Table 1

BACHELOR THESIS

VISUALIZING LAND USE CHANGE

The most frequently occuring correlations between land use change and demographic factors

Abstract

In collaboration with a parallel research developing weakly supervised Siamese Neural Networks for land use change detection, this project develops a number of exploratory interactive data visualizations with the aim of answering the following question: What are the most frequently occurring correlations between land use change and demographic factors? Change detection maps will be filtered and analyzed inside geospatial analysis tools like ArcGIS, and final results will attempt to show how, for example, population size correlates with the distribution and size of detected change. Finally, the graphs will be published on a webpage for the general public and policymakers, the main target group of this project.

Table of Contents

List of figures

Chapter 1 - Introduction

1.1 Objectives and Challenges

1.1.1 Objectives

The aim of this thesis project was to understand land use change, its dynamics, and where and how it occurs. There are multiple positive effects of the different land-use changes, which includes increases in resource use efficiency, well-being, wealth, and in agricultural production[7][5]. However, the negative effects of land-use and land-use change are, often, the source of significant climatic, economic, and sociopolitical perturbations, on a small and large scale [6]. Human intervention on ecosystems that lead to deforestation, soil degradation, rapid population growth, urbanization and industrialization, among other things, are associated with land use change [9]. Additionally, land use change is considered to be one of the biggest drivers of climate change and carbon dioxide emissions due to inefficient infrastructure usage and lack of high level planning [31]. Analyzing such land evolution, therefore, through recognition of patterns and of recurring behaviour come useful in bringing to light evidence about how, based on common demographic factors, the decision making of land use in a socioeconomic region transforms and affects the surrounding ecosystems. An awareness of such patterns, thus, is likely to allow the prediction of future land use behaviours, and a counteraction against nocive human intervention through eco sustainable urban and rural infrastructure planning.

The most important aspect of this project is to focus on visualizing the land use change, in different 2D and 3D settings. The visualization can be achieved through any medium or tool available, as soon as the visualization conveys the message in the most optimal way, or brings up some new evidence. In case of lack of appropriate frameworks, libraries, or visualization software, personal applications will be developed to increase the control over the variables. The visualizations are supposed to serve as a medium for revelation of hidden, small or big scale, reoccurring land use change patterns and drivers.

1.1.2 Challenges

There are a few challenges that can be addressed in this project. The main big challenges include the quality and quantity of data, the creation of advanced geospatial visualizations, time-series playback in visualizations, interactivity, and presentation to the public through means of a website.

Land use change is a highly complex process, influencing and influenced by many other factors[9]. The research in this field cannot exist without large availability of empirical data. Despite advances in sensing technology, uninterrupted time series of sufficient length to reflect social-ecological dynamics are lacking [16]. This poses a problem for the accuracy of the results, and limits the probability of finding insights on patterns in land use change. On the other side, conveying evidence through a visualization also represents a challenge. The main data source is in large part based on geographical values. This data will be represented in 2D or 3D environments, through use of multiple interactive visualization tools. The list of tools will, necessarily, include advanced geospatial analysis software like ArcGIS.

The complexity of the dataset required tweaking of lots of parameters in the visualization process before reaching significant visual evidence of recurrence in land use. Furthermore, the additional challenge (but not a strict requirement) was to present the findings to the public on a webpage. This type of delivery was meant to allow for more easily accessible results from any device and any time, to everyone. Additionally, the interactivity was developed in order to produce exploratory types of visualization. This was supposed to allow users to observe the data from different points of view interactively and autonomously.

1.2 Research Question

The central research question for this thesis was the following:

What are the most frequently occurring correlations between land use change and demographic factors?

The research project will branch into multiple sub-questions in order to facilitate the exploratory visualization of the data, while keeping the central focus on the researched topic. Below, the research sub questions are listed:

- What are the best ways to visualize land use change in space and time?
- How is time-series aerial imagery useful for detecting patterns in land use change?
- To what extent are demographic changes influencing land use change?
- To what extent are demographic factors relevant for the analysis of land use change?

1.3 Report Outline

This project thesis has a defined structure that will aid the reader to understand the process of reaching the answer to the above listed research question. Chapter 2 describes the current state of the art in the land use change research field and the required background information about the project. The state of the art research is based on a thorough literature review of selected articles, dated in the range of the past 5 years. It discusses what are the current findings in land use change, what are the common techniques used to analyze geospatial data, how to classify land use change, what evidence has been already found, and give a conclusion. Chapter 3 elaborates on the used techniques and methods adopted in this project. It includes the used methodology and introduces some of the used datasets and software tools. Chapter 4 will initiate the CreaTe Design Process with its first phase: ideation. Here the target user is defined, as well as the process of designing the final data visualizations. Later, the specifications are elaborated in Chapter 5, and the final visualizations are shown in Chapter 6. Finally, the evaluation part and the optimization of the work is executed, and the findings are recapped in the conclusion Chapter 8.

Chapter 2 - Background and State of the art

The aim of this section is to analyze recent literature of land use studies in different regions of the world (mostly small scale, focused in a specific geographical area), and literature of correlations between land use change and demographic factors. Both land use change and land cover change will be taken into account. The section will, thus, provide knowledge about the current state of the art in land use change and state of the art visualization techniques, as well as background information of this project.

2.1 Background

2.1.1 Background information

Land-use and land-use change are a branch of science studying the way the landscape is changing over time, due to natural processes or human intervention. With a growing urbanization pressure across the world in the past century [2], land use has become an important field of study. During the current century, global ecological changes are expected to have major impacts on almost all areas of human society, including ecological, social, economic, and political aspects [7]. Furthermore, big ecological changes (as for instance global warming), are fueling other ecological changes and landscape modifications [8], which shows a certain tendency toward chain reaction effects.

To avoid unpredictable and unexpected scenarios where the ecosystem experiences a dramatic collapse or damage, it becomes useful to analyze and study the driving forces behind land-use, which have a significant impact on ecosystems, and that is mainly caused by human intervention [7]. The assumption is that the awareness of the demographic factors with greatest influence on land-use, and of the land-use patterns, might be able to allow better decision making between those entities of human society, whose positions provide the power over decision making in land use.

Such an assumption requires a lot of supporting empirical data to have significant recognition [5]. In most studies related to land-use change, the acquired data is usually limited to a specific region. Thus, the goal is to analyze studies from different regions across the globe with the aim of finding common correlations between demographic factors, land-use change over time, and land-use patterns. What are the mostly occurring correlations between land use change and demographic factors of a specific socioeconomic region?

2.1.2 RISE

RISE is the first Research center in Cyprus focusing on Interactive media, Smart systems and Emerging technologies aiming to become a center of excellence empowering knowledge and technology transfer in the region. It is a joint venture between the three public universities of Cyprus (University of Cyprus, Cyprus University of Technology, Open University of Cyprus), the Municipality of Nicosia, and two renowned international partners, the Max Planck Institute for Informatics (MPI) from Germany and University College London (UCL) from the UK.RISE is designed to act as an integrator of academic research and industrial innovation, towards the sustainable fueling of scientific, technological and economic growth of Cyprus and Europe. RISE operates by embracing the motto: Inspired by Humans and designed for Humans, producing technologies to work “in the wild''. Such a focus, geared towards applications and going beyond traditional academic confines, brings direct tangible benefits to local society while ensuring an economically viable operation of the Centre. RISE has built a computer vision model which analyzes satellite images on a daily and/or weekly basis, recording land use change wherever and whenever it occurs. This change can be de/reforestation, construction works, transformation of land for agricultural purposes, etc. The model locates areas where such change occurs and collects the geo-coordinates together with the purpose/label of each change

2.1.3 Land cover and land use

It is important to create distinction between multiple land study factors existing in research. For example, confusion can arise when mentioning land cover and land use. By definition, land use involves the management and modification of an ecosystem/environment into a built environment, such as housing, agricultural activities, pastures, managed woods or semi-natural habitats [10]. Land cover, on the other hand, is the classification of different types of lands, and is directly affected by land use. These definitions are strictly related to the definition of land use change: a process by which human activities transform the natural landscape, usually emphasizing the functional role of land for economic activities.

2.1.4 Primary dataset

This project made use of two types of datasets to generate the final data visualizations. The two types, for clarity purpose, are categorized respectively as primary and secondary dataset. The primary dataset is the core of this thesis work. The parallel PhD research of I.Kalita et al. aims to produce a weakly supervised change detection method for land use change, with a state-of-the-art level of accuracy. The siamese network, first, generates a different image as a resulting output of a comparison of a satellite image pair. Following, through use of the combination of PCA and K-means algorithms, the change map is produced. As it can be observed in the figure below, the resulting change map is a binary image of size 256x256 pixels.

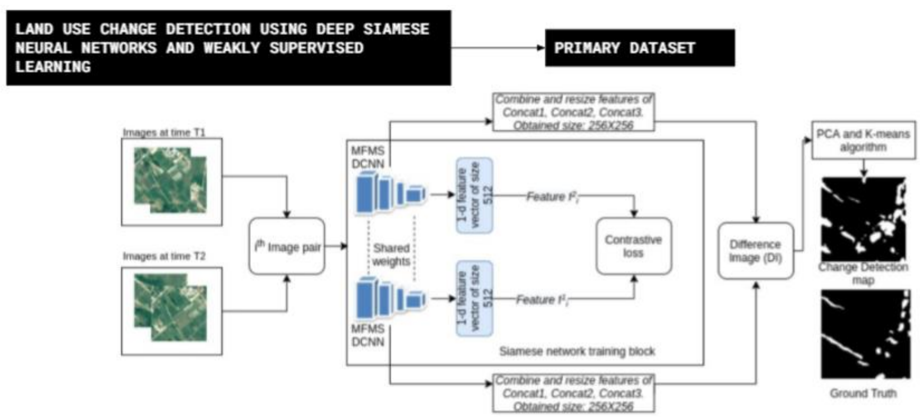

Figure 1. DNN comparing to images to generate change map

The satellite imagery is provided by Planet, a company which performs global monitoring and provides users with high quality aerial images of up to 50cm of resolution. The imagery used by this project covers a total area of 1500 square kilometers, located in between the two Cyprian cities Nicosia and Larnaca. The image dataset can not be fed in its entirety into the Deep Neural Network (DNN). Therefore, for this reason, the dataset is split in smaller patches of 256x256 pixels, size on which the output of the siamese DNN previously mentioned is based.

2.1.5 Secondary dataset

The second fraction of the used dataset includes any dataset that is not included in the change detection maps provided by the previously described DNN. More specifically, the secondary dataset is constituted by the land cover dataset provided by Copernicus Land Monitoring Service and by Michael Bauer Research datasets embedded into ESRI products.

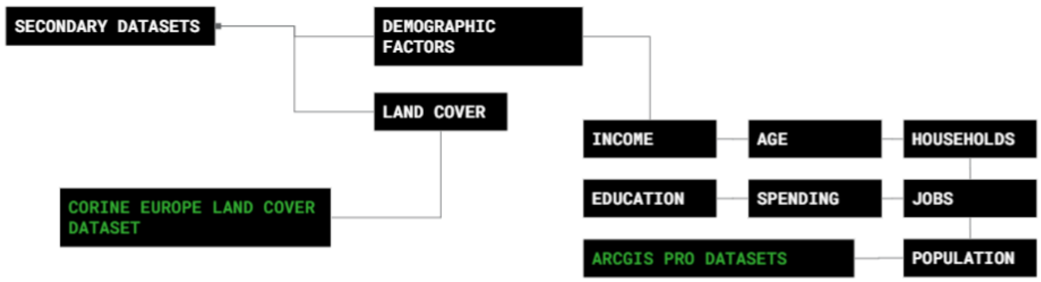

Figure 2. Overview of the secondary dataset

Copernicus Land Monitoring Service, or CORINE, is part of the European Copernicus Programme, in which data is collected by Earth observation satellites in combination with ground-based networks of sensors. The data collection provides raster and vector files covering the entirety of Europe, which contains recent information (up to 2019) about land cover under the form of over 44 classes (figure below). The dataset came in useful in the scope of this project due to its high detail and relevance with the land use change study.

Level 1	Level 2	Level 3
1 Artificial surfaces	11 Urban fabric	111 Continuous urban fabric 112 Discontinuous urban fabric
	12 Industrial, commercial and transport units	121 Industrial or commercial units 122 Road and rail networks and associated land 123 Port areas 124 Airports
	13 Mine, dump and construction sites	131 Mineral extraction sites 132 Dump sites 133 Construction sites
	14 Artificial, non-agricultural vegetated areas	141 Green urban areas 142 Sport and leisure facilities
2 Agricultural areas	21 Arable land	211 Non-irrigated arable land 212 Permanently irrigated land 213 Rice fields
	22 Permanent crops	221 Vineyards 222 Fruit trees and berry plantations 223 Olive groves
	23 Pastures	231 Pastures
	24 Heterogeneous agricultural areas	241 Annual crops associated with permanent crops 242 Complex cultivation patterns 243 Land principally occupied by agriculture, with significant areas of natural vegetation 244 Agro-forestry areas
3 Forest and semi natural areas	31 Forests	311 Broad-leaved forest 312 Coniferous forest 313 Mixed forest
	32 Scrub and/or herbaceous vegetation associations	321 Natural grasslands 322 Moors and heathland 323 Sclerophyllous vegetation 324 Transitional woodland-shrub
	33 Open spaces with little or no vegetation	331 Beaches, dunes, sands 332 Bare rocks 333 Sparsely vegetated areas 334 Burnt areas 335 Glaciers and perpetual snow
4 Wetlands	41 Inland wetlands	411 Inland marshes 412 Peat bogs
	42 Maritime wetlands	421 Salt marshes 422 Salines 423 Intertidal flats
5 Water bodies	51 Inland waters	511 Water courses 512 Water bodies
	52 Marine waters	521 Coastal lagoons 522 Estuaries 523 Sea and ocean

Figure 3. List of land cover classification inside CORINE's dataset

Micheal Bauer Research (MBR), on the other hand, provides users with recent demographic information.In this project, MBR datasets are accessed through ArcGIS directly. When performing analyses on the primary datasets, ArcGIS gives the possibility to enrich the data with NBR data, which works as a simple merge between two tables. In the final realization phase, 15 different demographic datasets are used, which include information about population, incomes, purchasing power, and age. The process of implementation of these dataset is described in greater detail later in the ideation and realization chapters.

2.2 State of the art on land use change

2.2.1 Demographic factors, land classification, and land use

To analyze the land use change literature, it is important to take a look at what are the main elements studied in the field. The first distinction that should be mentioned is the difference between land use and land cover. Land cover indicates the physical land type of the land. The main categories found in the literature covered in this review are: cropland (irrigated land, unirrigated farmland), forestland, grassland, water covered area, construction land, saline, bare land, and desert [7][10][11] Land use, on the other hand, documents how people are using the land. Land use can be categorized as a natural process or as human intervention. It is known that the biggest part of land use change is determined by human intervention [7]. Furthermore, according to Spitzer, “there are various forms of “indirect land use”, for example nature conservation, which may be included in multiple-use scenarios alongside the principal land-use types” [5]. This leads to the question of whether it makes sense to analyze more human behavior to understand which patterns in land use may occur systematically, why they occur, and how to use that knowledge to benefit ecologically friendly land decision making. This behavior could be classified as a set of demographic factors. Demographics include any statistical parameter that can describe the evolution of a population. There are, however, some particularly important factors, which include: population age structure, population size, population pressure, density, sex ratio, and mortality [3]. The purpose of this review is, as mentioned above, to determine whether there is, or there could be any correlation between the demographic factors and the way land use changes over time.

2.2.2 Resources and land use

When analysing land use change, understanding its underlying driving forces comes out to be a handful. The research about land use patterns can be expanded by researching what are the patterns in land use drivers. There are multiple factors to observe when a specific geographical area undergoes a period of intense human intervention on land. Previous research in isolated areas, located in China, Mexico, and Ethiopia [6][7][10], shows two dominating scenarios: in one case the main drive is related to the population growth and to the economic growth, while in the other case the main drive seems to relate to the immigration and resource scarcity.

Both scenarios are similar, yet, with some substantial differences. The first originates in an already settled social group in a specific region. Due to different reasons, including greater affluence of industries and increasing international economic interest, the socioeconomic region experiences an increasing rate of land expansion. This leads to the increasing population pressure and built area expansion. With the increasing number of population members, the need for resources increases accordingly: commercial agriculture and water become largely required. Additionally, food consumption changes, transportation infrastructure grows, and energy production facilities are built on the territory. These and other smaller factors sum up to the land use change in the observer region.
The second scenario, on the other hand, originates from the initial circumstances of resource scarcity. This means that the driver of land use change in a specific region that experiences intense land use change is its resource availability. The main resources that can be in many cases considered as the driving forces for land use change are water and arable land. Due to population immigration into the region, the growing population pressure is observed. In this scenario, thus, the mechanism of human intervention does not change when compared to the first scenario. In fact, it is mostly the same, and the main difference lies in the underlying reason that drives land use change. The two cases partially bring some light over what could be some of the patterns in land use change.

2.2.3 Demographics and land use change

This section leads to the center of this review. The current state of the art findings in land use change and demographic factors will be shown and analyzed. As mentioned in the introduction, lack of large scale empirical studies are lacking, or do not meet the quality standard for supporting evidence in a significant way. Thus, the analysis will mainly focus on smaller scale studies, ranging between town to region scale, and in multiple locations in different parts of the globe.

The analyzed geographical regions cover East Asia, Latin America, Africa and Europe. Each study located in different regions shows a different evidence of land use change. This is due to the fact that there are many factors that could be studied, as well as multiple points of view. For instance, some studies focused exclusively on the demographic factors in relation to land use change in newborn metropolitan areas, some others focused specifically on very large, growing urban areas. Population growth, science technology development and growth of the economy have been found to have an impact on land use [10]. The population size is highly related to land use change, and it varies in multiple ways. In some cases, the population size increases due to the flourishing economy or to the vast availability of resources in the area. In other cases, the population size increases due to immigration factors. These can originate from resource scarcity or national conflicts, and according to Martha Bonilla-Moheno, T. Mitchell Aide, and Matthew L. Clark, "case studies from Mexico have shown that national and international migrations have played a key role in determining patterns of land cover" [6]. It has been also proven that population pressure occurs in developing countries along with decreasing mortality and increasing fertility rates [6][7]. This increase leads to many other changes, as a chain reaction. The main immediate change that is observed is the requirement for higher supplies of food. This allows commercial agriculture and farming to expand dramatically. Both irrigated land and unirrigated farmlands take a big part in land use. The correlated exploitation of land, deforestation and irrigation systems, then, have detrimental effects on the environment [1][5].

Technological advancements inside a geographical area seem to be one of the main driving forces of the growing population. Medical technology, on one side, increases dramatically the lifespan of citizens, and decreases mortality rates [7]. Therefore, technology has an indirect effect on land use change, yet quite profound. This correlation, however, is relevant mainly in the transition between non-urban to urban areas. In fact, highly developed areas (i.e. megacities with population size greater than 10 million inhabitants) present a more complex demographic influence on land use change. For instance, if a developing region in Mozambique is compared with the capital city of Columbia, Bogota', some fundamental differences in land use change and driving forces of change can be observed. In the first case, urbanization takes place for the first time: a rural area grows in size, population, and infrastructure. There is a parallel increase in cultivated land, which is the main cause of the deforestation in the area. Furthermore, most of the cultivated land is taken care of by smallholder farmers, which affects the general decision making dynamic over land use. In the second case, the megacity Bogota' seems to be following different land use change patterns. The population growth is still at the base of the expansion. However, the overall land use change is related to a more complex set of factors. For instance, commercial agriculture becomes dominant, and greater amounts of acres are being managed by one organization. There's also a higher degree of artificial areas created. This includes building lands, artificial non-vegetated areas, mines, dumps, construction sites, industrial units, commercial units, and transport units. This change adapts to topographical factors, as well as to the neighboring municipalities' locations. The land use spread, in great part, is shaped by the infrastructural organization of the city itself [2][11]. As Claudia P. Romero states in her research on the megacity of Bogota', "a transition from agricultural or vegetated areas to artificial areas has mainly occurred, as expected, around the metropolitan area of Bogotá and also follow the spreading of urban areas along transport infrastructures and around secondary cities" [11].

The urban fabric seems to develop differently in different parts of the world, in different topographical conditions, as well as in different stages of demographic development. Every study agrees on the fact that human intervention is the main cause of land cover change, while almost every study agrees on the fact that land use change has a dramatic impact on ecosystems. This, mostly negative, impact leads to more societal implications that will, in turn, affect the land use change. The ecosystems are proven to be highly susceptible to human intervention[6][10][11], and the ecological damage decreases the quality of life of inhabitants of the region. Thus, protection policies may intervene to slow down the progressive damage. This intervention will be reflected in land use change under the form of reduced deforestation, reforestation, as well as less intensive agricultural practices with inferior soil exploitation.

Chapter 3 - Methods and Techniques

3.1 Subjects/participants

This project does partially rely on user's feedback and, thus, subjects involved in this project should be mentioned. The structure of this work includes, consequently, state of the art research, design, discussion, and conclusion. Design part, as it will be described in detail in the following sections, includes an evaluation phase. In this part the goal is to optimize the results for optimal user experience. Given the nature of this project, data visualizations should be tested and evaluated in terms of understandability, clarity, and webpage performance. Furthermore, in the specification phase of design, target users are defined. Therefore, the subjects should include individuals from such target groups. Due to time constraints and high reliance on state-of-the-art work executed in parallel, the evaluation for understandability is omitted, and, as it will be described later in the evaluation section, the focus is set on performance and optimization.

3.2 Instruments/measures/variables

3.2.1 Interactive visualization software

Due to the visual nature of this project, a set of visualization tools needs to be chosen and adopted through the entire length of the project. Especially in the initial stages of the project ideation and realization (phases adopted in the design process, described in the next section), the exploration of the possible software tools and their features is crucial.

The interactive visualization tools come in large numbers and with very different orientations and sets of features on the market. The best way to choose a tool or multiple tools is by experimenting. However multiple factors need to be kept into consideration. First, the duration of this project is relatively limited and short. Second, the target group is reached through web-based story telling. The final content needs to be delivered through means of a website and, thus, the produced visualizations need to be compatible with web platforms, as well as with a simple and long lasting embedding potential. Another example of a considered factor in the choice process is the aesthetics of the visualizations produced by the tool. This last factor is more of a personal choice rather than a purely functional approach. The reason for this is the idea that a well balanced and graphically designed visualization is more appealing and impactful for the general public and for any kind of observer. Given these factors, the choice of the tool becomes more of a systematic choice process, and the final choice can be clearly reached by satisfying the above-mentioned criteria.

Some of the explored tools include the following:

1. ArcMap:

   ArcMap makes part of the ESRI tools bundle and represents one of the most widely used geoprocessing tools by professionals and academics. The tool provides a very large amount of geoprocessing features and analysis plugins, which makes it possible to dig into large datasets and find evidence through many different pathways. The tool, however, presents some throwbacks on the processing time efficiency side, as well as on the web adaptability side. By experimenting with the tool, runtime tests were made on single patches of the primary dataset. The results showed a large amount of time necessary for most of the analyzes and plots. Considering the number of patches in the final dataset (2400+), such time performance was not optimal. Furthermore, integration with web platforms is relatively complex and time consuming, too. As such, this tool has been, in the end, not part of the toolset.

2. ArcGIS Online (AGO):

   ArcGIS Online is the online version of the ESRI bundle, provided by the same company. The tool presents an up-to-date user interface, ease of learning, a smaller number of analysis features compared to the local versions of ESRI bundle (ArcMap for example) which, however, is still a great number and optimal quality when compared to other visualization tools. Runtime in AGO presents similar magnitudes to ArcMap with one substantial difference: AGO calculations are not made locally but remotely, on the ESRI server. This is a strong advantage in terms of computer resources, usages, time, and reliability. Since the calculations are run on a server, the rendering of visualizations can be done in parallel with the other tasks that this project presents, without negative influences on the performance of the machine on which the project is made. The process is more time efficient to ArcMap because it is possible to run multiple analyzes and plots simultaneously with the same time performance as if those analyzes would run singularly. Reliability is also a key point in AGO. Any kind of local machine error that would interrupt the correct functionality of a program cannot affect the progress of the rendering of a visualization in AGO. All these points with, additionally, the good quality graphic design and easy web embedding make ArcGIS Online a perfect candidate for this project.

3. Kepler.gl:

   Created initially by and for Uber, Kepler.gl is a very simple yet powerful visualization tool. It is web based similarly to AGO, but with way less features. The offered functionality by Uber’s

Kepler is rather straightforward and highly optimized for the few visualization styles that it offers. The tool presents a very high-quality graphic design, very high data loading and plotting speeds, and extremely easy web embedding possibilities. Kepler.gl quickly became part of the tools stack of this project.

4. Tableau:

   Another interactive visualization tool with high potential. This software has an initial advantage of being part of a personal set of skills and experience. Tableau is different from all of the above-mentioned tools because it's main focus is on statistical studies rather than geospatial analysis. While it provides a few geographic plot features, it cannot be compared to Kepler.gl or AGO. Loading speeds in tableau are optimal, the aesthetics of the visualizations are of good quality, and the statistical orientation adds something that the previously mentioned tools do not present. Additionally, Tableau is very easily embeddable into web platforms. Thus, tableau is a good addition to the tools stack of the project.

5. Leaflet.js:

   Leaflet is a JavaScript framework, as the name suggests, and is a very broad set of features for, both, geospatial analysis and statistical studies. It combines elements of Kepler.gl and Tableau into a good quality, characteristic visual style. As all of the JavaScript frameworks, the system works closely with web development, which makes it perfect for the final delivery of this project. However, Leaflet.js presents some major throwbacks. In the first place, the tool requires a proper JavaScript proficiency in order to be executed properly. It is also time consuming to understand the various functionalities of the framework, as well as to implement the visualizations and the large amounts of data. The decision of not adopting Leaflet is primarily based on the high amount of time required to reach results that could be more easily reached through use of some of the above-mentioned tools.

6. Processing Framework:

   Java, in a similar way to Tableau, has the advantage of being part of a personal stack of skills and broad experience. Processing is a framework made for artists and designers and, as such, simplifies the process of drawing elements onto the screen. Unlike any of the previously mentioned tools (partly except Leaflet.js), Java and Processing are explored in this context without any geoprocessing or statistical features. Thus, the idea is to program a custom software that could give the final visualizations of this project a very personal touch. The framework is also used for exploring the different approaches that can be adopted to tackle the provided primary dataset. Generally, the features produced through this process are time efficient. However, Processing cannot be considered as part of the visualization tools stack specifically, but rather part of the general tools stack required by this project. This is due to the very time consuming nature of software writing, as well as to the uncertainty of realizing a working visualization that will, in the end, come into use for the final phase of this project.

3.2.2 Custom software

As it will be described in greater detail in the following chapters, a custom software has been developed for data filtering and simulation purposes. The programming language adopted is Java. The development environment, or IDE used is IntelliJ by JetBrains, and the used Java SDK is *openjdk-15*. The custom software made use of third-party libraries and algorithms:

- ➔ Simple Features GeoJSON Java (mil.nga.sf.geojson) - developed at the National Geospatial-Intelligence Agency (NGA) in collaboration with BIT Systems within the MIT license. The library provides an API for creation of GeoJSON files.
- ➔ OpenCV (4.5.2) - open source computer vision and machine learning software library used for contours detection and hierarchical separation of clusters of change.
- ➔ Concave Hull algorithm by Udo Schlegel v1.0 K-nearest neighbours approach for the computation of the region occupied by a set of points.
- ➔ Processing (framework) - PApplet imported into Java

3.3 Design

As a grounding framework for the design of this project, Design Process for Creative Technology by Mader & Eggink (2014) will be adopted. The framework aims to integrate design concepts from multiple disciplines including Interaction Design, Engineering Design, Human-Media Interaction and Industrial Design, and its underlying method is to combine the use of different technologies to develop a solution for a target audience. The Design Process for CreaTe defines design explicitly, and ranges between two different models: Divergence-Convergence and Spiral. For the scope of this project, the Divergence-Convergence model is used. Such a model consists of two phases. The first, Divergence phase, consists of generating as many ideas as possible and opening the design space. Here, this project will focus on exploring all the possible Interactive Visualization Software tools, brainstorming, sketching, doing exploratory visualizations. The phase will also focus on understanding the available data, and how it can be arranged and prepared for the future visualization. Subsequently, in the Convergence phase, the design space is reduced by making design decisions that bring the number of possible solutions to a lower amount. The Divergence-Convergence process is integrated in the Ideation, Specification, and Realization phases.

3.3.1 Ideation

In the ideation phase, the starting point is considered to me the requirement or the goal of the project. In the scope of this thesis the goal is to, primarily, visualize the land use change in time, and, additionally, to find possible correlations with some demographic factors. Thus, the ideation phase is mainly focused on the visualization. In this part the possibilities of visualizations are explored, as well as the tools used for creating them. As mentioned before, this phase includes both divergent and convergent parts. In the first, exploratory research of the fitting visualizations is made, possible target groups are being listed, as well as possible delivery modes. In the convergent phase, design choices are made to reduce the exploratory space. Hence, this phase makes the first statement of what is the destination group of this project. Thus, it is chosen who are the stakeholders of the project and, in regard to land use change visualization, it is decided who are the individuals who can make the most impact from the knowledge provided by this report. From this point, a transition to the specification phase is made.

3.3.2 Specification

Generally, this phase would require the use of multiple prototypes to explore the design space. Due to the digital nature of a data visualization, the process can be differentiated, but it still remains close to a physical prototyping process. The Creative Technology Design framework implies a continuous interplay between technology and user needs in this phase. Furthermore, some evaluation moments are made to make appropriate design choices. For a data visualization this phase would work as an iteration process where dummy visualizations are made and their understandability is evaluated. The evaluation of the understandability will lead to a new, adjusted visualization. The lack of the final dataset until the late stage of this project influences the specification phase in a way that data visualization prototypes are re-evaluated in every cycle according to the understanding, but also to the coherence with the current version of the dataset. For instance, the first versions of the change detection maps covered a time period of over 3 years, while the later versions were limited to 2 weeks time ranges. The resulting maps were radically different from each other in terms of amount of detected change.

3.3.3 Realization

In this phase the product is brought up together by following engineering design models like the Waterfall model or the V-model. The purpose of this phase is to meet the initially proposed requirements and, along with evaluations, bring the first viable product/s. In this phase the tools used are well established, visualization techniques sorted to match the user requirements, and the data stories in great part completed.

3.3.4 Evaluation

The final evaluation phase is rather straightforward and is expected to provide a final confirmation that requirements are met through user testing. Evaluation is integrated in all the previous phases to ensure the correct development of the data visualizations, and this allows for minimal margin of error and workload in this final development phase.

Chapter 4 - Ideation

In this section the process of creative orientation is described. Based on knowledge gathered through the state-of-the-art research, personal knowledge, and further web research of related works, the initial ideas on how to visualize geospatial phenomena are being introduced. Due to the fact that for this project a significant amount of time has been invested in data processing, this section will not only cover the process of design of visualizations and stories, but also the process of adapting the large amount of data to the limitations of web-based visualization tools, and the way to obtain classification maps built upon detection maps.

4.1 Data

This project is entirely based on geospatial data and is the very foundation of every visualization development. There are also secondary datasets accompanying the primary geospatial data, whose main aim is to bring evidence over some specific correlations between land-use change and demographics. The study of the provided data provides a clear scope of the study, and filters out obsolete visualization types in an early stage.

4.2 Concept design

The concept design in the very early stage was mainly based on exploration of related work, state-of-the-art concepts and of software tools. A collection of visualization tools is gathered and the features of such tools are explored broadly. This is meant for gaining an overview of what is, both, achievable and, considering the time constraint of this project, what is the most impactful and most pleasant to see data visualization. Immediately after, a clear design strategy needed to be adopted. Based on research, a useful design pattern emerged. The pattern, described in [14], allows for developing tools and methods for using and interpreting large volumes of geospatial data. The process consists of five steps: real world distribution of the studied phenomena, purpose of the map (or visualization) and the intended audience, collection of appropriate data, design and construct of the map (or visualization story), and final evaluation of user's satisfaction and understanding of the product. In the following sections, these design patterns steps are described in detail.

4.2.1 Real world phenomena

The very foundation of this project is to provide grounding for future research in environmental sustainability. The research question of this project is self-explanatory in terms of real world phenomena that are studied through the intended visualizations. The relation between land-use change and demographic factors implies that elements of both should be paired in a great amount of visualizations. The primary dataset shows how land is being used over time, and thus aims to deliver an overview of what are the most frequent human-activities done on the territory. Furthermore, the dataset shows the spatial scale of such activities. The classification of the activities allows for assuming what are the most environmentally intensive practices, and in which regions they tend to develop. Through attribution of socioeconomic factors and demographics, it is expected that land use development could be predicted for any socioeconomic region and not only the analyzed one, or that land use could be optimized and made more efficient on a high scale level.

4.2.2 Target group

The goal of this project is to create awareness of land-use behaviour in different regions and in different times. Such a necessity is created by climate change, as described in chapter 2. Thus, in order to choose the adequate target group/s for the data visualizations, it makes sense to select those entities, or socioeconomic influencers [17] who have the most influential power over land-use. However, the exploration of the possible target groups is not limited to the climate change urgency only. This section is mainly focusing on creating direct impact and on creating awareness.

According to [14], land is governed by formal and informal institutions. The local governments usually decide over detailed land uses, while the upper level governments provide planning system frameworks and entact environmental legislation. This means that despite the fact that high level governments usually do not have access to the detail of information over land use than do smaller entities, there is still a strong vertical relationship present. Affecting the frameworks and policies, in theory, would have a top down influence until the very specific type of land use. Therefore, it is reasonable to consider policy makers of higher levels of governments as one of the target groups of this design.

Another target group could be represented by the general public, citizens of large urban areas, and agricultural stakeholders. For this specific case, however, the project's aim would shift from direct impact to creating awareness. This process would have fundamentally similar dynamics, since in both cases we influence human individuals, policy makers and citizens. Awareness raising campaigns, generally, have the purpose of increasing concern, informing the targeted audience, as well as creating a positive image and attempting behavioural change[18]. Furthermore, according to [17], research shows that awareness raising is integral throughout the whole process of change, and not only at early stages of the process. Thus, with the great urgency that climate change poses, and the necessity for fast and radical change in patterns of special development this combines well.

4.2.3 Data collection

The data preparation phase of the ideation has a dualistic nature. This project is in great part based on data provided by a parallel postgraduate research of I. Kalita et al. This data, as mentioned before, contains the output of Deep Neural Networks (DNNs) and represents the change detection maps (CDs).

Land use, naturally, is best described by location based data. The collection of the data is, mainly, unrelated to the target group in this project. This is due to the fact that the availability of high quality geospatial data, which is detailed and collected periodically in time series, is hardly available, given the resources, time constraints, and the scope of this project. Hence, the process of data collection is, in part, rather random and explorative, rather than systematic and with a specific goal. The main two borderlines for the data research are:

➔ The information needs to represent a specific demographic phenomenon
➔ The information needs to be correlated to land use change.

In the scope of this project all the available data in the late phase of the development can be classified into two categories, primary and secondary datasets:

➔ The primary dataset includes all the data related to land use change. As mentioned in Chapter 2, any data related to land cover is not part of the land use related data. This class will be primarily composed of the information provided by I. Kalita et al. (Land Use Change Detection Using Deep Siamese Neural Networks and Weakly Supervised Learning ,2021) in the form of CDs.
➔ The secondary dataset included all the data not included in the primary dataset. This dataset is meant to enrich the primary dataset, and create correlation visualizations between land use and demographics. This class will contain a broader range of datatypes compared to the primary data.

Regarding the collection of the final data, this project can focus on open source data provided by different organizations, as well as copyrighted data which requires an economic contribution in order for the access to be granted.

4.2.4 Map construction

The establishment of the concepts is attained through different methods and techniques. These include individual brainstorming, moodboarding, storyboarding, as well as design schemes developed in the state-of-the-art field research. Initially, after a broad study of the provided data, the brainstorming gives the possibility to explore possible visualization possibilities, story combinations, as well as possible delivery kinds to the final target audience. Storyboarding and Moodboarding work complementary to brainstorming, and are intended to provide an extra dimension of inspiration through visual raffigurations.

To simplify the design process, it comes handy to assume the ranges of data types and visualization types. According to the geospatial visualization taxonomies studied by [19], the commonly proposed display modes include graph based techniques, geometric projections, pixel oriented techniques, hierarchical, icon based techniques. To further simplify, we can group all these techniques in maps, graphs, and tables.

The design approach, can be classified, as mentioned in [14], as following:

data-driven, where the technique is structured by the data types
representation-driven, where the balance between information and visual cluttering is researched
technique-driven, where the visualization coding is done through the grouping of common visualization techniques
Challenge-driven, where the visualizations are built around problems and challenges that specific techniques address.

For the scope of this project it seemed appropriate to focus on data-driven and representation-driven approaches. Firstly, as mentioned in the previous section, the selection of the data is rather exploratory. Thus, it comes in handy to explore which visualizations fit best the chosen data. This is relevant especially in the early stages of this project. In order to finalize the work, however, it seems reasonable to adopt a representation-driven approach. In such a way, the design process of the final visuation stories focuses on finding the right balance between what the user sees and what the user gets from a specific illustration. Due to the very dense and large nature of geospatial datasets, it is important to not cause visual or informational overload, and to keep the message as clear as possible for any of the chosen target groups.

Design classifications by interactivity are important to mention too. According to [19], classification by interaction is encountered in most of the geospatial visualization designs. This element fits appropriately in the representation-driven approach. While classification by data (or data-driven approach) is deeply rooted in the characteristics of the data, classification by interactivity depends mostly on the needs of the user and of the technology. In the end, the visualization is finalized by taking in consideration techniques, interaction, and clear storytelling.

4.2.5 Evaluation

The evaluation step is meant to finalize the design concept. This, however, does not imply a strictly sequential nature of it, but on the contrary happens throughout the whole design process and can be present at any stage. The evaluation phase focuses on the validation of specific choices, and comparison of different concepts. It allows us to narrow down ideas and to get closer to a final concept or a set of concepts. The evaluation phases might be done by user tests or performance tests. User tests will primarily focus on understanding where the interaction can be improved, as well as how to increase the understandability of the presented data. This comes in crucial in order to be able to create impact and to affect the opinion of the target audience. Performance tests can mainly be focused on loading time recordings or browser compatibility.

4.3 Interaction concepts

The construction of the visual spatial representation defines just a fraction of the final concept. Adding an extra dimension to the visualizations to improve visual exploration and depth of information sharing is an important factor in the design process of this project. Interaction is the additional layer that fits the scope of the project on both functional and non-functional aspects. The interaction ideation takes in consideration the extent of interaction, or how often it is used in the visualization, which weight does it have in the visualization, and how it affects user’s attention and understanding.

Interaction concepts can be developed based on the relationship with visual representations and expected user interaction. As [19] states, we can give a systematic approach to the design of interaction by assuming the relevant data types and display types (or visualizations). The most likely relevant data types in geospatial dataset are existential, location, and thematic. On the other side, the common display types include maps, graphs, and tables.

4.4 Ideation conclusion

This section concludes the ideation phase. At this point, the overview of the data, interaction and target audience are expressed.

Multiple possible target groups were specified in the previous sections, scientific community, public, and policy makers. The state-of-the-art nature of the topic and the materials covered in the visualization process of this project can well fit with all of the above mentioned target groups. By doing further research and based on the information mentioned in Chapter 2 in the state-of-the-art of the field research, stakeholders and related phenomena (i.e. Climate Change), policy makers result to be the best option. Given the urge of modern society to become more sustainable and less environmentally demanding, it brings the attention of this project to the group of individuals which can affect decisions that will end up affecting the sustainability development. It is important to mention that land use change visualization oriented towards the general public (creating and sustaining awareness) and the scientific community (further research in the field) could have, as well, an impact on the further sustainability developments. However, the magnitude of the influence of the policy makers in sustainability development strictly related to land use seems to be greater. An example of an important policy making entity to target could be the managers of large agricultural corporations. As can already be observed in the Corine dataset [24], which will be further discussed in the following chapter, a great part of land cover is constituted by Agricultural areas. Subsequently, improving land organization in such areas would theoretically lead to an improved environmental demand.

Chapter 5 - Specification

The following steps aim to narrow and to finalize the data visualization design concepts and stories. In particular, specific visualization types, target group, interaction, data and evaluation practices will be further defined. Subsequently, the final deployment and presentation modes will be discussed, as well as the data management. This section will also show how the expected outcome is expected to be reached, which kind of software tools are being used, and which sources are useful for the successful realization of this project. In other words, the specification section is meant to provide a solid grounding for the following realization phase.

5.1 Datasets

The specification phase aims to come to a final decision over data types, scales, formats, organization, et cetera. After testing multiple Interactive Visualization tools it comes clear on how data should be approached. Common pros and cons become clear and, based on the available visualization features of a software tool, the concept of the information data structure is finalized.

5.1.1 Change detection maps

Datasets composed by CDs are the primary source of data provided by the interested parties. Since the early stage of the project the design decisions are shaped by the structure of such a data collection. These maps, based on the research performed by I. Kalita et al. (2020), are supposed to be visualized and shown to the policy makers, which is the chosen target group. The maps consist of multiple formats of data, including images, textual files and very high resolution imagery. In this phase it is decided to discard the very high resolution imagery, since it doesn’t provide any value to this project.

Figure 6. Split_A file (left), Output file (middle), Split_B file (right)

The imagery representing an area of interest in time A and time B can be also classified as obsolete, at least for the first and middle stages of the realization of the concept. Such images could become handful in case of important evidence found at the very late stage of the development, where some strong evidence of a particular area could be easily shown to the target group with such pre-processed and cropped satellite images. The very core part of this dataset, the CDs, should be the main focus of the realization phase. More precisely, the most valuable part is the change detection map under numerical form, thus, contained in the respective textual files. The numerical data will need to be plotted and, possibly, visualized in a time-series sequence. Furthermore, any correlation between land use change and demographic factors will be primarily based on the information provided by this dataset. In the figure below, an example of a CD map retrieved on April 30th 2021 is shown. The region depicted by the data, in this case, includes part of urban and non-urban areas of Nicosia, Cyprus.

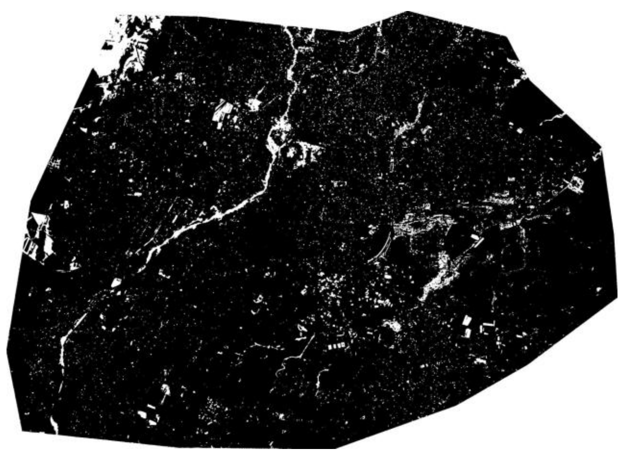

Figure 7. Change Detection map 30042021

Below are listed the key features of the dataset, while Figure 14 shows the combination of each single file from 2 different folders to illustrate the structure:

1. JPEG: Folder A: Contains the image taken at time T1 (Date: x1) --> Say imageA.jpg

Folder B: Contains the image taken at time T1 (Date: x2) --> Say imageB.jpg

split_A: Patches of imageA of size 256X256X3

split_B: Patches of imageB of size 256X256X3

2. TIFF: Folder A: Contains the image taken at time T1 (Date: 14/11/2017) --> Say imageA.tif

Folder B: Contains the image taken at time T1 (Date: 10/07/2020) --> Say imageB.tif

split_A: Patches of imageA.tif of size 252X252X3

split_B: Patches of imageB.tif of size 252X252X3

3. Lat_mask_entire: Each file contains the latitude, longitude, and date of each patch located at (TIFF->split_A folder)

4. Output: Contains the changed map. The change map is obtained based on the images available inside the JPEG->split_A and JPEG->split_B folders.

Note: The change map contains a lot of false-positive changes. It is an initial investigation only.

5. Lat_mask: Each file contains the latitude, longitude, date, and, class label of each change map image available inside the Output folder.

Additional Information:

6. JPEG and TIFF shared the same images with a different extension (jpg and tiff). The geological information is available on .tif extension images only.

7. Example of files located in lat_mask_entire:

Filename: Merge_14112017_0_0.txt: It contains the latitude and longitude information of the image Merge_14112017_0_0.jpg or Merge_14112017_0_0.tif

Information inside the file: <(35.18090331740593, 33.30699441910949, 0.0) 14112017> First two values represents latitude and longitude, ignore the value 0.0 and last value represent the date 14/11/2017

8. Example of files located in lat_mask:

Filename: Merge_14112017_0_1008.txt: It contains the latitude and longitude information of the change pixels available at Merge_14112017_0_1008.jpg. The change map is obtained by comparing the two images i.e. Merge_14112017_0_1008.jpg (available at JPEG->split_A) and Merge_10072020_0_1008.jpg (available at JPEG->split_B)

The information inside the file: <(35.146915772953236, 33.289444145504994, 0.0) 14112017 C> First two values represent latitude and longitude, ignore the value 0.0, the fourth value represents the date 14/11/2017, and the last one is the label of the class (C represents the label of the class)

Updates of the primary dataset containing information about change detection are provided through the entire duration of this project. Each update contains updated time frames, outputs of the DNN and geographical extensions (city and country scales). The structure of the dataset, however, is pre-planned and remains fixed until the end of the project. Thus, another requirement for this project is to develop a system to visualize batches of data not yet received. Such a working system requires high flexibility, and a high amount of simulation in order to be completed successfully on time. Based on the first batch of data received, it can be supposed that the main simulation activities should focus on generating change classification data, different time frames and different geographical scales. Classification data simulation can be randomized, where some common land uses are integrated for testing purposes. Another aspect of this dataset is the time variable. The dataset is divided into time sections representing the comparisons of two images in a two weeks range.

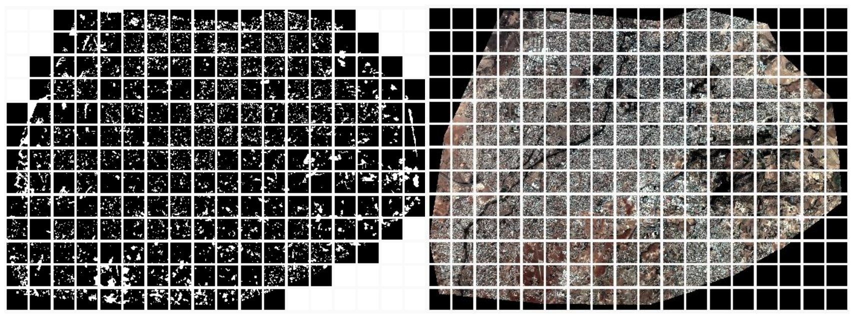

Figure 9. Assembly of 204 patches. Change maps (left), aerial imagery (right)

Given the dataset information above, it is possible to derive the possible visualization specifications. One point of focus is put on the visualization of the locations contained in the dataset, and, thus, the refiguration of where the land use change occurs. This is a crucial side of the entire project, where insight, first, is brought over the overall distribution of the change independent from any other variables. The visualization should be capable of showing clearly, thus, to effectively transmit the distribution of change in the area of interest, as well as the way this change transforms in time. Due to the high resolution of the data and the possible choice of web based presentation to the stakeholders (further discussed in section 5.1.3), detection change data in a raw state would most likely be impossible to visualize on a web page with an acceptable loading time, thus interfering with the overall experience. This leads to an additional specification regarding the performance. Data filtering and optimization is required for the successful completion of the project.

5.1.2 Demographic Factors

The secondary datasets can be considered as the ones containing any information other than the information about land use change.

As mentioned in the previous section, the efficiency requirement is in place. Thus, any required optimization related to the demographic factors data is also necessary. The filtering and optimization for the secondary data will in great part rely on the structure of the primary data. This means that for example, the scale, resolution, or time ranges will need to be taken into consideration. If the scale of the secondary dataset (i.e. pro capita expenditures power) covers entire regions of europe or multiple countries including Cyprus, while the change detection map covers only a specific region of Cyprus, the secondary dataset will need to be filter in a way that any data point that lats outside of the bounds determined by the extension of the primary dataset will be omitted.

The specification section also provides an indicational list of possible demographic factors that should be considered for the final data visualizations. Despite the fact that in the previous chapter it has been explained that the research of datasets related to demographic factors is done in a rather randomized way (data availability reasons), it is worth stating that there are some largely used groups of demographic data adopted in related research. Such common factors have been studies in chapter 2 and are listed below:

- ➔ Income per capita
- ➔ Population size
- ➔ Population growth
- ➔ Science technology development
- ➔ Immigration factors
- ➔ Mortality rates
- ➔ Fertility rate

Hence, an attempt to retrieve data related to any of the above listed fields is a specification for this project. In case of unavailability or poor quality of the data, the research of demographic factors can be branched into other fields.

5.1.3 Target group

This section will focus on discussing further the possible target group and how the design of the visualization concepts might be affected by it. As resulting from the ideation phase described in the previous chapter, the chosen target group for this project is the policy makers. Now it is the moment to reflect on how to approach this target group with a well fitting end product.

This process consists in defining specifications for the visual design, presentation style. As it has been mentioned in the contextualization section of the previous chapter, exploratory visualization seems to be the most appropriate visualization approach for the chosen target group. The approach consists in the process that involves a field expert creating maps based on relatively unknown data (in the case of this project state-of-the-art), and the purpose is to give the expert the possibility to solve a problem [24]. Research done by [24] developed a taxonomy related to the process of creating an exploratory visualization, as well as the definition of it. The further functionality specification related to this will be discussed in the next section.

On the other hand, the presentation style, the final delivery, is also affected by the specified target group. Other factors come into play here, as for example the available resources and professional network. Two are the main influencing points: interactivity as specification, and access to the research website of the supervisor of this project.

Chapter 6 - Realisation

6.1 Process

6.1.1 Data Proofreading

After the initial experimentation with the dummy data, the first drafts of the primary dataset became available. At this point, the project has reached a transitional point where the exploration of visualization designs shifted towards being more oriented towards the final primary dataset, instead of being fully explorative. The first drafts of the final data were incomplete and experimental. The information contained in the dataset was not useful for the final realization, however it was useful to understand the structure and the meaning of the dataset. This is due to the fact that all the drafts were based on a pre-planned structure which was meant to be kept until the latest development stages.

Given this information, the data check was performed. Due to the algorithmic nature of the generated data in the primary dataset, possible errors needed to be found and fixed. Without this process and by omitting small errors in the structure of the data, with the increasing scale of the dataset there would be an increasing magnitude of error, too. The data check is performed on a structural level, with a focal point placed over the coherence of the data. It is important to mention that the correctness check is not performed over accuracy levels and correctness of change detection (content of the primary dataset) since it is not within the scope of this project. Furthermore, such an accuracy check would mainly need to be verified on the very Deep Neural Network generating the data. Naturally, due to the very high resolution of the data, the only way to verify the coherence of the data is to perform visual plots. After performing some simple data visualizations and by plotting raw data onto a geographical map, some inconsistencies have arisen. More specifically, every dataset part of the Lat_mask_entire (according to the dataset structure shown in the previous Chapter), presented incorrect coordinates. The positions were misrepresenting the original change detection map.

Figure 10. Example of an incongruence found in the data, on the left the plot of the raw data, on the right the change map (ground truth)

As shown in the figure above, the plot of the raw detection change data on the left is not matching the file of the Output folder on the right, which is to be considered as the ground truth. The raw coordinates appeared to be mirrored and rotated by 90 degrees. To tackle this problem, top down investigation was made in order to spot the step in which the generation of the data was compromised. As shown in I. Kalita et al. (Land Use Change Detection Using Deep Siamese Neural Networks and Weakly Supervised Learning ,2021), each pair of images is compared and the outcome is initially produced under the form of a pixel coordinates set. The following step is the conversion to geographical coordinates. In order to debug the misrepresentation issue, the pixel mask (the first output of the Deep Neural Network) has been requested in order to perform further analysis. After observing the plot of the pixel mask, no difference from the raw coordinates in the dataset has been found. Thus, it became clear that there was a bug, a typo in the code function dedicated to coordinate writing. Indeed, it ended up being the case that the software generating coordinates persistently inverted the x-axis with y-axis, and, therefore, produced mirrored results. The second inconsistency was noticed in the file names of the primary dataset. As mentioned in the previous chapter, the naming of the files has a specific formatting. All the names in the dataset contain two parts that make each file uniquely distinguishable, the date and the upper left coordinate (the first coordinate) of the region that the dataset is covering. Each image in the Split folders and, thus, every Mask file has a total range of 256x256 pixels or data points. This is reflected in the names' first pixel coordinate, where all the numbers are in a right to left, top to bottom order increased by 256 on x-axis or y-axis. For example, the first file in the dataset contains 30042021_0_0 elements, where the first number represents the date, while the second and the third values represent the coordinate. The following file, then, should be 30042021_0_256, 30042021, etcetera. After the right bound of the total analyzed area is reached, the first name of the second row will be 30042021_256_0, and so on. Hence, the name is highly important to identify with precision the analyzed sub region. The inconsistency was in the fact that the data contained in the Mask files was not part of the correct file. For clearance, the figure below illustrates the inconsistency, where the left side is the correct scenario, while the right side represents the experienced scenario. The issues have been solved by the creator of the content, and no other issues have been spotted.

Figure 11. Example of an incongruence found in the data,on the left the expected order, on the right the actual order

6.1.2 Software

Data visualization, given all the modern software available on the market, is commonly done on a high level, in so-called interactive data visualization software. On the other hand, such software, sometimes, can be limited, each software can present different functionalities, and the learning curve can be high. Programming custom software to access, manipulate, and visualize data can come in handy. Such software can give the possibility to alter the design process. Such change occurs due to the fact that it is possible to think about the possible design concept unchained from the possibilities that the tools on the market offer. Thus, it can result in a more creative and open ideation.

For the scope of this project, the custom software is written mostly in Java. The main focus of the development phase is put on high flexibility rather than on performance. The custom tool has as a main purpose the manipulation of data meant to optimize the interaction in large dataset visualization done, later, in the tools taken from the market, like Tableau, ArcGIS, and Kepler.gl. This approach, ultimately, gives a low level control over each part of the data. Some of the main features developed in the custom software include:

- Conversion of .txt files into .csv files - most of the interactive visualization tools on the market do not support .txt files. The primary data used in this project is provided in a .txt file. Hence, the conversion needs to be made, along with some additional formatting and labeling for correct functionality inside commercial software.
- Formatting of the columns, coordinates retrieval, and data simulation - data formatting is done in order to make multiple sources of data more uniform and consistent between each other. Furthermore, multiple techniques adopted in the project require for back and forth transformation between pixel coordinates and geographical coordinates. Lastly, due to the incongruent delivery of the final primary dataset containing the change detection maps, as well as to the time limitation that this project includes, a lot of simulation data needs to be simulated before receiving the final dataset. This process ensures that the requirements of the project are reached even in absence of the final information, and the visualizations need to be on data received in the very late stage of the development.
- Data funneling into individual datasets - the primary data covers large areas which are easier to analyze by fractions. Different datasets also present a different structure and extension. Also, procedures like data extraction (will be discussed more in detail later in this section), splitting of change detection maps and clustering through computer vision requires high flexibility in terms of data writing.
- Data simplification for GeoJSON formats
- GeoJSON file generation
- Cluster identification
- Pixel Coordinates to Geographical Coordinates conversion

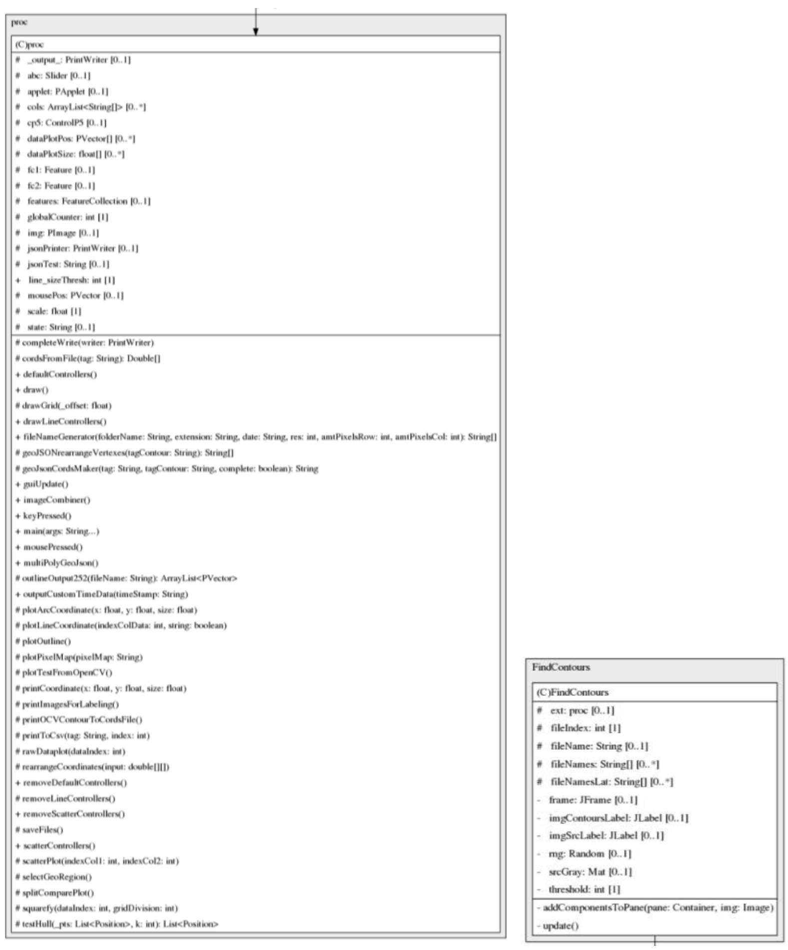

Figure 13. Class diagram of the custom software

6.1.3 Geometry

During the ideation and specification phases, the efficiency issues were inevitably discussed. Such issues, usually, would not arise in a simple, pre-rendered visualization. In a static visualization, the data remains the same, the calculation of all data points and related expressions is performed once, and there is no need to update the pixels in which the data points are being visualized. For this project, however, all the previous factors come into play due to interaction features. Thus, the visualizations won't be static, and will change based on the user's input. Furthermore, the data stories are destined to be published on a web platform and, subsequently, require a certain degree of efficiency and speed of loading.

The high resolution Change Detection maps data present a data point every 3 meters. For a section of the Nicosia city, the CD maps measured in a time lapse of 2 weeks present multiple dozens of millions of data points with classified change, and such number multiplies for areas at a regional or country scale. By implementing the interaction features and time-series animations, and without doing any data optimizations, the web platform would need to load millions of points for every time shift and for every scaling/zoom action. The most intuitive solution to lower the resolution of the dataset while maintaining the main visual features is to group the individual clusters of points into one polygon composed of border points..

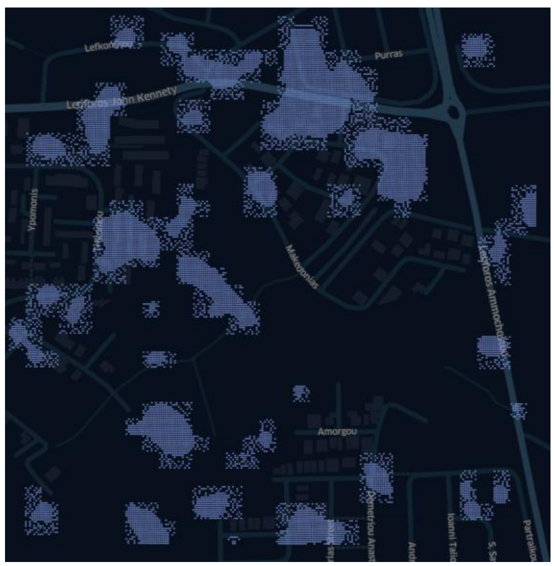

Figure 14. Raw plot of a single patch

The solution adopted for this task consists in applying computer vision to the files in the Output folder. Such a solution is made possible because of the way the data of change detection is generated. As mentioned in the background section, the change map is produced on a pixel level. Each image of the folder Split_A is compared with the corresponding image in Split_B, where each of the *256* x *256* pixels are compared, and the output is written in a pixel mask dataset. Hence, by applying computer vision on the Output folder contents, an image of the same resolution is generated, and the output data is placed in a new pixel mask file. This pixel mask, same as the one used at the early stage of Change Detection map generation, includes all the cartesian pixel coordinates which, subsequently, are used to retrieve geographical coordinates from the Lat_mask_entire folder containing *65537* data points, which is equivalent to 256^2. The data is filled in horizontally, from left to right, from top to bottom. We can use all the previous information to derive the following formula:

`int currentIndex = ((int) currentPos.getY() * res) + (int) currentPos.getX();` where currentIndex = current coordinate in pixel mask file

This calculation uses the cartesian coordinate of the data point in a pixel mask as the index of the array of data points of Lat_mask_entire files. The used computer vision algorithm is provided by the open source OpenCV library:

```
cv2.findContours(cannyOutput, contours, hierarchy, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_NONE);
```

The library's contour detection algorithm offers multiple contour retrieval modes. In order to generate simplified geometry data, RETR_EXTERNAL is the most appropriate. This mode retrieves only the extreme outer contours. It sets `hierarchy[i][2]=hierarchy[i][3]=-1` for all the contours. This means that only the biggest layer in the hierarchy is kept, while other possibly overlapping contours are eliminated. The figure below illustrates the process expressed until now.

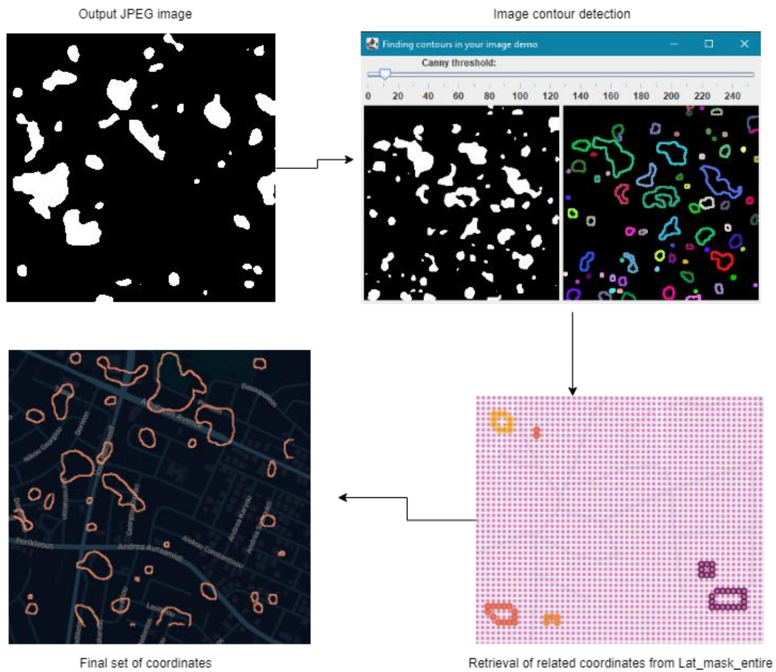

Figure 15. Data simplification process

The following step is to complete the data simplification procedure by generating complete polygons. THe final goal for this is to generate correctly a GeoJSON which can later be plotted in Kepler.gl, ArcGIS, or Tableau. Some new issues arise in this phase, where the process of creating polygons is compromised by a faulty ordering of polygon vertices, or too distant contour points which are not being connected, which results in an open polygon that cannot be filled with color in the visualization tools. This can be seen in the bottom left corner of the figure above, where the vertices are not being filled correctly.

Figure 16. Example of correct (right) and incorrect (left) polygons.

Hence, the goal now is to write a GeoJSON file with correct order and full connectivity. For this task the Concave Hull algorithm is applied. The algorithm takes an array of points and outputs a new array of points which are ordered counterclockwise on the cartesian matrix, where all the necessary points for a full closed polygon are provided. The algorithm uses the K-NEAREST approach for the computation of the region occupied by a set of points. The algorithm and the source are described in the Appendix. Each set of polygons generated by openCV findContours is separated by a delimiter in the pixel mask, which is useful for separating each cluster from the others.

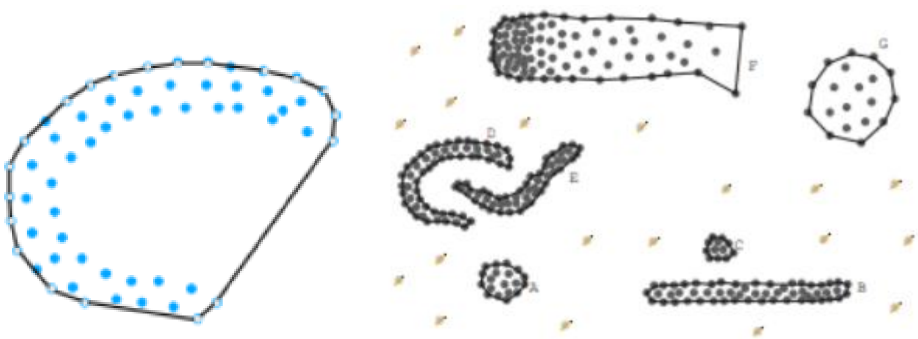

Figure 18. Convex Hull on the left (not useful), Concave Hull on the right (adopted for this project)

Finally, the final result is reached and needs to be written in a GeoJSON format. This format offers many different features which allow for a multitude of different drawing styles, shapes, and feature structures. For the scope of this task, however, only FeatureCollection, Feature, and Polygon will be used. Each feature will contain a single polygon containing a specific type of change as a property, and each feature will be stored in a feature collection. The exported generated file should have the following structure:

```
String polygonFeatureJson = "{
"    "type": "Feature",
"    "properties": {    “OBJECTID”:0;
            “CLUSTER_CHANGE_CLASS”:reforestation
          },
"    "geometry": {
"      "type": "Polygon",
"      "coordinates": [
"        [
"          [
"            -20.7421875,
"            38.8225909761771
"          ],
"          [
"            -22.8515625,
"            -36.03133177633187
```

```
"       ]
"      ]
"    ]
"   }
"  }";
```

The resulting GeoJSON file ends up being 10 to 20 times less memory expensive than the original CD map. Now, the data covering large areas can be visualized efficiently in Kepler.gl and Tableau. It is important to mention, however, that by grouping all the clusters in separate polygons it is assumed that each polygon represents one specific class of change. For example, it is assumed that a cluster of points in the Lat_mask folder cannot contain both reforestation and construction classes. The assumption seems to be reasonable due to the very limited size of each cluster.

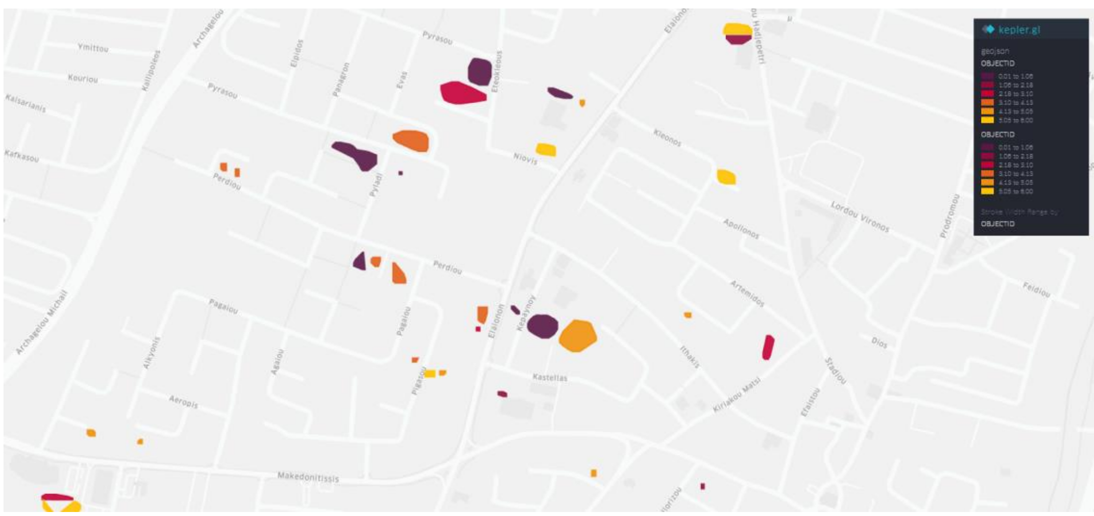

Figure 19. Final geometry plot of a single patch

On the other hand, some of the found secondary datasets related to demographic factors were provided by the organizations in an already prepared geoJSON format. For instance, during the development of the visualizations and further research into supporting datasets for the change detection maps, Corine dataset seemed to fit elegantly with the scope of the visualizations of this project [24]. The datafile is originally offered in 3 different data formats: raster, geodatabase, and SQLite. After downloading the open source data, a plot in ArcGIS is performed. One of the features offered by ArcGIS is to export the featured layers in multiple formats, which include geoJSON. This method of work creates a certain consistency between the different datasets used for the final visualizations.

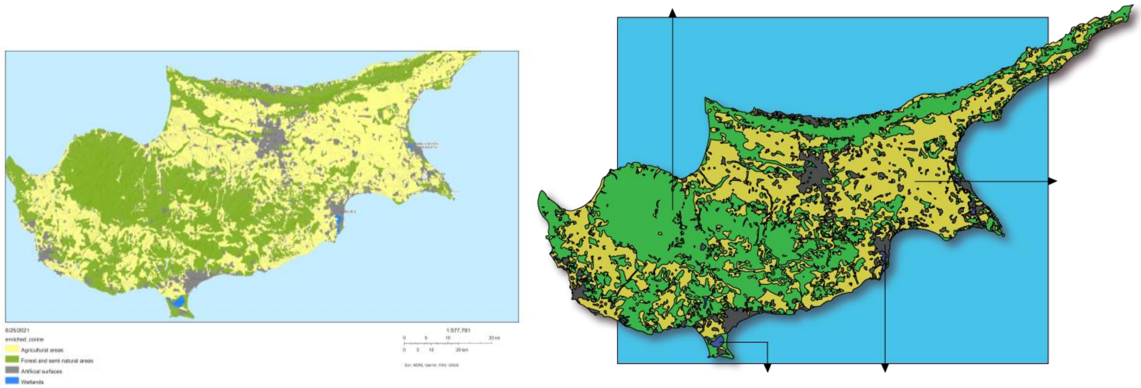

Figure 21. CORINE dataset filtered and plotted in ArcGIS on the left, processed image for the presentations on the right

6.1.4 Data extraction

Depending on the time frame in which the land use change has been analyzed and the geographical scale of the dataset, just plotting the CD map on a geographical map results in very low visual detail. Each cluster of detected change can, sometimes, be not bigger than a single pixel. Hence, it can be hard to visually make sense of the data. A useful solution adopted for the visualization concept of this project is to isolate relatively small areas of interest. This can be useful for, for example, visualizing relationships between land use of multiple cities, or land use differences between rural and urban areas, or cities and coastlines. The data extraction feature offered by ArcGIS is a perfect fit for the scope of this project. The tool allows for manual selection of an area of interest, which will later be used to include all the geographical coordinates situated inside of the selected area. Finally, the result is exported as a new CSV file.

Figure 22. Data extraction

6.1.5 Interpolation

In the case of CD maps, the depth and detail of the map is very granular. The data points of such a dataset, as mentioned above, has a resolution of approximately 3 meters. Furthermore, the nature of such data is very straightforward, which is showing exactly where the change occurred, and due to the fact that such an output is based on satellite imagery, there are no areas that remain uncovered. On the other hand, datasets related to demographic factors are radically different from CD maps. Demographics datasets are different from each other. Every dataset is produced by different entities who make use of different monitoring techniques and systems, different locations, as well as different scales and precision ranges. In general, such datasets are way less detailed than CD maps, and are limited to specific regions or districts. This leads to some areas that are less or not at all covered. The problem, thus, of imprecise comparison between high resolution CD data and demographics data arises. Here, Interpolations come useful for filling such monitoring gaps. The interpolation tool available in ArcGIS allows to predict values at new locations based on measurements found in a collection of points. THe tool takes point data with values at each point and returns areas classified by predicted values. Such tools can be used, for example, to predict electricity usage across a selected region based on measurements taken at individual monitoring stations.

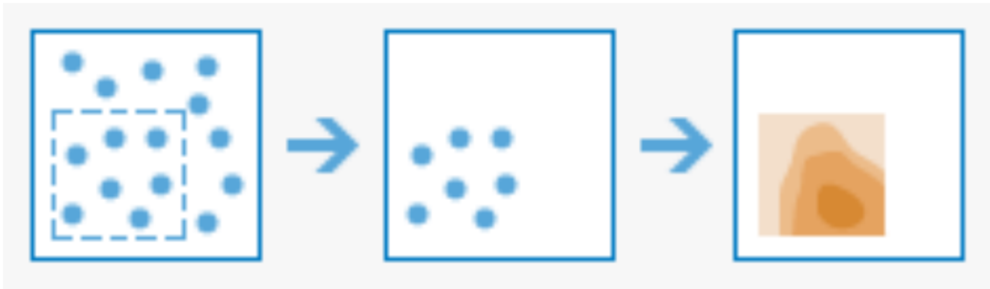

Figure 23. Interpolation illustration

The interpolation tool presents a small fallback, where it approximates large areas with a small amount of data to the few data points present in the area. To eliminate such hazardous predictions the outcome of the interpolation is filtered. Depending on the area, the range of the filter (expressed in square kilometers) can be adjusted, and big areas with little amount of data can be eliminated. In the figures below, the result of interpolation of Nicosia region with and without filter is shown. For this specific case, a large area of 69 square kilometers has been estimated by the interpolation based on less than 10 data points.

Figure 25. Interpolation execution

6.1.6 Hotspot analysis

In a large dataset presenting hundreds of thousands of locations, naked eye cannot manage to see the trends and subtle patterns that are taking place in that specific situation. Thus, similarly to Interpolation, hotspot analysis comes into play to help the data scientist to bring hidden patterns to the surface.

ArcGIS offers such an analysis feature, and it is adopted in the scope of this project. Hotspot analysis is highly question-oriented. This means that the Input Field, which can be raw values or raw incident counts, will determine whether there are lots of incidents, or there high/low values for a particular attribute cluster spatially [27]. The potential applications of hotspot analysis can be found in epidemiology, voting pattern analysis, retail analysis, disease outbreak concentrations, and geospatial analysis.

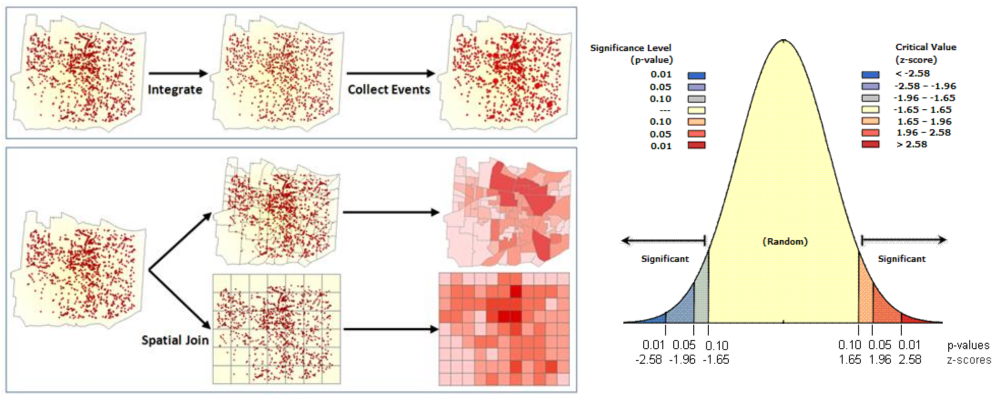

Figure 27. Hotspot analysis on the left, normal distribution with z-scores and p-values

As with most feature tools provided by ArcGIS, hotspot analysis can exclusively work with point and line data. Therefore, before being able to successfully complete the process of hotspot clustering, the simplified geoJSON spatial dataset needs to be converted into a point dataset. After the dataset is compliant, the hotspot analysis can begin. The method adopted within this project is th Getis-Ord GI* method. As such, the program will look at each feature in the available data within the context of neighbouring features in the same datasets [26]. In order for analyzed spots to be considered as significant hotspots, the feature should have a high value and should be surrounded by other features with a high value. The value, in the case of this project, is determined by the amount of locations present in a specific cell range. The analysis is performed in conformity to a generated Fishnet grid, which determines the spatial organization of the analysis and, finally, returns a z-score and a p-value. In the figure above (right), the distribution of p-values and z-scores can be observed. The p-value represents a probability of a certain event occurring. For pattern analysis tools, a small p-value is a statistical confirmation that an event is very unlikely to be the result of random processes. Z-scores, on the other hand, are the amounts of standard deviations, where a very low negative z-score is associated with very small p-value, as well as very high positive ones. Thus, very high or very low z-scores are indicators of low probability of an event being the result of a random process. Hotspot analysis, therefore, is strictly bounded to normal distribution.

6.1.7 Tools

In this section a list of all the used software tools is shown, and the choice explanation is provided. Great part of the used tools is composed of interactive visualization tools, as well as geospatial analysis-dedicated tools. The tools have been chosen based on multiple criteria. The criteria, strength, and weaknesses have been described in chapter 3. The final tool stack used for the realization of the data visualizations include Tableau, Kepler.gl, and ArcGIS Online. For embedding Tableau content, Tableau Public server is used, while Kepler.gl requires access to a dropbox account for storing and sharing map content online.

6.2 Results

This section will illustrate the final outcomes of this project, where the final visualizations will be analyzed and discussed. The data visualizations will, on one side, illustrate the dynamics and the geographical distribution of land use change in space and time, and, on the other side, will show if there are any correlations between certain demographic factors and land use. More specifically, an overall study of the geographical distribution of land use change is illustrated, first. Then, the CORINE information about land cover is overlaid over land use distribution and some further discussion is done. Finally, the exploratory visualization takes over to allow the user to explore by himself the possible researched correlations. The last part is characterized, thus, by a high grade of interactivity and self-gaining knowledge.

All the results shown below are snapshots from the visualizations published on a website, and the codes can be easily embedded into any web domain. The final primary dataset used for the realization of the data visualization include 5 different 2-week comparisons and 2 different scales. The first scale covers an area composed of 204 patches, while the second scale is composed of 2491 patches (table below).

Dataset nr	**Date 1/2**	**Date 2/2**	**Nr of patches**

1	2021-04-30	2021-05-01	204
2	2021-05-29	2021-06-13	204
3	2021-02-15	2021-03-05	2491
4	2021-03-05	2021-03-30	2491
5	2021-03-30	2021-04-13	2491

Table 2. Contents of the final primary dataset

6.2.1 Land use change raw distribution

After receiving the final dataset, first raw plots are made. First, the data is being pre-processed and simplified in order to allow web-based visualization tools to cope with the upload size, as well as any visualization software with loading speed. After the dataset is simplified and optimized, the raw plot is made. Such plots are useful to get an overview of what the studied data presents, and it gives an indication of what are the focal points of the study. For example, after the first raw plot, the first thing that jumps into attention is the difference between rural and urban land use change. The pattern repeats itself for both of the cities included (partially) in the provided dataset.

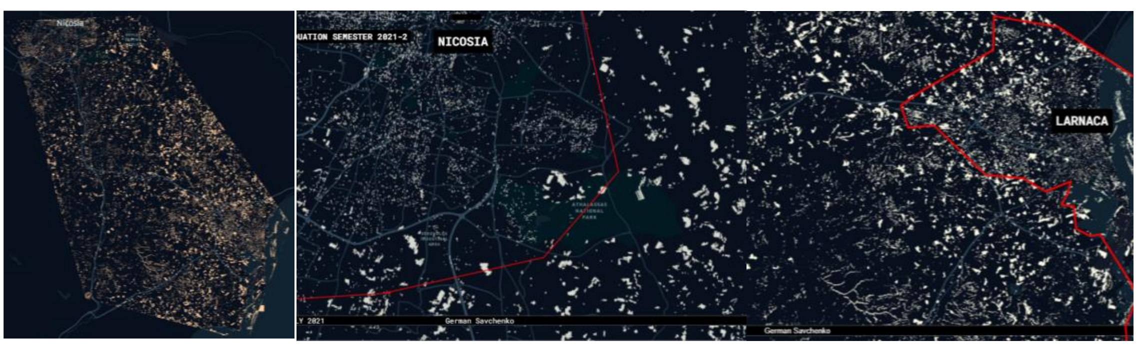

Figure 30. Raw plots, entire analyzed area (left), Nicosia city (middle), Larnaca (right)

More specifically, it can be seen that rural areas present way larger areas of detected land use change, while in urban areas the magnitude of detected change is rather small but more frequent. As shown in the figures above, both Nicosia and Larnaca areas present a granular, and similar to each other change detections. By observing the plot over satellite view, it can be clearly seen that the underlying reason for this is the fact that artificial areas present a higher amount of construction sites in residential areas, for example. On the other hand, the rural areas outside of the cities are mainly constituted by crops and farm lands covering greater areas. This plot represents, also, a first validation of the overall high accuracy and potential of the field. The dataset, undoubtedly, at the moment of the writing of this report still contains a large amount of false positives. However, the plot still manages to maintain the expected coherency, or, in other words, it confirms the fact that rural areas are, indeed, larger than most of the elements in the city areas and thus, present greater land use change areas.

The following step is to perform a raw plot of the dataset filtered by time frame. The final primary dataset provided contains, in total, 5 time frames of which only 3 extend to up to 1500 sq. km, while the remaining 2 are limited to the Nicosia region. The choice, then, of focusing on the first 3 time frames is made for a couple of reasons. Firstly, more relevant evidence comes out when larger areas are analyzed. Secondly, it seems to be impractical to mix time-series data based on different coverage areas, and such incongruencies might be too distracting for the user. After the selection of the time frames is made, the plot is rendered.

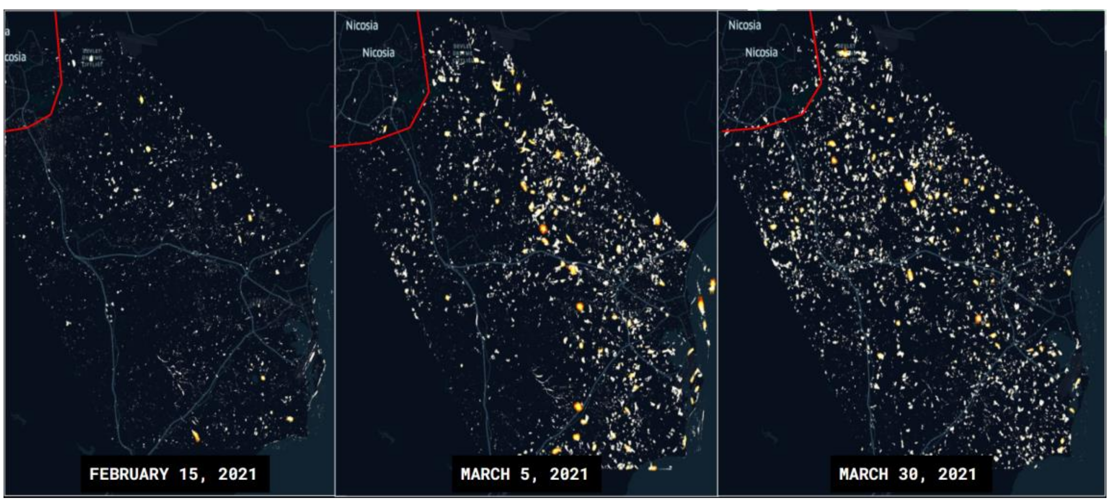

Figure 31. Plot of the 3 time comparisons composed by 2491 patches (datasets 1 and 2 excluded)

In a similar way to the first general first raw plot of the complete dataset, interesting details come to the attention instantaneously. As can be observed in the figure above, the 3 time frames ranging from February 15th to March 30th of the year 2021, a clearly observable change is occurring. Firstly, the change detection map seems to be coherent, and seems to present a specific pattern. More specifically, the change occurring in the second comparison presents a great amount of detected change appearing in the eastern area included between the two cities Nicosia and Larnaca. In the third comparison, it can be seen that the pattern proceeds and the detected land use change extends from east to west through the, roughly, central part of the analyzed region and up to Nicosia. In order to facilitate the overview of the changing land use, the use of Kepler.gl's playback has been adopted. Such feature offers the possibility to apply a dynamic filter on a date variable of the dataset, and, subsequently, to automatically change the date filter in a chronological order. Such a feature is not uncommon between interactive visualization tools on the market. However, Kepler.gl offers a dramatically higher speed of loading of the data. This is due to the fact that the web application is built upon Deck.gl and, as such, makes use of the WebGL high performance capabilities.

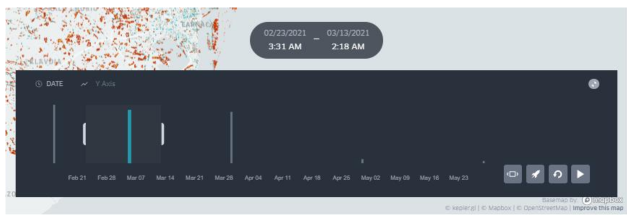

Figure 32. TIme playback in Kepler.gl

After analyzing the raw plots of the distribution of land use change across the studied region, further technical analysis is necessary to confirm the visual evidence. Until now, Kepler.gl has been used for quickly plotting and iterating the data. All the above show figures of this section are the resulting outcomes of the Kepler.gl visualization tool. Such a platform, however, does not offer in depth analysis and statistical study features. As mentioned in Chapter 3, ArcGIS and Tableau are the tools offering high functionality in terms of analysis and visualization as well. More specifically, ArcGIS is highly specialized in geospatial analysis and brings to users very specific geoprocessing tools.

The next steps, thus, are focusing on how to provide specific statistical proof of some of the evidence previously observed. Since the primary dataset containing change detection maps has, as the main and only feature, locations of the detected change, the analysis is based exclusively on geographical locations. Naturally, the analysis focuses on analysis where a higher amount of detected change is occurring. Subsequently, Hotspot analysis offered by ArcGIS came into play. As explained in the section dedicated to Hotspot analysis, the adopted simplified data with geoJSON format was not compliant with the required geoprocessing feature. This is due to the fact that geoJSON is a geometry file, while the required features work exclusively with point and line data. Thus, point data needs to be extracted from the spatial geometry data. The extraction of centroids, another feature offered by ArcGIS, gives the possibility to approximate the distribution of change to a further step. By obtaining the point data of the centers of each single cluster, the hotspot analysis can operate. In the figure below, a snapshot of the resulting visualization is provided. The snapshot is taken directly from the website on which it is published, and, in other words, it is a real time interactive dashboard which allows the user to interactively compare the hotspot analysis performed on data filtered by 3 different time frames. As mentioned before, a red spot represents a high incident, a blue spot represents a low incident, while a white spot represents a statistically not significant incident.

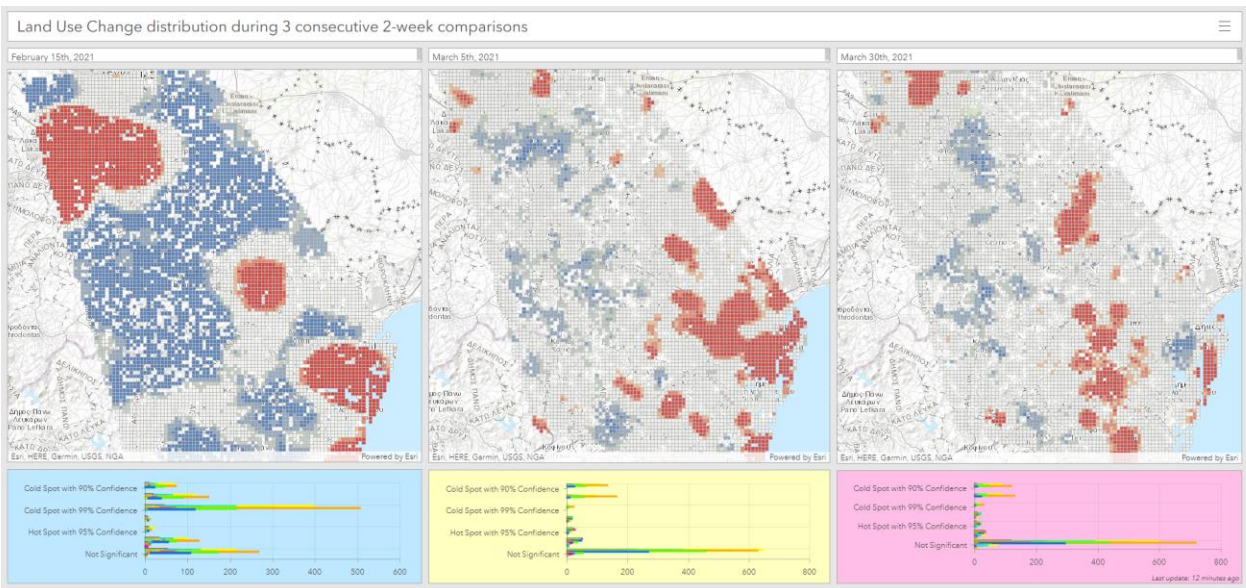

Figure 33. Hotspot analyses and bar charts with z-score distributions (bottom)

The visualizations, respectively, illustrate changes occured in the comparisons dated February 15th, March 5th, and March 30th of the year 2021. Bar charts representing the count of detected change in each of the significance regions is provided for each time frame to summarize visually the distribution of the areas. In the first plot on the left, large areas with low incidents can be observed. This means that the occurrence of change in those areas is rather low or rarified. It can be observed that a drastic transition occurs in the February to March transition. The entire month of March seems, unlike the end of February, to present a significant amount of change across most of the analyzed region. This can be also seen by the supporting bar charts, where the cold spots are dominant only in the month of February, while remaining relatively low in the following time frames. ArcGIS labels areas as not significant whenever the amount of detected change lays halfway in between cold spots and high spots.

These areas are mainly covering agricultural areas or forests, where the change occurs in large quantities but rather distant from each other. Overall, the rate of detected land use change seems to follow the expected patterns, where parts of urban areas are labeled as hotspots.

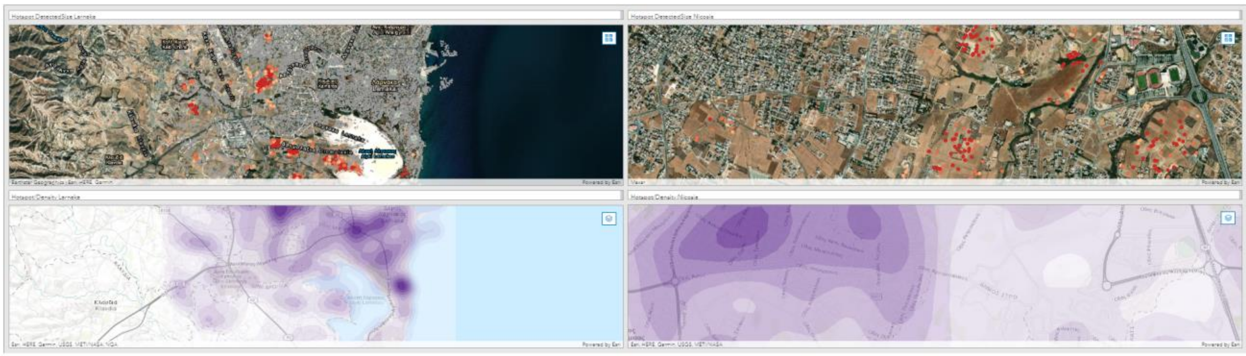

Figure 34. Hotspot analysis change size-based (top), Hotspot analysis location-based (bottom)

Further visualizations were focused on the urban fabrics. The data delimited to the regions of Nicosia and Larnaca have been extracted and exported as new, distinct datasets. After re-loading them into ArcGIS, the same procedures described before in the hotspot analysis have been applied. In this case, however, the visualizations try to explore both of the primary dataset dimensions: location and size of detected change. In the figure above it can be seen that 2 pairs of visualizations are rendered. The left part raffigures Larnaca regions, while the right part raffigures Nicosia. The top side represents the hotspot analysis plots configured to analyze the size of the detected change. Therefore, as it was expected from the observations previously discussed, the cells with higher z-score were concentrated in the agricultural areas around the city, while the city itself was entirely considered as a not significant region and colored in grey. The bottom part is composed of density plots and hotspot plots. The user, in the published visualization, can interactively switch from density to hotspot analysis. In the case of the figure above, the density plot is chosen, and covers the same regions shown in the top part. These analyses were configured for location analysis and, thus, denser areas indicate greeted amounts of detected change. This, once again, has the aim of highlighting the different dynamics of land use change in urban and agricultural areas.

A conclusive visualization for the distribution study is made at the end. The graph represents a hotspot analysis executed on all the time frames together with the aim of showing the longer-term trend of land use change. The hotspot analysis is configured for location analysis. Therefore, it can be observed that in a 6 weeks range, most of che changes occur inside urban areas and extending towards more rural areas along rivers (Nicosia), or roughly along highways (Larnaca).

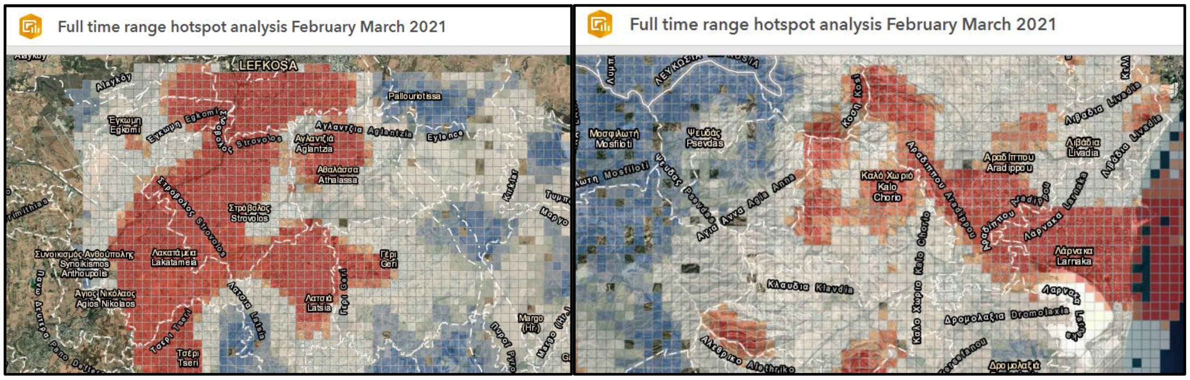

Figure 36. Hotspot analysis of all the time comparisons combined

6.2.2 Land Cover and Land Use

This section will describe the adoption of the Copernicus land dataset in the scope of this project. The idea is to offer to the user the possibility to filter detected change based on the land cover of a specific region. This process can work as an assumption for the classification of the detected change, where it is assumed that, for example, a detected cluster of data laying over a confirmed agricultural area represents a human intervention on a crop. It is also important to specify that the most recent version of the CORINE dataset (used for the following plots) are dated 2019. Therefore, there is an approximately 2 years difference between the primary dataset and the CORINE dataset. This means that the following visualizations cannot, for example, be used to predict the development of the urban fabric, nor to find significant evidence supporting such development. The visualizations are meant for the user's understanding and exploration of the distribution of land use change across verified land cover types.

The following render represents what has been described above. For the purposes of this visualization, Kepler.gl fits the task once again. The same type of visualization, however, is done in both Kepler and Tableau. The reason for this is the fact that Kepler is highly pleasing graphically and reactive to interaction, but does not allow the deployment of a simple control panel which is intuitive to the user.

Figure 37.

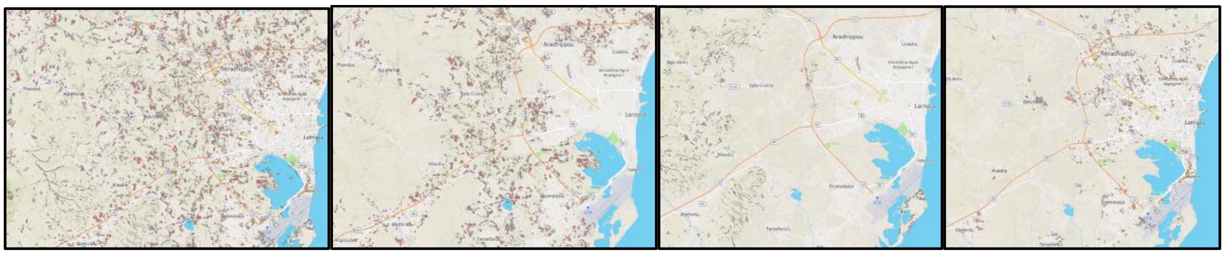

Figure 41. CORINE comparison with land use change, Kepler.gl (top), Tableau (bottom)

The local version of Kepler.gl that allows for developers to have a higher grade of manipulation of the software is the solution to the program, and a control panel can be created through the use of secondary plugins (i.e. Python GUI libraries). Due to the time limitations of this project, however, such a solution is not adopted. The interactive filtering is still possible inside Kepler.gl, the only limitation is that it is not intuitive to the user. The user, in this case, would be required to access the control panel, to find the correct filter to modify, and, finally, to alter the filter and see the results on the screen. For this reason two visualizations are created in two different interactive visualization tools, where Tableau presents the optimal result including an intuitive filter panel but slower performance.

Figure 42. D spatial visualization of the distribution of change based on CORINE and size of change

Until now only the geographical locations of the detected change has been taken into account. There is, however, an additional dimension that can be considered. After the simplification process and the clusters hierarchy buildup, earlier described in this chapter, a new data feature could be taken into account: area of the detected change. Such data can be extracted by modifying the algorithm composing the geoJSON file in the custom JAVA program. The solution consists in adding an extra property to each of the polygons generated, which contains the count of points that that specific polygon is composed of. This gives an approximate measure of the size of the area of each detected cluster of land use change. In the figure above, a 3D plot of a dataset enriched with area values and CORINE classes is shown. The plot is executed in Kepler.gl, which offers the possibility to create elevation of the available geometry. In this case, the color of the geometries has been assigned based on the CORINE land cover class, while the elevation is determined by the area of the detected change. The visualization gives an additional overview of how the land is distributed in terms of land class and magnitude of size across the analyzed regions, and bigger spikes, as expected, are present in agricultural areas between Larnaca and Nicosia.

Given this information, the process of visualizing land use in correlation to demographic factors can be described in the following section. Before starting however

6.2.3 Land Use and Demographic factors

This final section of the results will aim to discuss the findings and relevant evidence observed in the data visualizations. The visualizations done in this phase are exploratory and attempt to meet the requirements of this project. Main focus here is to allow the user to explore the different combinations of demographic factors in relation to the detected land use change autonomously, and to come up with possible conclusions. Interactivity allows for both data points overview, as well as drop down control panels that link chosen demographics datasets to the land use change data. More specifically, the two land use change variables considered here are geospatial locations, size of detected change, and Z-Score. This last variable is the result of the previously executed hotspot analysis. The Z-Score ranges between, roughly, -5.5 and +22, where values below -3 are considered as cold spots with 99% confidence level, above +3 as hotspots with 99% confidence level, and the values between -2 and +2 are considered as statistically insignificant. It is important to specify, however, that areas falling into statistically insignificant range are, in great part, representing land use change detected in agricultural areas, while the cold spots are the areas where the amount of change detected is low. For clarity reasons, a fixed set of demographic variables is taken into account, which includes:

1. Population
2. Expenditures on food and non-alcoholic beverages
3. Expenditures on clothing
4. Expenditures on Electronics
5. Population aged 0-14
6. Population aged 60+
7. Primary education
8. Secondary education
9. University education
10. Purchasing power Total
11. Purchasing power per capita
12. Total households
13. Average household size
14. Unemployment rate

This list of dataset is entirely provided by Michael Bauer Research GmbH in collaboration with ESRI, which provides global coverage of detailed and high quality sociodemographic data for analysis purposes.

Initially, some general plots are made in order to understand the best way to represent the connection between land use and demographic factors. Population and cluster point count, for example, are being chosen initially, and plotted through means of different visualization techniques, like scatter plots or bar charts. After an interesting and relevant visualization is obtained, the following step is to create interactivity. This step will allow the user to interactively go through all of the above mentioned datasets and to compare the differences or the patterns. One example is shown in the figure below, where the z-score and Population variable is taken into account. The two variables are plotted into a scatter plot, where, subsequently, an interesting trend emerges. It can be seen that the population tends to concentrate more extensively in hot spot areas. An increasing population size is driven by an increasing z-score, and the z-score plots shown previously show that hot spots are in large part located around and within urban areas. Thus, an assumption of population pressure being one of the land use change drivers can be established.

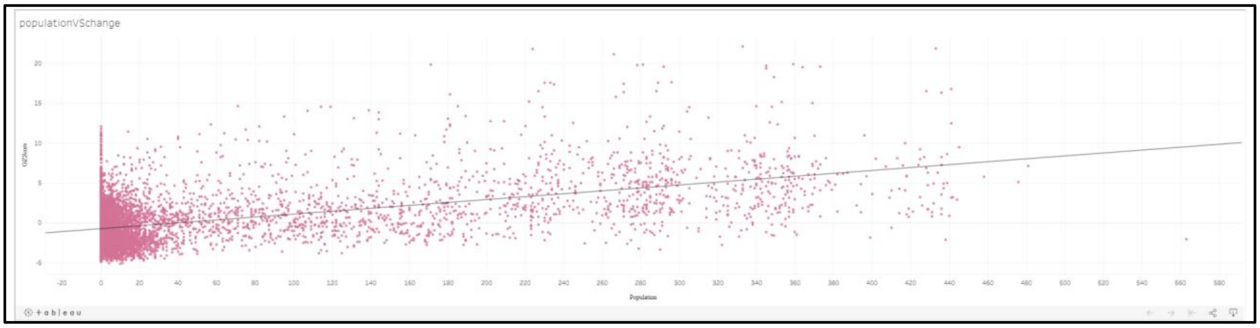

Figure 43. Scatter plot showing relationship between z-score and population

The plot is done in Tableau, which is highly optimized for some of the main plot types, like scatter plot, as well as for interactivity. Now that an interesting pattern is found for a single pair of variables, the interactivity can be implemented and further analysis can take place. For clarity purposes, some data is filtered out inside Tableau. This filtering includes all the spots which are below the 99% confidence level. Such filtering, additionally, creates a visual separation between the three areas, while eliminating just a small amount of data points. The result can be observed in the figure below. Furthermore, similarly to the first scatter plot of population in relation to z-score, trend lines are deployed for guiding the user's perception of the overall distribution of the data.

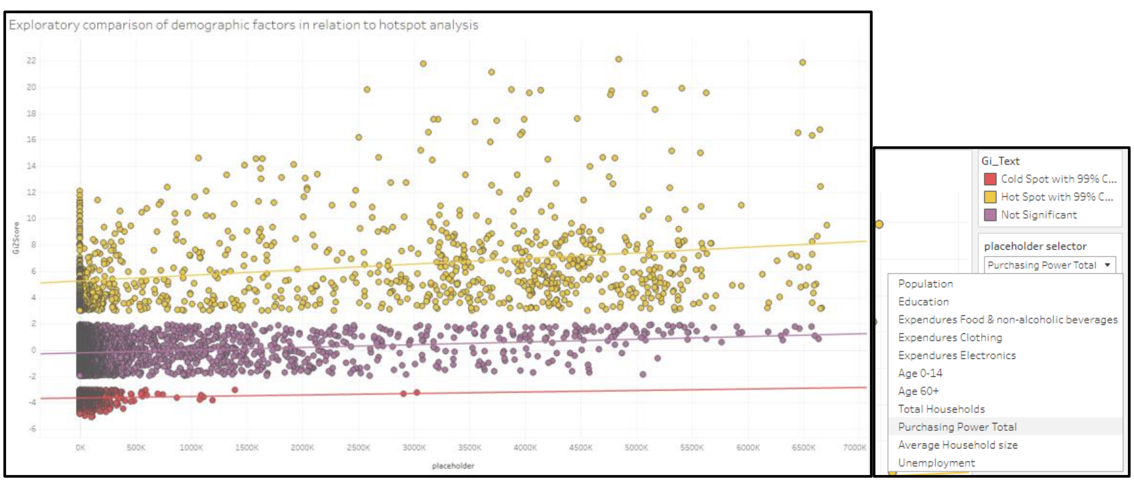

Figure 45. Scatter plot showing the relationship between z-score and multiple demographic factors (purchasing power total currently selected)

The following visualization is a dual comparison of band areas. While comparing multiple demographic factors with the distribution of land use change, no interesting patterns were noted. However, interesting behavior came into play when, instead of using the geographical distribution of land use change, the size of detected change was taken into account, or cluster point count. Furthermore, the land cover subdivision provided by the CORINE dataset is implemented. In this visualization it can be clearly observed, in a similar way to the previous scatter plots, that the demographic factors are very uniform and follow much similar patterns. Additionally, some factors present relatively different proportions when compared to others. One example can be observed in the visualization below. The average household size (bottom), in this case, maintains the same pattern but with a radically different distribution in terms of land cover.

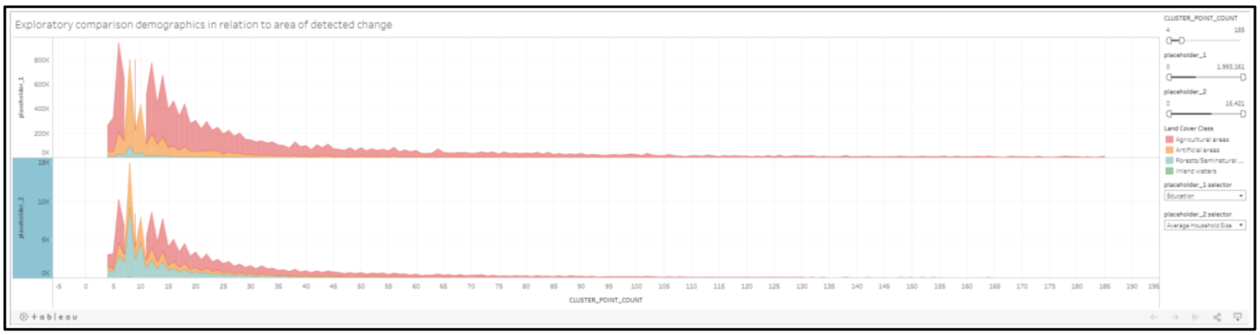

Figure 46. Area band plots showing the relationship between Area of the detected change and multiple demographic factors (education and average household size currently selected)

Some preliminary analysis can lead to multiple hypotheses. For instance, it could be assumed that education is not a driver of land use change in forests and semi-natural areas, or the average household size could be assumed to not be a driver of change, since it's behaviour is equally distributed across the territory.

Next visualization is a linear plot where the two different demographic variables affect the values on the x-axis and the values of the size of the plot elements. Furthermore, the visualization is divided into different z-score ranges once again. The purpose of such visualization is to observe cross relation between multiple demographic variables, observe their conformity to each other, as well as their relation to the overall distribution of change. The main evidence here can be found in the irregularities of size of the elements. Mostly, all of the analyzed variables present a roughly similar x-axis dispersion, which doesn't allow for any significant conclusion other than that most demographic variables taken into consideration are similarly distributed across the analyzed region and in relation to the detected land use change. On the other hand, as it can be observed in the figure below, the combination of unemployment and purchasing power per capita present irregularities unlike most of the other comparisons. This process gives the user clues of where to focus and what to research further. The result, thus, might indicate that there are irregularities in the cross interaction between the two demographic factors, as well as irregularities in terms of correlation with detected land use change.

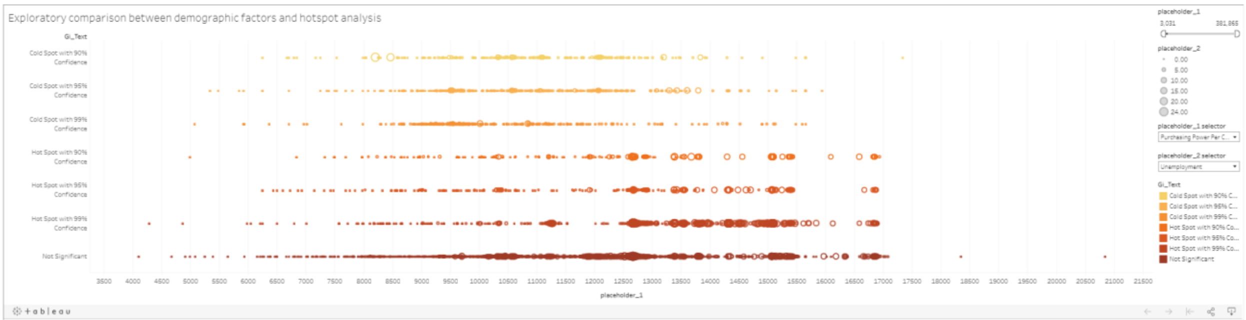

Figure 47. Linear plot showing the relationship between the z-score, demographic factor of choice 1, and demographic factor of choice 2 (purchasing power per capita and unemployment rate currently selected)

Finally, an example of visualization based on standard deviation is produced. The following figure represents a triple comparison between 3 datasets related to Education, where the trend line is deployed in order to guide the user's perception towards the subtle trend of the dataset, and the color of the elements is determined by the calculation of the standard deviation. This allows for further discussion over how the data tends to distribute, and raffigures elements in white when the values are considered to be included in the range of statistically normal range, and blue when elements are considered to be outliers. In the visualization below it can be seen, for example, that secondary education is the most influential in terms of correlation with land use change. The trend line reveals that the tendency is to have higher amounts of population with a secondary degree inside hotspots of change, while the peaks of all of the education levels are located in regions edging hot spots and statistically not significant spots, which given the previous observations are represented by agricultural areas (statistically not significant) and urban areas (hot spots).

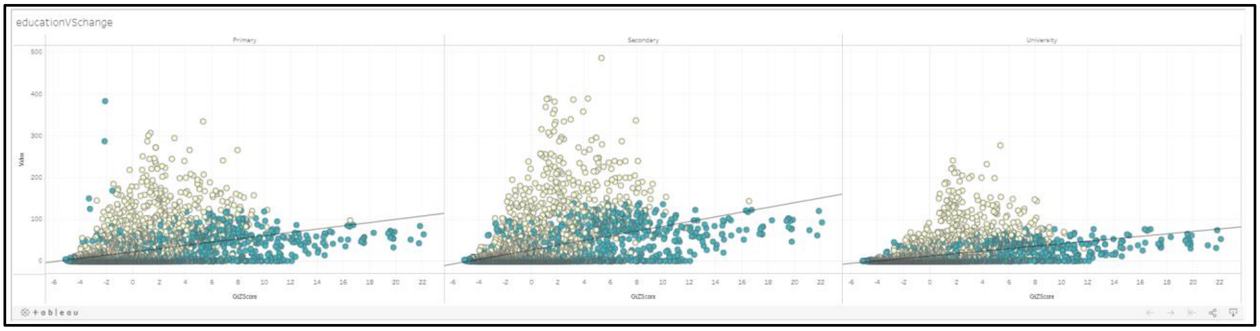

Figure 48. Scatter plots showing the relationship between education levels and z-score

Chapter 7 - Evaluation

After the realization phase of the project, the content is loaded on a web page with the aim of presenting it to the public and to the policy makers. The evaluation phase should focus, then, on mainly two factors: understandability and performance. The first should aim to gather information from testing users about the ease of learning and understanding the content of the data visualization, while the second should aim to gain an overview of how the content loads on a web page. Due to time constraints and major delays in the delivery of the final primary data, the scope of this section will be narrowed down to performance testing, evaluation, and optimization.

7.1 Performance

The evaluation of performance consists of verifying whether the visualizations and the data are optimized to the right point. This phase is also, therefore, useful for testing whether the simplification algorithms applied to the primary data used in this project work as expected and allow for the most possible fluency in user experience while maintaining the maximum amount of original detail. First, the testing method is researched and defined. Here, possible tools and methods are researched online and possible options are brought into consideration. Additionally, metrics and variables are described, which will allow for precise systematic measurement of efficiency. Next, the testing is performed and the results are discussed. In this step design decisions are reviewed and, if necessary, an optimization plan is described. Lastly, the optimization is performed and the results are discussed once more. The results should be able to reach an acceptable level of usage, and eventual tradeoffs are listed. Finally, in the last section of this chapter, the recap of the evaluation discussion is presented.

7.1.1 Testing

Performance testing is a broad field and can cover a large amount of different factors, as well as focus on different areas of software and web elements. Web testing, specifically, can cover a long list of factors before being considered as complete. Some of the elements that are commonly included in website testing include mobile-friendly testing, functional testing, usability testing, security testing, UI testing, performance testing, and compatibility testing [30]. It is important to consider, however, that the main focus of this project is to create data visualizations for policy makers and the general public. Thus, the website publication can be considered as of secondary importance, and the testing process narrowed down. The simplification of the web testing process is also determined by the fact that all of the visualizations produced by this project are embedded through the means of the web applications that produced them (i.e. Tableau Public, Kepler.GL, ArcGIS Online). In other words, this means that the content produced with the chosen software has been tested and optimized before the release and through the entire period of the software's market presence. Given this information, UI testing, compatibility testing, security testing, mobile-friendly testing and functional testing are, in theory, irrelevant in the scope of this project. Usability testing is, also, partially tested by the respective companies, while the custom content of the data visualizations produced by this project are the remaining elements requiring testing. As mentioned above, however, due to time limitations this phase will be skipped and the main focus of the testing is on performance.

Performance testing includes stress testing, load testing, stability testing, volume testing, concurrency testing, page load speed testing, and endurance testing. Some of the tools for such testing include Apache, JMeter, HP, and LoadRunner. Here some elements can also be considered as obsolete in the scope of this project. For example, it is irrelevant to perform stress testing because the aim of this project is to create content and to publish it online, and it is also irrelevant to perform stability and load testing due to the fact that the resulting content of this project is published on a third party server. A simpler tool, however, is chosen for the scope of this project. PageSpeed Insights (PSI) by Google reports on performance of a page on both mobile and desktop devices, and provides suggestions on how a page can be improved. PSI fits greatly the requirements of this testing and, as such, can be considered as the testing environment of this evaluation phase.

The page performance is summarized by a score provided by PSI. A score of 90 or above is considered to be good, 50 to 90 indicates that improvements need to be made, and a score below 50 is considered to be poor. The score is provided through use of Lighthouse, which calculates the score on a basis of a set of metrics. These include First Contentful Paint (FCP), First Meaningful Pain (FMP), Speed Index (SI), First CPU Idle , Time to Interactive, Max Potential First Input Delay (FID), Total Blocking Time, Largest Contentful Paint (LCP).

FCP marks the time at which the first text or image is painted, SI shows how quickly the contents of a page are visibly populated, Time to Interactive is the amount of time it takes for the page to become fully interactive, Total Blocking Time is a sum of all time periods between FCP and Time to Interactive, LCP marks the time at which the largest text or image is painted, and CLS measures the movement of visible elements within the viewport. Given the information above, the test can be executed and the results discussed

7.1.2 Evaluation

After running the test, a score of 43 is obtained. Such a score is a clear indication that improvements should be made in order to improve the user experience. The test has been done on a single webpage containing approximately 15 visualizations. These visualizations are embedded in the main 3 interactive visualization tools composing the tool stack of this project: Kepler.gl, Tableau, and ArcGIS Online. ArcGIS, specifically, resulted to be the most memory intensive. As it can be seen from the snapshot taken from the testing environment (figure below), the main drivers of the low score are Time to Interactive, SI, and Total Blocking Time. High Time to Interactive time is the first indication that the elements are not loading fast enough, while SI suggests that there are too many elements cross interfering with each other, which compromises each individual loading speed. By further expanding the tree map it is also possible to see the statistic regarding the used memory. This feature also provides the possibility to see how much memory is not being used and, therefore, slowing down the overall Time to Interactive. The diagnostics suggest, additionally, to reduce unused JavaScript, eliminate render-blocking resources, limit third-party listeners, lazy load third-party resources with facades, and reduce unused CSS.

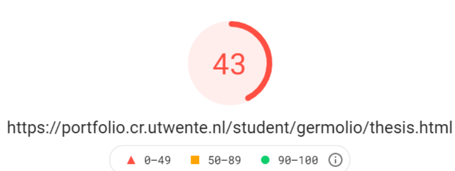

Figure 53. Webpage score before optimization (top left), webpage score after optimization (top right), memory usage (bottom)

7.1.3 Optimization

Given the information in the previous section, the optimization process is quite straightforward. Content of the page needs to be limited and the webpage code needs to be cleaned. The first solution adopted is to split the content of a webpage. The 15 data visualizations need to be separated in 3 or more different tabs. This not only diminishes the load over one single page, but is also relevant for the final data stories. In the realization chapter the results have been discussed in 3 different subchapters: raw distribution, correlation with land cover, and correlation with demographics. Therefore, the graphs are sorted and placed in new HTML files. Next step is to reduce unused JavaScript. During the diagnostic, unused JavaScript resulted to be one of the most delaying factors in loading speed. A general check of the code brought to light some badly written code, unused libraries, and unnecessary animations. The score at this point reached a value of 51, which is still surprisingly low. As mentioned in the list of diagnostics in the previous section, another fix was necessary. PSI suggested lazy load third-party resources with facades in order to avoid unnecessary loading of elements that are not visible to the user. Subsequently, lazy loading has been set for each data visualization. After running the performance testing tool again, the score reached up to the value of 77. The score is still not optimal and not in the 90-100 range.

However, by looking at the metrics in figure 53, Time to Interactive drastically improved by going from 14.1s to 2.9s. It can be considered to be in the generally accepted delay range, whereas delays above 3 seconds are considered the threshold representing the moment in which 40% of the internet users leave the page. Speed index has also improved from 8s to 1.7s. On average, the loading speed became 4 to 5 times faster than before. Considerable fractions of memory were still labeled as unused. This is due to the fact that the embedded visualizations do not show all of the features on loading moment. When interaction between the user and the data visualization occurs, stored background information about data becomes available. For example, if the cursor is placed above a data point in a visualization made in Tableau, information about that data point pops up, or if the user decides to switch the dataset used in the dual comparisons in the demographics section, the new datasets are being used only from that specific moment. The source of the 'waste' is in the visualizations themselves, and highly dependent on the embedding settings provided by the respective visualization tools. By following the embedding approach adopted by this project, it is not really possible, nor necessary, to improve further on this. Slight improvements can be made by deleting unnecessary legends, control panels, or menus.

7.2 Evaluation conclusion

The performance testing process has shown that there are some flaws in how the final results of this project are organized, and that the content was not ready yet to be presented to the target audience. After performing optimizations and following the suggestions provided by the diagnostics of PSI, an acceptable performance score has been reached without any major trade offs.

Chapter 8 - Conclusion

The finalization of the project consists in reviewing the key findings and bringing conclusive final evidence, answers to research questions and giving recommendations for future work. The four sub-research questions are answered, and the answer to the main research question is derived.

What are the best ways to visualize land use change in space and time?

In the realization phase, a number of visualizations have been rendered. Before each visualization was considered as complete, lots of tests and experimenting with plots took place. Such processes sorted out, naturally, the visualizations that did not fit with the context of this project or that did not communicate the message properly. Therefore, the list of final visualizations can be considered as the list of best ways to visualize land use change in space in time. The final results included scatterplots, area bands, linear plots, density maps, hotspot maps, geographical plots in 2D and 3D. More specifically, each visualization has been useful for a specific task. For example, scatter plots and area charts have been useful to visualize interaction between land use change and demographic factors, while geographical plots were optimal for visualizing the distribution of land use change in space and time, as well as analyzing the land use change in comparison with land cover data.

How time-series aerial imagery is useful for detecting patterns in land use change?

This project can be considered as a confirmation of the potential of the aerial imagery combined with DNNs detecting land use change. The results have shown that even at this preliminary stage of land use change research in combination with neural networks, the possibility to find useful information for large scale land use planning is great. The data resources made available for the realization of this project were relatively limited. However, with greater amounts of larger scale and longer time period comparisons data it can be definitely possible to find insightful patterns and dynamics in land use evolution.

To what extent are demographic changes influence land use change?

It is unclear, from the results, if there are demographic factors that significantly influence land use change. It was also uncertain whether demographic factors are influencing land use change or are being influenced by it. Most visualizations showed a uniform distribution of demographic factors in general, which made it difficult to spot correlations. The data visualizations showed that populations tend to concentrate into hotspot areas, or where the change occurs most frequently. Furthermore, another evidence has shown that land use change has a high magnitude of transformation in the transition between winter and spring. Overall, there's a lack of larger scale datasets and classification data of land use change. Due to this, no concrete assumptions were made.

To what extent demographic factors are relevant for the analysis of land use change?

As mentioned above, no clear correlations between land use change and demographic factors have been found. By observing the resulting visualizations, however, it seems that most significant findings were discovered by the simple observation of the distribution of change. For example, the high granularity of detected change of cities has been spotted, while the change in rural areas clearly presented clusters of greater size. In some visualizations it was possible to compare how the change distribution transforms in time, where rural areas seemed to follow a subtle pattern of change, while cities seemed to indicate the direction of development (In Larnaka case, large amount of detected change seemed to concentrate in the direction of Aradippou village). Larger scale dataset with multiple years of 2-weeks comparisons could, most likely, reveal a greater amount of evidence and trends in human intervention on land. At this point of progress, it seems that simple land use change distribution evolution in time is more useful in studying land use change than searching for driving demographic factors. Land use classification data and higher amounts of data could also, however, reveal the opposite. Thus, no significant assumptions are made here.

What are the most frequently occurring correlations between land use change and demographic factors?

Some information mentioned in chapter 2 seems to match the final findings of this project. Population, for example, can be considered as one of the demographic variables that changes in relation with land use change. The fact that highly populated areas are characterized by a more granular type of land use is one of the pieces of evidence that appeared in the visualizations. Purchasing power per capita, Level of education, expenditures and age seem to follow roughly the same patterns as the population when compared to land use change distribution. The size of detected change also does not appear to be affected by demographic factors. The only evidence regarding the magnitude of detected change is the fact that there is an inverse relationship with population size: a greater area of change, usually, equals to a lower amount of population when compared to the population size and size of detected change in the cities. This mainly shows that most of the change occurs in agricultural areas, where the population is limited, while all the other demographic variables remain roughly the same. Thus, population size seems to be the only demographic factor to be significantly correlated with land use change in the scope of the analyzed data.

Chapter 9 - Future Work

9.1 Large scale data

One of the biggest downfalls of this project was the lack of a higher amount of data. The provided primary dataset covered a total of 1500 sq. kilometers, that, in the end, resulted in good results but still was limiting. The analyzed data included a fraction of the city of Nicosia, Larnaka, and the area in between the two. It's been useful to understand how land use change tends to develop inside urban fabrics and rural areas. It is clear, however, that a dataset covering the entirety of Cyprus could have delivered much more different results overall. A country-scale data is definitely a key for finding big change patterns in human intervention over ecosystems. Such information availability could be useful in multiple ways. For example, a dataset that covers tens of different relatively large cities might confirm the patterns in urban fabric evolution, or it might reveal different types of patterns depending on the city's population or topological location. Additionally, country-scale data could give the possibility to create an evaluation map of the current extent of ecosystem degradation, and to carefully counter-act such degradation by exploiting less areas with high levels of human intervention and by distributing such intervention more equally across the territory. This could also be planned in terms of routes and highways to improve the overall country infrastructure usage.

9.2 Longer time frames and classification

In combination with a larger scale, the research on land use change would greatly benefit from having larger amounts of time comparisons. Time-series, specifically, are crucial for determining the trends and are the only way to better understand the involvement of demographic factors in land use change. By gaining an overview of the evolution of land use change it would be possible to create statistics over what is the extent of land degradation, is it deteriorating further, and which are the areas where counter action is highly required. A good example of the potential of longer time series and larger scale dataset is data acquired before, during and after the Covid pandemic waves. Such a time-series would provide a clear overview of how land use change operates during normal times, and how it is being affected after quarantine measures are being implemented and reduce human activity in general. Also, the data would show how the change is re-igniting after the countermeasures are being lifted. Additionally to time-series, classification of the detected change is also key to deeper understanding of land use change. Land use classification is the process of attributing a type class to each of the detected change clusters, which may include, for example, activities like construction sites, crops, deforestation, reforestation, and many others. This classification would add an extra dimension to the data which could greatly improve the study of demographic correlations with land use change. For instance, demographic factors could be attributed to the different distributions of a specific change type, change types could be studied more on a geographical level and systematic weights could be attributed to each of the land use classes. These weights could indicate the level of influence on an ecosystem, its average duration, or the level of resource dependency. Subsequently, a lot more factors could contribute to finding evidence and patterns in both land use change and demographic factors, as well as their cross interaction.

9.3 Land use change specialized software

An additional suggestion for improving the study is to develop a highly specialized software for exclusively land use change analysis. The software should be able to load large amounts of geospatial data, to simplify the data by generating geometry or centroids, and to efficiently visualize the data in space and time. Such a tool should provide the user with a set of geospatial analysis tools for pattern and distribution studies. The performance of the tool could be optimized by making use of GLSL or WebGL APIs. GLSL or OpenGL could be taken advantage of to allocate most of the calculations to the graphics card. After the release of this software as an opensource tool, it could act as an open ground for average users and scientists to develop the field of land use change analysis.

References

[1] G. K. Heilig, “Anthropogenic Factors in Land-Use Change in China,” Population and Development Review, vol. 23, no. 1, p. 139, Mar. 1997.

[2] E. Filipe and S. Norfolk, *Understanding changing land issues for the rural poor in Mozambique*. London: International Institute for Environment and Development (IIED), 2017.

[3] U. Horvat and I. Žiberna, “The correlation between demographic development and land-use changes in Slovenia,” AGS, vol. 60, no. 2, pp. 33–55, 2020.

[4] Vesterby M, Heimlich RE. Land use and demographic change: results from fast-growth counties. Land Econ. 1991.

[5] Hoffmann, “Demographic Change and Land Use,” in Human-Environment Interactions, Springer International Publishing, pp. 63–74, 2020.

[6] Kolb, M., Galicia, L. Scenarios and story lines: drivers of land use change in southern Mexico. *Environ Dev Sustain* 20, 681–702, 2018.

[7] Tolessa, T., Dechassa, C., Simane, B. *et al.* Land use/land cover dynamics in response to various driving forces in Didessa sub-basin, Ethiopia. *GeoJournal* 85, 747–760, 2020.

[8] V. H. Dale, “THE RELATIONSHIP BETWEEN LAND-USE CHANGE AND CLIMATE CHANGE,” Ecological Applications, vol. 7, no. 3, pp. 753–769, 1997.

[9] D. Lu Correspond, P. Mausel, E. Brondizio, and E. Moran, “Change detection techniques,” International Journal of Remote Sensing, vol. 25, no. 12, pp. 2365–2401, 2004.

[10] Y. Shi, R. Wang, L. Fan, J. Li, and D. Yang, “Analysis on Land-use Change and Its Demographic Factors in the Original-stream Watershed of Tarim River Based on GIS and Statistic,” Procedia Environmental Sciences, vol. 2, pp. 175–184, 2010.

[11] C. P. Romero, A. García-Arias, C. Dondeynaz, and F. Francés, “Assessing Anthropogenic Dynamics in Megacities from the Characterization of Land Use/Land Cover Changes: The Bogotá Study Case,” Sustainability, vol. 12, no. 9, p. 3884, 2020.

[12]T. van Willigen, “Measuring the user experience of data visualization”, Master thesis, Dept. EEMCS, University of Twente, Enschede, NL, 2018.

[13] V. Rautenbach and S. Coetzee, “MapFactory – Towards a mapping design pattern for big geospatial data,” Proc. Int. Cartogr. Assoc., vol. 1, pp. 1–6, May 2018

[14] S. Schöttler et al., “Visualizing and Interacting with Geospatial Networks: A Survey and Design Space,” Computer Graphics Forum, Apr. 2021

[15] T. Tahi Hamonangan Tambunan, “Identifying Stakeholders in Land Use Management Process and Its Critical Factors in ASEAN,” JIM, vol. 4, no. 5, p. 237, 2015[Online]

[16] T.Krawchenko, 2017. “The governance of land use” OECD, p.3-26

[17] Influencers of Land Users’ Perceptions of Land Resources: The Case of Mphosong-Bolahla Catchment, Lesotho

[18] Climate ADAPT (2016 ,June 07). Awareness campaigns for behavioural change [Online]. Available:https://climate-adapt.eea.europa.eu/metadata/adaptation-options/awareness-campaigns-for-behavioural-change

[19] B. H. Sibolla et al., “TOWARDS THE DEVELOPMENT OF A TAXONOMY FOR VISUALISATION OF STREAMED GEOSPATIAL DATA,” ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., vol. III–2, pp. 129–136, Jun. 2016[Online]. Availablehttp://dx.doi.org/10.5194/isprsannals-III-2-129-2016.

[20] Mader, A. H., & Eggink, W. (2014). A Design Process for Creative Technology. In E. Bohemia, A. Eger, W. Eggink, A. Kovacevic, B. Parkinson, & W. Wits (Eds.), Proceedings of the 16th International conference on Engineering and Product Design, E&PDE 2014 (pp. 568-573).

[21] bized.aacsb.edu. (2019, Dec. 29). The ethics of data visualization [Online] Available: https://bized.aacsb.edu/articles/2020/january/the-ethics-of-data-visualization

[23] CORINE Land Cover nomenclature conversion to Land Cover Classification system (2010, Dec. 10) [Online] Available: https://land.copernicus.eu/eagle/files/eagle-related-projects/pt_clc-conversion-to-fao-lccs3_dec2010.

[25] M.-J. Kraak, “Exploratory Visualization,” in Encyclopedia of GIS, Springer US, 2008, pp. 301–307 [Online]. Available http://dx.doi.org/10.1007/978-0-387-35973-1_397.

[26] Geospatiality. (2016, Jan 21). “What is hotspot analysis?” [Online] Available: https://glenbambrick.com/2016/01/21/what-is-hotspot-analysis/

[27] ESRI. “How Hot Spot Analysis (Getis-Ord Gi*) works” [Online] Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm

[28] ESRI. “What is a z-score and what is a p-value” [Online] Available: https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/what-is-a-z-score-what-is-a-p-value.htm

[30] EasyQA. “How to test a website” [Online] Available: https://geteasyqa.com/qa/test-website/

[31] D. Domingo et al., "Effect of zoning plans on urban land-use change: A multi-scenario simulation for supporting sustainable urban growth," Sustainable Cities and Society, vol. 69, p. 102833, Jun. 2021[Online].