A picture containing clipart

Description automatically generated

**THE UNIVERSITY OF EDINBURGH**

**SCHOOL OF GEOSCIENCES**

**TITLE IN CAPITALS GOES HERE**

*BY*

**Isabelle Rich**

in partial fulfilment of the requirement for the

Degree of BSc with Honours in

Ecological and Environmental Sciences

April 2019

**Abstract**

**Table of contents**

Abstract

Table of contents

Acknowledgements

List of abbreviations

1. Introduction……………………………………………………………… 1
   1. Objectives and rationale…………………………………... 2
   2. Research questions, hypotheses and predictions…….... 2
2. Methods
3. Results
   1. Q1
   2. Q2
   3. Q3
4. Discussion
   1. Key findings
   2. Q1
   3. Q2
   4. Q3
   5. Limitations
   6. Future research
5. Conclusion
6. Reference list
7. Appendices

**Acknowledgements**

I firstly would like to thank Dr Isla Myers-Smith, not only for her continuous support, wisdom and advice that kept me grounded during dissertation work, but for all the opportunities working with her has given me: I can safely say that assisting with her 2017 field season changed my life for the better. I am grateful for PhD student Gergana Daskalova for sharing code and resources, and her dedication, knowledge and encouragement throughout this process and always. Without her, I would have never found my love for data science. I would like to also thank the rest of the Team Shrub research group at the University of Edinburgh: my coursemate Cameron Cosgrove, Dr Sandra Algers-Blondin and Mariana Garcia Criadio for helping me plan and achieve the goals of my dissertation.

I would like to thank all the additional academics I reached out to that graciously provided advice: Sam Harrison (University of Edinburgh, PhD candidate), Keiko Nomura (University of Edinburgh, PhD candidate) and Dr Alemayehu Midekisa (Geospatial research specialist, University of California, San Francisco), thank you for solving some seemingly unsolvable problems. I am also grateful for the vast amount of resources on the internet for assisting me within seconds.

I must acknowledge my large and wonderful support network. Thank you to my coursemates for sharing their wisdom, time and company during our long days at the library. I am beyond grateful for my flatmates. Without these ladies, I would not have the confidence, strength and passion to complete this dissertation. A final thanks goes to my family: thank you for supporting my decision to travel across the pond for university. I can confidently say it was the best decision I’ve ever made.

**List of abbreviations**

EU European Union

GEE Google Earth Engine

LUC Land use change

LUCAS Land Use and Coverage Area frame Survey

SPE Socio-politico-economic

SUC Soviet Union collapse

**Introduction**

Land use, as defined by human use of land (Meyer and Turner, 1992), is undoubtedly an

important part of all civilisations due to the provision of natural resources (Foley et al., 2005;

Turner et al., 2007). Human-driven land use change through urbanisation, deforestation and

agricultural expansion has placed pressure on the functioning of several ecological processes

such as carbon cycling, as well as ecosystems themselves (Foley et al., 2005; Turner et al.,

2007). Since 1850, roughly 35% of anthropogenic carbon dioxide (CO2) emissions have

resulted directly from human land use, altering the global carbon cycle (Foley et al., 2005;

Turner et al., 2007). Natural habitat destruction through land conversion is also one of the

largest threats to terrestrial biodiversity, causing extinctions and range reductions (Foley et

al., 2005; Jetz et al., 2007). However, habitat loss such as forest loss and habitat

fragmentation have both proven to also have possible positive effects, including increased

population size (Fahrig, 2017; Daskalova et al., 2018).

Habitat fragmentation and destruction has primarily occurred through changes in agricultural

practices (Foley and Ramankutty, 1999), with croplands and pastures covering over 40% of

Earth’s land surface (Foley et al., 2005). Expansion is made possible through technologies

produced during the ‘Green Revolution,’ an agricultural revolution during the mid-twentieth

century that increased global food production (Foley et al., 2005). However, modern practices

may be compromising long-term ecosystem services (e.g. air quality and nutrient cycling) for

short-term yield increases (Foley et al., 2005). Scientists are therefore concerned with

mitigating against the negative effects of land use change (Foley and Ramankutty, 1999).

Countries appear to follow similar trajectories of changing land use regimes, moving from

subsistence to intensive agriculture at differing rates, depending on their socio-economic

contexts (Lambin et al., 2001; Foley et al., 2005). However, a study in Ethiopia indicates that

not all countries follow this pattern, as Ethiopia experienced deintensification within a changing

socio-economic environment (Reid et al., 2000). Rapid socio-economic changes are said to

accelerate land use change, with land abandonment rates high with regulation change and

the establishment of new institutions (Prishchepov et al., 2013). Agricultural abandonment,

defined as the cessation of agricultural activities, is linked with a shift towards more intensive

agriculture, with smaller farms more likely to be abandoned (Prishchepov et al., 2013). With

rapid shifts in the socio-economic environment, Latvia proves as an ample study site to

examine the common land use trajectory.

Satellite imagery has often been used in studies aiming to quantify influence of socioeconomic

events on land use change (Reid et al., 2000; Prishchepov et al., 2012). However,

satellite imagery cannot show land use specifically, instead depicting land cover, which

indicates solely the type of land (e.g. water, forest etc.). Algorithms must therefore be

developed to effectively categorise land use types. Such studies (Reid et al., 2000;

Prishchepov et al., 2012) only consider the impacts of one socio-economic event, rather than

several over time. Analysing if the signature of multiple socio-economic shifts can be detected

through land cover change could shed light into the importance of socio-economic events as

drivers of agricultural transitions on a country-scale.

In this study, I will focus on Latvia due to its quick-changing socio-economic status, making it

an appropriate case study to examine if land use change can be linked to socio-economic

events. The two events I will examine are (1) the Soviet Union collapse in 1991 and (2) the

addition of Latvia to the European Union (EU) in 2004. After the Soviet Union, there was an

increase in abandoned land, tree cutting and percent coverage of protected areas

(Prishchepov et al., 2013). After joining the EU, the share of large farms (intensive) increased,

while the share in small farms (extensive) decreased (Csaki and Jambor, 2009). Ultimately,

this type of analysis could be replicated for other countries to outline the impacts of shifting

economic status on land use and thus, have implications for wider aspects such as ecosystem

services, the economy and human movement and urbanisation across Europe and globally.

**Objectives**

This study aims to investigate the importance of SPE events as drivers of land use

change in Latvia through the use of satellite imagery. Although the importance of SPE

events on land use change is acknowledged (Prishchepov *et al.*, 2012), it remains

unclear whether a recognisable, country-scale signature is left on the landscape. Using

satellite imagery, pixel-scale analysis can be completed to determine specific land cover

transitions over time, potentially unveiling a link between socio-economic events and land use

change. My findings will give insight into the homogeneity, or lack thereof, of the effects of

socio-economic events on a country-scale. Results will reveal the transition patterns between

each land use type, including extensive, intensive and abandoned land. Ultimately, my study

will uncover the importance of socio-economic events as a driver of land use change in Latvia,

permitting predictions about land use under changing socio-economic conditions to be made.

**2. Methods**

To answer my three main research questions, I constructed a classification of land use change in Latvia between 1989 to 2011 using GEE (Gorelick *et al.*, 2017). My workflow diagram, depicting the key steps to this process, is shown in Figure 1.

**Figure 2.1** – Workflow diagram, created on Microsoft PowerPoint.

**Image processing**

Landsat 5 Thematic Mapper satellite imagery (1985-2011) covers my study period well and is commonly used in similar classification studies (Prishchepov *et al.*, 2012; Fonji and Taff, 2014; Sidhu *et al.*, 2018). I selected Landsat 5 Surface Reflectance imagery (30 metre resolution), which is atmospherically corrected, preventing the occurrence of clouds and shadows in the imagery (Zanter, 2018). For each study year (1989-2011), I employed an additional cloud mask to remove any remaining pixels containing clouds or shadows, as well as any edge pixels that do not contain all bands of interest. The additional mask increased average classification accuracy across all study years by 4%. I chose summer images, depicting the growing season, to best characterise the spectral signatures of my different classes (Fonji and Taff, 2014). I took the median of each year’s image collection to an annual composite for each study year (Pimple *et al.*, 2018). I cropped each year’s image to the border of Latvia and placed a grid over the area in which to extract data from, both increasing the speed of my analysis. Lastly, I selected blue, green, red, near-infrared and shortwave infrared bands, each with corresponding wavelengths, for my classification (Pimple *et al.*, 2018). My chosen bands will act as predictor variables for my classification.

**Classification background**

Classification approaches create categorical datasets, such as land use (Horning, 2010). Classifications aim to investigate the relationships within a group of objects to determine if the data can be summarised into classes (Gordon, 1999). There are three key steps to a classification: train, validate and test (Suthaharan, 2016).

In a supervised classification, relationships are discovered using data of known values, also called training points (Albalate and Minker, 2013). Training points train the classifier to infer prediction rules that form a decision tree (Albalate and Minker, 2013). Decision trees are composed of binary questions which group data by specific characteristics, such as bandwidth (Strobl *et al.*, 2009). With each step of the tree, the classifier asks whether a criterium of a specific class is met. After the classifier is trained, it can be applied across data of unknown classes. Test data of known values, normally obtained in the training phase, are then passed through the classifier to determine classifier accuracy and error (Suthaharan, 2016).

Random forest classifications are a supervised learning technique which is composed of numerous decision trees, creating a forest of trees (Horning, 2010). As multiple trees decide the class of each object, a voting process begins, where the final class assigned is the one that is predicted most (Horning, 2010). Using multiple trees increases classification accuracy (Suthaharan, 2016), as there is decreased overfitting, where noise is used to predict classes (Horning, 2010). A separate validation set, where overfitting is accounted for and parameters are pruned, is therefore not required (Suthaharan, 2016).

A close up of a sign

Description automatically generated

**Figure 2.2** – Random forest conceptual diagram, created with Microsoft PowerPoint.

**Training data**

My training data represent the known locations of each land use type in 2012. Field data results in high certainty in the training set (Millard and Richardson, 2015).

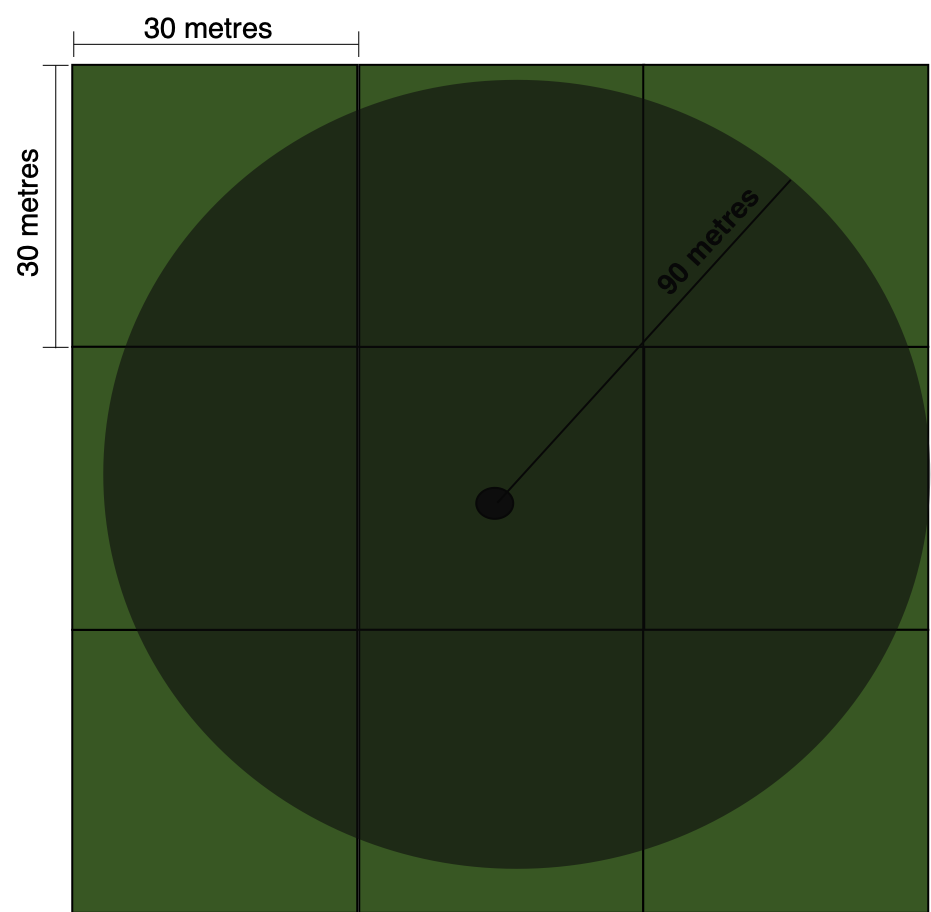
I used the LUCAS dataset (Eurostat, 2013), which contains GPS coordinates of both land use, as defined by the socioeconomic activities and land cover for European countries, for every three years since 2009. I chose to use 2012 data, as it is the earliest dataset that clearly separates fallow and abandoned agricultural land. Within R 3.5.3 (R Core Team, 2013), I filtered data to represent each of my land use classes: abandoned land and extensive and intensive agriculture. I also included four additional classes (artificial land, wetlands, water and forestry), as a larger training dataset improves classification stability and accuracy (Millard and Richardson, 2015). This collection of classes is supported by a similar study conducted by Fonji and Taff (2014).

To select which data would form each of my classes, I used the LUCAS 2012 Technical Reference Document (Eurostat, 2013), which contains descriptions of each land use and land cover category. Table 2.1 displays the criteria that define each of my key classes. The criteria to determine additional class can be found in Appendix X.

**Table 2.1 –** Criteria chosen to represent each key land use class.

|  |  |
| --- | --- |
| **Land use** | **Criteria** |
| **Abandoned** | 1. Abandoned agricultural land. 2. Abandoned land, filtered to include cropland, woodland, grassland and bare land. 3. Semi-natural and natural areas not in use. 4. Unused, spontaneously revegetated land. |
| **Extensive** | Heterogeneous crops planted mainly for own consumption produced in kitchen gardens or allotments, filtered to include cropland, woodland, grassland and bare land (natural areas). |
| **Intensive** | Industrial agriculture, filtered to include cropland, woodland, grassland and bare land. |

Following filtering, I had the coordinates of a total of 4013 data points, 1434 of which are from my three key classes (Table 2.2; Appendix X). I transformed these coordinates into a spatial file by setting the projection to that of GEE: Pseudo-Mercator WGS 84 (EPSG: 3857).



**Figure 2.3** – Circular buffer with 90 metre radius around data point, created with Microsoft PowerPoint. Square boxes represent pixels.

Once imported into GEE, I created a 90-metre buffer around each point, turning point data into circular polygons (Figure 2.3). As each point represents an area or polygon of land, a buffer was necessary to compensate for GPS precision (Wulder *et al.*, 2005) and ensure the surrounding pixels were included. I chose a 90-metre buffer, creating a circular polygon with an area of 25446.9 square metres. This buffer size was used by Sader *et al.* (1995) for their land use classification. It is further supported by the average farm size in Latvia, with more than 91% of farms being this size or larger in 2003 (Zdanovskis and Pilvere, 2015).

To set up my training data for my classification, I divided my data randomly by class into two groups: training (80%) and testing (20%), also known as a holdout approach (Suthaharan, 2016). The holdout approach permits assessment of how the trained classifier works with unseen data (testing group).

**Table 2.2** – Total number of points obtained from LUCAS and the number used for training (80%) and testing (20%) for key classes. Total includes four additional classes.

|  |  |  |  |
| --- | --- | --- | --- |
| **Land type** | **Number of LUCAS points** | **Number of points in training sample** | **Number of points in testing sample** |
| **Abandoned** | 112 | 90 | 22 |
| **Extensive** | 27 | 18 | 9 |
| **Intensive** | 1295 | 1038 | 257 |
| **Total** | 4013 | 3223 | 817 |

I then sampled the imagery to collect the bandwidths across the locations of my training set. The bandwidths correspond to the unique spectral signature of each land type. As the training set refers to 2012 data, it would be most suitable to sample 2012 imagery. However, Landsat 5 only contains imagery until the beginning of 2012 and does not cover the summer months that my data corresponds to. Landsat 7 could have been used, however, 22% of images are missing due to instrument failure (Alexandridis *et al.*, 2013). When examining error, it was clear that this failure had an effect on images covering Latvia. I therefore chose to substitute 2012 images with 2011 images. Consequently, my training points will not precisely match up with the imagery. However, large-scale land use change is not expected to occur between 2011 and 2012, as this period was a stable time for Latvia’s economy and therefore, land use was kept constant (Skribane and Jekabsone, 2013). For instance, abandoned area coverage changed less than 0.1% between 2011 and 2012 (Arika and Mazure, 2017). Upon retrieving the unique bands, the classifier was ready to be constructed and trained.

**Random forest classification**

There are two main inputs when constructing a random forest classifier: the number of decision trees and number of predictor variables per node of each decision tree. I set the number of decision trees to 30 with consideration of the computational burden on GEE (Nomura and Mitchard, 2018). The default number of predictor variables used at each node to form the binary rules within the classifier is commonly set to the square root of the number of input variables (Gislason *et al.*, 2006). Here, the number of input variables is the number of bands (6) used to classify the image. Limiting the number of predictor variables reduces the computational complexity, as well as the correlation between decision trees (Gislason *et al.*, 2006).

I trained the classifier using the bandwidths gathered from sampling the input imagery and applied this classification to the image. This process, from sampling the input imagery to training and applying the classification, was completed for each of my study years. The output

**Data collection**

**Classification accuracy and error**

**Statistical analyses**

**Reference list**

Albalate, A. & Minker, W. (2013) *Semi-Supervised and Unsupervised Machine Learning: Novel Strategies*, Hoboken, NJ, USA: John Wiley & Sons, Inc.

Alexandridis, T.K., Cherif, I., Kalogeropoulos, C., Monachou, S., Eskridge, K. & Silleos, N. (2013) ‘Rapid error assessment for quantitative estimations from Landsat 7 gap-filled images’, *Remote Sensing Letters*, 4, 920–928.

Arika, E. & Mazure, G. (2017) ‘PROBLEMS AND SOLUTIONS FOR ABANDONMENT OF UTILISED AGRICULTURAL AREAS IN LATVIA’, *Economic Science for Rural Development*, 307–314.

Eurostat, E.C. (2013) ‘LUCAS 2012 (Land Use / Cover Area Frame Survey)Technical Reference Document: C-3 Land use and Land Cover Classification (revised)’.

Fonji, S. & Taff, G.N. (2014) ‘Using satellite data to monitor land-use land-cover change in North-eastern Latvia’, *SpringerPlus*, 3, 61.

Gislason, P.O., Benediktsson, J.A. & Sveinsson, J.R. (2006) ‘Random Forests for land cover classification’, *Pattern Recognition Letters*, 27, 294–300.

Gordon, A.D. (1999) *Classification*. 2nd ed., Chapman & Hall/CRC (Monographs on statistics and applied probability, 82).

Gorelick, N., Hancher, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017) ‘Google Earth Engine: Planetary-scale geospatial analysis for everyone’, *Remote Sensing of Environment*, 202, 18–27.

Horning, N. (2010) ‘Random Forests : An algorithm for image classification and generation of continuous fields data sets’, *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*.

Millard, K. & Richardson, M. (2015) ‘On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping’, *Remote Sensing*, 8489–8515.

Nomura, K. & Mitchard, E. (2018) ‘More Than Meets the Eye: Using Sentinel-2 to Map Small Plantations in Complex Forest Landscapes’, *Remote Sensing*, 10, 1693.

Pimple, U., Simonetti, D., Sitthi, A., Pungkul, S., Leadprathom, K., Skupek, H., Som-ard, J., Gond, V. & Towprayoon, S. (2018) ‘Google Earth Engine Based Three Decadal Landsat Imagery Analysis for Mapping of Mangrove Forests and Its Surroundings in the Trat Province of Thailand’, *Journal of Computer and Communications*, 06, 247–264.

Prishchepov, A.V., Radeloff, V.C., Baumann, M., Kuemmerle, T. & Müller, D. (2012) ‘Effects of institutional changes on land use: agricultural land abandonment during the transition from state-command to market-driven economies in post-Soviet Eastern Europe’, *Environmental Research Letters*, 7, 024021.

R Core Team (2013) *R: A language and environment for statistical computing*, Vienna (R Foundation for Statistical Computing).

Sader, S.A., Ahl, D. & Liou, W.-S. (1995) ‘Accuracy of Landsat-TM and GIS Rule-Based Methods for Forest Wetland Classification in Maine’, *Remote Sensing of Environment*, 53, 133–144.

Sidhu, N., Pebesma, E. & Câmara, G. (2018) ‘Using Google Earth Engine to detect land cover change: Singapore as a use case’, *European Journal of Remote Sensing*, 51, 486–500.

Skribane, I. & Jekabsone, S. (2013) ‘Structural Changes in the Economy of Latvia after it Joined the European Union’, *Intellectual Economics*, 7, 29–41.

Strobl, C., Malley, J. & Tutz, G. (2009) ‘An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests’, *Psychological methods*, 14, 323–348.

Suthaharan, S. (2016) *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, (Integrated Series in Information Systems).

Wulder, M.A., Skakun, R.S., Dymond, C.C., Kurz, W.A. & White, J.C. (2005) ‘Characterization of the diminishing accuracy in detecting forest insect damage over time’, *Canadian Journal of Remote Sensing*, 31, 421–431.

Zanter, K. (2018) ‘Landsat 4-7 Surface Reflectance (LEDAPS) Product Guide’, US Geological Survey.

Zdanovskis, K. & Pilvere, I. (2015) ‘AGRICULTURAL DEVELOPMENT IN LATVIA AFTER JOINING’, *V OLUME*, 8.