.3 [24.1, 189.4]	92.5 ± 74.0 [15.6, 269.3]	89.9 ± 49.6 [32.2, 170.3]	114.8 ± 61.3 [33.2, 216.8]	$F_{(5,54)} = 0.5$, $p = 0.81$
		pH	7.5 ± 0.5 [6.8, 8.3]	7.4 ± 1.1 [6.0, 9.2]	7.2 ± 0.5 [6.6, 8.1]	7.4 ± 0.6 [6.9, 8.8]	7.6 ± 0.6 [6.9, 8.3]	7.0 ± 0.3 [6.6, 7.5]	$F_{(5,54)} = 1.3$, $p = 0.30$
		DO (mg/L)	3.5 ± 1.2 [1.9, 6.1]	4 ± 1.5 [1.4, 6.0]	3.2 ± 0.5 [2.6, 4.0]	4.1 ± 1.2 [2.8, 6.5]	3.8 ± 1.2 [1.5, 5.7]	2.4 ± 0.5 [1.5, 3.1]	$F_{(5,54)} = 3.6$, $p < 0.01$

Table 9. Qualitative description of cluster characteristics. TURB = turbidity

Cluster	Description of characteristics					
	Median(RED)	TURB	Land cover (100m)	Land cover (1km)	Size range	Spatial distribution
1	Low	Low	Very highly vegetated	Very highly vegetated	Ponds to largest lakes	Suburban & peri-urban
2	Med	Low	Highly vegetated	Highly vegetated	Ponds to largest lakes	Suburban & peri-urban
3	Low	Low	Very highly vegetated	Highly vegetated	Ponds to small lakes	All
4	High	High	Highly impervious	Highly vegetated	Ponds to small lakes	Peri-urban
5	High	High	Highly impervious	Highly impervious	Ponds to large lakes	Urban core & suburban
6	Med	High	Highly vegetated	Highly impervious	Ponds to medium lakes	Urban core & suburban

Table 10. Number of ponds and lakes in Kuala Lumpur by cluster and district.

	Cluster							
District	1	2	3	4	5	6	Total	Urban structure
Kuala Lumpur	0	2	20	4	23	38	87	Urban core
Petaling	0	10	8	0	14	12	44	Suburban
Subang	6	7	14	2	18	25	72	Suburban
Ampang	10	3	1	0	3	3	20	Suburban
Shah Alam	28	23	38	15	60	35	199	Suburban
Putrajaya	4	1	7	1	1	2	16	Suburban
Klang	3	2	4	5	30	11	55	Suburban
Sepang	48	46	23	37	15	14	183	Peri-urban
Kajang	43	21	25	23	26	20	158	Peri-urban
Selayang	83	34	20	24	12	6	179	Peri-urban
Total	225	149	160	111	202	166		

Table 11. Summary of Pearson's correlations between fieldwork chemical variables and remotely-sensed physical variables. ** indicates moderate correlations (r = 0.3 to 0.7) and * indicates weak correlations (r < 0.3). Values in **Bold** indicate statistically significant results ($p < 0.05$).

(Pearson's *r*, *p*-value)	Area.log	Shape_Idx.log	mean(NEAR3).log	median(RED)
Turbidity	-0.08, 0.56	-0.1, 0.43	-0.07, 0.61	**0.32****
Conductivity	0.1, 0.47	-0.03, 0.81	0.02, 0.89	-0.15, 0.24
pH	**0.39****	**0.34****	-0.21, 0.11	**-0.31****
DO	0.21, 0.1	0.11, 0.38	**-0.27***	-0.11, 0.42

Table 12. SIMPER analysis on cluster pairs for a 4-cluster k-means model on all variables, showing cumulative contributions of most influential variable. TURB = turbidity, COND = conductivity and PMS = percentage of pond margin shaded.

Cluster pairs	Cumulative contributions of most influential variable	
1 vs 2	TURB (0.48)	COND (0.90)
1 vs 3	COND (0.73)	
1 vs 4	COND (0.46)	TURB (0.71)
2 vs 3	TURB (0.68)	PMS (0.82)
2 vs 4	TURB (0.54)	COND (0.87)
3 vs 4	COND (0.59)	PMS (0.77)

Table 13. Mean ± SD of pond variables for each cluster of sampled ponds (n = 60). Statistically significant results for one-way ANOVA between the 6 clusters are indicated by *. TURB = turbidity, COND = conductivity.

Variable	Mean ± SD of ponds in cluster [min, max]				One-way ANOVA
	1	2	3	4	
median(RED)	0.111 ± 0.038 [0.083, 0.184]	0.090 ± 0.026 [0.053, 0.153]	0.074 ± 0.014 [0.058, 0.099]	0.089 ± 0.022 [0.059, 0.129]	$F_{(3,56)} = 2.9$, $p = 0.045$*
TURB (NTU)	118.7 ± 53.4 [83.5, 224.6]	24.2 ± 16.0 [4.2, 67.4]	24.7 ± 16.7 [9.9, 66.6]	29.2 ± 18.1 [6.3, 61.1]	$F_{(3,56)} = 29.8$, $p < 0.001$*
COND (µS/cm)	113.9 ± 46.2 [76.6, 184.1]	42.8 ± 16.0 [15.6, 67.2]	188.6 ± 35.0 [148.1, 269.3]	94.6 ± 16.7 [69.9, 128.2]	$F_{(3,56)} = 84.6$, $p < 0.001$*
Description	High TURB, med COND	Low TURB, low COND	Low TURB, high COND	Low TURB, med COND	

Table 14. SIMPER analysis on cluster pairs for a 3-cluster hierarchical model on remotely-sensed variables for sampled ponds (*n*=60), showing cumulative contributions of most influential variables.

Cluster pairs	Cumulative contributions of most influential variables		
1 vs 2	pca100 (0.55)	pca1k (0.75)	
1 vs 3	pca1k (0.38)	pca100 (0.68)	Area.log (0.87)
2 vs 3	pca100 (0.45)	pca1k (0.69)	Area.log (0.84)

Table 15. SIMPER analysis on cluster pairs for a 3-cluster hierarchical model on remotely-sensed variables for sampled ponds (*n*=60), showing cumulative contributions of most influential variables. TURB = turbidity, COND = conductivity and PMS = percentage of pond margin shaded.

Cluster pairs	Cumulative contributions of most influential variable	
1 vs 2	TURB (0.50)	COND (0.87)
1 vs 3	TURB (0.60)	COND (0.83)
1 vs 4	TURB (0.44)	COND (0.85)
2 vs 3	COND (0.53)	PMS (0.76)
2 vs 4	TURB (0.75)	
3 vs 4	COND (0.64)	PMS (0.84)

Figure captions

Fig. 1. Correlation matrix and plots of pond metrics

Fig. 2. Greater Kuala Lumpur land cover map, indicating local authority districts. Inset indicates location of Greater Kuala Lumpur study area in Peninsular Malaysia.

Fig. 3. Ponds and lakes in Greater Kuala Lumpur (n = 1013) (a) and the number of ponds undetected by remote sensing classification in 24 sampled areas (2 km radius) across GKL, determined by manual identification from very high resolution Google Earth imagery (b).

Fig. 4. PCA ordination biplot of remotely-sensed variables of all ponds (n = 1,013). Colour indicates cluster generated by a 6-cluster PAM model. The variables 'pca100' and 'pca1k' refer to land cover within 100 m and 1 km respectively reduced to the first principal component.

Fig. 5. Number of ponds/lakes from each cluster per km^2 in Kuala Lumpur by local authority district (a) and the number of pond/lake (wetland) types per km^2 in Kuala Lumpur by local authority district (b).

Fig. 6. Co-inertia analysis of PCA ordinations of remotely-sensed variables and fieldwork variables (n = 60) for sampled wetlands.

Fig. 7. PCA ordination biplot of all sampled variables (n = 60) generated by a (a) PAM model (colour indicates cluster generated by the 6-cluster PAM model from remotely-sensed physical variables) and (b) a 4-cluster k-means model (colour indicates cluster generated by a 4-cluster k-means model from all variables). The variables 'pca100' and 'pca1k' refer to land cover within 100 m and 1 km respectively reduced to the first principal component; 'pca_mac' and 'pca_bank' refer to macrophyte coverage and bank material respectively reduced to the first principal component.

Fig. 8. PCA ordination biplot for sampled ponds (n = 60) generated for (a) remotely-sensed variables. (colour indicates cluster generated by a 3-cluster hierarchical model from remotely-sensed variables) and (b) fieldwork variables (colour indicates cluster generated by a 3-cluster hierarchical model from fieldwork variables). The variables 'pca100' and 'pca1k' refer to land cover within 100 m and 1 km respectively reduced to the first principal component; 'pca_mac' and 'pca_bank' refer to macrophyte coverage and bank material respectively reduced to the first principal component.

Fig. 1

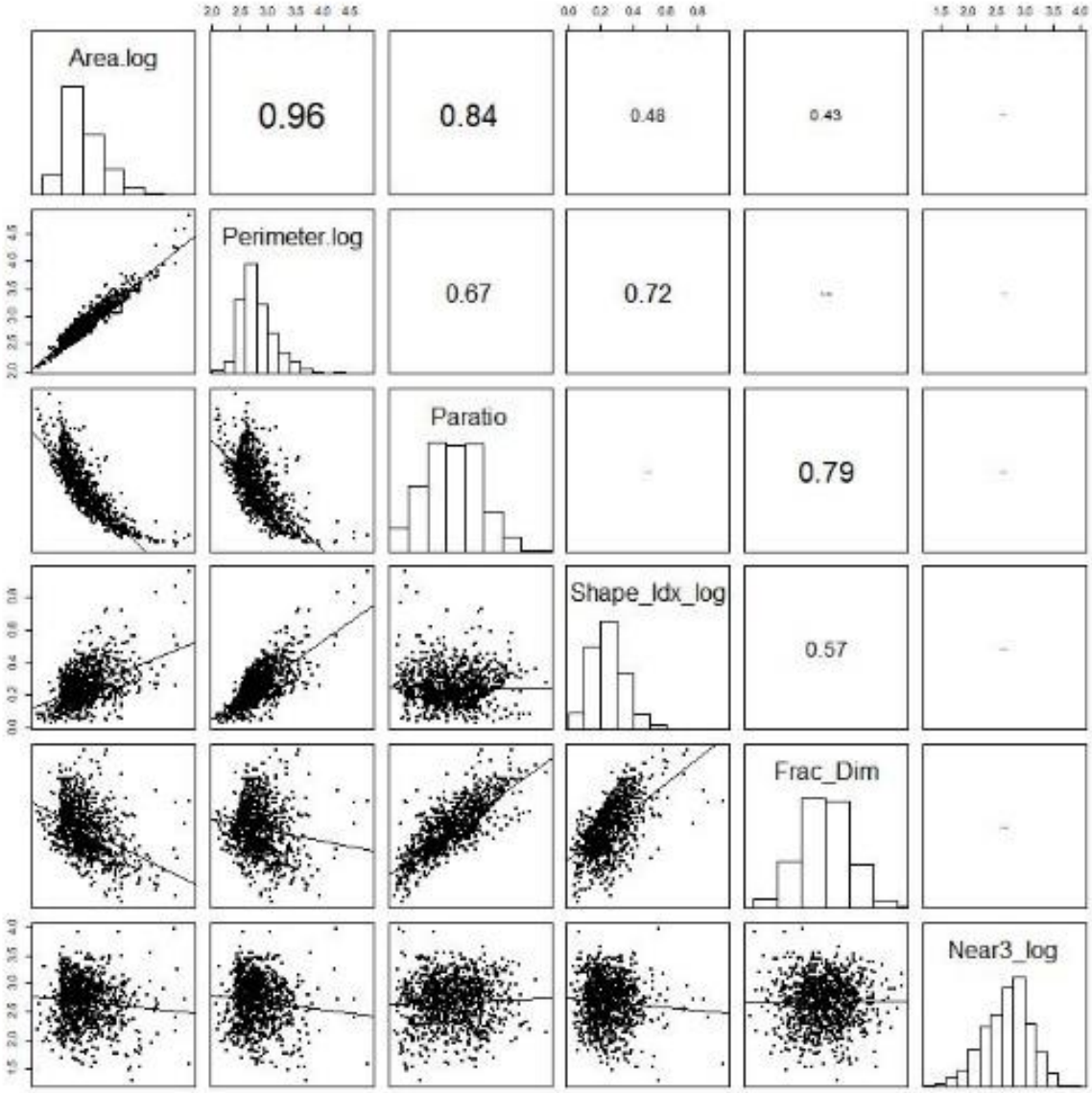

Area.log
0.96
0.84
0.48
0.43
Perimeter.log
0.67
0.72
Paratio
0.79
Shape_Idx_log
0.57
Frac_Dim
Near3_log

Fig. 2

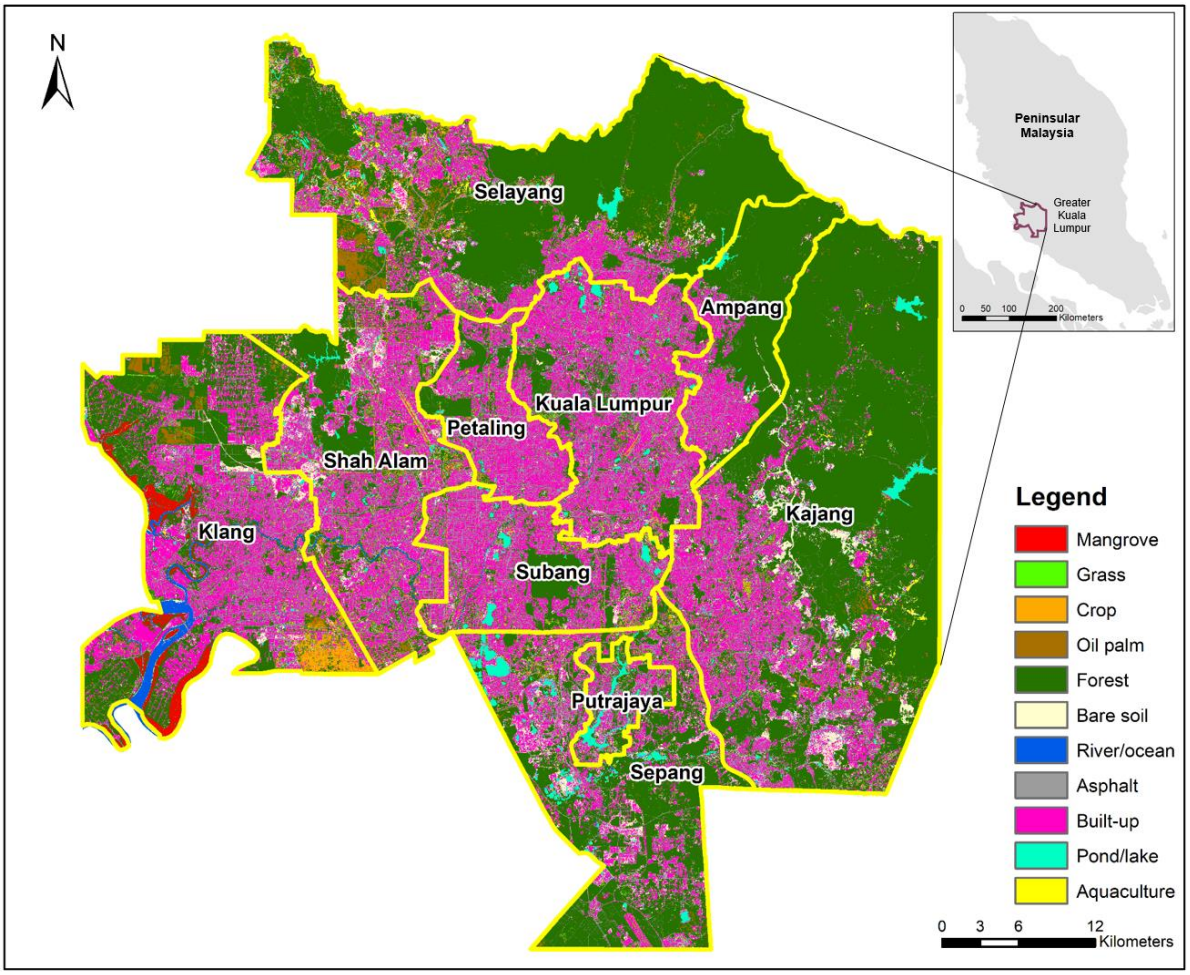

N
Selayang
Ampang
Kuala Lumpur
Petaling
Shah Alam
Klang
Kajang
Subang
Putrajaya
Sepang
Peninsular Malaysia
Greater Kuala Lumpur
0 50 100 200 Kilometers
Legend
Mangrove
Grass
Crop
Oil palm
Forest
Bare soil
River/ocean
Asphalt
Built-up
Pond/lake
Aquaculture
0 3 6 12 Kilometers

Fig. 3

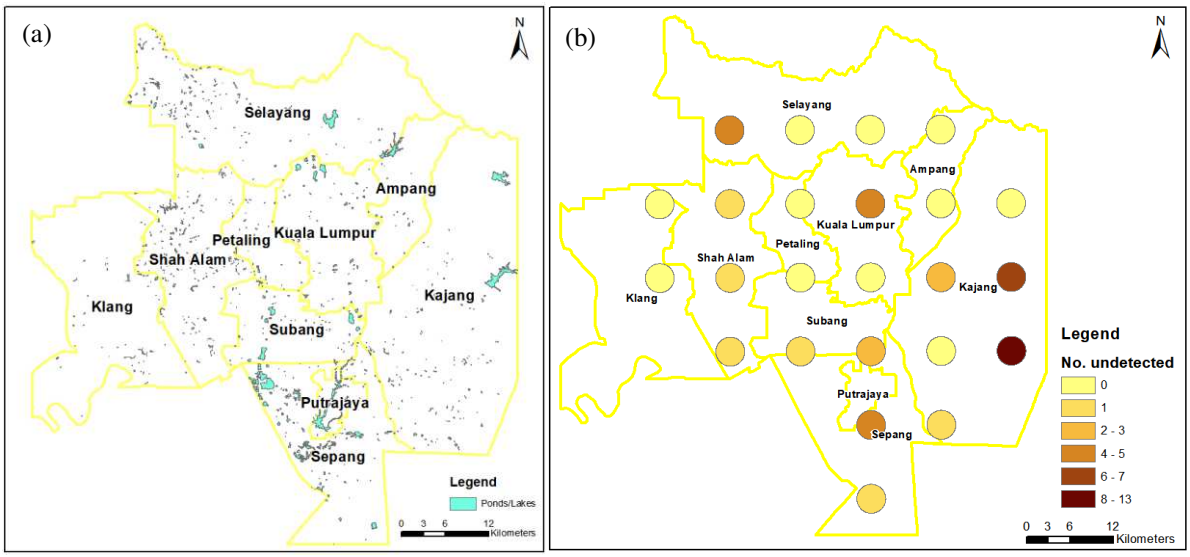

(a)
N
Selayang
Ampang
Petaling
Kuala Lumpur
Shah Alam
Klang
Kajang
Subang
Putrajaya
Sepang
Legend
Ponds/Lakes
0 3 6 12
Kilometers
(b)
N
Selayang
Ampang
Kuala Lumpur
Petaling
Shah Alam
Klang
Kajang
Subang
Putrajaya
Sepang
Legend
No. undetected
0
1
2 - 3
4 - 5
6 - 7
8 - 13
0 3 6 12
Kilometers

Fig. 4.

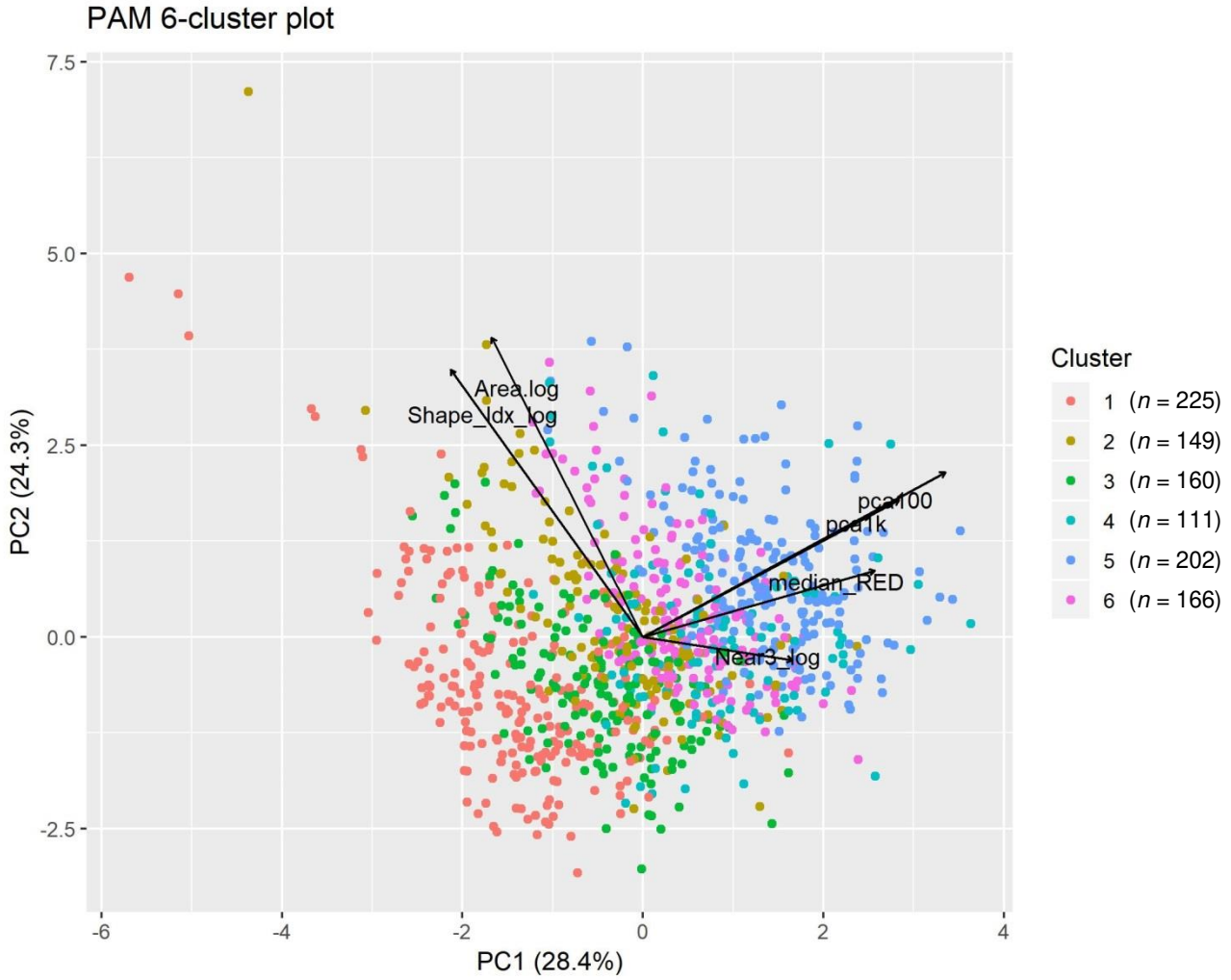
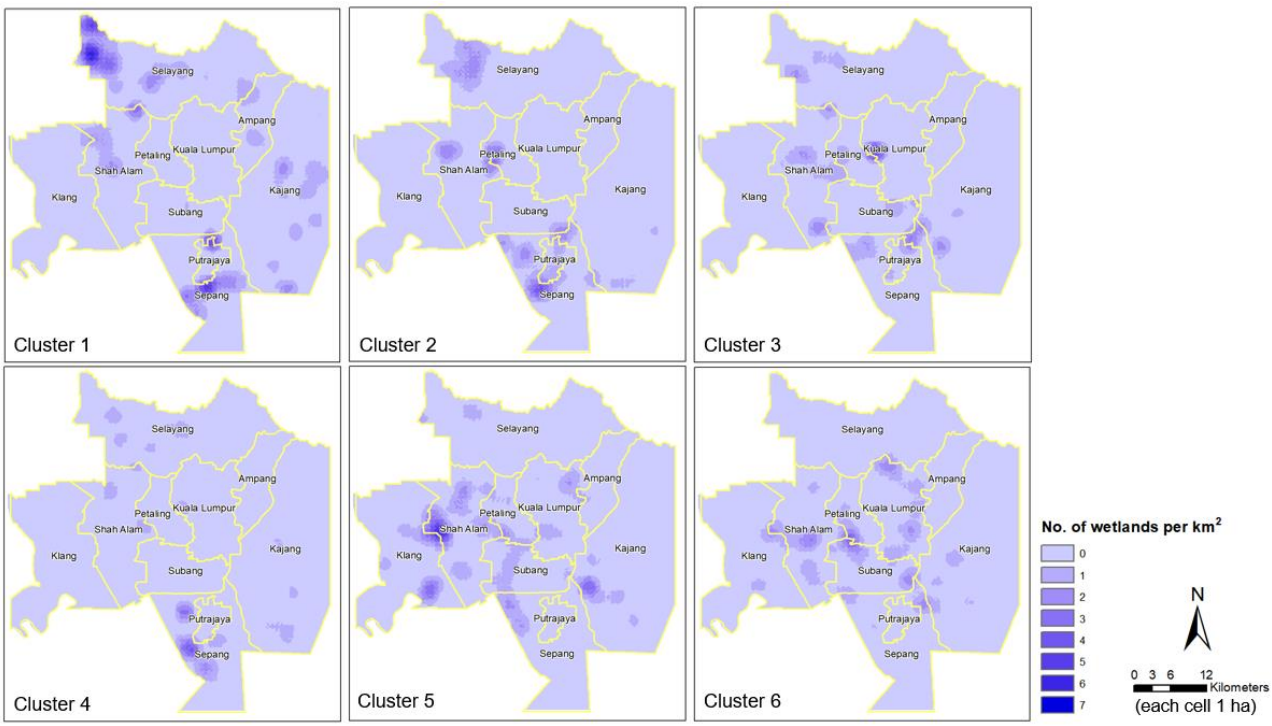

PAM 6-cluster plot
7.5
5.0
2.5
0.0
-2.5
PC2 (24.3%)
-6
-4
-2
0
2
4
PC1 (28.4%)
Area.log
Shape_Idx_log
pca100
pca1k
median_RED
Near3_log
Cluster
1 (n = 225)
2 (n = 149)
3 (n = 160)
4 (n = 111)
5 (n = 202)
6 (n = 166)

Fig. 5.

(a)


Selayang
Ampang
Kuala Lumpur
Petaling
Shah Alam
Klang
Kajang
Subang
Putrajaya
Sepang
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
No. of wetlands per km²
0
1
2
3
4
5
6
7
N
0 3 6 12
Kilometers
(each cell 1 ha)

(b)

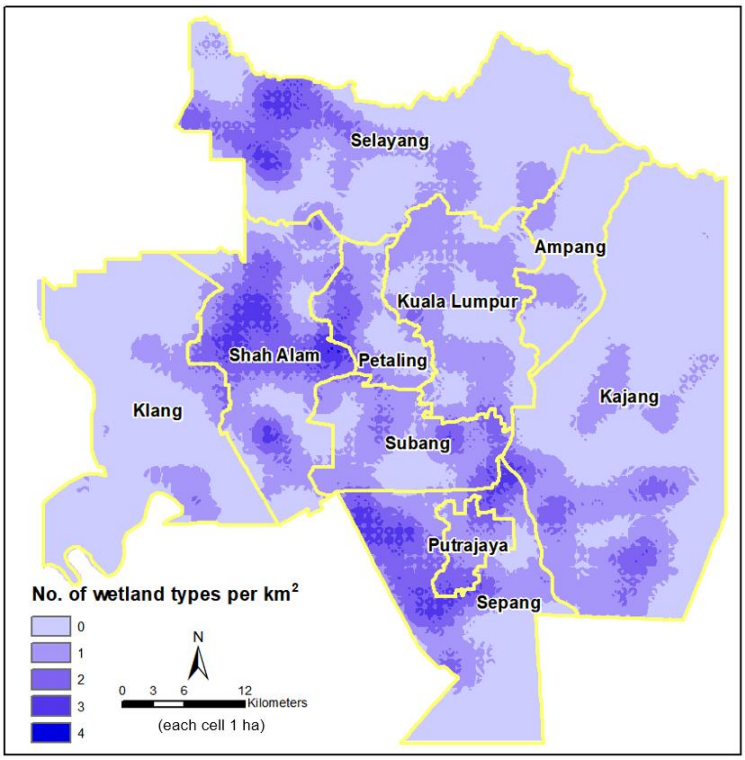

Selayang
Ampang
Kuala Lumpur
Shah Alam
Petaling
Klang
Kajang
Subang
Putrajaya
Sepang
No. of wetland types per km²
0
1
2
3
4
N
0 3 6 12
Kilometers
(each cell 1 ha)

Fig. 6.

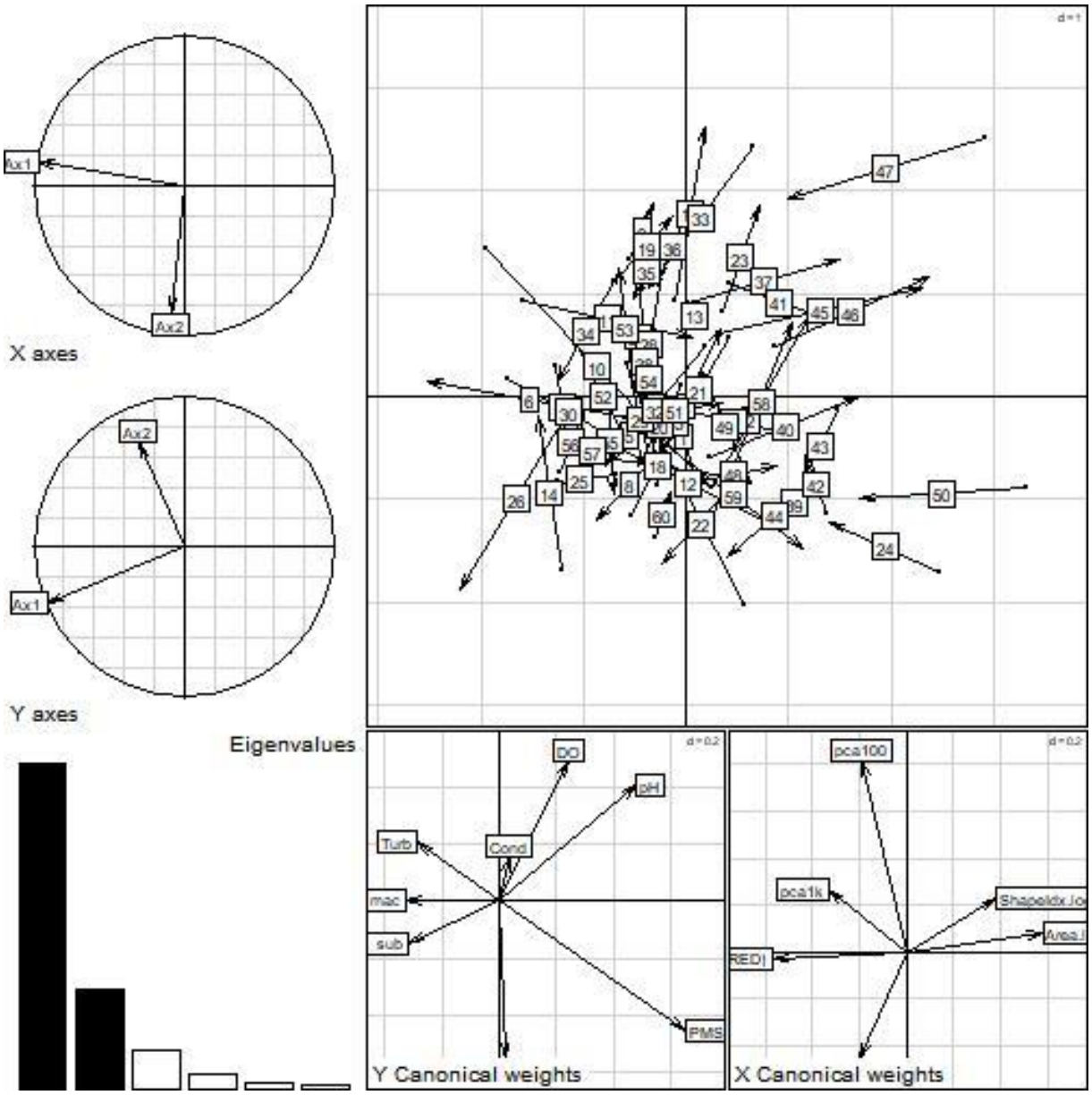
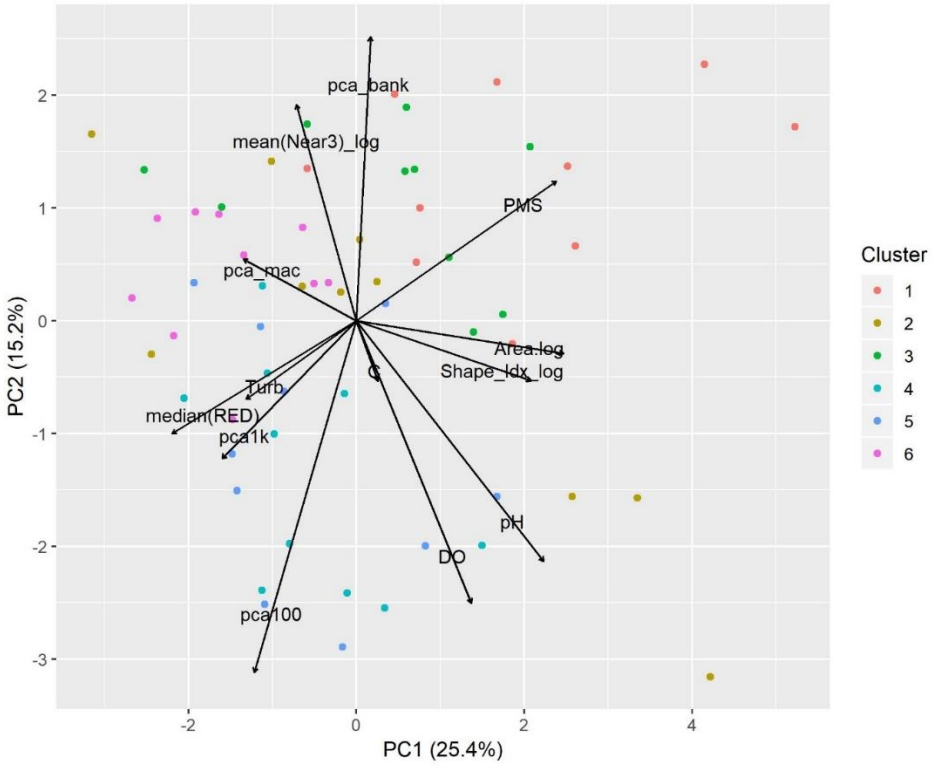

Ax1
Ax2
X axes
Ax2
Ax1
Y axes
Eigenvalues
DO
pH
Turb
Cond
mac
sub
PMS
Y Canonical weights
pca100
pca1k
REDI
X Canonical weights

Fig. 7.

(a)


pca_bank
mean(Near3)_log
PMS
pca_mac
Area.log
Shape_Idx_log
C
Turb
median(RED)
pca1k
pH
DO
pca100
PC2 (15.2%)
PC1 (25.4%)
Cluster
1
2
3
4
5
6

(b)

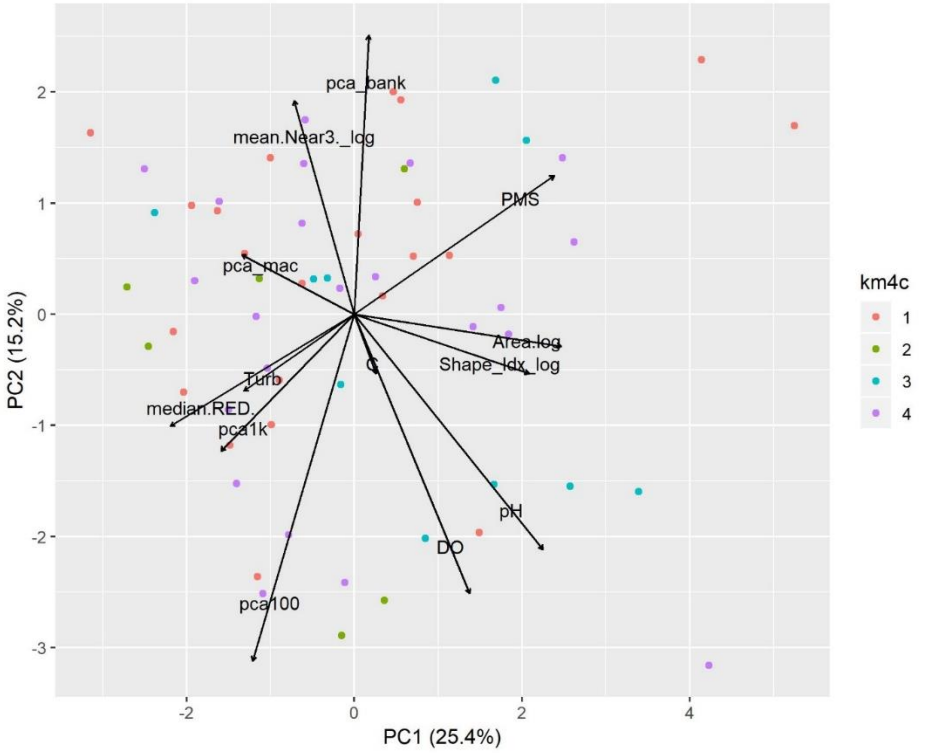
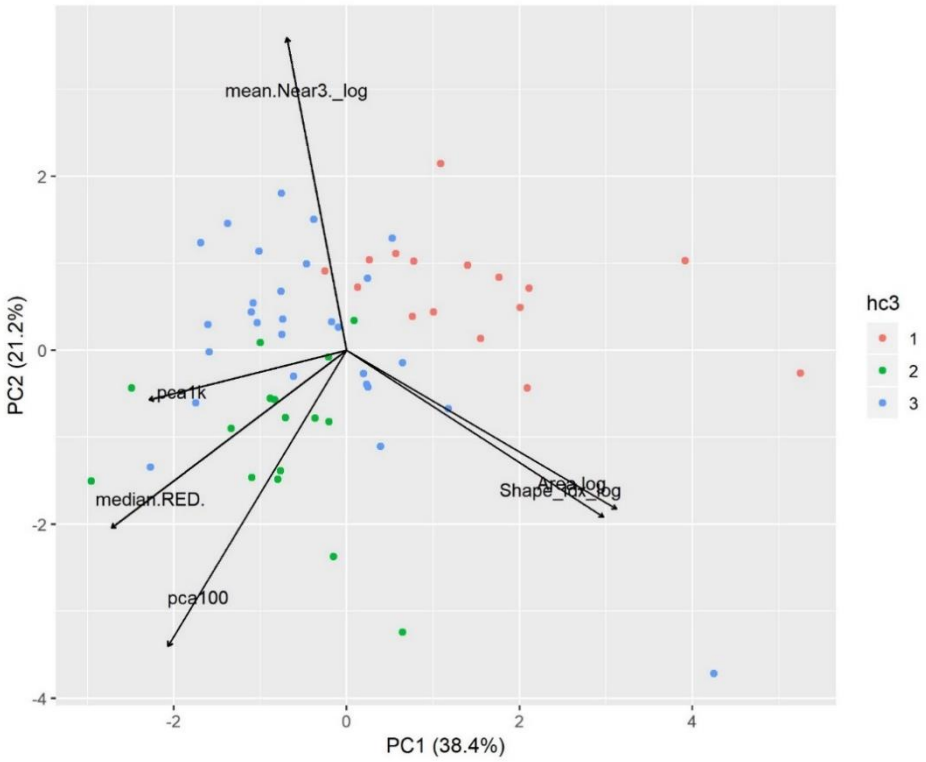
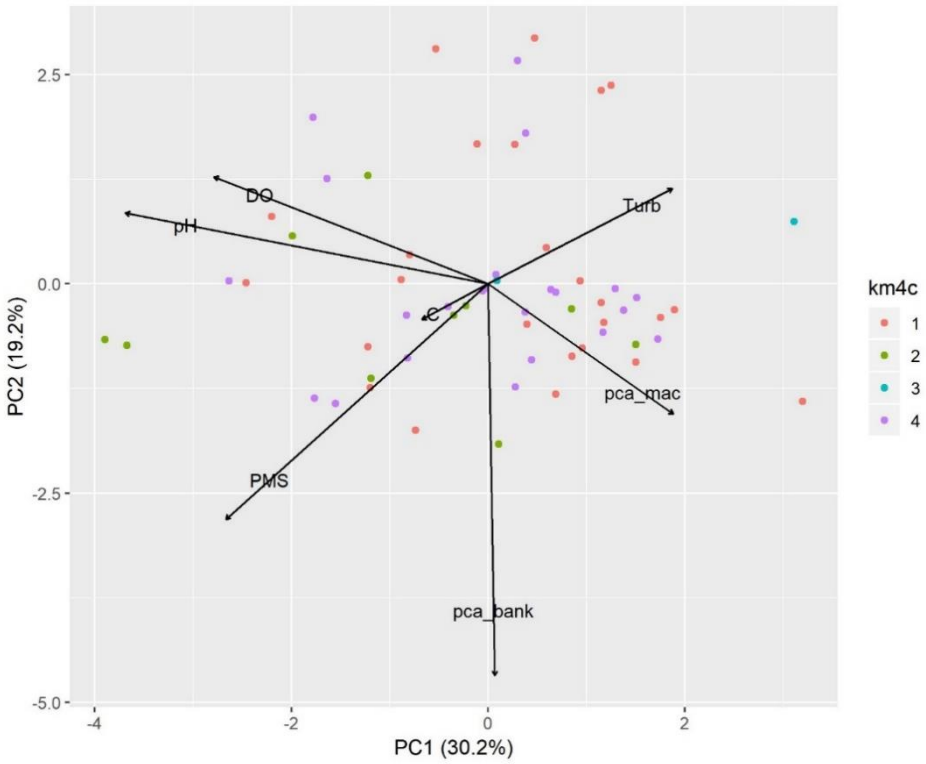

pca_bank
mean.Near3._log
PMS
pca_mac
Area.log
Shape_Idx_log
C
Turb
median.RED.
pca1k
pH
DO
pca100
PC2 (15.2%)
PC1 (25.4%)
km4c
1
2
3
4

Fig. 8.

(a)


mean.Near3._log
pca1k
median.RED.
pca100
Area_log
Shape_idx_log
PC2 (21.2%)
PC1 (38.4%)
hc3
1
2
3

(b)


DO
pH
Turb
C
pca_mac
PMS
pca_bank
PC2 (19.2%)
PC1 (30.2%)
km4c
1
2
3
4

Figures

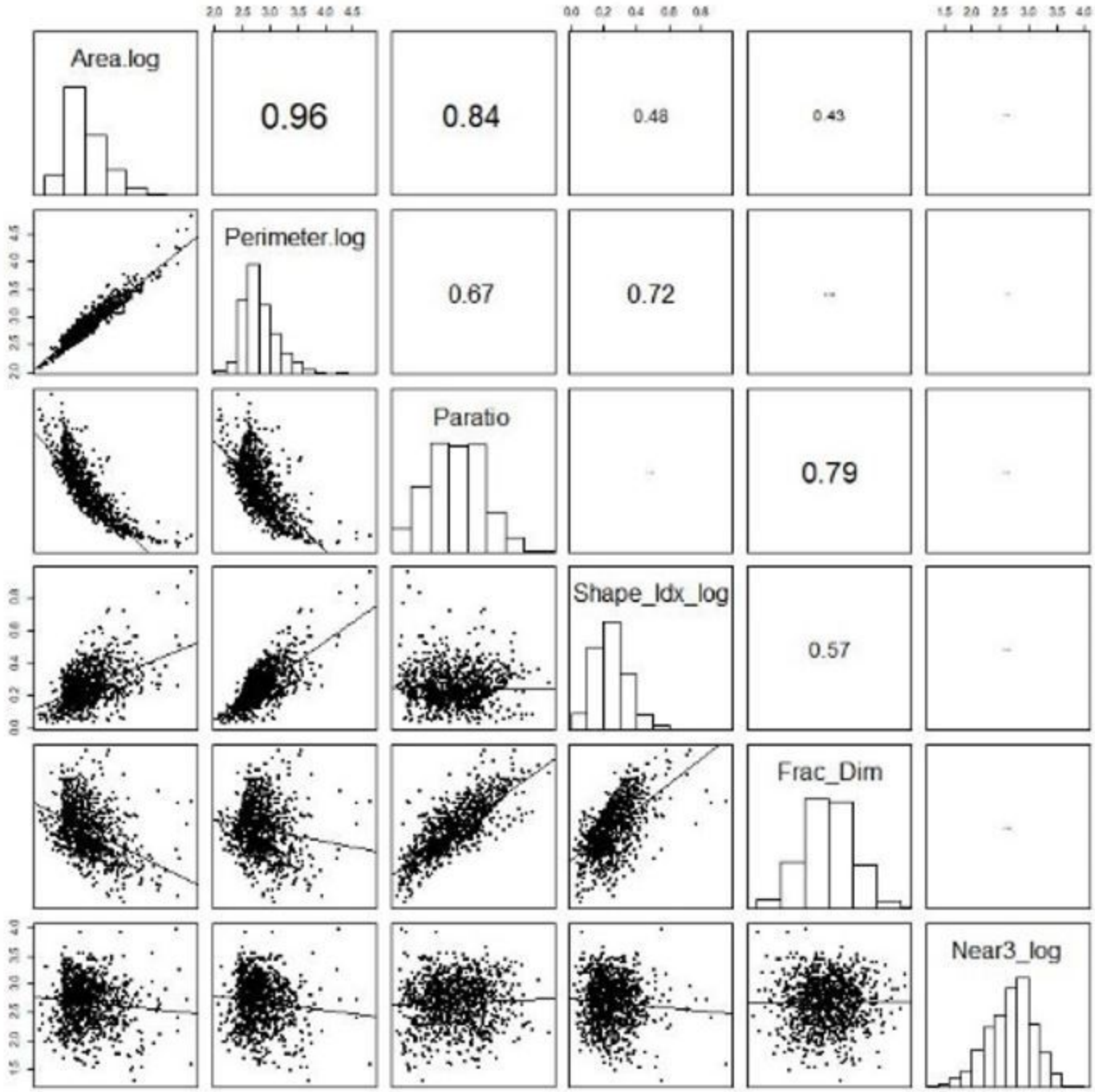


Figure 1

Correlation matrix and plots of pond metrics

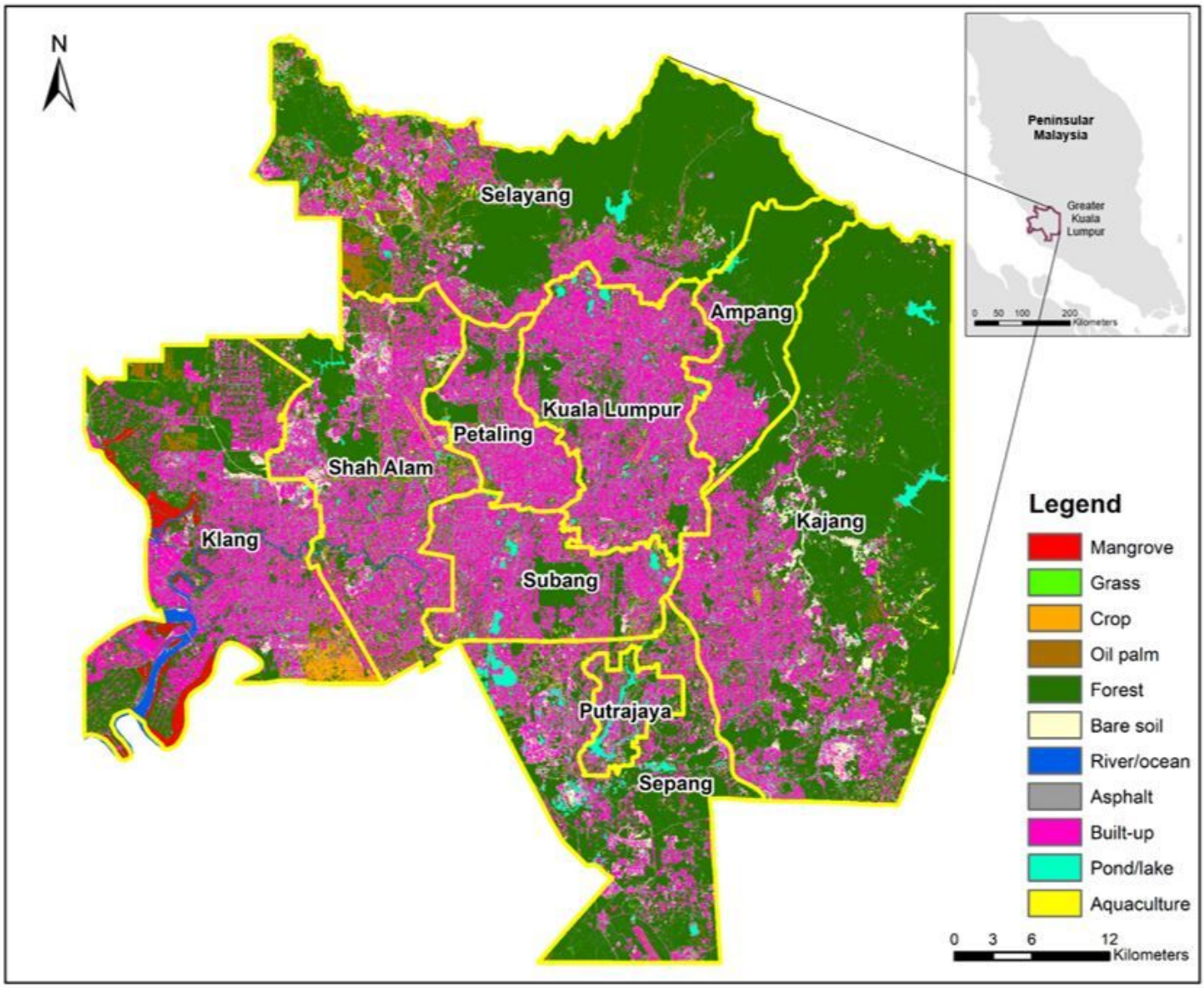


Figure 2

Greater Kuala Lumpur land cover map, indicating local authority districts. Inset indicates location of Greater Kuala Lumpur study area in Peninsular Malaysia. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

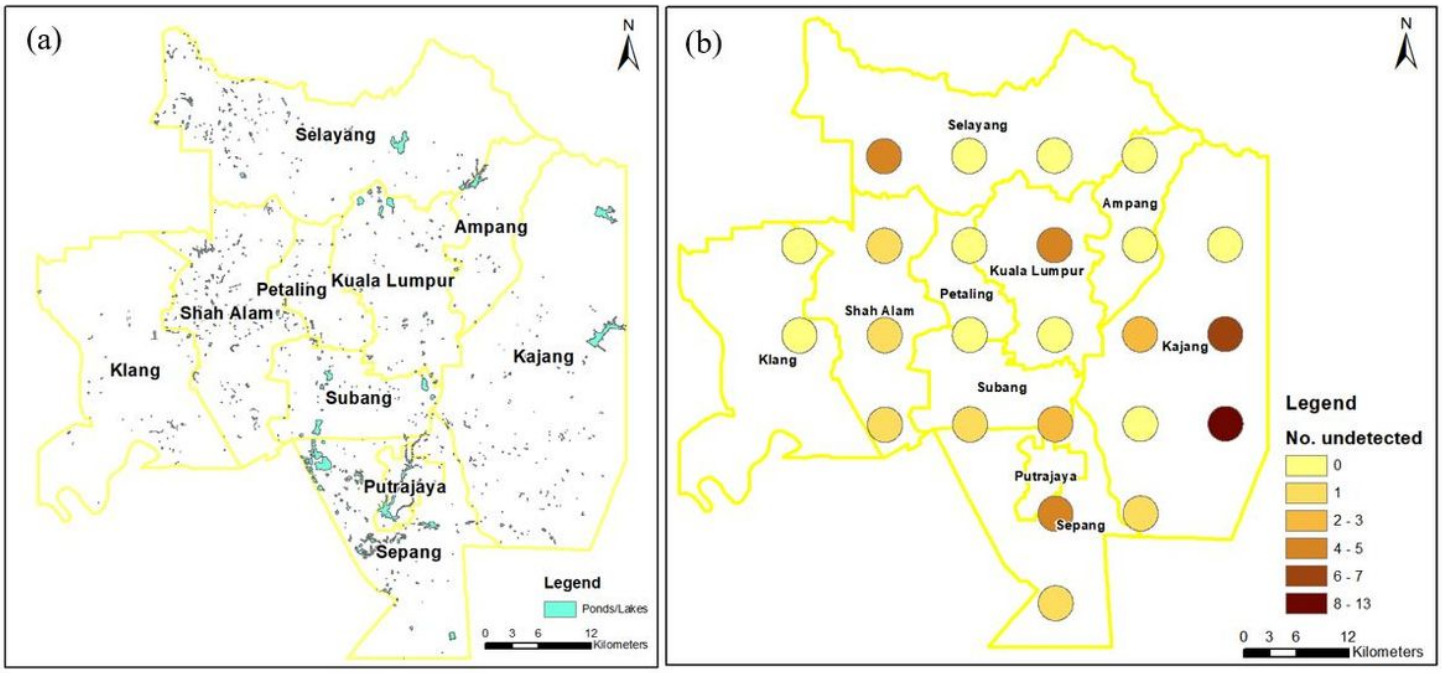


Figure 3

Ponds and lakes in Greater Kuala Lumpur (n = 1013) (a) and the number of ponds undetected by remote sensing classification in 24 sampled areas (2 km radius) across GKL, determined by manual identification from very high resolution Google Earth imagery (b). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

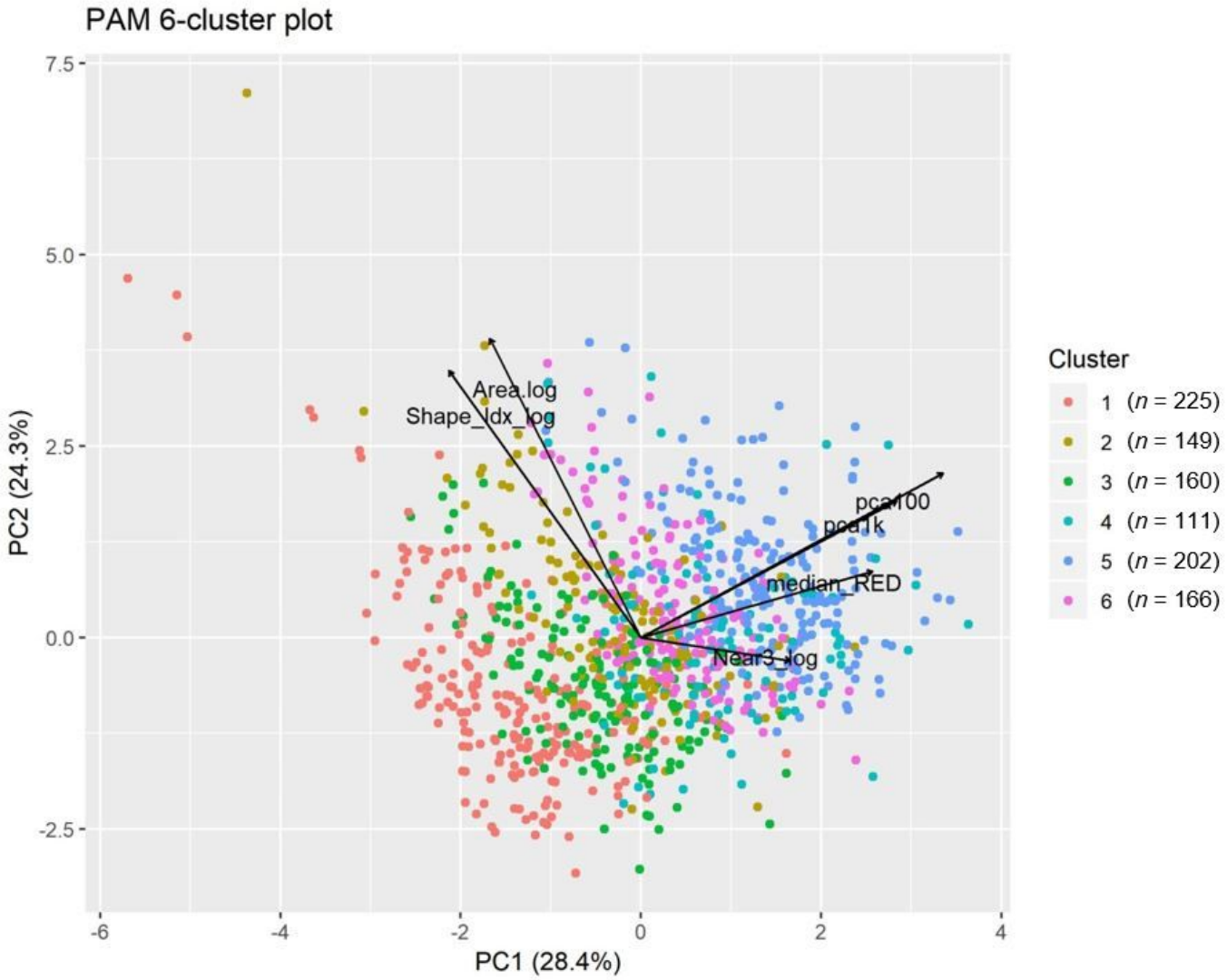


Figure 4

PCA ordination biplot of remotely-sensed variables of all ponds (n = 1,013). Colour indicates cluster generated by a 6-cluster PAM model. The variables 'pca100' and 'pca1k' refer to land cover within 100 m and 1 km respectively reduced to the first principal component.

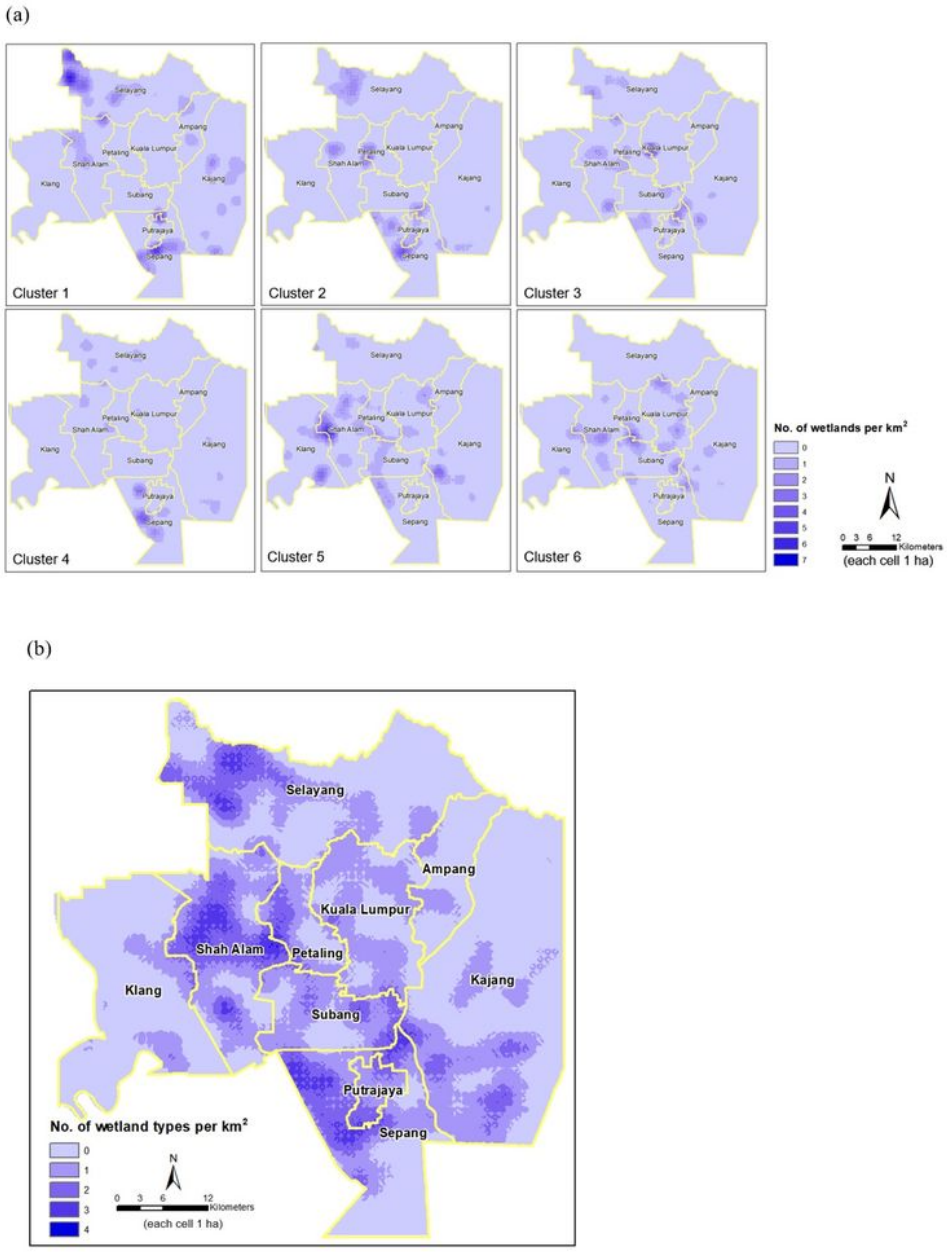


Figure 5

Number of ponds/lakes from each cluster per km2 in Kuala Lumpur by local authority district (a) and the number of pond/lake (wetland) types per km2 in Kuala Lumpur by local authority district (b). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country,

territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

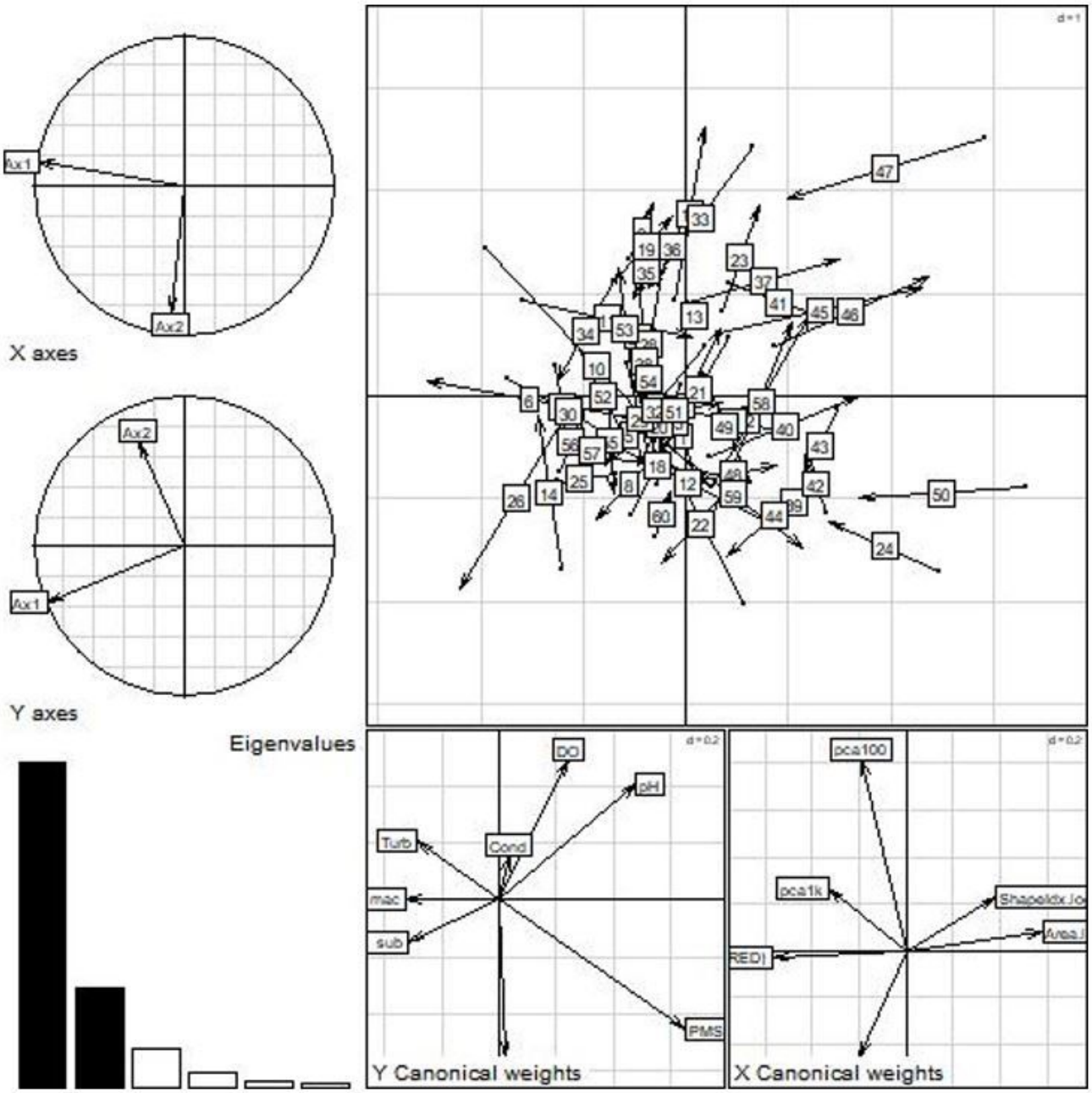


Figure 6

Co-inertia analysis of PCA ordinations of remotely-sensed variables and fieldwork variables (n = 60) for sampled wetlands.

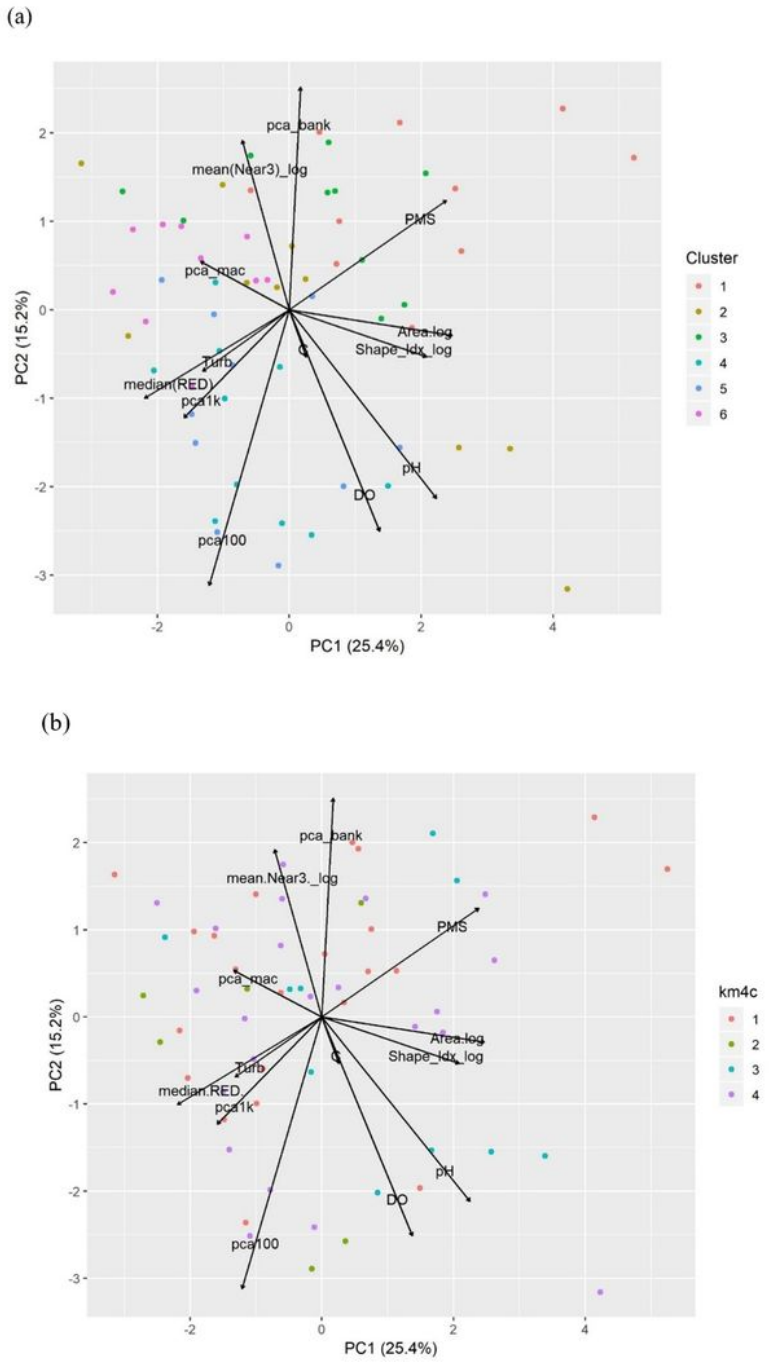


Figure 7

PCA ordination biplot of all sampled variables (n = 60) generated by a (a) PAM model (colour indicates cluster generated by the 6-cluster PAM model from remotely-sensed physical variables) and (b) a 4-cluster k-means model (colour indicates cluster generated by a 4-cluster k-means model from all variables). The variables ‘pca100’ and ‘pca1k’ refer to land cover within 100 m and 1 km respectively

reduced to the first principal component; ‘pca_mac’ and ‘pca_bank’ refer to macrophyte coverage and bank material respectively reduced to the first principal component.

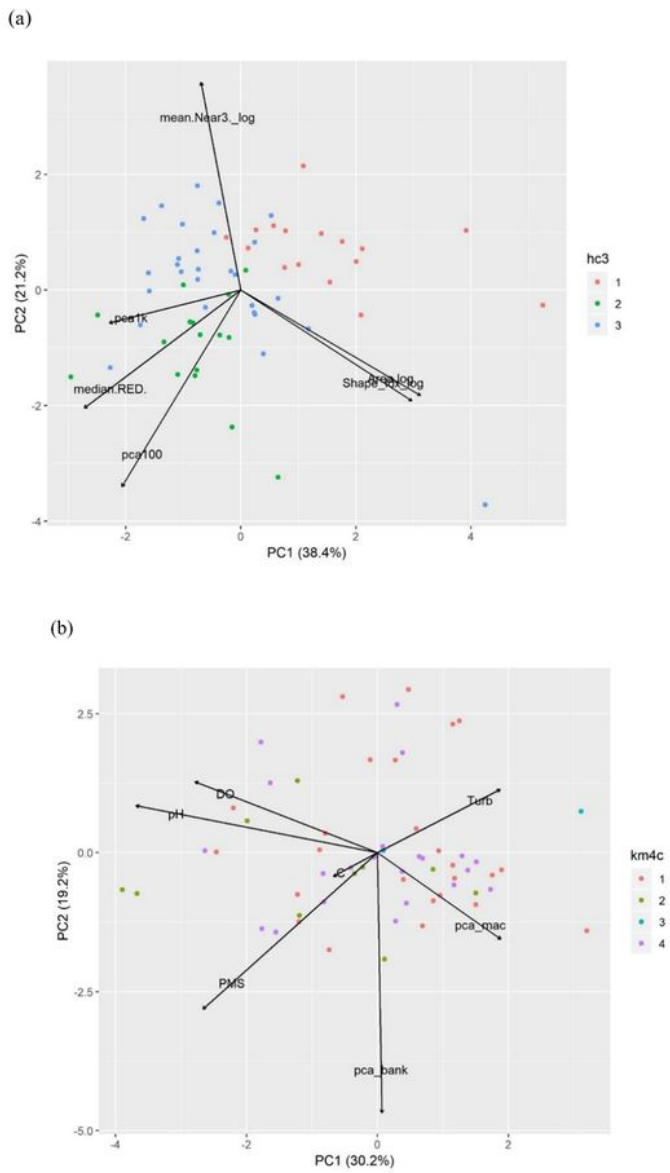


Figure 8

PCA ordination biplot for sampled ponds (n = 60) generated for (a) remotely-sensed variables. (colour indicates cluster generated by a 3-cluster hierarchical model from remotely-sensed variables) and (b) fieldwork variables (colour indicates cluster generated by a 3-cluster hierarchical model from fieldwork

variables). The variables ‘pca100’ and ‘pca1k’ refer to land cover within 100 m and 1 km respectively reduced to the first principal component; ‘pca_mac’ and ‘pca_bank’ refer to macrophyte coverage and bank material respectively reduced to the first principal component.