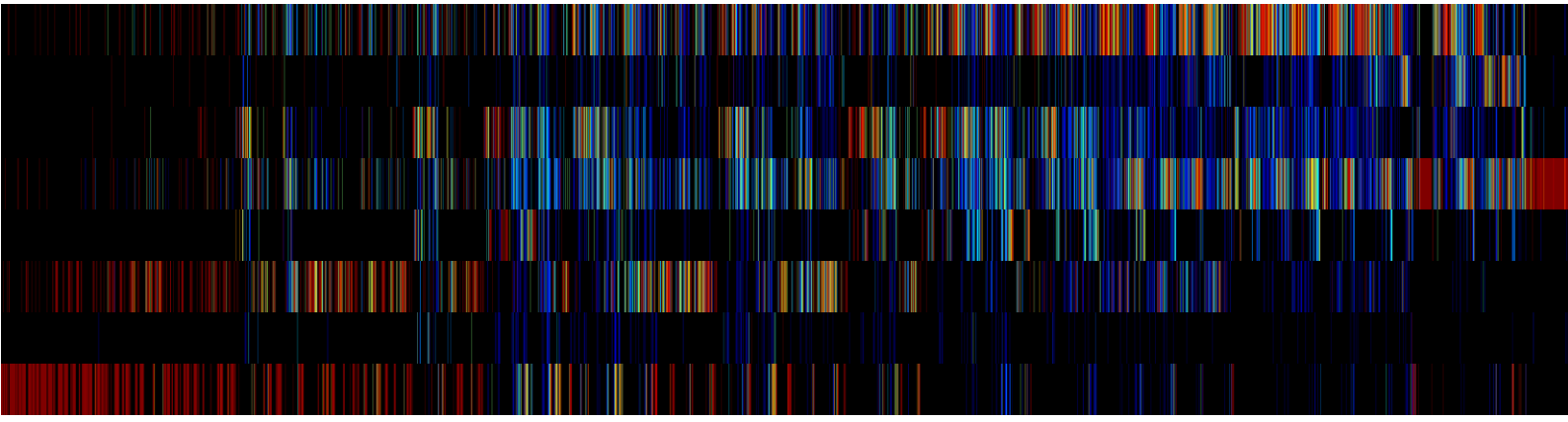
European Commission

JRC TECHNICAL REPORTS

A statistical analysis of the relationship between landscape heterogeneity and the quantization of remote sensing data

Antonia Degen, Christina Corbane, Martino Pesaresi, Thomas Kemper

2018



Joint Research Centre

EUR 29340 EN

Contact information
Name: Christina Corbane
Address: European Commission, Joint Research Centre, Space, Security and Migration (Ispra), Disaster Risk Management (JRC.E.1)
E-mail: christina.corban@ec.europa.eu
Tel.: +39 0332 78 3545

JRC Science Hub
https://ec.europa.eu/jrc

JRC111818
EUR 29340 EN

ISBN 978-92-79-92990-8
ISSN 1831-9424
doi:10.2760/731774

How to cite: Degen, A., Corbane, C., Pesaresi, M. and Kemper, T., A statistical analysis of the relationship between landscape heterogeneity and the quantization of remote sensing data, EUR 29340 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-92990-8 (pdf), doi:10.2760/731774 (online), JRC111818.

Abstract

This report addresses the investigation of the relationship between the landscape heterogeneity and the sequencing of remote sensing imagery for the purpose of better understanding the parameters of the Symbolic Machine Learning developed within the Global Human Settlement Layer project. To address this issue statistical regression analysis was conducted between the sequences derived from the Landsat satellite data and different landscape metrics derived from land cover maps. The results show that only the Relative Patch Richness influences the number of sequences for different levels of image reduction levels. The Shannon Landscape Diversity Index seems to be related to the Number of Sequences in the image until a certain Level of Quantization that may be an indicator of the optimal parameter for the sequencing of the input satellite data. These results represent a good step forward in the attempt to automatize the parameters set of the Symbolic Machine Learning classifier.

Table of contents

1. Introduction

This report addresses the investigation of landscape heterogeneity and its influence on the data reduction process and the sequencing of remote sensing imagery for the purpose of better understanding the parameters of the Symbolic Machine Learning (SML) developed within the Global Human Settlement Layer (GHSL) project.

The SML classifier for the detection of build-up areas for the GHSL is a product of the Joint Research Centre of the European Commission. It involves in an initial step a data reduction or image quantization. The quantization focuses on the reduction of the satellite imagery input. Quantization is the process of converting a continuous range of values into a finite range of discrete values. This is necessary because remote sensing data is available in high temporal, spatial and spectral resolutions. Therefore quantization is required in order to produce compact features that can be used in retrieval and classification systems. The number of levels selected for quantizing the image is often determined through trial and error and through experimentation. Determining an appropriate quantization requires an understanding of the heterogeneity of image and the amount of information contained within the satellite imagery.

Due to the actual heterogeneity of land surfaces, the satellite detects heterogeneous signals in each band. From this assumption follows that an image pixel has a high spectral heterogeneity when the landscape is heterogeneous. The analysis performed in this report is based on the following assumptions:

- There is a relationship between the quantization of remote sensing data and landscape heterogeneity.
- There is a relationship between the sequences derived from the quantization followed by the encoding of the information into vector of image features and landscape heterogeneity.

The landscape heterogeneity is usually measured with different indices and metrics in the scientific field of ecology these metrics can give information about the landscapes diversity, fragmentation and configuration.

To investigate these assumptions, a set of experiments has been developed using as input data Landsat 8 imagery and a land cover map. The search of a statistical regression analysis with the given input data is due to the attributes of the data, e.g. distribution, scale etc., not a trivial task. After many attempts, the Generalized Linear Model (GLM) with an inverse Gaussian family was selected. This model gives good results for the Number of Sequences and the selected landscape metrics.

2. Overview of the SML classifier for the classification of remote sensing data

2.1 Framework and aim of the SML

The SML is the base of the GHSL workflow. It was designed to handle remote sensing big data. The term "Symbolic" refers to the data reduction. The quantization and data reduction take place in the first step. The image data is translated into sequences, which are used in the second step of the Association Analysis (Pesaresi et al., 2016b).

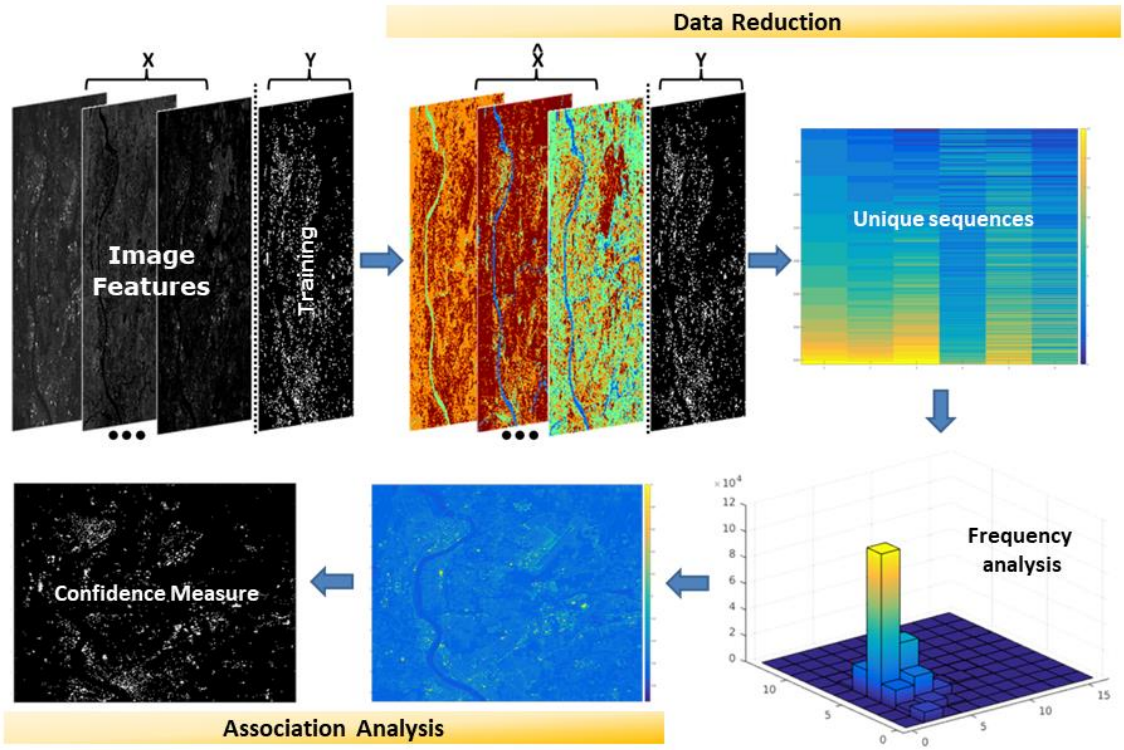


Fig. 1: Symbolic Machine Learning (SML) (Pesaresi et al., 2016b).

The Association Analysis is an automatic inferential engine constructing the rules that are associating the data sequences with a giving learning set (Corbane et al., 2017). Therefore the association between two different parts of the data, the X (Input data) and the Y (Known class membership, e.g. Settlement) is evaluated with a frequency-based supervised classification. The ENDI (Evidence-based normalized differential index), is a confidence measure for the data-abstraction association which is in the continuous $[-1, 1]$ range (Pesaresi et al., 2016b). This measure scores the data sequences in X according to the number of their occurrences in each reference class in the Y data. To obtain the classification results, thresholds are derived through the analysis of the ENDI distribution (Pesaresi et al., 2016a).

The focus of the SML lays on the discovery of associated rules that describe the full data range and not only sample data (Pesaresi et al., 2016b).

2.2 Quantization of radiometric features

The quantization is the first part of the SML and is the focus of this report. Main goal of the quantization is the data reduction. This goal is achieved by a translation of radiometric information into sequences. It is important to note that quantization should not be confused with classification.

D_{m*n*F} be a dataset with $m * n = mn$ spatial samples or pixels and F features or descriptors such as bands or derived image features. Let X_{m*n*F} be a two-dimensional data matrix, $X = [x_1, x_2, x_i, \dots, x_F]$ with F expression the number of used features and $x_i \in Z_+^{mn}$. Let $\hat{X}$ be the set of all of unique data instances of X, having a cardinality of $|\hat{X}| \leq mn$; this magnitude depends in the specific number of symbol is s_i used to encode the x_i values and on the number of features F. With these assumptions the Average Support of $\hat{X}$ can be estimated as:

$$supp_{\mu} = {|X|}/{|\hat{X}|}$$

The $|\hat{X}|$ influences the generalization and computational performances of the SML classifier. When the $supp_\mu$ value is too small it may lead to over-fitting in the association analysis of the SML (Pesaresi et al., 2016b). When the number of features F is given, the control of the $supp_\mu$ can be accomplished by controlling the Number of Sequences used for the encoding of X by the quantization:

$$X_{q_iF} = \left[\frac{\lfloor x + 0.5 \rfloor}{q_i : x \in \boldsymbol{x_i}}\right]_{mn*F}$$

With the Level of Quantization $q_i = max(\boldsymbol{x_i})/s_i$ and $i = 1, \ldots, F$, respectively. The Level of Quantization is empirically determined so that the $supp_\mu$ is in the range of 10^3 to 10^4. The number of sequences is denoted by $\hat{X}_{q_iF}$ (Pesaresi et al., 2016a).

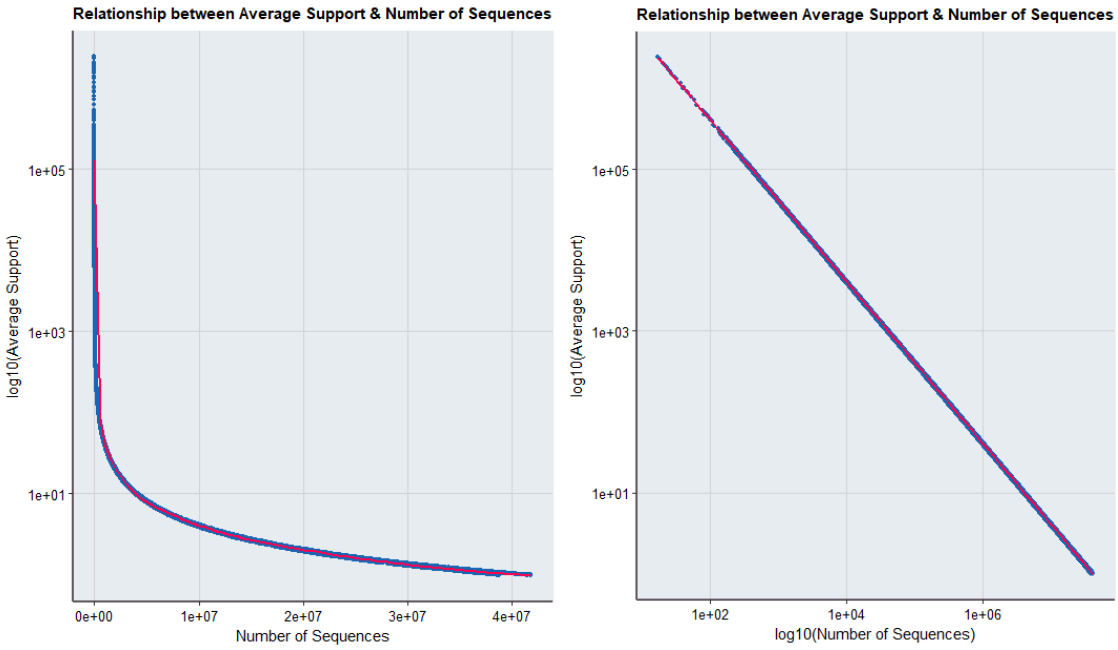


Fig. 2: Relationship between the Average Support and the Number of Sequences of 1148 Scenes and eight different Quantization Levels (32, 64, 128, 256, 512, 1024, 2048, 4096), a total of 9184 data points. Left: y-axis is logarithmic scaled; Right: both axis are logarithmic scaled.

The Average Support and the Number of Sequences are therefore highly linked, see fig. 2 (Pesaresi et al., 2016a).

For the SML, Pesaresi et al. (2016b) introduced the interestingness measure evidence-based normalized differential index (ENDI). It is a generalization of a measure with four main properties: firstly, it is an objective measure; secondly it is algorithmically fast; thirdly it is a descriptive measure that does not vary with the cardinality expansion; and fourthly it belongs to the measures of deviation from equilibrium, taking an equal number of positive and negative examples. The ENDI scores the data sequences $\hat{X}_{q_iF}$, according to the number of their occurrence in each reference class, and is therefore a necessary property for the decision or prediction making within the application of the SML (Pesaresi et al., 2016a):

$$\Phi_E^{ab} = \frac{\Phi_E^a + \Phi_E^b}{2}$$

With Φ_E^a , where the frequencies of the joint occurrences among X data instances and the positive and negative references instances respectively (Y^+ and Y^-):

$$\Phi_E^a(X, Y^+, Y^-) = \frac{f_{pos} - f_{neg}}{f_{pos} - f_{neg}}$$

With Φ_E^b , where probabilities are calculated as $p_{pos} = \frac{f_{pos}}{N_{pos}}$ and $p_{neg} = \frac{f_{neg}}{N_{neg}}$ where N_{pos} and N_{neg} denote to the total of the positive and negative elements of the reference set, respectively:

$$\Phi_E^b(X, Y^+, Y^-) = \frac{p_{pos} - p_{neg}}{p_{pos} - p_{neg}}$$

The ENDI values are ranging between -1 and +1, the best threshold is at zero level, where results are being balanced between commission and omission error (Pesaresi et al., 2016a). The more the ENDIs distribution differs, the easier the land cover class are separable based on the sequence data set (Pesaresi et al., 2016b).

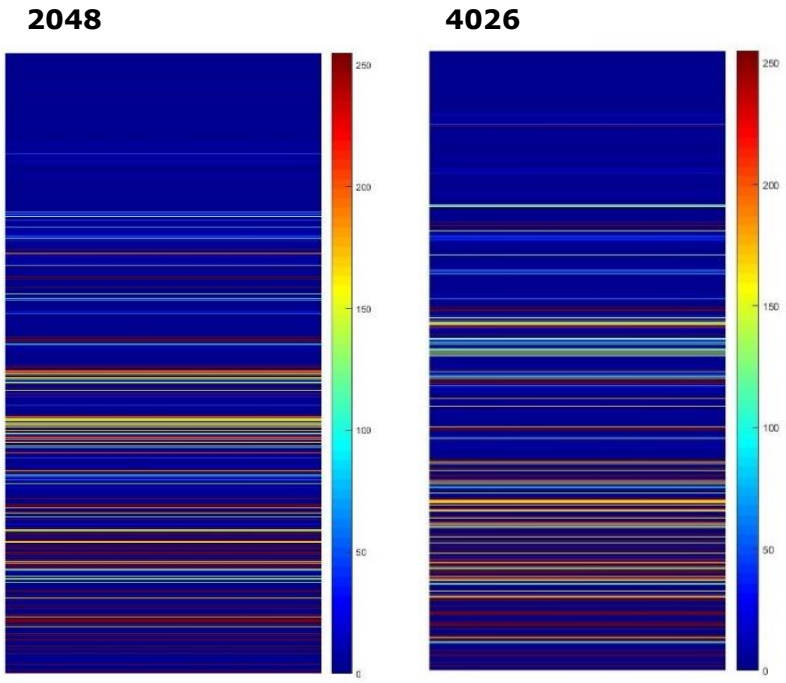


Fig. 3: Two examples of the ENDI of "artificial land" class. The left image corresponds to a Quantization Level of 2048, the right image corresponds to a Quantization Level of 4026. Both results are derived from the same input data.

As fig. 3 shows, different Quantization Levels lead to different results in the sequencing approach. When a small level is selected, the sequencing is very fine while a higher level leads to a generalising result (Pesaresi et al., 2016a). Choosing a Quantization Level has therefore a great influence on the sequencing, the actual data reduction and hence on the ENDI outputs (I.e. the classification results of the SML).

3. Landscape heterogeneity and its relation with image radiometry

The analysis of spatial patterns and heterogeneity is a central scope of geographic research and is often tackled with remote sensing data (Herold et al., 2005). Remote sensing offers the opportunity to observe land surface globally in a high spatial and temporal resolution. Hence, changes and patterns can be monitored and linked to other spatial data, e.g. population data, climate data.

The spatial resolution of remote sensing data is often rather moderate compared to the actual diversity within the landscape; the radiometric signal, which reaches the sensor, is getting various signals due to the patterns in the landscape since many objects in the landscape are smaller than the sensors resolution. The result is intra-pixel heterogeneity (Garrigues et al., 2006).

Spatial heterogeneity can be defined as the surfaces attributes and its varying measurements in space. The variability and the structure of these patterns are depending on its observational scale, therefore the geographical extent and the resolution of the data. Different types of landscape have different spatial variability. Garrigues et al., (2006) used variogram models, based on Normalized Difference Vegetation Index (NDVI) data, which were calculated from different sensors and different resolutions to quantify the spatial heterogeneity at landscape scale. They found that cropland is very heterogenic while forests, bare land and sparse vegetation are somewhat homogenous. The pixel size must be small enough to capture the spatial heterogeneity and minimize the spectral variability within the pixel. A coarse resolution leads to a loss of detectable heterogeneity in the observed landscape. These results show that the heterogeneity of a given landscape is depending on its spatial variability and on the resolution of the imagery.

Tuanmu and Jetz, (2015) developed heterogeneity metrics, which are based on textural features from Enhanced Vegetation Index (EVI) data from the Moderate Resolution Imaging Spectroradiometer (MODIS). The EVI input data has a 250 meter resolution and was captured during a 16-day composites between 2001 and 2005. These metrics cover the evenness, the contrast, entropy and homogeneity, but also conventional metrics like the Shannon index (Fig. 5) on a 1 km resolution.

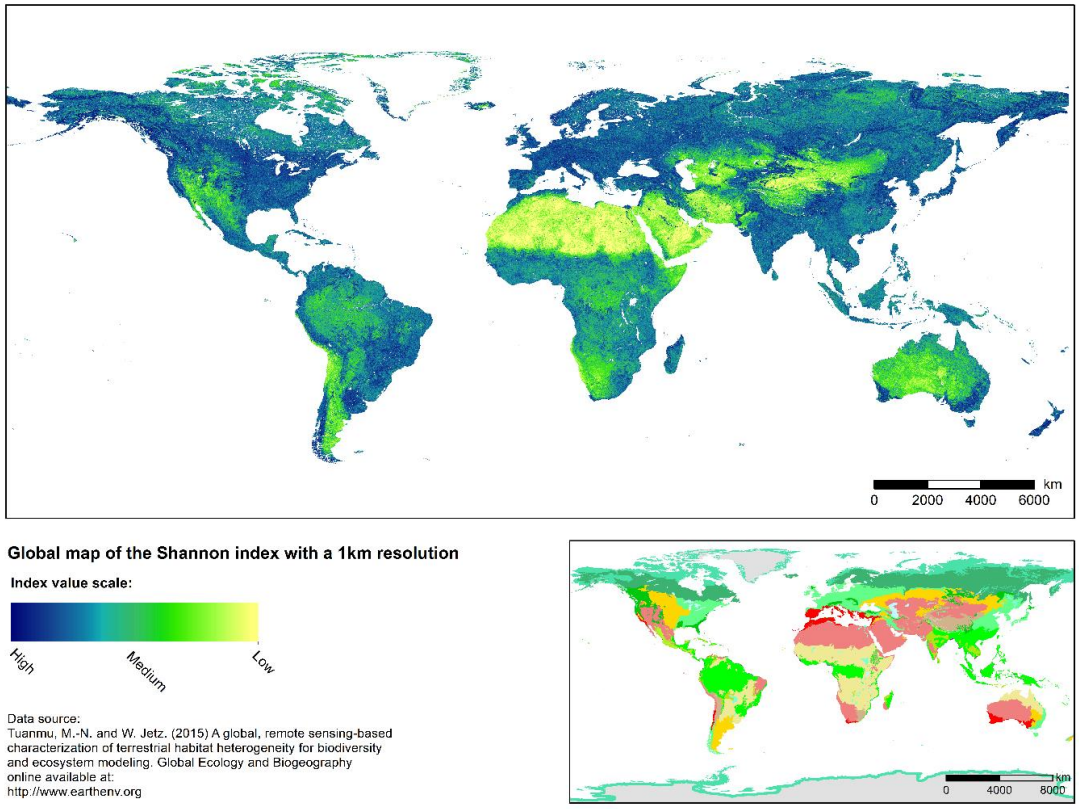


Fig. 4: Global map of the Shannon index with a 1 km resolution, and a side map with the global ecoregions (Tuanmu and Jetz, 2015).

When comparing the main map and the map of the global ecoregions, it becomes clear that the results of Tuanmu and Jetz (2015) also confirm the results of Garrigues et al. (2006); The type of landscape is a fundamental attribute for the spatial heterogeneity. One of the advantages of calculating heterogeneity from remote sensing data or relevant indices like the vegetation index, is that the data has a continuous scale. It is therefore possible to address the heterogeneity within one land cover class (cropland, forest etc.) as well. In addition, are changes in spatial heterogeneity over a timespan easier to monitor (Tuanmu and Jetz, 2015).

Nonetheless, the classification of remote sensing data is useful if the research focus lies on heterogeneity (Herold et al., 2005). Spectral but also spatial heterogeneity has a major impact on the success of a classification approach due to the coarse resolution of the imagery compared to the real landscape an intra-pixel heterogeneity is given. The overall approach of a classification the finding of similarities in the spectral attributes of different pixels and separated it into class. There is a variety of methods, which can be used to conduct a classification with remote sensing data. The classification scheme is essential for landscape analysis, because its quality and resolution have a major impact on the stability of metrics (Huang et al., 2006).

In the framework of this report, the focus sets on the determination of heterogeneity or homogeneity, while using classified satellite data. An example of classified remote sensing data is shown in fig. 5.

Neighbouring pixels of the same class become patches. The patch represents a relatively homogeneous area with clear boundaries towards its surrounding. These patches resemble structures like e.g. forest or cultivated land in the real landscape, but much more generalized.

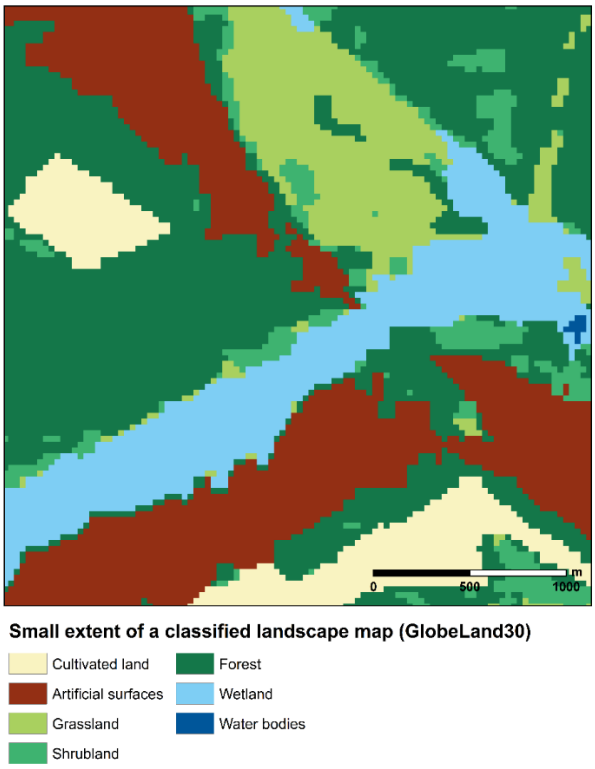


Fig. 5: Example for a landscape map derived from the classification of Landsat satellite imagery (GlobeLand30).

Additionally, it is possible to characterize patches by their size, compactness and other attributes, this provides much richer information than pixel based analysis. The example given in fig. 5 shows a close up view of a classified land cover product with the resolution of 30 m. The size of patches are highly varying, the smallest unit is the single pixel.

A landscape can be defined as a composition of different patches with different characteristics. The outlines of a landscape however are not easily described; from an ecologic perspective, a landscape may be a species habitat or an ecosystem. A landscape can also be defined by its size and its evenness. Nevertheless, the interaction of the patch pattern or its heterogeneity, respectively, play an important role on the definition of landscape (McGarigal, 2015).

Landscape metrics were developed to measure the landscapes heterogeneity and diversity and to compare different types of landscapes with each other (Herold et al., 2005; Plexida et al., 2014). Landscapes contain a spatial mosaic of patches of different classes. A landscape therefore exists of various elements, the configuration of these elements define the pattern of the landscape (McGarigal, 2015).

The landscapes diversity, its configuration and fragmentation serve as indices for landscape heterogeneity (McGarigal, 2015). The size and resolution of the observed landscape data, is a stabilizing factor for the calculation of the landscape metrics (McGarigal, 2015; Plexida et al., 2014). Within the landscape, the focus lays on the patches composition and spatial configuration to determine the landscapes heterogeneity. Spatial configuration is more difficult to quantify, e.g. patch isolation, patch shape or core area (McGarigal, 2015).

In this research only those metrics, which describing the composition or diversity of the observed landscape were assessed. The spatial configuration e.g. patch isolation, patch shape, which is more difficult to quantify, is not considered in this report because of its strong sensitivity to the spatial resolution of the input satellite data used in the land cover classification (McGarigal, 2015; Herold et al., 2005).

4. Input data for the experiment

4.1 Landsat imagery

For this study a sample of 1148, cloud free Landsat 8 scenes was used. The resolution of those images is 30m and comprise seven bands (Coastal/ Aerosol, Blue, Green, Red, NIR, Short Wave IR 1, Short Wave IR 2).

The images were selected from the original 9442 image dataset used as a baseline for the GHSL built-up layer extraction from the Landsat 2014 data collection (Corbane et al., 2017). The images were chosen because they were cloud free. In fig. 7 the footprints of the sample images and their global distribution are shown.

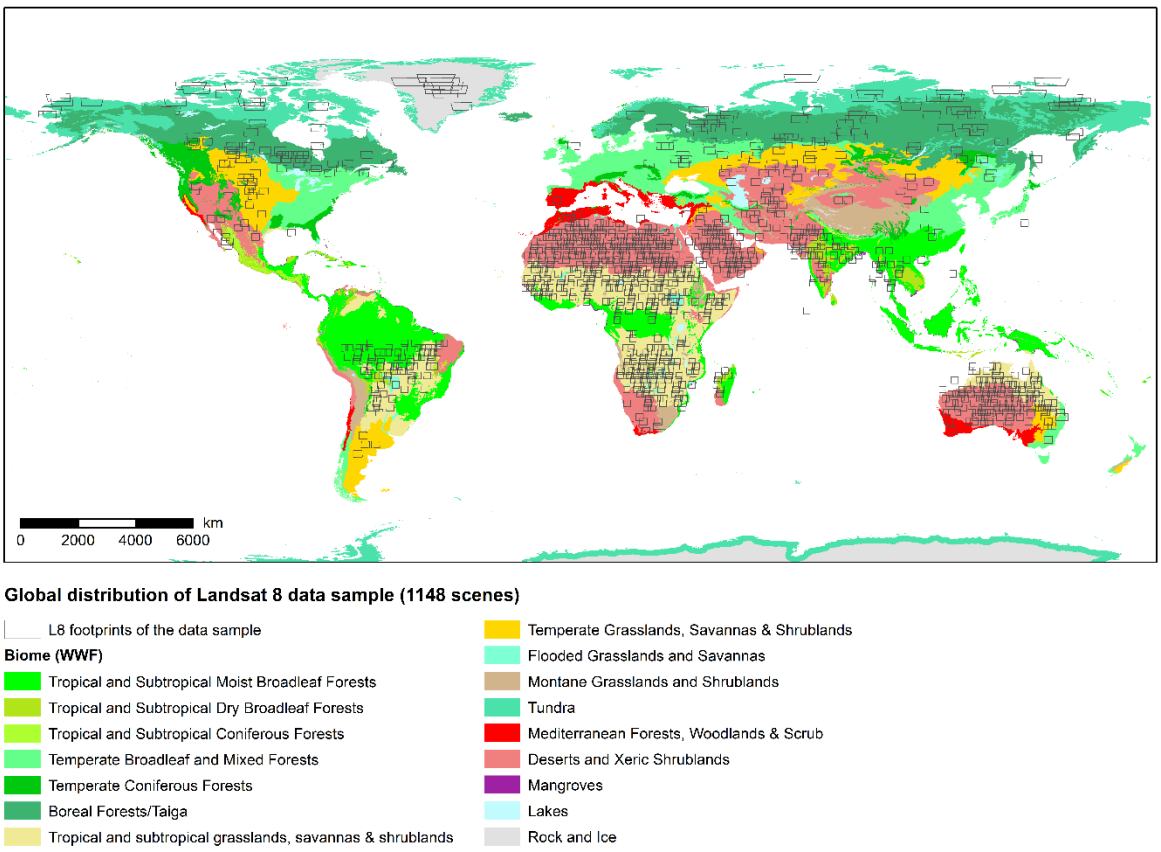


Fig. 6: Global distribution of Landsat 8 data sample overlaid on the global map of biomes.

Fig. 6 shows that the data sample is evenly distributed around the world and across different eco regions. Therefore is the dataset considered as a representative sample of the original data that has a global coverage.

4.2 GlobeLand30

A classification dataset serves as a proxy for the quantization; it withholds the information about the landscape heterogeneity. The classification approach classifies patterns or patches, respectively, within the landscape. These describe the landscape and therefore are fundamental for the estimation of the heterogeneity. It is necessary that the dataset is globally available. Two global land cover products were suitable: The Climate Change Initiative-land-cover (CCI) and the GlobeLand30. In fig. 8 both these products are shown. Even though the CCI-land-cover has a number of 22 unique land cover classes, the resolution of 300 meters is rather low. In comparison has the GlobeLand30 only 10 unique land cover classes but a higher resolution of 30m. This is particularly well visible in the urban area of Milan (Red). On the left map the urban structure is much more detailed than on the right map. In the framework of this research, the GlobeLand30 was selected due to its high resolution, which is also identical to the resolution of the Landsat 8 imagery.

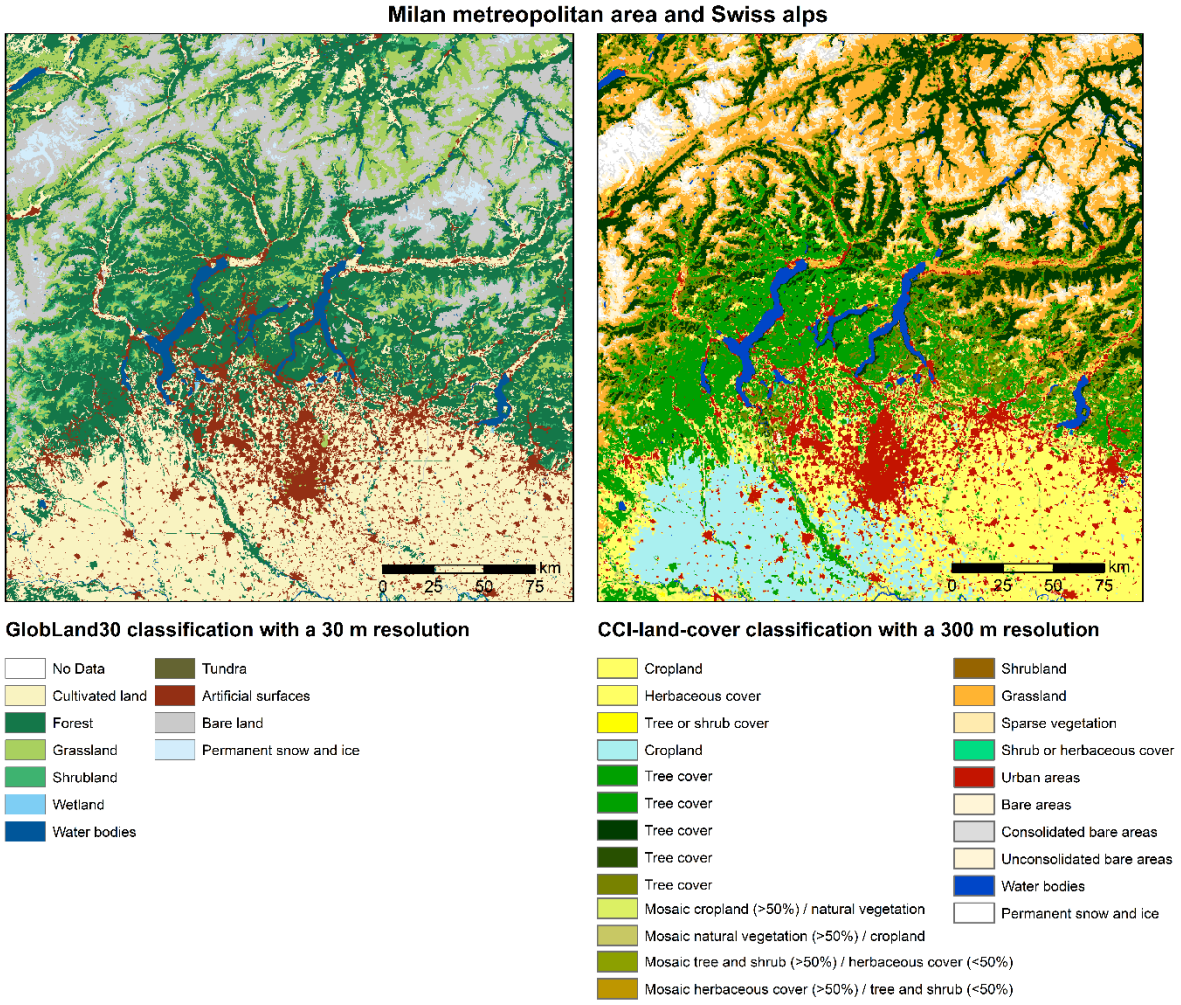


Fig. 7: Two different classification maps of the Milan Metropolitan Area and Swiss Alps (Left: GlobeCover30, right: CCI-land-cover).

GlobeLand30 was created to support the research on global change by the Ministry of Science and Technology of China (National Geomatics Center of China, 2014). The product represents the first global open access land cover product at 30 meter resolution and is based on the fine scale Landsat imagery acquired in 2010 (Li et al., 2015).

Visual analysis techniques were used to derive a hybrid pixel- and objected-oriented approach. The map includes ten thematic classes with one associated to "artificial surfaces", this class was used as a references for GHSL built-up layer classification (Corbane et al., 2017).

Table 1: GlobeLand30 classification scheme, list taken from the (National Geomatics Center of China, 2014)

Code	Type	Content
10	Cultivated land	Lands used for agriculture, horticulture and gardens, including paddy fields, irrigated and dry farmland, vegetation and fruit gardens, etc.
20	Forest	Land covered with trees, with vegetation cover over 30%, including deciduous and coniferous forests and sparse woodland with cover 10 – 30%, etc.

30	Grassland	Lands covered by natural grass with cover over 10%, etc.
40	Shrubland	Lands covered with shrubs with cover over 30%, including deciduous and evergreen shrubs and desert steppe with cover over 10%, etc.
50	Wetland	Lands covered with wetland plants and water bodies, including inland marsh, lake marsh, river floodplain wetland, forest/shrub wetland, peat bogs, mangrove and salt marsh, etc.
60	Water bodies	Water bodies in the land area, including river, lake reservoir, fish pond, etc.
70	Tundra	Lands covered by lichen, moss, hardy perennial herb and shrubs in the polar regions, including shrub tundra, herbaceous tundra, wet tundra and barren tundra, etc.
80	Artificial surfaces	Lands modified by human activities, including all kind of habitation, industrial and mining area, transportation facilities and interior, urban green zones and water bodies, etc.
90	Bareland	Lands with vegetation cover lower than 10%, including desert, sandy fields, Gobi, bare rocks, saline and alkaline lands, etc.
100	Permanent snow and ice	Lands covered by permanent snow, glacier and icecap.

In the maps (Fig. 8), are two examples shown of the GlobeLand30 classified land cover map. One subset is taken from North America, the other from South Africa. The small overview map shows were both extent are located exactly. The land cover in both scenes is very different in terms of types and number of classes.

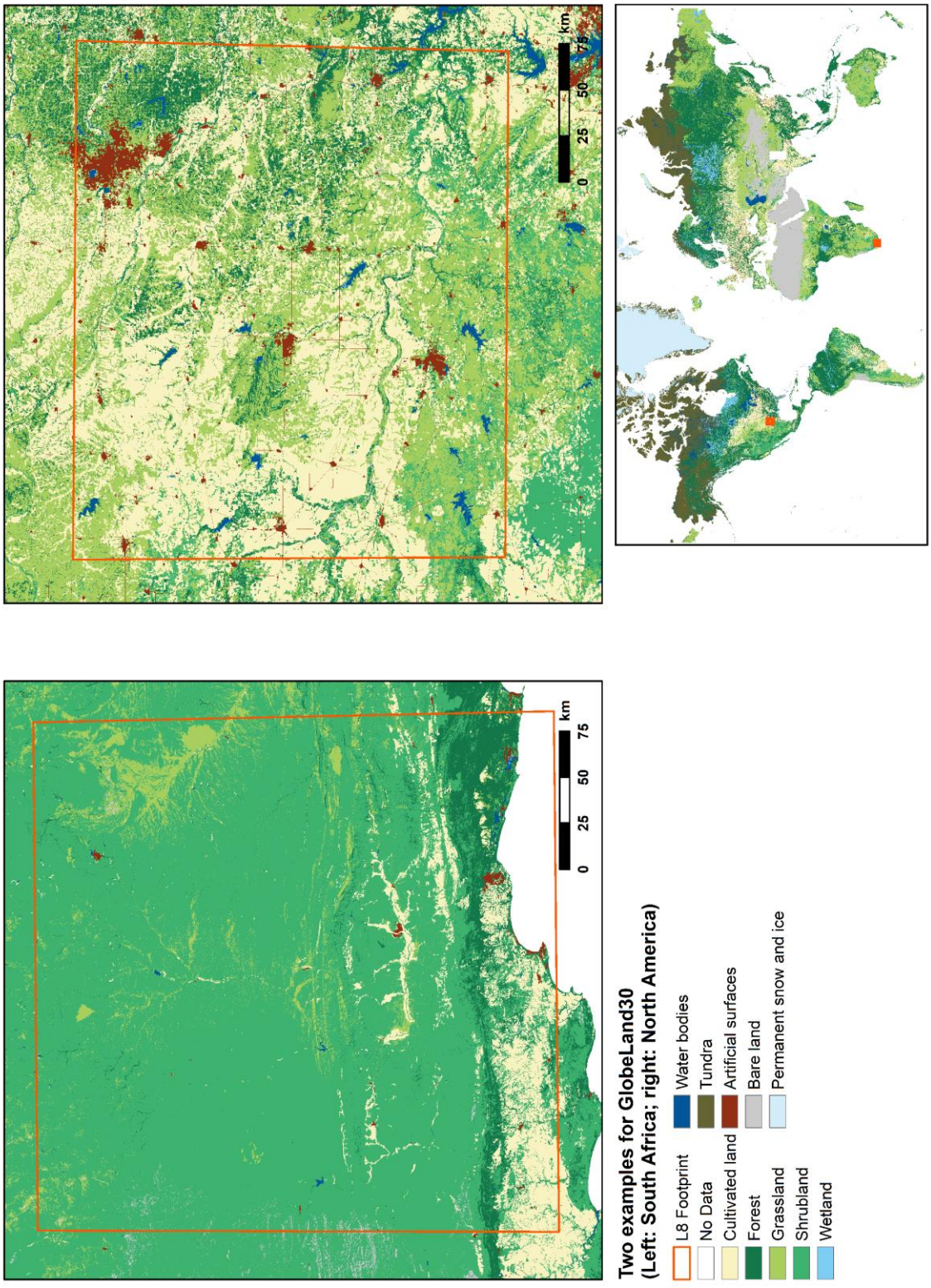


Fig. 8: Two examples for GlobeLand30, one example from North America (Right) and South Africa (Left) and a small overview map. The red square represents the footprint of Landsat 8 scene from the data sample.

5. Statistical analysis of quantized satellite data and landscape metrics

5.1 Estimation and selection of landscape metrics

Landscape metrics were developed to descriptively and statistically measure the heterogeneity of a landscape (Herold et al., 2005). To simplify these calculations, McGarigal (2015) developed an open source software tool, which allows its user to calculate a great variety of metrics. Within the software documentation, every metric and index is defined with a formula and an explanation. From this catalogue, nine metrics that fitted the data type best and are also frequently used in literature for the measurement of heterogeneity, were selected (Garrigues et al., 2006; Huang et al., 2006; Plexida et al., 2014; Tuanmu and Jetz, 2015).

Table 2: Landscape metrics and their formula based on McGarigal (2015).

Index	**Formula**	**Description**	**Unit**
Total Area	$TA = A\left(\frac{1}{10000}\right)$	Total area of the landscape with A for Area.	Hectar
Number of Classes	-	Total number of classes within the landscape	Count
Number of patches	-	Number of patches in the landscape of a class i	Count
Simpson Index	$SIDI = 1 - \sum_{i=1}^{m} P_i^2$	P_i is the proportion of the landscape occupied by class i. With m as number of classes.	None with a range of 0 to 1.
Shannon Index	$SHDI = \sum_{i=1}^{m} (P_i * lnP_i)$	P_i is the proportion of the landscape occupied by class i. With m as number of classes.	Non with a range of 0 to ∞.
Patch Richness Density	$PRD = \frac{m}{A}(10000) * (100)$	m equals the number of classes (patch types), and A is the total landscape area	Number per 100 hectares
Relative Patch Richness	$RPR = \frac{m}{m_{max}} * (100)$	m equals the number of classes in the landscape and m_{max} the total number of possible classes.	Percent

The calculation of the metrics was conducted in Matlab. The code for this process is available in annex A. Before the metrics were calculated the GlobeLand30 raster file, which served as the landscape data, was clipped to the extents of the 1148 Landsat 8 scenes (See fig. 7 in chapter 4.1). These newly created datasets, which have now the same extent as the original Landsat 8 scenes, serve as the input for the calculation of the landscape metrics listed in table 2.

5.2 Quantization of Landsat data

The quantization of the Landsat data is driven by Quantization Level (q_i). The level is usually selected experimentally using the Average Support as a guide for defining the threshold. According to Pesaresi et al. (2016b), an appropriate Quantization Level is defined so that the average support ($supp_\mu$) is in the range 10^3-10^4. These $supp_\mu$ orders of magnitude were tested as satisfactory for classification exercises incorporating noisy training sets.

For the purpose of built-up areas extraction from Landsat 8 data in the framework of GHSL, a Quantization Level of 512 was used. For this study, to analyse the relationship between the quantization and the heterogeneity of the landscape, we tested different Quantization Levels ranging from 32 to 4096 levels.

Levels of Quantization							
32	64	128	256	512	1024	2048	4096

The quantization was conducted in Matlab. The sample Landsat 8 dataset was implemented in the script and the first seven 30 meter bands (Coastal/ Aerosol, Blue, Green, Red, Near Infra-Red, Short Wave Infra-Red 1, Short Wave Infra-Red 2) were used as input for the quantization. For each scene and every Quantization Level the Average Support ($supp_\mu$), the Number of Sequences ($\hat{X}_{q_iF}$) and the Number of Levels (s_i) were calculated.

6. Results

6.1 Selected landscape metrics

For this analysis, a total of seven landscape metrics were calculated (See chapter 3). Prior to analysing the relationship between the independent variables and the image quantization metrics (Dependent variables), was it necessary to identify and exclude highly correlated variables that may lead to unstable models.

Fig. 9 shows a graphical correlation matrix. The darker the shade of red, the higher the Pearson correlation coefficient. Three pairs of variables were highly correlated with each other: The Shannon Index and the Simpson Index, the Patch Richness Density (PRD) and the Total Number of Patches (Patches), the Relative Patch Richness (RPR) and the Total Number of Classes (Classes).

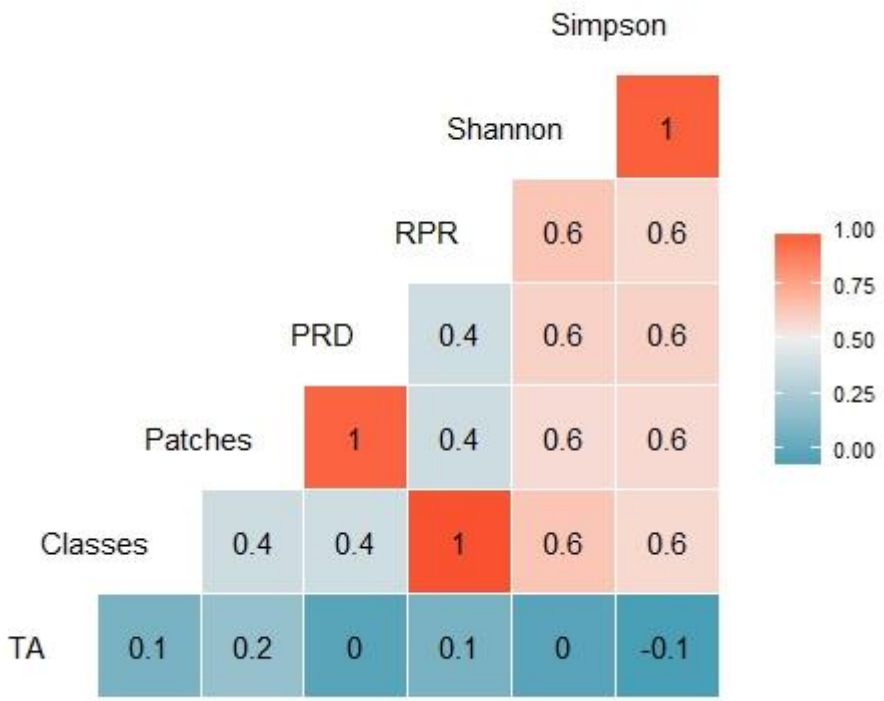
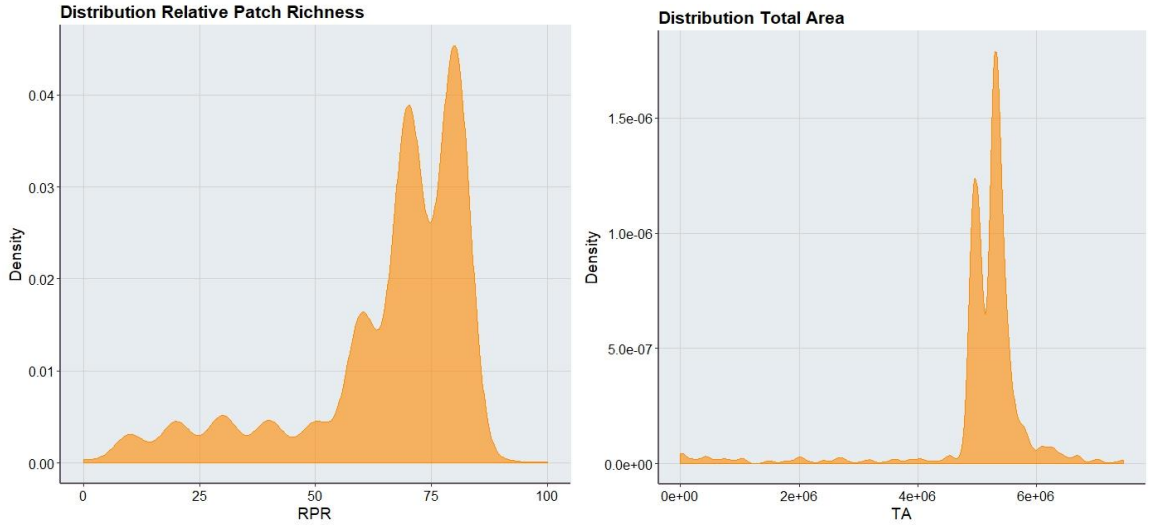


Fig. 9: Correlation plot of the eight landscape metrics.

Total Area (TA), PRD, RPR and the Shannon Index where used as explanatory variables in the subsequent statistical analysis because they are non-correlated.

A normal distribution is desirable for the regression analysis. In the figures below the distributions of the four retained landscape metrics are being displayed.



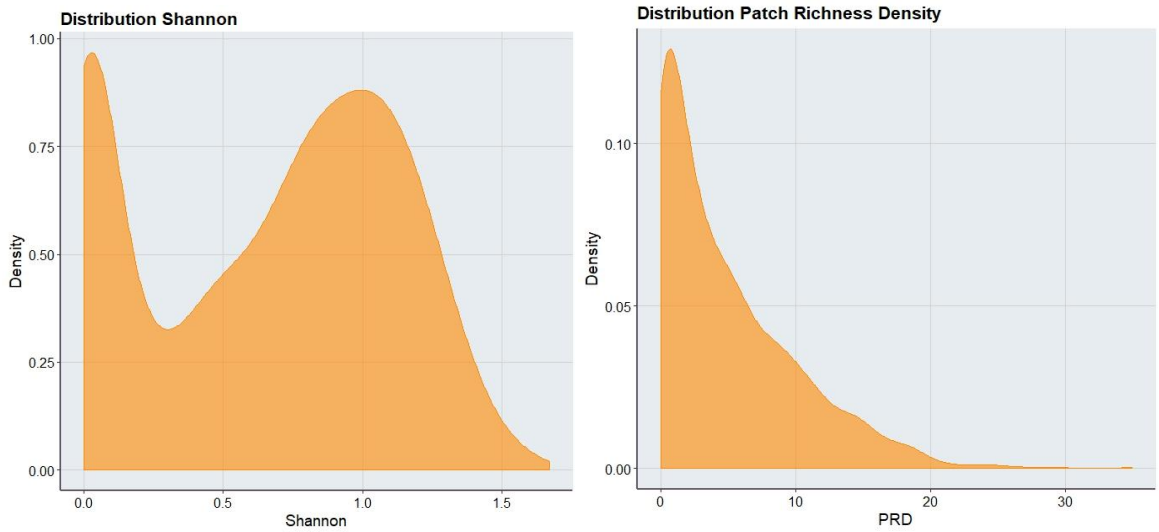


Fig. 10: Distribution plots of the Relative Patch Richness, the Shannon Index, the Patch Richness Density and the Total Area within the data subset.

The distribution plots in fig. 10 show, that the four metrics are not normally distributed. Therefore the assumption of normally distributed data, which is required for linear regression analysis is violated. This needs to be considered in the regression analysis, since not all regression models are robust against unevenly distributed explanatory data.

6.2 Statistical distribution of quantized Landsat data

Similar to the selection process of the landscape metrics, the three variables of the quantization: the Quantization Level, the Average Support and the Number of Sequences, have been tested for correlation. Table 3 shows the correlation coefficients of the three quantization variables.

Table 3: Cross correlation (Pearson) of Quantization Level, Average Support and Number of Sequences.

	Quantization Level	**Average Support**	**Number of Sequences**
Quantization Level	1	0.3024651	-0.4507505
Average Support	0.3024651	1	-0.1206494
Number of Sequences	-0.4507505	-0.1206494	1

To investigate the relationship further, the Quantization Level has been plotted against the Average Support and the Number of Sequences, respectively (See fig. 2 in Chapter 2.2 for a plot of the relationship between Number of Sequences and Average Support).

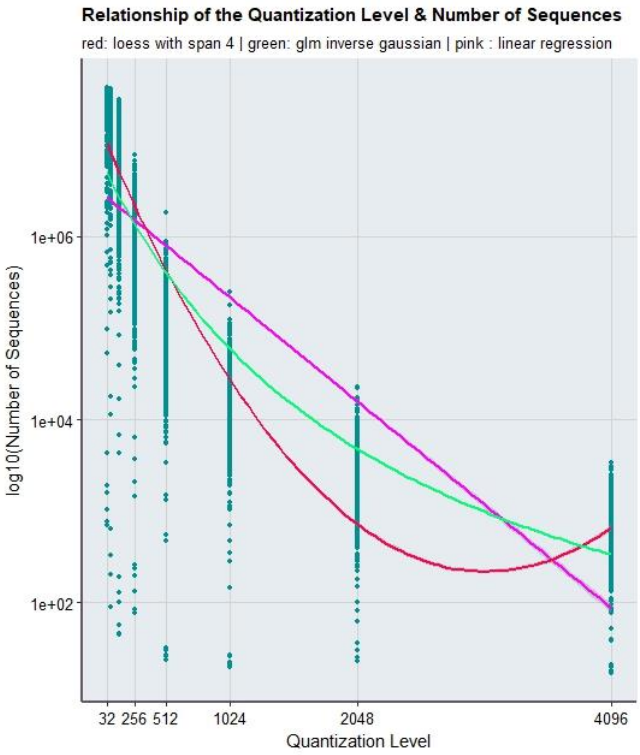


Fig. 11: Relationship of the Quantization Level and the Number of Sequences.

Fig. 11 presents the relationship between the Number of Sequences (Log10 scale) and the Quantization Levels. The graph shows clearly, that a smaller level leads averagely to a higher sequences count, while a high level leads typically to smaller number of sequences. This phenomena makes sense, since the Quantization Level implies how detailed the input data is being sequenced. The smaller the level, the higher the number of sequences. The three curves in the graphic are supposed to explain this relationship. The linear regression (pink) clearly does not fit the data. The loess curve (Red) seems to fit the data better but unfortunately needs to be smoothed drastically to fit the data. Loess referees to the local regression based on k-nearest-neighbour. It therefore subsets the data cloud and fits each data point locally. By smoothing the curve, a perfect but all over bad fitting curve is given. Its results are therefore not representing the data well. The best fitting curve seems to be the Generalized Linear Model (GLM) with inverse Gaussian (Green).

In the figure below (Fig. 13), the relationship of the Quantization Level and the Average Support is visualized. Here, the linear regression curve, is fitting the data much better than in the plot above. Again, the smoothed loess curve is following the data distribution nicely. The GLM inverse Gaussian, however, does not fit the data.

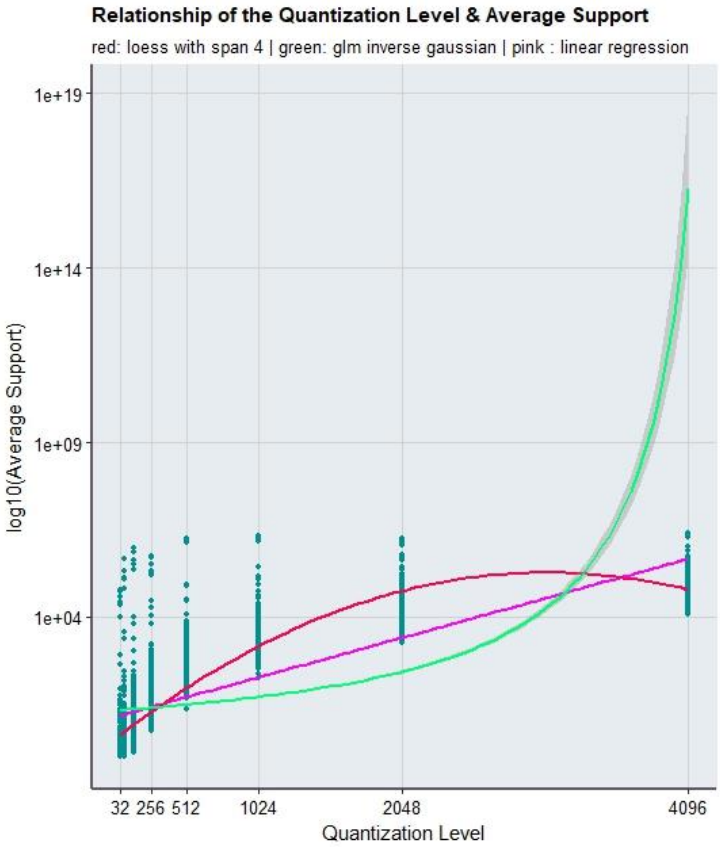
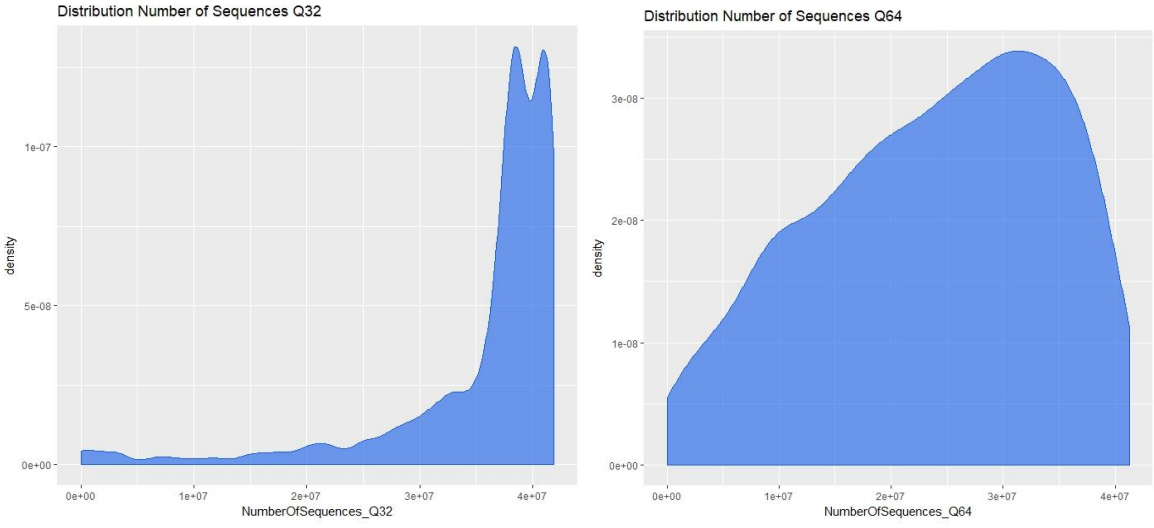


Fig. 12: Relationship of the Quantization Level and the Average Support.

In a next step the distribution of the Number of Sequences and the Average Support were looked at. Data distribution is very important for the selection of a statistical model.



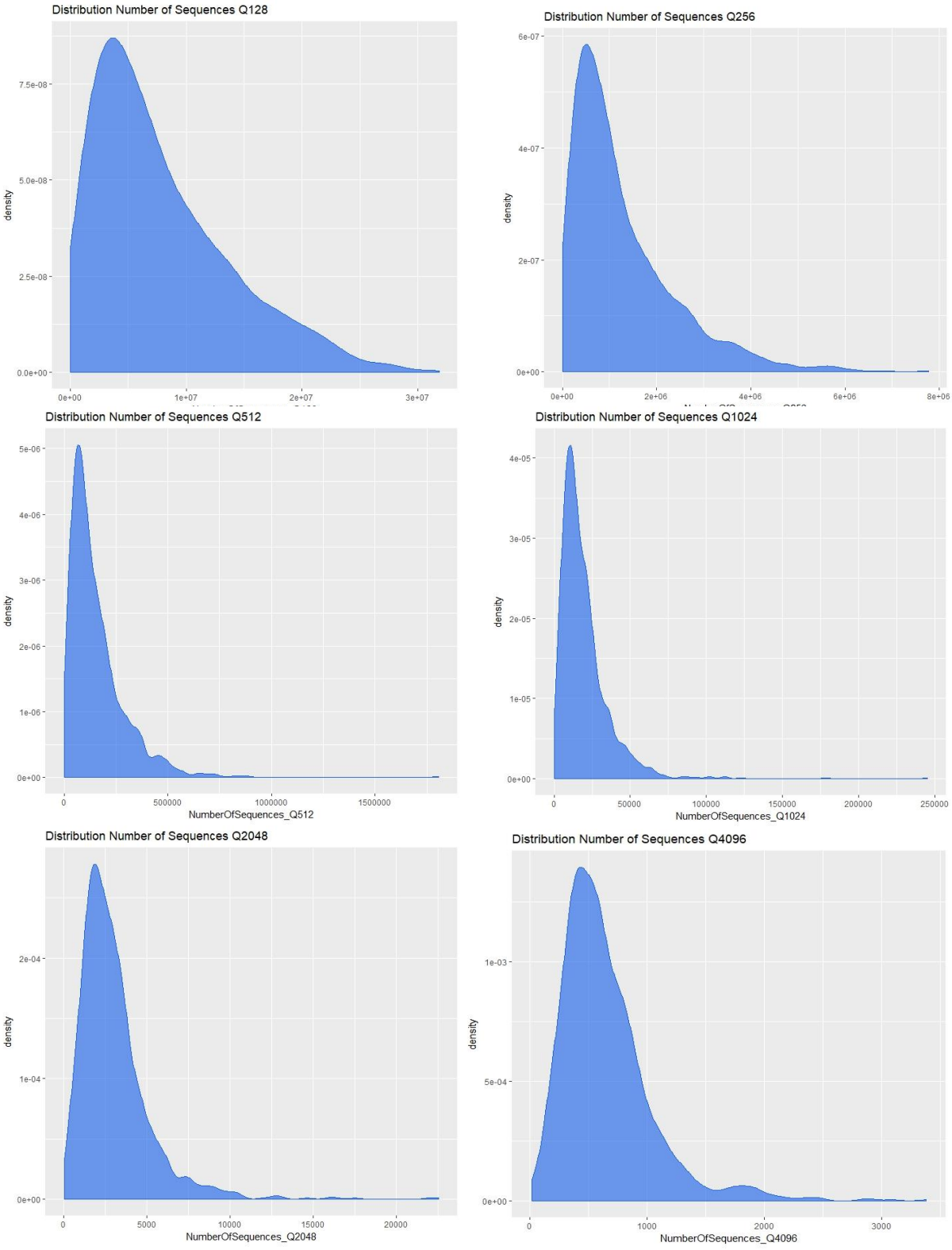


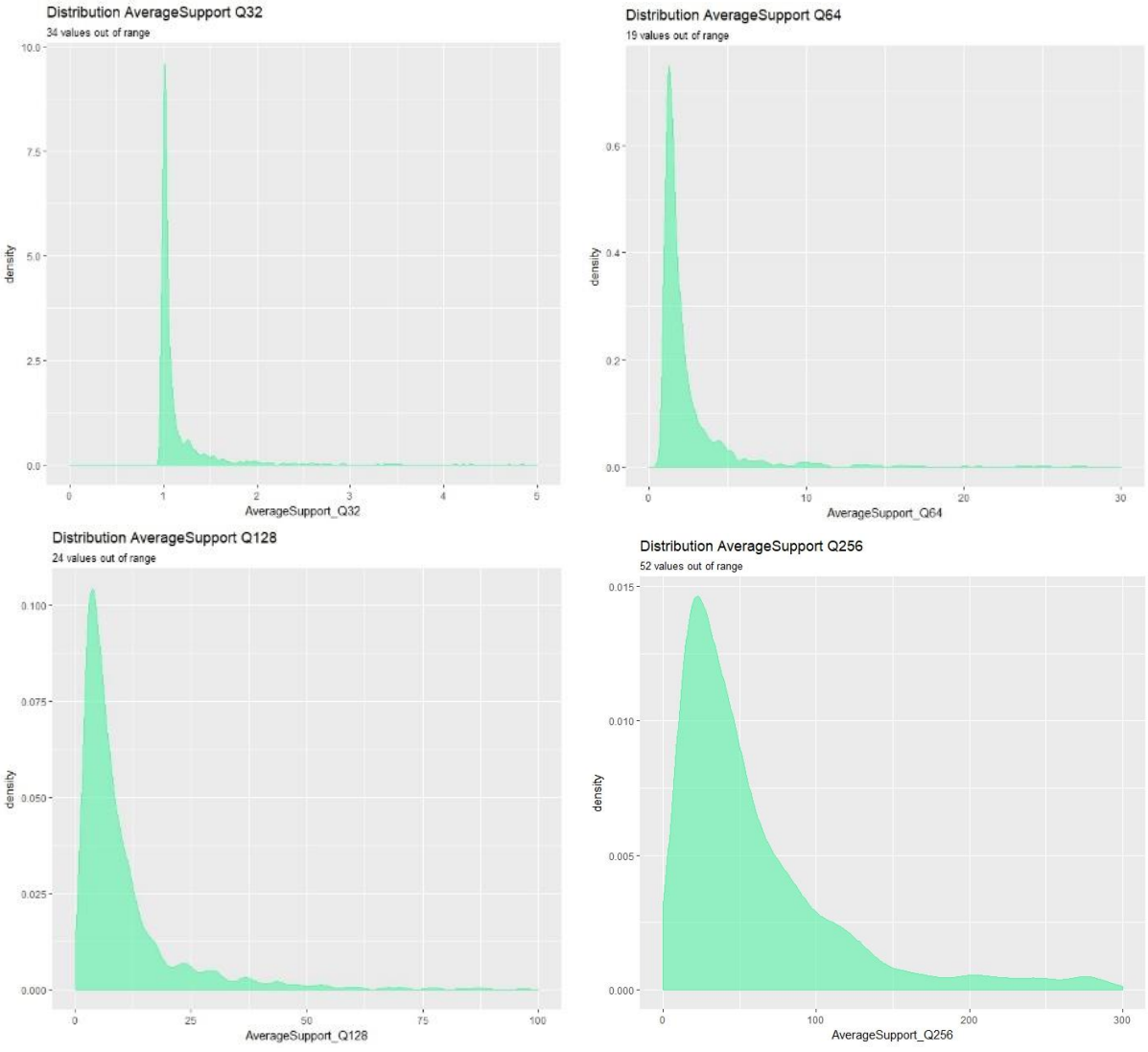
Fig. 13: The eight plots show the distribution of the Number of Sequences at each Quantization Level.

Looking at the distribution plots, it becomes clear that the shape of the distribution curve of the Number of Sequences at Q32 and Q64, differs a lot from the other plots. The first, might be explained by the fact that a small Quantization Level leads to an almost classification-like sequencing process with a high number of sequences.

The remaining density plots are all quiet evenly distributed and similar, only the range of the Number of Sequences varies. For those six plots a shape of an inverse Gaussian distribution is noticeable (Fox, 2016).

The key assumption for the investigation of the relationship between the quantization parameters and the landscape metrics is that the primer is depending on the latter. This relationship is best being described with a multivariate regression.

Nested regression models always look for the best fitting variables and strive towards simplicity. When no explanatory variable is needed, the null hypothesis is fulfilled; this means the data does not fit the model. It is therefore desirable that one of the conditions of the alternative hypothesis are met (James et al., 2013).



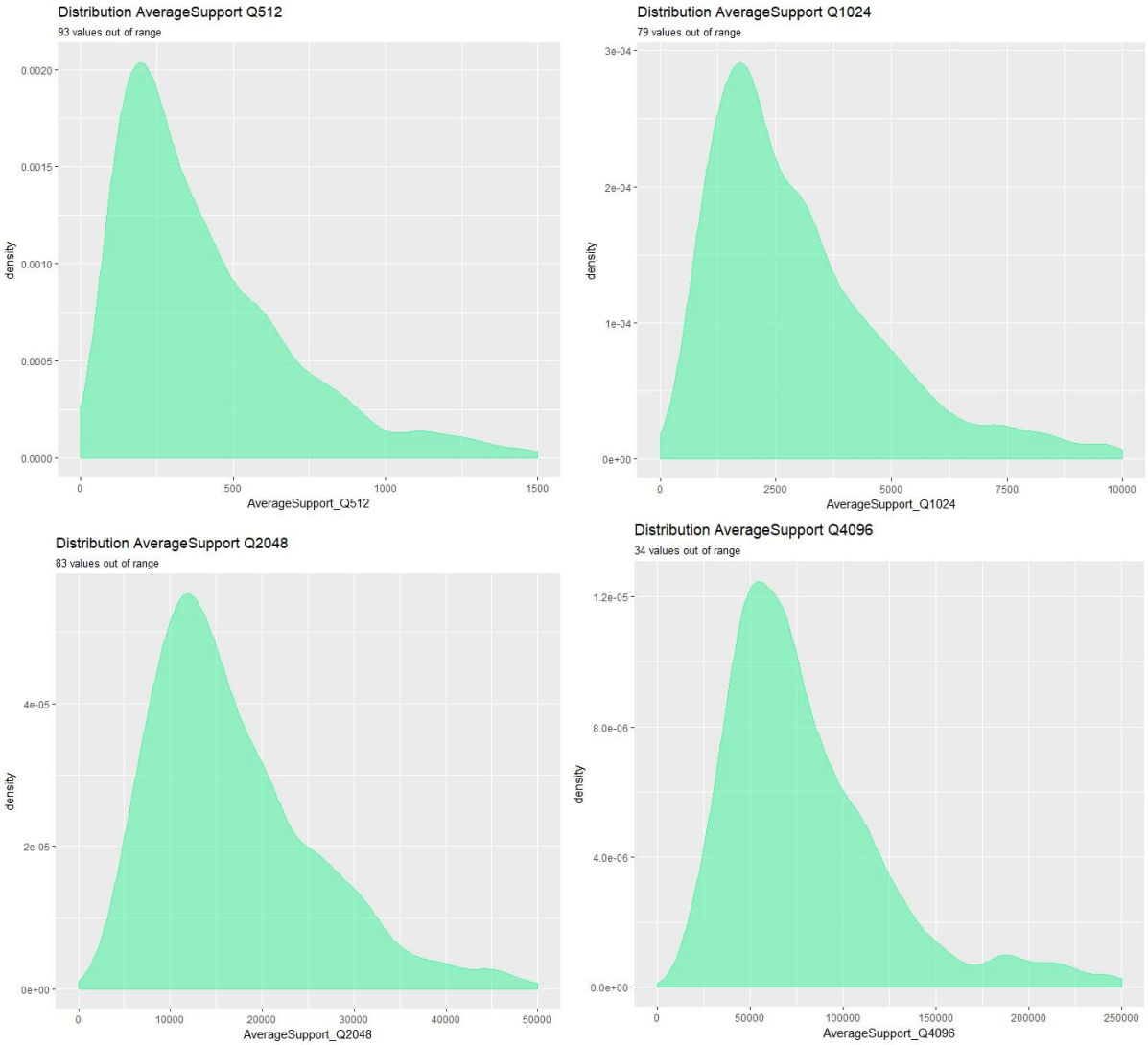


Fig. 14: The eight plots show the distribution of the Average Support at each Quantization Level. Due to the wide range of values on the x-axis, the highest values needed to be cut of due to a better visualization.

Similar to the distribution plots of the Number of Sequences, the results of the Average Support at Q32 and Q64 also differ from the other density plots. Additionally the distribution at Q128 is also varying from the densities related to the other Quantization Levels. A clear distribution cannot be identified clearly. For this reason, the regression is only conducted with the Number of Sequences as depending variable.

6.3 Results of the GLM

After a wide range of different regression models has been tested, the Generalized Linear Model (GLM) was chosen as best fitting model. The advantages of the GLM are that these models can be modified regarding to the input data. The data distribution of the depending variable is considered and GLMs are multivariate regression models and therefore have more than one explanatory variable. The model consists of three components (Fox, 2016):

1. The **random component**: This refers as the depending variable Y_i, this variable has to be a member of an exponential family.
2. A **linear predictor**, therefore a number of explanatory variables:

$$\eta_i = \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

Each X_{ik} resamples a specified function of each of the explanatory variables.

3. A **link function**, $\mu_i \equiv E(Y_i)$, which transforms the expectations of the depending variable Y_i to the linear predictor X_{ik}:

$$g(\mu_i) = \eta_i = \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

The link function is invertible; the equitation can therefore also be written as:

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})$$

The inverse link $g^{-1}()$ is also called the mean function.

A property of distribution is that the conditional variance of Y_i is a function of its mean $V(\mu_i)$ and the dispersion parameter Φ, depending on the exponential family the dispersion parameter is set to be fixed.

In contrast to other linear regression models, the advantage of GLM is that, the transformation of the explanatory variables is partially detached from the depending variables distribution. In fig. 15 three different family functions and the responding link function, are used to display the relationship of the Number of Sequences and the RPR. The green curves, seems to describe the data the best. This was confirmed in further analysis.

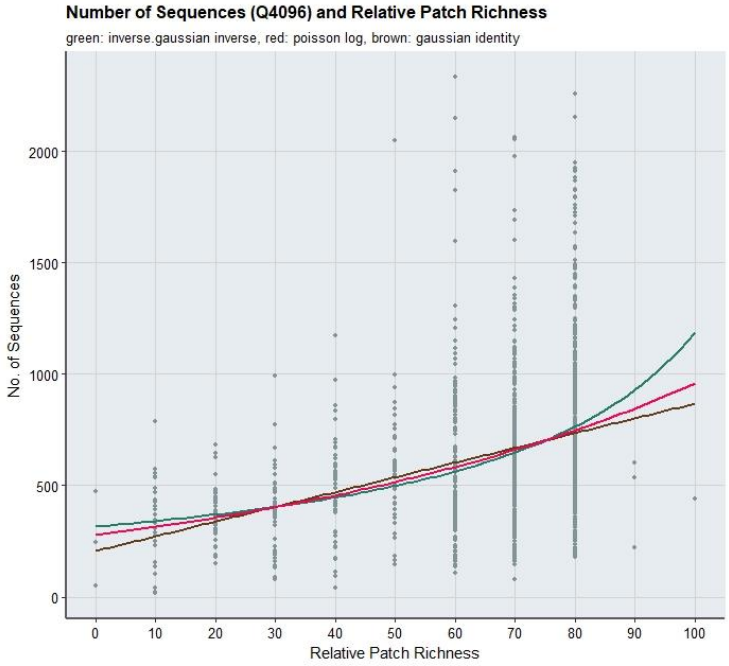


Fig. 15: GLM with different family and link function, for the Number of Sequences and the RPR.

Therefore, in this particular analysis, the inverse-Gaussian exponential family was selected. It is particular useful for the modelling with positive continuous data. This family has two relevant parameters, μ and λ (Inverse dispersion parameter).

$$p(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left[-\frac{\lambda(y-\mu)^2}{2y\mu^2}\right] for\ y > 0$$

The expectation and variance of Y are $E(Y) \equiv \mu$ and $V(Y) = \mu^3/\lambda$. Hence, the variance of the inverse-Gaussian distribution increases with its mean at a rapid rate. The skewness also increases with the value of μ or decreases with the value of λ, respectively (Fox, 2016).

The following formula was implemented in R for the estimation of the significances of the explanatory variables:

$$Number\ of\ Sequences_{QL_i} \sim RPR_i + TA_i + Shannon_i + PRD_i$$

Summary tables were calculated as an output for each Quantization Level (Annex C). The following table shows only the p-values which resemble the significance of each explanatory variable as well as the intercept of the Y-axis, for each GLM.

Table 4: P-values for each explanatory variable and the intercept for each GLM with the depending variable Number of Sequences for single Quantization Level.

	Q32	Q64	Q128	Q256	Q512	Q1024	Q2048	Q4096
intercept	$2.73e^{-201}$	$8.44e^{-102}$	$7.65e^{-46}$	$2.37e^{-30}$	$9.22e^{-27}$	$9.22e^{-27}$	$8.43e^{-44}$	$1.03e^{-62}$
RPR	$2.77e^{-08}$	$7.64e^{-16}$	$2.08e^{-14}$	$1.56e^{-11}$	$4.31e^{-13}$	$4.31e^{-13}$	$2.45e^{-25}$	$9.89e^{-31}$
TA	$5.15e^{-115}$	$1.36e^{-39}$	$4.83e^{-08}$	0.029	0.53	0.53	0.002	0.01
Shannon	$2.28e^{-07}$	$2.34e^{-06}$	0.011	0.219	0.363	0.363	0.724	0.129
PRD	0.027	0.202	0.779	0.588	0.106	0.106	0.0051	0.003

This table shows that the RPR is highly significant for the Number of Sequences at all corresponding Quantization Levels. The TA also shows high significances for the Quantization Levels of 32, 54 and 128, and a minor significance at the level 2048. The Shannon shows only high levels of significance at the Quantization Levels 32 and 64 while the PRD is show minor significances at Q2048 and Q4096.

The p-values show that the RPR variable is highly significant explanatory for the Number of Sequences at all Quantization Levels. The RPR is equals the division of the number of classes in the landscape divided by the total number of possible classes multiplied by 100. The following figure shows a plot of regression with the inverse Gaussian function of the Number of Sequences at Quantization Level 512 and the RPR. The data points are deviating from the main curve, but are generally following the curve. A trend is visible.

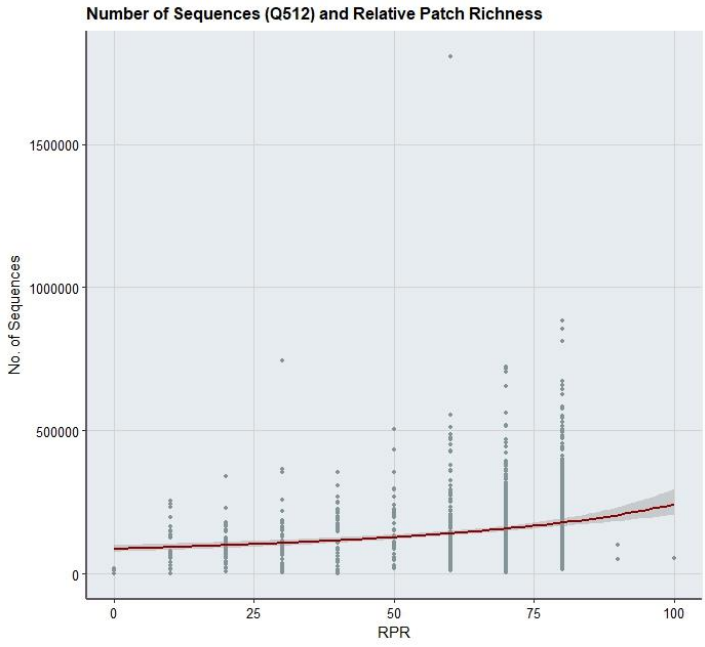


Fig. 16: Regression plot of Number of Sequences and RPR at Q512.

The variable TA only shows good p-values at the small Quantization Levels. In the following figure it becomes clear that this is due to the small Number of Sequences in the higher Quantization Levels. Generally is it questionable if the usage of TA is useful in this context, because the Landsat Scenes are always the same size. Only in some cases where there is Ocean or some other No data pixels in the image are present, TA can be considered useful. When landscapes with natural borders (E.g. river, street) are being compared and they have different total areas, the metric TA makes sense. In this case however, the extent of the scene resembles the landscape border, therefore almost all of the landscapes have the same size, except the ones with sea or ocean in it.

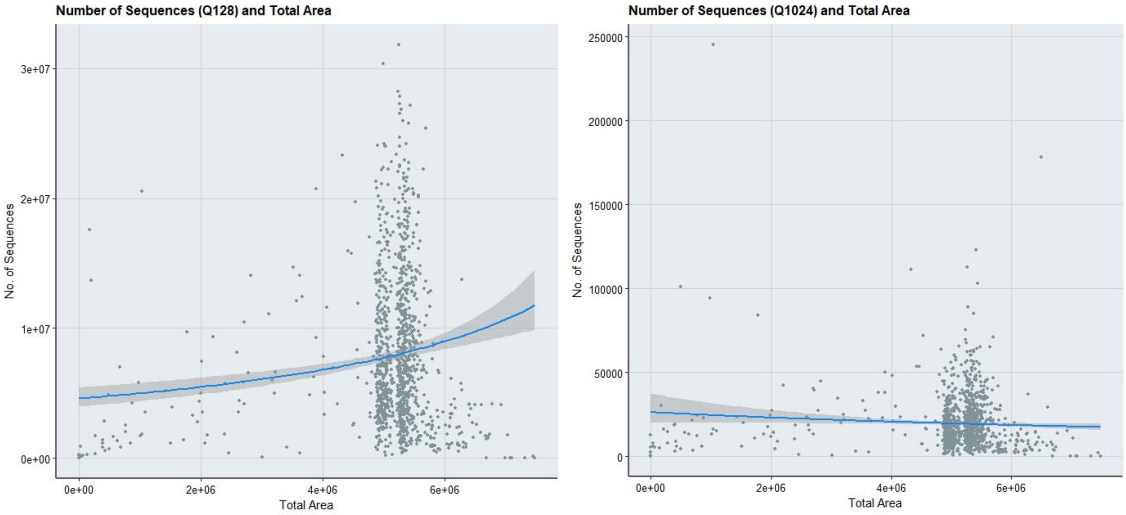


Fig. 17: Regression plot of Number of Sequences and Total Area at Q128 and Q1024.

The Shannon Index has only high significance levels at very low Quantization Levels, as well. Because the Index puts a number on the diversity of a landscape, it is the most forward metric to describe heterogeneity. Because of the higher Number of Sequences, and the higher accuracy of the sequencing with small Quantization Levels, the Shannon Index is significant for the Q32 and Q64. The Shannon Index is dwarfed at Q128, through the generalization of the sequencing process. The Shannon Index seems to be related to the Number of Sequences until a certain Level of Quantization. This may be an indicator that the index is the optimal parameter for the sequencing of the input satellite data.

7. Conclusion and Discussion

The results of the GLM show that only the RPR is significant at all Quantization Levels with the depending variable Number of Sequences. The analysis for the Average Support as a depending variable was not possible. The scopes, which were addressed in the introduction, were therefore not fully met. However, it was made clear that landscape heterogeneity influences the sequencing approach. Further investigation is needed to fully understand this relationship. With different data it might be even possible to derive a ruleset which can be implemented in the SML. In the following paragraph a discussion of the methods presented in the previous chapters are given. Additionally some suggestions are made on how this topic can be approached differently in the future.

Landscape heterogeneity is approachable with different methods. In this study the focus laid on diversity. If we compare the maps in fig. 8, we can see how different the fragmentation of the landscape is. However the less complex landscape of South Africa has higher values in the landscape metrics which were included in this research. Even the Shannon Index is higher for the South African landscape. This brings the question up, if it would not be better to estimate landscape heterogeneity with metrics, which are

focusing on the fragmentation. Even though fragmentation is harder to distinguish with remote sensing data, especially within the pixel itself. Further are the proportions of the different classes, how they are spread over the landscape etc. relevant for diversity assessment. Mixed signals will rather be dominated by a large patch, even though the rest of the image is highly diverse. This is visible on scene level North American landscape, which is much more irregular as the South African scene, even though it has lesser classes. For further investigations of the landscape heterogeneity, the fragmentation of the landscape needs to be taken into account.

A major problem of the GLM family is the lack of a measure for the models goodness-of-fit in generally. For linear regression models there is always the R-square, this is not the case for GLM but only the deviance on which basis a pseudo-R-square can be calculated. This is however not a reliable measure; the goodness-of–fit can only be estimated on behalf of the p-values or other tests like the Likelihood-Ratio-Test or the Rao Score Test. These tests however are only suitable for the comparison between two models; the model with the explanatory variable A is compared to the model with variable B. However, there is no information on how the model with A and B is preforming. For this issue a solution needs to be found because the total quality of the model needs to be evaluated.

If the heterogeneity of the landscape should be implemented into the workflow of the SML, the focus of further experiments needs to lie on patches of the built-up class. This would suggest that the distribution within the landscape and therefore other landscape metrics needs to be evaluated.

This study showed however, that the Shannon Index might be a significant index for the choice of Quantization Level. The dwarfing of significance of the Shannon at Q128 shows that the index is sensitive towards the Level of Quantization. The higher the Level of Quantization and the generalisation respectively, the more insignificant the Shannon becomes. This may be an indicator that the index is the optimal parameter for the sequencing of the input satellite data because it gives a threshold for the Level of Quantization. These findings need further consideration and carrying on analysis.

Bibliography

Corbane, C., Pesaresi, M., Politis, P., Syrris, V., Florczyk, A.J., Soille, P., Maffenini, L., Burger, A., Vasilev, V., Rodriguez, D., Sabo, F., Dijkstra, L., Kemper, T., 2017. Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping. Big Earth Data 1, 118–144. https://doi.org/10.1080/20964471.2017.1397899

Fox, J., 2016. Applied regression analysis and generalized linear models, Third Edition. ed. SAGE, Los Angeles.

Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Quantifying spatial heterogeneity at the landscape scale using variogram models. Remote Sens. Environ. 103, 81–96. https://doi.org/10.1016/j.rse.2006.03.013

Herold, M., Couclelis, H., Clarke, K.C., 2005. The role of spatial metrics in the analysis and modeling of urban land use change. Comput. Environ. Urban Syst. 29, 369–399. https://doi.org/10.1016/j.compenvurbsys.2003.12.001

Huang, C., Geiger, E.L., Kupfer, J.A., 2006. Sensitivity of landscape metrics to classification scheme. Int. J. Remote Sens. 27, 2927–2948. https://doi.org/10.1080/01431160600554330

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, Springer Texts in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-7138-7

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J. Stat. Softw. 28, 1–26.

Li, X., Gong, P., Liang, L., 2015. A 30-year (1984-2013) record of annual urban dynamics of Beijing City derived from Landsat data. Remote Sens. Environ. 166, 78–90.

McGarigal, K., 2015. FRAGSTATS HELP.

Pesaresi, M., Corbane, C., Julea, A., Florczyk, A., Syrris, V., Soille, P., 2016a. Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. Remote Sens. 8, 299. https://doi.org/10.3390/rs8040299

Pesaresi, M., Syrris, V., Julea, A., 2016b. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. Remote Sens. 8, 399. https://doi.org/10.3390/rs8050399

Plexida, S.G., Sfougaris, A.I., Ispikoudis, I.P., Papanastasis, V.P., 2014. Selecting landscape metrics as indicators of spatial heterogeneity—A comparison among Greek landscapes. Int. J. Appl. Earth Obs. Geoinformation 26, 26–35. https://doi.org/10.1016/j.jag.2013.05.001

Tuanmu, M.-N., Jetz, W., 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling: Global habitat heterogeneity. Glob. Ecol. Biogeogr. 24, 1329–1339. https://doi.org/10.1111/geb.12365

List of figures

List of tables

Annex A

Matlab code for the calculation of landscape metrics:

```
function fragstats(start, stop)


warning off;

%// Check OS
inpath = [eos '[FILE PATH NAME]'];
product_list = [processing '[FILE PATH NAME]'];
LU_set = [data '[FILE PATH NAME]'];
NoData = 255;
outpath = [processing '[FILE PATH NAME]'];

% GLC Classes
% 10: Cultivated land
% 20: Forest
% 30: Grassland
% 40: Shrubland
% 50: Wetland
% 60: Water bodies
% 70: Tundra
% 80: Artificial surfaces
% 90: Bare land
% 100: Permanent snow and ice
% 255: NoData


[~,~,x_list] = xlsread(product_list);
L = length(x_list);

for i = start:stop
    tic;
    product = char(x_list(i));
    disp([num2str(i) '. Computing fragstats for product ' product '...']);
    filepath = [inpath filesep product];
    image = dir([filepath filesep '*_B1.TIF']);
    filenames = {image.name};
    pathnames = {image.folder};
    [~,scene_id,~] = fileparts(char(filenames));
    scene_id = scene_id(1:end-4);
    image_path = [filepath filesep filenames{1}];
    geoinfo = geoiminfo(image_path);
    LU = geoimwarpfromfile(LU_set, geoinfo);
    LU(LU == NoData) = 0;
    % geoimwrite(uint8(LU), [outpath filesep product '.tif'], geoinfo);

    % L3 - Total Area (TA)
    L3_TA = sum(sum(LU>0)).*(geoinfo.GeoTransform(2)^2)./10000; % in Hectares

    % L1 - Total number of classes - Patch Richness (PR)
    [L1_PR_c,ia,ic] = unique(LU);
    if min(min(LU))>0
        L1_PR = length(L1_PR_c);
    else
        L1_PR = length(L1_PR_c) - 1;
        L1_PR_c = L1_PR_c(2:end);
    end

    % % L7 - Total number of Patches
    % L7_NP = bwconncomp(LU, 8);
    % % NP = regionprops('table', NP);
    % if min(min(LU))>0
```

```
    % L7_NP = L7_NP.NumObjects;
% else
    % L7_NP = L7_NP.NumObjects - 1;
% end

PatchesPerClass = cell(10, 5);
for i=1:10
    classes = [10 20 30 40 50 60 70 80 90 100];
    LU_class = classes(i);
    LU_c = ismember(LU, LU_class);
    proportion = sum(LU_c(:))/sum(sum(LU>0));
    CC = bwconncomp(LU_c);
    L7_NP = CC.NumObjects;
    PatchesPerClass{i,1} = LU_class;
    PatchesPerClass{i,2} = L7_NP;
    PatchesPerClass{i,3} = sum(LU_c(:));
    PatchesPerClass{i,4} = proportion.*(log(proportion));
    PatchesPerClass{i,5} = proportion^2;
end

valid_data = sum(cell2mat(PatchesPerClass(:,3)));
L7_NP = sum(cell2mat(PatchesPerClass(:,2)));

% L2 - Patch Richness Density
L2_PRD = (L7_NP.*10000.*100)./(sum(sum(LU>0)).*(geoinfo.GeoTransform(2)^2));

% L3 - Relative Patch Richness
L3_RPR = (L1_PR/10).*100;

% L4 - Shannon's Diversity Index
L4_SHDI = -(nansum(cell2mat(PatchesPerClass(:,4))));

% L5 - Simpson's Diversity Index
L5_SIDI = 1-(nansum(cell2mat(PatchesPerClass(:,5))));

outrec.Product_id = product;

outrec.Total_Area = L3_TA;
outrec.Number_of_Classes = L1_PR;
outrec.Number_of_Patches = L7_NP;
outrec.Patch_Richness_Density = L2_PRD;
outrec.Relative_Patch_Richness = L3_RPR;
outrec.Shannons_Diversity_Index = L4_SHDI;
outrec.Simpsons_Diversity_Index = L5_SIDI;
if valid_data > 0
    outrec.class_10_dens = (PatchesPerClass{1,3}./valid_data).*100;
    outrec.class_20_dens = (PatchesPerClass{2,3}./valid_data).*100;
    outrec.class_30_dens = (PatchesPerClass{3,3}./valid_data).*100;
    outrec.class_40_dens = (PatchesPerClass{4,3}./valid_data).*100;
    outrec.class_50_dens = (PatchesPerClass{5,3}./valid_data).*100;
    outrec.class_60_dens = (PatchesPerClass{6,3}./valid_data).*100;
    outrec.class_70_dens = (PatchesPerClass{7,3}./valid_data).*100;
    outrec.class_80_dens = (PatchesPerClass{8,3}./valid_data).*100;
    outrec.class_90_dens = (PatchesPerClass{9,3}./valid_data).*100;
    outrec.class_100_dens = (PatchesPerClass{10,3}./valid_data).*100;
else
    outrec.class_10_dens = 0;
    outrec.class_20_dens = 0;
    outrec.class_30_dens = 0;
    outrec.class_40_dens = 0;
    outrec.class_50_dens = 0;
    outrec.class_60_dens = 0;
    outrec.class_70_dens = 0;
    outrec.class_80_dens = 0;
    outrec.class_90_dens = 0;
    outrec.class_100_dens = 0;
```

```
    end
    outrec.class_10_NoP = PatchesPerClass{1,2};
    outrec.class_20_NoP = PatchesPerClass{2,2};
    outrec.class_30_NoP = PatchesPerClass{3,2};
    outrec.class_40_NoP = PatchesPerClass{4,2};
    outrec.class_50_NoP = PatchesPerClass{5,2};
    outrec.class_60_NoP = PatchesPerClass{6,2};
    outrec.class_70_NoP = PatchesPerClass{7,2};
    outrec.class_80_NoP = PatchesPerClass{8,2};
    outrec.class_90_NoP = PatchesPerClass{9,2};
    outrec.class_100_NoP = PatchesPerClass{10,2};
    outrec_tab = struct2table(outrec);
    writetable(outrec_tab, [outpath filesep product '.csv']);
    toc
end
```

Annex B

R code for the Generalized Linear Model:

```
#Load and prepare Data

mcsv <-read.table("[FILE PATH NAME]", sep = ",", header = TRUE)

ld <-read.table("[FILE PATH NAME]", header = TRUE, sep = ",")


#Select relevant columns

select<- c("scene_id", "NumberOfSequences_Q32", "NumberOfSequences_Q64",
"NumberOfSequences_Q128", "NumberOfSequences_Q256", "NumberOfSequences_Q512",
"NumberOfSequences_Q1024", "NumberOfSequences_Q2048", "NumberOfSequences_Q4096")

my <-mcsv [select]


#Get rid of outliners

NoS <-my[-c(18, 41, 69, 256, 384, 463,245,255, 1140), ]


#Merging

ndata <- merge(ld, NoS, all.x = FALSE, all.y = FALSE, sort = TRUE, by.x =
"Scene_id", by.y = "scene_id", all = TRUE)

select<- c("Scene_id", "RPR", "TA", "PRD", "Shannon", "NumberOfSequences_Q32",
"NumberOfSequences_Q64", "NumberOfSequences_Q128", "NumberOfSequences_Q256",
"NumberOfSequences_Q512", "NumberOfSequences_Q1024", "NumberOfSequences_Q2048",
"NumberOfSequences_Q4096")

ndata <-ndata [select]


#GLM model function for all Number of Sequences at all eight Quantization Level

funnos <- function(x) glm((x) ~  RPR + TA + Shannon + PRD , data = ndata, family =
inverse.gaussian(link = "inverse"))

runf <- lapply(ndata[,c(7:14)],funnos)

Sumnos<- lapply(runf, function(x) (coef(summary(x))[,1:4]))

write.table(Sumnos, "[FILE PATH NAME]", sep = ",", dec = ".")
```

Annex C

Results of the multivariate analysis for Number of Sequences at the eight different Quantization Levels.

NumberOfSequences_Q32	Estimate	Std. Error	t-value	p-value
(Intercept)	7.66E-08	2.04E-09	37.56419	2.73E-201
RPR	-1.00E-10	1.79E-11	-5.59481	2.77E-08
TA	-8.51E-15	3.31E-16	-25.6802	5.15E-115
Shannon	4.34E-09	8.33E-10	5.207137	2.28E-07
PRD	-1.36E-10	6.14E-11	-2.20748	0.02748

NumberOfSequences_Q64	Estimate	Std. Error	t-value	p-value
(Intercept)	1.21E-07	5.08E-09	23.78894	8.44E-102
RPR	-3.78E-10	4.63E-11	-8.17842	7.64E-16
TA	-1.14E-14	8.31E-16	-13.6896	1.36E-39
Shannon	9.68E-09	2.04E-09	4.74612	2.34E-06
PRD	-1.94E-10	1.52E-10	-1.27611	0.202178

NumberOfSequences_Q128	Estimate	Std. Error	t-value	p-value
(Intercept)	3.34E-07	2.24E-08	14.87325	7.65E-46
RPR	-1.73E-09	2.24E-10	-7.74679	2.08E-14
TA	-1.99E-14	3.63E-15	-5.49476	4.83E-08
Shannon	2.38E-08	9.38E-09	2.534234	0.011403
PRD	-1.97E-10	7.03E-10	-0.27966	0.779786

NumberOfSequences_Q256	Estimate	Std. Error	t-value	p-value
(Intercept)	1.73E-06	1.47E-07	11.79107	2.37E-30
RPR	-1.08E-08	1.59E-09	-6.81183	1.56E-11
TA	-5.10E-14	2.34E-14	-2.17936	0.02951
Shannon	8.12E-08	6.60E-08	1.22928	0.219222
PRD	2.69E-09	4.98E-09	0.540607	0.588885

NumberOfSequences_Q512	Estimate	Std. Error	t-value	p-value
(Intercept)	1.18E-05	1.07E-06	10.98829	9.22E-27
RPR	-9.20E-08	1.26E-08	-7.33166	4.31E-13
TA	1.04E-13	1.66E-13	0.627051	0.530752
Shannon	4.72E-07	5.20E-07	0.908824	0.363636
PRD	6.44E-08	3.99E-08	1.614182	0.106766

NumberOfSequences_Q1024	Estimate	Std. Error	t-value	p-value
(Intercept)	8.48E-05	7.29E-06	11.64062	9.22E-27
RPR	-7.86E-07	9.00E-08	-8.73382	4.31E-13
TA	3.26E-12	1.10E-12	2.969341	0.530752
Shannon	3.52E-06	3.70E-06	0.950086	0.363636
PRD	6.62E-07	2.86E-07	2.315715	0.106766

NumberOfSequences_Q2048	Estimate	Std. Error	t-value	p-value
(Intercept)	0.000581	4.01E-05	14.49383	8.43E-44
RPR	-5.26E-06	4.94E-07	-10.6568	2.45E-25
TA	1.84E-11	6.01E-12	3.064506	0.002232
Shannon	7.07E-06	2.00E-05	0.35284	0.724274
PRD	4.32E-06	1.54E-06	2.803559	0.00514

NumberOfSequences_Q4096	Estimate	Std. Error	t-value	p-value
(Intercept)	0.002861	0.000161	17.80304	1.03E-62
RPR	-2.34E-05	1.97E-06	-11.8734	9.89E-31
TA	6.21E-11	2.41E-11	2.578405	0.010051
Shannon	-0.00012	7.91E-05	-1.5176	0.129395
PRD	1.78E-05	5.97E-06	2.974138	0.003

Europe Direct is a service to help you find answers to your questions about the European Union
Free phone number (*): 00 800 6 7 8 9 10 11
(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu

How to obtain EU publications

Our publications are available from EU Bookshop (http://bookshop.europa.eu),
where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.
You can obtain their contact details by sending a fax to (352) 29 29-42758.

KJ-NA-29340-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/731774

ISBN 978-92-79-92990-8