International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

Climbing up the hills: expansion of agriculture around the Ruma National Park, Kenya

Valeska Scharsich, Dennis Ochuodho Otieno & Christina Bogner

To cite this article: Valeska Scharsich, Dennis Ochuodho Otieno & Christina Bogner (2019) Climbing up the hills: expansion of agriculture around the Ruma National Park, Kenya, International Journal of Remote Sensing, 40:17, 6720-6736, DOI: 10.1080/01431161.2019.1591647

To link to this article: https://doi.org/10.1080/01431161.2019.1591647

INTERNATIONAL JOURNAL OF REMOTE SENSING
2019, VOL. 40, NO. 17, 6720–6736
https://doi.org/10.1080/01431161.2019.1591647

Taylor & Francis
Taylor & Francis Group

Climbing up the hills: expansion of agriculture around the Ruma National Park, Kenya

Valeska Scharsich[a], Dennis Ochuodho Otieno[b,c] and Christina Bogner [a]

[a]Ecological Modelling, BayCEER, University of Bayreuth, Bayreuth, Germany; [b]Department of Plant Ecology, BayCEER, University of Bayreuth, Bayreuth, Germany; [c]Department of Botany, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

ABSTRACT

A major factor driving changes in land use and land cover (LULC) is the human population growth associated with an expanded agricultural production. In the Lambwe valley in Homabay County, Kenya, the most important reason for accelerated population growth in the last decades was the control of the tsetse fly, the biological vector of trypanosomiasis. The goal of our study is to quantify the changes of LULC in the Lambwe valley in the last 30 years, giving special attention to the Ruma National Park. We classified three Landsat images of the Lambwe valley from 1984, 2002, and 2014 by Random Forests. The Ruma National Park itself showed a diverse composition probably supported by frequent fires that lead to a short-term reduction of savanna. Nevertheless, the national park is well protected, and no profound changes could be observed. Outside the national park, agricultural area increased by about 12%, savanna and the dense forest, that used to grow at higher altitudes, decreased by about 8% and 6%, respectively. In particular, agriculture expands towards higher altitudes with steeper slopes thus leading to a larger risk of soil erosion.

ARTICLE HISTORY
Received 5 July 2018
Accepted 30 December 2018

1. Introduction

Worldwide, ecosystems are changed by humans. An essential factor contributing to this change is the growing human population associated with the increased need for food and agricultural land to produce it. Since land suitable for food production is limited, agriculture is often expanded into areas less favourable for cultivation which can lead to land degradation.

Not only terrestrial but also aquatic ecosystems like lakes can be affected by expanding agricultural activities. Indeed, an increasing input of sediments, nitrogen, and phosphorous originating from agriculture can lead to eutrophication. Worldwide, numerous lakes are endangered; one example being Africa's largest lake, the Lake Victoria. Its basin is one of the most densely populated areas in Africa with a population growth rate of 3–4% in 2015 (United Nations, Department of Economic

CONTACT Christina Bogner christina.bogner@uni-bayreuth.de Ecological Modelling, BayCEER, University of Bayreuth, Bayreuth, Germany

and Social Affairs, Population Division 2017). In the lake, the enormous growth of invasive aquatic plants points to the evolving eutrophication and is an important factor for a change of the lake's ecosystem (Juma, Wang, and Li. 2014; ILEC 2005).

Protected areas situated in regions with a rapidly growing population, like the Ruma National Park in Lambwe valley, adjacent to Lake Victoria, are under particular pressure. Land use intensification in the surroundings can lead to conflicts between local communities and the protection goal (DeFries et al. 2007). An important step to develop useful strategies to reduce such conflicts is the detailed knowledge of the changes of land use and land cover (LULC) in the surrounding area.

One method to detect changes in LULC is the analysis of time series of satellite images, which allows to identify breakpoints or gradual changes. Yet, in many areas, this method cannot be used for extended time spans because of lack of data. Indeed, the only (now freely available) long time series of higher resolution satellite images, namely the Landsat imagery, show gaps between 1984 and 1999 all over the world, especially in Africa. This is caused by the commercialization of the Landsat data during this period resulting in many missed observations due to the lack of 'obvious and immediate buyers' (NASA n.d.).

A straight forward method to detect changes between satellite images is the post-classification comparison, where single images instead of time series are compared after being classified individually (e.g. Bayarsaikhan et al. 2009). One main advantage of this approach is the reduced influence of atmospheric and environmental differences between the images (Lu et al. 2004) since the classification into different LULC classes depends on the single image only. This is especially important when using satellites with different sensor types like the Landsat imagery. The reliability of the post-classification comparison depends essentially on the classification accuracy of individual images.

Although some older studies based on aerial photographs exist (Muriuki et al. 2005), recent systematic assessments of LULC changes in Lambwe valley are missing. The rate and patterns of land use change are good indicators of long-term conditions of natural environments (Fox et al. 1995) and provide crucial information regarding ecological processes and ecosystem stability. This is an invaluable information needed for a sustainable management of natural resources. In this study, we investigate the effects of the increasing human population on the LULC during the last 30 years, since Ruma obtained its status as a national park. In particular, we want to know which parts of the valley changed the most. Thereby, we evaluate whether the need for food production is getting in conflict with the need of protection, namely the protection against soil erosion on steeper slopes and the protection of biodiversity in the Ruma National Park.

2. Material and methods

2.1. *Study area*

Our study area is located in Lambwe valley at the Lake Victoria in southwestern Kenya (027′–047′ S and 3406′–3424 'E) at an elevation of 1,107 m above sea level. It covers an area of 561 km^2 and was originally a dense forest, at least on the hills (Allsopp and Baldry 1972; Njoka et al. 2003). It lies between the Kanyamwa escarpment and Gwasi and Gembe hills. The Ruma National Park, fenced in 1971, occupies the centre of the valley and covers an

area of approximately 120 km^2 (Allsopp and Baldry 1972). Human settlements with domestic livestock are scattered around the park. To reflect different land uses in our study area we divided it into four subregions, namely the national park, the western, northern, and southern regions (Fig. 2.1). The national park shows a sharp boundary to its surroundings, and the western region contains the Gwasi hills and their foothills.

The park was established in 1966 as a game reserve to protect the last sanctuary of the Roan antelope (*Hippotragus equinus*) in Kenya. In 1983 it got the status of a National Park (Muriuki et al. 2005; Awange, Aseto, and Ong'ang'a 2004). Until the 1960s, the tsetse fly (*Glossina*) infestation kept humans and grazing livestock out of the valley, allowing the forest and its rich biodiversity to flourish. The tsetse fly is a vector of trypanosomiasis, the sleeping sickness, which affects humans and livestock. In humans, it can be fatal if not treated and treatment requires specifically skilled staff (World Health Organization n.d.). In livestock, the disease leads to reduced birth and increased mortality rates (Muriuki et al. 2005). Starting in the 1960s, the tsetse control programs using traps and selective bush clearing were initiated by the Kenyan government and the World Health Organization and have opened up the area for human settlement (Muriuki et al. 2005). As a consequence, between 1960 and 1970 the population more than tripled (Wellde et al. 1989).

2.2. *Landsat images and their preprocessing*

For our study, we used three Landsat surface reflectance images downloaded from the EarthExplorer (U.S. Geological Survey n.d.) on 17 May 2016 with a resolution of 30 m $\times$ 30 m. They were acquired by L5 TM on 10 September 1984, by L7 ETM on 4 September 2002 and by L8 OLI/TIRS on 15 September 2014. The time intervals are quite wide and unequal because (1) only a few images are available for the 1990s and (2) we chose images with a minimal cloud contamination and avoided Landsat 7 images collected after 31 May 2003 that are affected by the scan line error. In addition, we restricted the analysis to images that were acquired in approximately the same period of the year to enhance comparability.

The blue band of the Landsat image of 2014 shows some artefacts. They are most obvious in the western part, outside our study area (see linear boundaries of colour changes in Figure 1), yet they are also found inside the study area. Therefore, we decided to exclude the blue band in all images from further analysis.

The surface reflectance data produced by the Earth Resources Observation and Science (EROS) Center are already atmospherically corrected. We applied three further corrections before analysis, namely (1) masking of water and clouds using the cfmask band, (2) image registration and (3) topographical correction because the study area is hilly and this affects the illumination of the Earth surface. For the last step, we chose the Minnaert method with slope correction (Goslee et al. 2011) and used the ASTER digital elevation model (ASTGTM2 S01E034), which we resampled to match the resolution of the Landsat images.

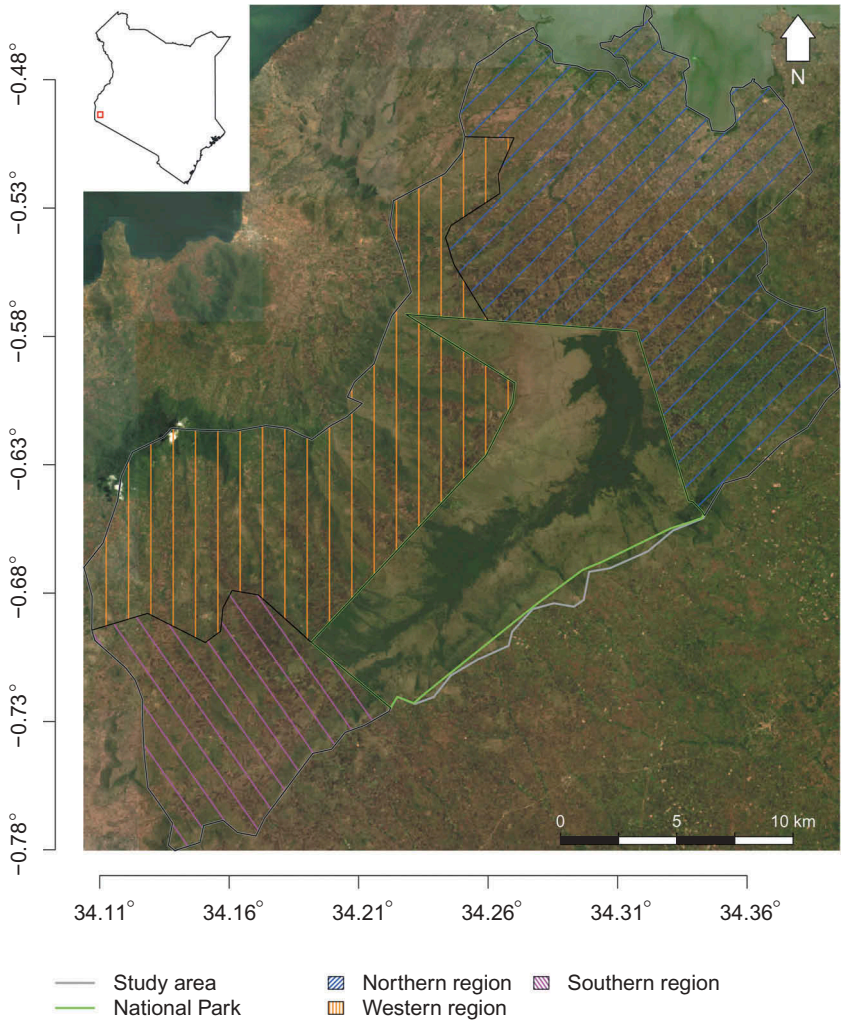

Figure 1. In left upper area: location of the Lambwe valley in Kenya. Study area with the Landsat image of 2014 in the background, showing the subregions used in the analysis.

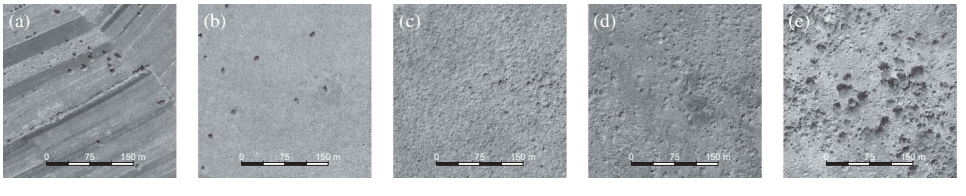

Figure 2. Examples of 300 m × 300 m patches of LULC classes. (a): *agriculture*, (b): *savanna*, (c): *dense shrub*, (d): *light forest,* and (e): *dense forest*. WorldView-1 images, courtesy of DigitalGlobe Foundation.

2.3. *Choosing the training data*

In this study, we classified LULC with Random Forests (RF) (Breiman 2001), a classifier often used in remote sensing. It has been reported to perform best over a wide variety of data sets (121 data sets, with RF surpassing 90% of the maximum accuracy in 84% of the data sets) (Fernández-Delgado et al. 2014). A major challenge in classification is the acquisition of reliable training data. In particular for historical images, it is not always possible to get direct information about past LULC. To overcome this issue, we inferred past LULC classes from unchanged areas in recent satellite images. For defining unchanged areas we used the change-vector analysis (Bovolo, Marchesi, and Bruzzone 2012). We have already tested this approach in the Matobo National Park in Zimbabwe (Scharsich et al. 2017).

For 2014 we could identify training data from WorldView-1 high-resolution images (0.5 m × 0.5 m) provided by the DigitalGlobe Foundation and from own field visits. The training data for 2002 and 1984 were derived by transferring class labels of unchanged pixels (see below for details) from 2014 to the corresponding pixels in the older images.

2.3.1. *Training data for 2014*

To infer the training data for 2014 we identified five main LULC classes in the Lambwe valley based on field visits and visual interpretation of WorldView-1 high-resolution images:

agriculture contains fallow land and cropped land
savanna dominant in the National Park, far from the river
dense shrub dominant in the National Park in the vicinity of the river
light forest dominant on the hills
dense forest exists only on the top of the hills

For each of these five classes, we identified in the WorldView-1 high-resolution images 1000 pixels and used them as training data for the classification of the Landsat image of 2014.

2.3.2. *Selection of predictors*

Our first classification results (data not shown) suggested that the spectral information contained in the Landsat images (bands 2–5 and 7 for Landsat 5, and 7 and 3–7 for Landsat 8) could be insufficient to classify LULC in the study area accurately. Indeed, the agricultural land is highly fragmented and fallow fields covered by grass can be easily confounded with savanna. Therefore, we derived further predictors for the classification, namely two vegetation indices, NDVI and EVI, altitude, slope, aspect, and features derived by mathematical morphology.

Since we decided to avoid the blue band in Landsat images (contaminated by artefacts, see section 2.2) we calculated the two-band EVI as follows:

$$\mathrm{EVI} = 2.5\frac{N - R}{N + 2.4R + 1}, \tag{1}$$

where N is the near-infrared band and R the red band, respectively. Jiang et al. (2008) showed this to be an appropriate way to replace the information of the blue band and obtain results comparable to EVI values calculated with the blue band.

The altitude, slope, and aspect were calculated from the resampled aster digital elevation model. This information should facilitate the classification of *dense forest* and *light forest* since they are dominant at a higher elevation.

Predictors derived by mathematical morphology allow to take neighbourhood relationships of pixels into account and find specific shapes in an image. Because the agricultural area consists of rectangular fields, we chose a 3×7 rectangle as a fundamental structural element and applied a closing operation to Landsat bands, NDVI and EVI:

$$\text{closing}(b(p)) = \text{erosion}(\text{dilate}(b(p)), \tag{2}$$

where

$$\text{dilate}(b(p)) = \max_{p_i \in w_{3\times7}(p)} b(p_i)$$

$$\text{erosion}(b(p)) = \min_{p_i \in w_{3\times7}(p)} b(p_i),$$

where $w_{3\times7}(p)$ is the set of pixels contained in a 3×7 pixels window (structural element) around a pixel p and $b(p)$ is the corresponding band with the surface reflectance or vegetation index value. We used closing to separate *savanna* from *agriculture*, in particular, the fallow land, effectively. In summary, we used 17 predictors for classification, namely five surface reflectance bands, EVI, NDVI, altitude, slope, aspect, and seven predictors by applying mathematical morphology to the five surface reflectance bands, NDVI and EVI.

2.3.3. *Training data for 2002 and 1984*

We obtained reliable training data by transferring the LULC class labels between pixels of a recent image to older ones under two conditions. First, the LULC class in 2014 should be certain. Second, the pixel's class should remain the same between images. The first condition can be verified by using the probabilities that the RF classifier attributed to different LULC classes. For the second condition we measured the degree of change by calculating the length of the change vector ρ:

$$\rho = \left(\sum_{b=1}^{B} (X_{b,2} - X_{b,1})^2 \right)^{1/2}, \tag{3}$$

where $X_{b,i}$ is the surface reflectance of band b in image i and B is the number of bands compared in the two Landsat images (Bovolo, Marchesi, and Bruzzone 2012). We compared images in 1984 with 2014 and in 2002 with 2014.

We defined 1000 training pixels for each LULC class in 1984 and 2002 in several steps and took pixels satisfying the following criteria:

(1) The classifier voted with at least 90% for the LULC class.
(2) The change vector ρ is smaller than the 10%-quantile of the distribution of all change vectors.

If at least 1000 pixels satisfied these criteria, we selected 1000 pixels randomly. If not, we selected 1000 pixels satisfying:

(1) The classifier voted with at least 60% for the LULC class.
(2) The change vector ρ is smaller than the 25%-quantile of the distribution of all change vectors.

The similarity of the training data calculated by the proximity measure of RF can be used as a marker for the quality of the training data. The proximity is calculated for each pair of training samples by counting the frequency they both end up in the same node of the trees of RF (Breiman 2003). In our training data, the similarities within each LULC class are two order of magnitudes higher than the similarities between different LULC classes (see Supplementary Material, Section 2.1) for each of the three years. This shows the suitability of our training data sets. Only data points belonging to the same LULC class show strong similarities, whereas data points of different origin are dissimilar.

The training data for 1984 and 2002 were used for classification with RF in the same way as the training data for 2014.

2.4. *Classification of land use and land cover*

We used 1000 training pixel for each LULC class to train an RF classifier for each year separately. The trained RF was used to classify all remaining pixels (about 610,000) of the corresponding images. To run the RF classifier it is necessary to set three parameters, namely the number of predictors sampled randomly at each tree node from the set of possible predictors (*mtry*), the minimal size of the terminal nodes (*nodesize*) and the number of trees in the forest (*ntree*). We used for all parameters the standard values, which were $mtry = 4$ (i.e. the square root of possible features) (Breiman 2003) $ntree = 1000$ and $nodesize = 1$. An overfitting of RF is hardly possible and, as long as *ntree* is large enough, the variability of its performance is quite low for low dimensional data (Breiman 2001; Strobl, Malley, and Tutz 2009).

Subsequently, we used the classified images for a post-classification comparison. This method allows us to understand specific changes in different subregions of our study area pixel by pixel. The interpretation of the results of this method is quite intuitive and easy to communicate.

2.5. *Determining the accuracy of the classification*

To assess the accuracy of our classification we used a fivefold stratified cross-validation that we repeated 20 times. Stratification means that in each sampled subset the proportions of LULC classes were equally distributed. For this purpose, we divided the training data into five parts. With each part once left out as a test set, the other four parts were used to train the RF. Thus, we obtained the classification error for each pixel and could calculate the accuracy. By repeating it 20 times we could infer a median classification accuracy (1-classification error) and its standard deviation. The classification accuracy (c) was calculated by:

$$c = \frac{\sum_{j=1}^{5} n_{j,j}}{\sum_{k,j=1}^{5} n_{k,j}} \tag{4}$$

where $\mathbf{n} = (n_{k,j})$ is the (5 × 5)-confusion matrix. This matrix provides the number of pixels in predicted classes (rows) versus given classes (columns). This means, the diagonal elements display correctly classified pixels, and the rest of the matrix shows misclassified ones.

In addition, we calculated the median of the per-class effectiveness (e.g. Sokolova and Lapalme 2009) to evaluate whether some LULC classes were better identified than others. It is defined as

$$E_i = \frac{n_{i,i} + \sum\limits_{\substack{k,j=1 \\ k,j \neq i}}^{5} n_{k,j}}{\sum\limits_{k,j=1}^{5} n_{kj}} \tag{5}$$

where E_i is the per-class effectiveness for one of the five LULC classes i (i = 1, ..., 5), and $\mathbf{n} = (n_{k,j})$ the confusion matrix as in Equation (4).

Following Pontius and Millones (2011) we decided to avoid the commonly used κ, which is more 'redundant and misleading' than helpful, and instead, calculated the median of the two error measures 'allocation disagreement'

$$A = \sum_{g=1}^{J} \min\left[\left(\sum_{i=1}^{J} p_{i,g}\right) - p_{g,g}, \left(\sum_{j=1}^{J} p_{g,j}\right) - p_{g,g}\right] \tag{6}$$

and 'quantity disagreement'

$$Q = \frac{1}{2}\sum_{g=1}^{J}\left|\left(\sum_{i=1}^{J} p_{i,g}\right) - \left(\sum_{j=1}^{J} p_{g,j}\right)\right| \tag{7}$$

where $\mathbf{p} = (p_{i,j})$ is the estimated population matrix. Because we sampled each class equally, it can be calculated as

$$p_{i,j} = \frac{n_{i,j}}{J\sum_{j=1}^{J} n_{i,j}}, \tag{8}$$

where J is the number of determined clusters ($J = 5$ in our case) and $\mathbf{n} = (n_{k,j})$ the confusion matrix as in Equation (4).

These error measures sum up to the error of the classification and describe its different aspects. The quantity disagreement explains the error caused by an under- or overestimation of the total number of pixels belonging to the LULC classes, whereas the allocation disagreement sets the portion of LULC classes as given and evaluates only the error by suboptimal assignment of the LULC classes to the pixels. With these two different error measurements, it is possible to decide separately, whether single pixel-to-pixel changes and/or changes of portions of LULC classes are classified reliably.

In addition to the classification accuracy, we calculated the (b-ary) Shannon Entropy H (Shannon 1948) for each pixel x to derive a measure for the internal model accuracy as used by Hengl et al. (2017), for example. It is calculated by:

$$H(x) = -\sum_{k=1}^{5} p_k(x) \times \log_5(p_k(x)), \tag{9}$$

where p_k is the probability of the pixel x to belong to LULC class k, which corresponds to the frequency of votes in RF. The logarithm to base 5 is used to constrain $H \in [0, 1]$. H indicates how certain the RF classifier was about the attributed LULC class. In case of a very sure classification, all trees of RF voted for the same LULC class and the Shannon Entropy Index is equal to zero. In case of an equal distribution of the votes among all LULC classes, H equals 1 indicating a maximum of uncertainty for the attributed LULC class.

All calculations were done in R (R Core Team 2016). For georeferencing and topographical correction we used the add-on package landsat (Goslee et al. 2011); for resampling the add-on package raster (Hijmans 2016); for the methods of mathematical morphology the add-on package mmand (Clayden 2017) and for classification the add-on packages randomForest (Liaw and Wiener 2002) and caret (Kuhn 2016).

3. Results

3.1. *Classification accuracy*

The classification results for all three Landsat images are reliable since the median classification accuracy for each of them is high (>97%) and its standard deviation low (<0.2%) (Table 1). In addition, for each LULC class, the per-class effectiveness is over 99%, except for *agriculture* and *savanna* in the year 1984 (about 97%). This points to the problems we had in differentiating these two LULC classes (see Section 2.3.2).

Since the classification accuracy is high, the errors quantity disagreement Q and allocation disagreement A are low. In any of the three years A represents at least 70% of the error of the classification, so the error of assigning single pixels to a wrong LULC class, given the proportions of LULC classes, is more prominent than the error concerning the prediction of the proportions itself.

Table 1. Median classification accuracy from the fivefold cross validation (20 repetitions) with standard deviations in parenthesis. E_i: per-class effectiveness for *agriculture* ($i = 1$), *dense forest* ($i = 2$), *savanna* ($i = 3$), *dense shrub* ($i = 4$) and *light forest* ($i = 5$), c: classification accuracy, Q: quantity disagreement and A: allocation disagreement.

	1984	2002	2014
c (%)	97.91 (0.11)	99.28 (0.04)	99.72 (0.02)
E_1 (%)	97.98 (0.16)	99.65 (0.03)	99.58 (0.03)
E_2 (%)	99.39 (0.06)	99.39 (0.06)	100 (0.01)
E_3 (%)	97.02 (0.15)	99.61 (0.02)	99.64 (0.04)
E_4 (%)	99.37 (0.07)	99.31 (0.06)	99.99 (0.01)
E_5 (%)	99.71 (0.07)	99.85 (0.01)	99.89 (0.03)
Q (%)	0.62 (0.07)	0.18 (0.03)	0.04 (0.02)
A (%)	1.46 (0.12)	0.54 (0.04)	0.24 (0.02)

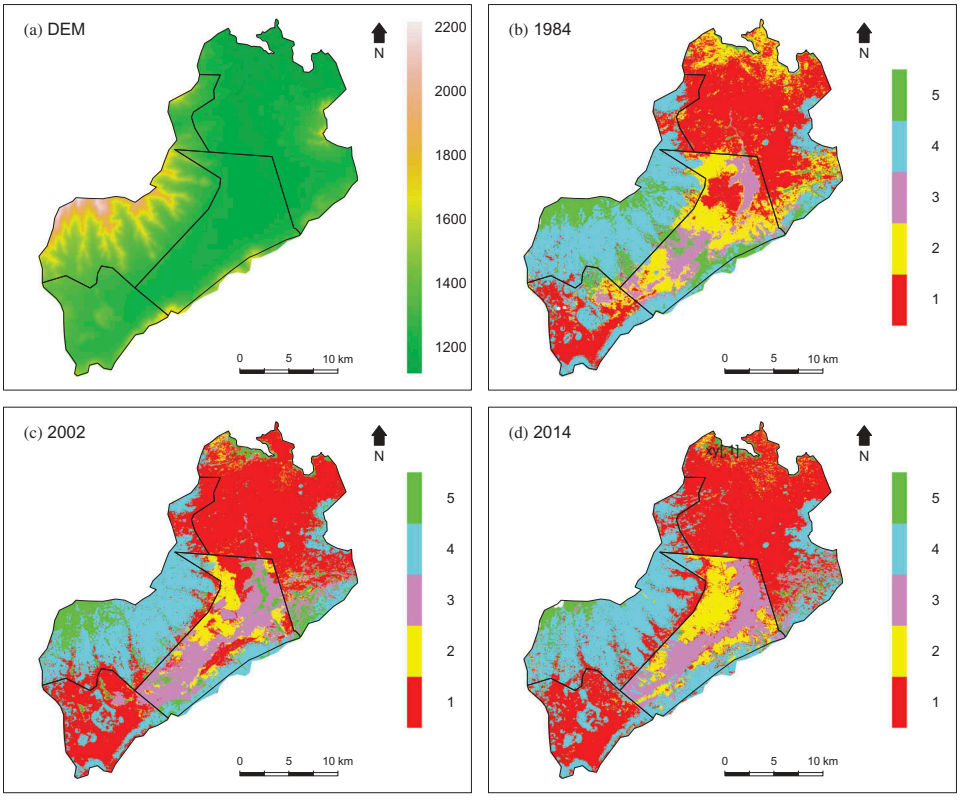

Figure 3. Study area in 0°27'–0°47' S and 34°06'–34°24' 'E; (a): Elevation data (aster) in m, (b)–(d): Spatial distribution of five LULC classes in the study area for the years 1984, 2002 and 2014. The corresponding LULC classes are 1: *agriculture*, 2: *savanna*, 3: *dense shrub*, 4: *light forest* and 5: *dense forest*. The black lines show the boundaries of the national park and the subregions.

3.2. *Land use and land cover maps*

We found that certain LULC classes were prominent at specific elevations (Figure 3) as already indicated in their description (Section 2.3.1). Whereas the *dense forest* prevails at the highest altitude, the *light forest* is prominent on the lower hills. Outside the national park, the *agriculture* dominates the lower part of the valley. In the national park, *savanna* and *dense shrub* are the most frequent LULC classes. The latter can be found especially along the river.

To give an impression of how certain the RF classifier was about the LULC class in each pixel, we calculated the Shannon Entropy Index H for all three years (Figure 4). Note that H differs from the accuracy of the classification and reflects the distribution of the votes of the single decision trees inside RF. For a detailed look on the probabilities of RF for each pixel, see the Supplementary Material, Section 1.1.

Comparing Figures 3 and 4 we observe large values of H especially in regions with a highly structured LULC. These are not only regions with mosaic-like patterns of different LULC classes but also the transitions between LULC classes, for example from *dense shrub* to *savanna* in the national park. Often, these regions coincide with areas of

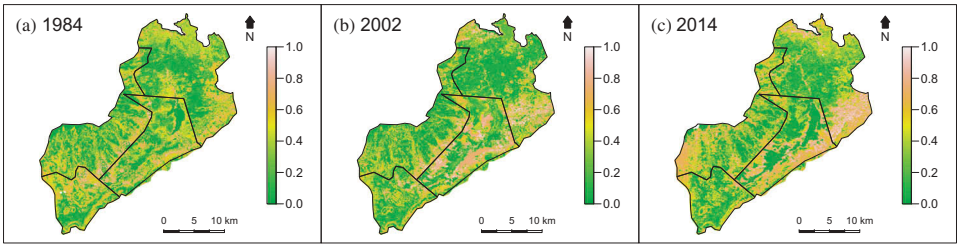

Figure 4. Study area in 0°27′–0°47′ S and 34°06'–34°24' E; Shannon Entropy Index for the years 1984 (a), 2002 (b), and 2014 (c).

LULC change. This indicates that a changing area or an area of transition might contain characteristics of different LULC classes. Thus, the RF classifier that bases its final decision for a LULC class on a majority vote could still contain information on the minor LULC classes in the distribution of the votes of single decision trees.

3.3. *Changes in Ruma National Park and subregions*

The general change patterns in the Lambwe valley can be inferred from Figure 3 already. One of the prominent changes is the decrease of *dense forest* in the western region where it was pushed to the highest elevations. Similarly, *light forest* also retreated to higher altitude. *Savanna* and *dense shrub* disappeared in the southern region that is now dominated by two LULC classes, namely *agriculture* and *light forest*. The same change can be observed along the shore, where *savanna* almost vanished until 2014. In contrast, the changes in the Ruma National Park are quite diverse and do not follow a clear pattern.

To get a more detailed overview of the changes we calculated the differences between the proportions of LULC classes for all three possible combinations of the LULC maps (Figure 5). The exact numbers can be found in the Supplemental Material, Section 2.2. The national park stands out with its highly dynamic LULC. The main reason for the variation is probably fires occurring frequently. Indeed, 30 fires were detected in

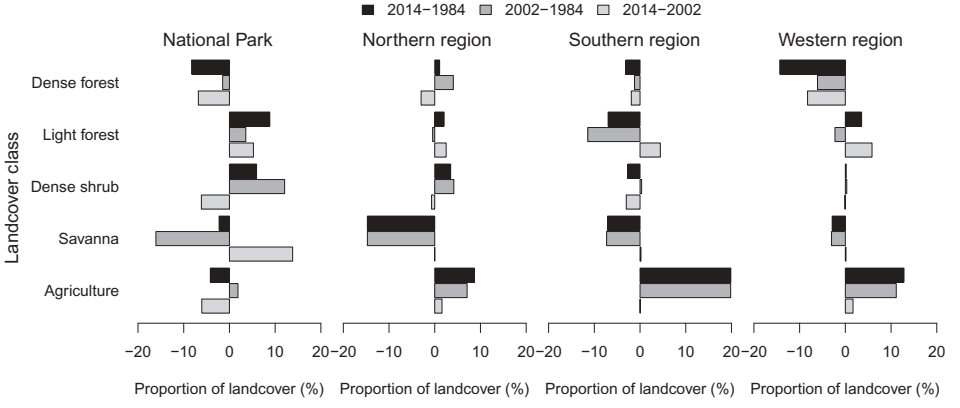

Figure 5. Differences of proportions of LULC classes between the three years 1984, 2002 and 2014 for the national park and the three subregions.

2008 and 57 in 2009 (Kariuki et al. 2012). The LULC class *agriculture*, detected in the national park, probably shows a misclassification of former burned areas. We examined the original Landsat images (bands: blue, green, red) visually in the areas classified as *agriculture* and they fit the expected black impression of burned land. The origin of the fires is diverse; some of them are accidental and caused by farmers who prepare their fields before the planting season by burning the farm litter, for example. However, some fires are started on purpose, mainly by poachers, who know about the subsequent green flush and its attractiveness to grazers (Kariuki et al. 2012). The only stable development in the national park seems to be the increasing proportion of *light forest* (1984–2014: $+8.8\%$) and the decrease of *dense forest* (1984–2014: -8.3%).

In contrast to the national park, all subregions surrounding the park have one main change in common, namely an increase of *agriculture* and a decrease of *savanna*. This development occurred between 1984 and 2002 mainly, probably caused by the strong population growth in this time. The northern regions were already dominated by *agriculture* and *savanna* in 1984. Thus, the main changes there concerned these two LULC classes ($+8.7\%$ and -14.7% for *agriculture* and *savanna*, respectively). By contrast, in the southern and the western subregions, the proportions of the other LULC classes also changed.

The changes in the southern subregion occurred between 1984 and 2002 mainly and consist of an additional decrease of *light forest* (-11.4%). The loss of *light forest* is concentrated on the central hill of the subregion, whereas the loss of *savanna* (1984–2002: -7.3%) is located in the southern end of the national park. In 1984 the vegetation types of the national park continued outside the borders, following the river. In 2002 this vegetation, mainly *dense shrub*, still existed in parts. By contrast, in 2014 hardly anything remained and the limits of the national park constitute a sharp border (Figure 3).

For the western subregion, *dense forest* is an additional important factor. While parts of the new *agriculture* area (1984–2002: $+11.1\%$) is gained by *savanna* losses (1984–2002: -3%), the main part is contributed by decreasing *dense forest* (1984–2002: -6.1%). Between 2002 and 2014 we could find an additional development. Although changes in *agriculture* and *savanna* are rather negligible, there is a tendency of *dense forest* (2002–2014: -8.3%) being replaced by *light forest* (2002–2014: $+5.8\%$). We attribute this to the proximity of *dense forest* to intensively used areas (i.e. *agriculture*). Therefore, the influence on the forest by collecting firewood, for example, increases and results in a conversion to *light forest*. In summary, we see an increase of *agriculture* starting in the lower part of Lambwe valley where it replaces *savanna*, and continuing by climbing up the hills where it causes deforestation in *light forest*. These changes are accompanied by a conversion of *dense forest* to *light forest*.

4. Discussion

Today's population in the Lambwe valley is around 60,000 (The World Vision, personal communication). Comparable historical numbers are lacking because the Lambwe valley does not represent a defined administrative district and calculations based on censuses are not comparable. However, the published population growth rates can be used to give an impression of the development in this region (Muriuki et al. 2005). In the 1960s

the population grew fastest (7.2%), mainly induced by a settlement initiative of the Kenya Government and initiatives to control the tsetse fly as reported by the same authors. This rapid population increase, rich soils and climatic conditions favourable for agriculture in the valley have promoted intensive agricultural practices. Indiscriminate clearing of forests to create land for settlement, crop production and cattle grazing has followed, leading to serious environmental degradation and species losses (Njoka et al. 2003). Thus, the successful control of the tsetse fly opened up an area with fertile soil; however, it also led to an increased pressure on the whole valley.

An overview of the conversion from natural to anthropogenically dominated ecosystems is depicted in Figure 6. The vegetation of the Ruma National Park can be seen as the former main land cover type in the valley, whereas the forest with diverse tree species represents the original vegetation on the hills. By cattle grazing and forest burning the land changes and is now often threatened by erosion, leading to a higher sediment load in the river and contributing to water quality issues in Lake Victoria.

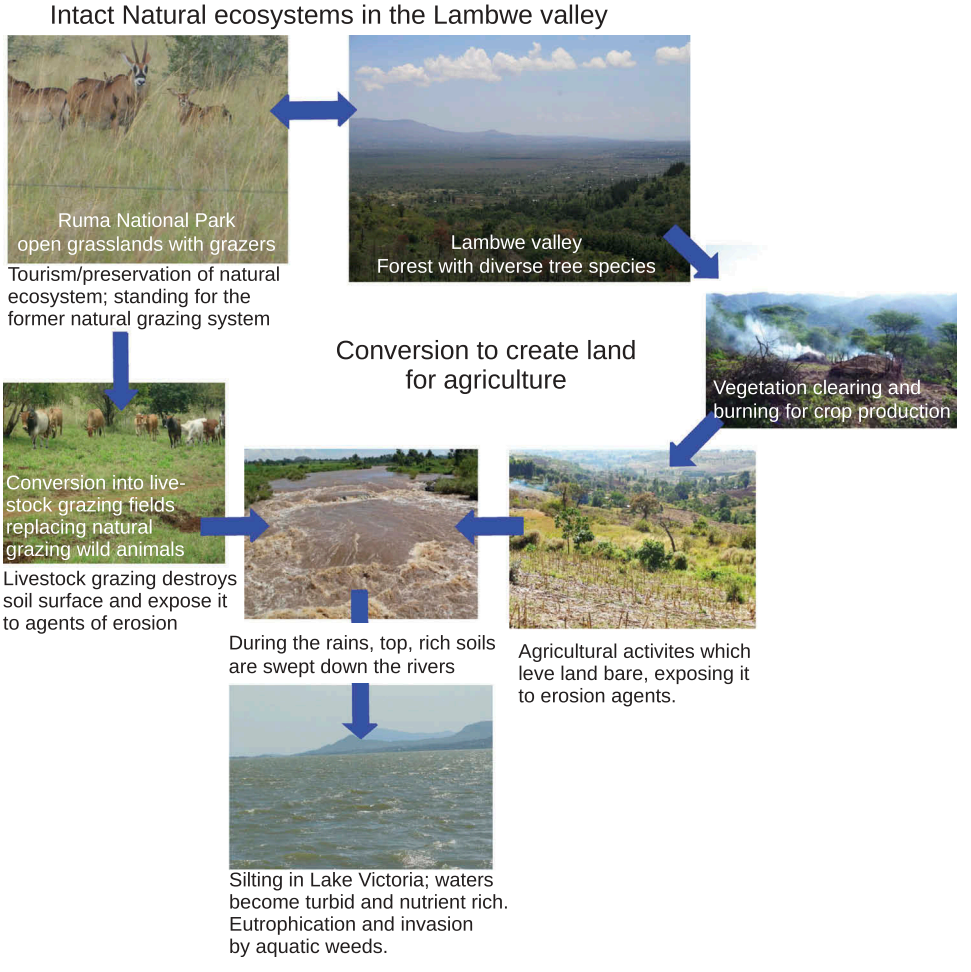

Figure 6. Overview of some ecological processes occurring in Lambwe valley.

Muriuki et al. (2005) reported that between 1948 and 1993, land under cultivation increased by about 22% to reach about 32%, whereas grasslands and woodlands decreased. Although we defined our LULC classes differently from their study, the overall increase of agricultural areas seems to continue. Indeed, we observed a decrease of *savanna* (-8%) and *dense forest* (-6%) in favour of *agriculture* ($+12\%$) between 1984 and 2014 outside the national park.

Ruma National Park is situated in the middle of the Lambwe valley covering about 120 km^2 and, therefore, occupying potential land for cropping. This land use conflict decreases the acceptance of the park by the local communities. Thus, there have been attempts to degazette the park and to convert it to settlement and agricultural land (Kariuki et al. 2012). Further reasons for conflicts are the prohibition of collecting fire wood and hunting game in the national park. In addition, the park acts as a habitat and dispersal centre for tsetse flies, as well as a refuge for trypanosomiasis among the wild animals, from where tsetse flies spread and attack livestock (Muriuki et al. 2005).

There is no evidence of illegal agricultural activities in the national park, however, poaching and illegal collection of fire wood still occur despite the fact that the park is almost entirely fenced (Kariuki et al. 2012). The growing population could also have an indirect impact on the park. With extending agricultural activities towards the park borders (Figure 3), the frequency of fires in the national park could increase, given that farmers often rely on fire to clear litter and debris in preparation for cropping. To reduce accidental fires, the park management regularly installs fire breaks along the park boundary, in addition to increasing awareness through education among the local communities (Kariuki et al. 2012). The goal of the park management is to follow a predefined burning schedule, which sustains the open grasslands and also promotes the encroachment of woody vegetation and bushland. The latter is important for several grazer communities, especially the roan antelope, for which the Ruma National Park is the last sanctuary in Kenya. Roan antelope prefers the boundary between savanna and shrubland/woodland as habitat (De 1974).

In general, an increase in the agricultural area and in particular the expansion of agriculture (including pasture) towards steeper hill slopes can lead to soil and vegetation degradation (through overgrazing, for example) and soil erosion. During the rainy periods, large volumes of rich top soil can be eroded and washed into the rivers and eventually into Lake Victoria (Figure 6). For the Lambwe valley, the consequences are already noticed by local communities, who observe an increase in the degraded areas that are now unsuitable for any agricultural use. This is in line with our observations. Indeed, we found locations with bare land, probably caused by soil erosion, on the WorldView-1 images. Turbidity and influx of nutrient loads affect the water quality in Lake Victoria and its biodiversity and increases the pressure on the endemic species and the ecological structure in the lake. Eutrophication as a result of influx of nitrogen- and phosphorus-rich soils is one of the consequences, while invasive species such as water hyacinth (*Eichornia crassipes*) have taken advantage of the fertile lake waters (ILEC 2005).

Njoka et al. (2003) reported that outside Ruma National Park, essential ecosystem services such as water availability, soil quality and biodiversity have decreased significantly. Also, valuable tree species have vanished from the area due to over-exploitation, while soil erosion, soil degradation and river siltation were on the rise. Our study

quantifies the recent LULC changes and shows that agricultural expansion continues on steeper slopes. Ruma National Park seems to be well protected and probably still fulfils its purpose of protecting biodiversity despite the lack of acceptance among the neighbouring communities. This negative attitude towards the park could flame conflicts in the future, particularly, if the pressure to find cultivable land continues to increase.

5. Summary and conclusions

In this study, we investigated changes of LULC in and around Ruma National Park in Kenya situated in Lambwe Valley near Lake Victoria. We classified three Landsat images with a time span of 30 years with a supervised classification algorithm using high-resolution images of WorldView-1 to create training data. Inside the national park, we did not observe any change of LULC that could be attributed to illegal practices. This is likely because of the perimeter fence around the park and the effective surveillance by the park rangers. Outside the national park, LULC changes point towards a decrease in dense forest and savanna and an increase in agricultural areas that are now expanding up the hill slopes. We attribute these changes to the population growth in the Lambwe valley and thus to a rising demand for food production. Cultivation of steeper slopes increases the risk of soil erosion, while the ongoing deforestation may lead to a complete loss of forests at the higher altitudes through conversion into cropping areas and pasture.

Acknowledgments

This research was supported by the German Federal Ministry of Education and Research (grant number FKZ 01UC1201). The first author acknowledges the support by the German Research Foundation. The data ASTER GDEM is a product of METI and NASA. The Landsat surface reflectance data were downloaded from earthexplorer.usgs.gov and are courtesy of the U.S. Geological Survey. The WorldView-1 images were provided by the DigitalGlobe Foundation. We thank Luke Lukaria Kanyi for his support concerning local details about Ruma National Park.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [HU 636/16-2] and by the Bundesministerium für Bildung und Forschung [FKZ 01UC1201].

ORCID

Christina Bogner http://orcid.org/0000-0003-4495-0676

References

Allsopp, R., and D. A. Baldry. 1972. "A General Description of the Lambwe Valley Area of South Nyanza District, Kenya." *Bulletin of the World Health Organization* 47 (6): 691–697.

Awange, J. L., O. Aseto, and O. Ong'ang'a. 2004. "A Case Study on the Impact of Giraffes in Ruma National Park in Kenya." *Journal of Wildlife Rehabilitation* 27 (2): 16–21.

Bayarsaikhan, U., B. Boldgiv, K.-R. Kim, K.-A. Park, and D. Lee. 2009. "Change Detection and Classification of Land Cover at Hustai National Park in Mongolia." *International Journal of Applied Earth Observation and Geoinformation* 11 (4): 273–280. doi:10.1016/j.jag.2009.03.004.

Bovolo, F., S. Marchesi, and L. Bruzzone. 2012. "A Framework for Automatic and Unsupervised Detection of Multiple Changes in Multitemporal Images." *IEEE Transactions on Geoscience and Remote Sensing* 50 (6): 2196–2212. doi:10.1109/TGRS.2011.2171493.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Breiman, L. 2003. *Manual–Setting Up, Using, and Understanding Random Forests V4.0*.

Clayden, J. 2017. "Mmand: Mathematical Morphology in Any Number of Dimensions." *R package version 1.5.2*. https://cran.r-project.org/package=mmand

De, P. U. 1974. "Habitat-Preference in South African Antelope Species and Its Significance in Natural and Artificial Distribution Patterns." *Koedoe* 17: 1.

DeFries, R., A. Hansen, B. L. Turner, R. Reid, and J. Liu. 2007. "Land Use Change around Protected Areas: Management to Balance Human Needs and Ecological Function." *Ecological Applications* 17 (4): 1031–1038. doi:10.1890/05-1111.

Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim. 2014. "Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?" *Journal of Machine Learning Research* 15: 3133–3181.

Fox, J., J. Krummel, Y. Sanay, E. Methi, and N. Pogre. 1995. "Land Use and Lanscape Dynamics in Nothern Theiland: Assessing Change in Three Upland Watersheds." *Ambio* 24 (6): 328–334.

Goslee, Sarah C. 2011. "Analyzing Remote Sensing Data in R: The Landsat Package." *Journal of Statistical Software* 43 (4): 1–25. doi:10.18637/jss.v043.i04.

Hengl, T., J. M. de Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, et al. 2017. "SoilGrids250m: Global Gridded Soil Information Based on Machine Learning." *PLOS ONE* 12 (2): 1–40. doi:10.1371/journal.pone.0169748.

Hijmans, R. J. 2016. "Raster: Geographic Data Analysis and Modeling." *R package version 2.5-8*. https://CRAN.R-project.org/package=raster

ILEC. 2005. *Managing Lakes and their Basins for Sustainable Use: A Report for Lake Basin Managers and Stakeholders*. International Lake Environment Committee Foundation: Kusatsu, Japan.

Jiang, Z., A. R. Huete, K. Didan, and T. Miura. 2008. "Development of a Two-Band Enhanced Vegetation Index without a Blue Band." *Remote Sensing of Environment* 112 (10): 3833–3845. doi:10.1016/j.rse.2008.06.006.

Juma, D. W., H. Wang, and F. Li. 2014. "Impacts of Population Growth and Economic Development on Water Quality of a Lake: Case Study of Lake Victoria Kenya Water." *Environmental Science and Pollution Research* 21 (8): 5737–5746. doi:10.1007/s11356-014-2524-5.

Kariuki, A., J. Wambua, F. Lala, S. Ngene, T. Ikime, I. Makau, and K. L. Rukaria. 2012. "Ruma National Park, Management Plan, 2012-2017." In *Kenya Wildlife Service*.

Kuhn, M. 2016. *Classification and Regression Training*. R package.

Liaw, A., and M. Wiener. 2002. "Classification and Regression by randomForest." *R News* 2 (3): 18–22.

Lu, D., P. Mausel, E. Brondízio, and E. Moran. 2004. "Change Detection Techniques." *International Journal of Remote Sensing* 25 (12): 2365–2401. doi:10.1080/0143116031000139863.

Muriuki, G. W., T. J. Njoka, R. S. Reid, and D. M. Nyariki. 2005. "Tsetse Control and Land-Use Change in Lambwe Valley, South-Western Kenya." *Agriculture Ecosystems & Environment* 106 (1): 99–107. doi:10.1016/j.agee.2004.04.005.

NASA. n.d. "Landsat Science. Landsat 5." Accessed 3 February 2019. https://landsat.gsfc.nasa.gov/landsat-5/

Njoka, T. J., G. Muriuki, R. Reid, and D. Nyariki. 2003. "The Use of Sociological Methods to Assess Land-Use Change: A Case Study of Lambwe Valley, Kenya." *International Journal of Sociology* 7: 181–185.

Pontius, R. G., and M. Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing* 32: 4407–4429. doi:10.1080/01431161.2011.552923.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Scharsich, V., K. Mtata, M. Hauhs, H. Lange, and C. Bogner. 2017. "Analysing Land Cover and Land Use Change in the Matobo National Park and Surroundings in Zimbabwe." *Remote Sensing of Environment* 194: 278–286. doi:10.1016/j.rse.2017.03.037.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (3): 379–423. doi:10.1002/bltj.1948.27.issue-3.

Sokolova, M., and G. Lapalme. 2009. "A Systematic Analysis of Performance Measures for Classification Tasks." *Information Processing & Management* 45 (4): 427–437. doi:10.1016/j.ipm.2009.03.002.

Strobl, C., J. Malley, and G. Tutz. 2009. "An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests." *Psycological Methods* 14 (4): 323–348. doi:10.1037/a0016973.

U.S. Geological Survey. n.d. "EarthExplorer." https://earthexplorer.usgs.gov/

United Nations, Department of Economic and Social Affairs, Population Division. 2017. *World Population Prospects: The 2017 Revision*.

Wellde, B. T., M. J. Reardon, D. A. Chumo, D. Waema, D. H. Smith, D. Koech, T. Siongok, and L. Wanyama. 1989. "Demographic Characteristics of the Lambwe Valley Population." *Ann Trop Med Parasitol* 83 (Suppl 1): 29–42. doi:10.1080/00034983.1989.11812408.

World Health Organization. n.d. "Trypanosomiasis, Human African (Sleeping Sickness)." Accessed 3 February 2019. https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness)