MASTER THESIS (30 ECTS)

University of Utrecht, Faculty of Geosciences

Sustainable Development (MSc), Track: Energy & Materials

A proof of concept for mapping environmental change in areas of resource extraction at the source of energy transition supply chains.

Using open access satellite data and cloud computing to create remote sensing time series.

Student: 5371058, Simon Teichtmann

s.c.teichtmann@students.uu.nl

Submission: 19/07/2022

Word count: 16,175

Supervisor: Britta Ricker (PhD) and Emilinah Namaganda (PhD candidate)

2nd reader: Dr. ir. C.G.M. Kees Klein Goldewijk

Abstract

The availability of cloud computing and open access satellite data provides new opportunities for the application of earth observation (EO) based environmental monitoring. Mapping the environmental change resulting from resource extraction at the source of energy transition supply chains represents a particularly interesting application for these services. The energy transition fosters a high resource demand, and the extraction of resources goes hand in hand with the appropriation and exploitation of land, often conducted with little or no environmental concerns. Using Google Earth Engine (GEE), Landsat and Sentinel imagery, I developed an example of a workflow model to create remote sensing time series to measure land cover and vegetation change. The model is based on scripts using the programming language JavaScript. To test the workflow, it was applied to two case studies, mapping change resulting from graphite and natural gas extraction in Cabo Delgado, Mozambique. The case studies were investigated in time series between 2005 and 2021. The analyses shows that extraction related facilities and infrastructure replace, to a large areal extent, natural areas and cause a significant increase in unvegetated areas. Thereby, the workflow model yielded high classification accuracies of > 90 %. Testing the model in both case studies proved it reproducible and scalable under the requisite that specific parameters, such as the region of interest and time series intervals are provided as inputs to the scripts. The use of open access satellite imagery and the GEE platform make the model applicable in circumstances of low financial and technical means, as data is free and no costly hard- and software is required. Therefore, the proposed workflow can be particularly beneficial to monitor change in areas of resource extraction which commonly occur in vulnerable, low-income regions, where Environmental Impact Assessments (EIAs) and their follow ups are often weak or being neglected. Hence, my workflow model provides a reliable solution to map change at the source of energy transition supply chains in an effort for a more sustainable energy transition. By identifying and quantifying the change, private and non-profit decision makers can develop and enhance plans to preserve, manage and restore adjacent land and nature and protect local communities.

Key concepts: remote sensing, energy transition, time series, open access, resource extraction.

Table of contents

Table of Figures

Table of Tables

Abbreviations

BGP…….……Balama graphite project

CART……Classification and Regression Trees

DRC…….Democratic Republic of Congo

EIA…….……Environmental Impact Assessment

EO……...…..Earth Observation

GEE…….…….Google Earth Engine

GIS…….……Geoinformationsystem

HSV…….…….……..Hue, Saturation, and Value

IEA……...International Energy Agency

LNGP…….…….…….Mozambique LNG project

MNDWI……...….Modified Normalised Difference Water Index

NDBI…….…Normalised Difference Built-up Index

NDVI……...…….Normalised Difference Vegetation Index

NIR…….……..Near-infrared

PAC……...…..Project-Affected Community

RF……..…Random Forest

SVM…….……..Support Vector Machine

TOA…….……..…..Top of atmosphere

USGS……..……United States Geological Survey

1. Introduction

The energy sector is the largest source of anthropogenic greenhouse gas emissions (IEA, 2021a). Anthropogenic greenhouse gas emissions and namely carbon, are the primary cause of global warming (Lamb et al., 2021). To mitigate global warming, a transition from a fossil-based energy supply towards a carbon free energy supply is crucial. This calls for an extensive decarbonisation of the energy sector, a transformation widely known as the energy transition. The energy transition triggers an increasing demand for low carbon fuels such as natural gas and hydrogen and low carbon technologies like storage systems and solar panels (IEA, 2021c). To satisfy this demand and manufacture the required technologies, a significant amount of resources is needed. Due to studies from the International Energy Agency (IEA) (2021b) the energy transition could cause a fourfold increase in mineral demand by 2040. Over 20 materials are affected by this trend, and most significant are the growth rates for lithium and graphite. The demand for lithium is expected to grow by over 40 times and graphite by 25 times between 2020 and 2040. Hund et al. (2020) expect an increase in lithium and graphite demand of 500 per cent by 2050 compared to 2018. Lithium and graphite are just two of the most common materials for lithium-ion batteries, which require many more minerals such as cobalt and nickel (Gallo et al., 2016). Lithium-ion batteries are a major storage technology for the energy transition and make up a substantial share of the mineral demand. Moreover, natural gas is a key bridging fuel towards a net zero energy sector. It is used to substitute the more emission intensive energy production from coal or oil (Stephenson et al., 2012). Even though, global demand projections for natural gas are not as severe as for graphite and lithium, an annual growth rate of 1.5% from 2019 to 2025 is projected (IEA, 2020). Current demand projections will have to be updated and are likely to be surpassed in light of the Russian invasion of Ukraine in February 2022. As a result of comprehensive sanctions against gas from Russia, the global gas production and infrastructure are currently altering and expanding to an unexpected extent, resulting in the need for new gas suppliers and new extraction projects (Höhne et al., 2022). However, to decarbonise the energy sector and mitigate global warming the increasing demand for low carbon fuels and technologies needs to be satisfied.

To provide the required resources for the energy transition, global extraction rates are expected to increase rapidly (Gielen, 2021). The anticipated pace, scale, and intensity of the energy transition-induced resource extraction pose significant risks. Resource extraction goes hand in hand with the appropriation and exploitation of land, causes extensive land use change and is often the source of environmental degradation and negative social implications such as the displacement of people (Lèbre et al., 2020). This introduces a contradiction in the global efforts to mitigate and adapt to climate change, achieve a sustainable energy transition, and raises justice concerns in the supply chains of energy systems (Huber & Steininger, 2022). Furthermore, a high share of the reserves of the energy transition materials is located in vulnerable states with high measures of fragility and corruption, where governments struggle to safeguard against the negative consequences, and financial benefits of the resource extraction remain largely with elites and foreign investors (Church & Crawford, 2018). For example, the Democratic Republic of Congo (DRC) is home to large cobalt reserves, Brazil and Mozambique for graphite, and Zimbabwe for lithium. For decades already, cobalt extraction in the DRC has been conducted with little or no environmental concerns, "…leaving wastelands of disused mines" (Banza Lubaba Nkulu et al., 2018, p. 499). The graphite extraction in Mozambique has caused resettlements which resulted in the loss of farming livelihoods which have not yet been adequately compensated (Khassab, 2021; Wiegink & Kronenburg García, 2022). The global lithium extraction has caused severe environmental degradation and the abandonment of ancestral settlements (Agusdinata et al., 2018). Hence, considerable implications by extractive activities can be observed on the land cover and environment but also the local communities.

The rapidly growing demand trends for energy transition materials estimated by Hund et al. (2020) let the authors conclude that the vulnerabilities and risks in resource-rich, low-income countries will intensify with increasing extraction rates. Consequently, monitoring change in areas of resource extraction is vital to mitigate potential impacts. To accomplish that, Environmental Impact Assessments (EIA) for extraction projects provide an important contribution. An EIA provides information about likely environmental impacts in the pre-decision phase of a planned project. EIAs are conducted or commissioned by the extractive companies themselves and need to be approved by the local governments before production. During the production phase, the EIA follow-up aims at regularly monitoring the anticipated implications. This way, resulting environmental implications can be better detected and governed. However, even though EIAs do exist and are binding for the extractive industries, they are often weak or being neglected, most notably in low-income countries (Edwards et al., 2014; UNDP & UN Environment, 2018). Particularly, the follow-up phase is rarely conducted as it requires substantial resources in terms of money, time and expertise (Arts & Morrison-Saunders, 2012; Marshall et al., 2005). Yet, despite the difficulties, monitoring potential implications resulting from the extraction of resources is crucial. Gathering this information provides a better understanding of what is happening in the local surroundings. Based on this knowledge public, private and non-profit decision makers can develop and enhance plans to preserve, manage and restore adjacent land and nature but also protect local communities. Hence, this knowledge is also key to proceed with the energy transition in a sustainable manner and to protect the planet and the people in the future. However, to achieve that, improved and more cost-efficient methods are urgently needed. Thus, the barrier for extractive companies but also other stakeholders such as concerned parties, public and non-profit organisations to monitor implications of extractive activities can be reduced.

The application of earth observation (EO) and geographic information systems (GIS) has been proven to be practical and cost-efficient to monitor social and environmental impacts resulting from extractive activities (Charou et al., 2010; Legg, 1994; Werner et al., 2019). Additionally, the proliferation of open access satellite data and the advent of cloud computing have broadened the range of applications for using geospatial tools (Hird et al., 2017). Open access satellite data is free to use as it is commonly provided by state funded programmes. Most popular are the Landsat and Sentinel imagery archives. The former was released by the United States Geological Survey (USGS) and the latter by the Copernicus Programme from the European Space Agency (Belward & Skøien, 2015). Advances in cloud computing with platforms such as Google Earth Engine (GEE) reduce the requirement of high computing power to download and manipulate geospatial data. Therefore, current advances in satellite data collection and analysis as well as the availability of cloud computing have made the use and processing of EO data more cost-efficient and accessible. However, cloud-based data systems are not broadly used in earth data science yet, but increasingly attract the interest of users, particularly due to ever growing data volumes and limited processing capacities (Wagemann et al., 2021). Considering these developments, using earth observation based on cloud-computing and open access satellite data, to monitor, quantify and geovisualise environmental impacts and land cover change resulting from extractive activities seems sensible. This accounts particularly to stakeholders in vulnerable and remote regions where EIAs are primarily neglected due to a lack of financial means or political support. By exploring this possibility, a contribution to a more sustainable energy transition can be made.

1.1. Research objectives.

The purpose of this thesis is to develop a proof of concept to map and identify change resulting from extractive activities at the source of energy transition supply chains. Remote sensing data has been identified as a suitable, cost-effective way to map change in areas of resource extraction. At the same time, the use of open access satellite data and cloud computing is under-utilised in current research on mapping change in these areas. Therefore, I investigated the use of open access satellite

data and cloud computing platforms to build time series and formulated two research objectives to approach a proof of concept solution.

The first objective of my thesis was to create a reproducible, scalable, and accessible workflow model to map land cover change and environmental change in areas of resource extraction. The model is based on earth observation data analyses, using solely open access satellite imagery and GEE as a cloud computing platform. The second objective was to apply and test the model, to monitor and reveal change resulting from extractive activities at the source of energy transition supply chains. In focussing on the extraction of resources at the source of energy transition supply chains, I aim to emphasise a rapidly expanding extraction frontier and the potential complications for a sustainable energy transition.

To meet the objectives, two case studies in the province of Cabo Delgado in northern Mozambique were conducted. Cabo Delgado is subject to a vastly expanding energy transition-induced extraction frontier. The workflow model was tested in both case study areas and the results obtained provide insights into the effects of extractive activities on their surroundings. The terms workflow and model will be used interchangeably throughout the research.

1.2. Research questions.

How may open access satellite data and cloud computing be used to monitor change in areas of resource extraction at the source of energy transition supply chains?

I. What is an example of a reproducible, scalable model to map change in areas of resource extraction using open access satellite data and cloud computing?
II. What extent of change can be monitored in the case study areas over time by applying the workflow model?
III. How reliable are the results obtained by combining satellite data and cloud computing in the monitoring process?

2. Conceptual Background

2.1. Monitoring the impacts of extractive activities on their surrounding by using GIS and remote sensing.

GIS is seen as a crucial tool for project-specific EIAs and risk mitigation as it enhances the processes with valuable mapping strategies and visualisations (Gharehbaghi & Scott-Young, 2018). Conducting EIAs with the application of GIS techniques is considered essential towards a sustainable development of modern societies (Abbas & Ukoje, 2009). Larger extraction companies do make use of these tools to conduct EIAs and are even provided with specifically tailored software applications for their needs (ESRI, 2018). However, the weakness or neglect of mandatory EIAs is a recurring issue, particular in vulnerable regions facing vastly expanding extraction frontiers (Edwards et al., 2014). Werner et al. (2019) have conducted a literature review on the application of GIS and remote sensing in the context of mining activities, outside the mining industry. Valuable contributions to assess the impacts on water, land, society and economy were found in academia and from civil society organisations. Yet, the authors also criticised a strong focus on solely project-specific assessments and a weak attention to cumulative impacts. By considering multiple sites in one particular region or of one particular commodity in various extraction sites, cumulative impacts could be better assessed and provide new insights on the impacts of extractive activities. Evaluating the environmental impacts of mining by classifying the land cover and observing the change over time has been proven a cost-effective method, particularly in remote locations which are difficult to access (Paull et al., 2008). Charou et al. (2010) have used multi-temporal satellite images to conduct automated land cover classifications and analyse the land change. In doing so, the authors assessed the impacts of mining

activities on land and water and emphasised on the ability to cover large areas as well as the low costs of using remote sensing. To emphasise environmental degradation resulting from lithium mining in Chile the use of vegetation indices and temperature was tested successfully by Liu et al., (2019). Adding on the land change and environmental degradation aspects, Lechner et al. (2019) underscored land cover classes such as settlements and infrastructure to represent socio-environmental impacts in the GIS based research on mining activities. When examining the literature many more valuable contributions can be found such as Firozjaei et al. (2021), who developed a model to not just examine historical impacts of mining activities but also to predict likely land cover change in mining areas. Altogether, most contributions analyse land change in the surroundings of extraction locations by creating remote sensing time series and subsequently deduce environmental but also societal implications. Whereas the use of open access satellite data such as Landsat imagery is quite common in academic research on mining activities, the use of cloud computing or the combination of both is still weak. Nevertheless, Hird et al., (2017) or Huang et al. (2017) already focussed on that combination to map land cover dynamics in wetland and urban areas. Therefore, the combination of remote sensing with open access satellite data and cloud computing to investigate land cover change and environmental degradation resulting from extractive activities constitutes a gap in current literature. With my thesis I filled this gap and developed a reproducible, scalable workflow. By having a particular focus on the cost and accessibility benefits the workflow is supposed to be particularly relevant for concerned parties in vulnerable regions, facing vastly expanding extraction frontiers due to increasing demand figures of energy transition materials. Thereby, a contribution to a more sustainable energy transition can be made, as also Lèbre et al. (2020) call for a greater focus and more in-depth interrogations on “The synergies and trade-offs at the source of energy transition material supply chains…” (p.4).

2.2. Remote sensing time series for monitoring land surface dynamics.

Time series analysis is an analytical approach in remote sensing and often applied to reveal land surface dynamics by creating land cover maps (Gómez et al., 2016). To reveal land surface dynamics, satellite images are spectrally analysed over a defined period of time and within a defined geographic area. Spectral analysis is the quantitative or qualitative investigation of reflectance properties of the land surface obtained from satellite imagery (Mustard & Sunshine, 1999). By investigating spectral reflectance researchers can see beyond visible light to monitor and measure phenomena like plant health, moisture extent and more. To construct time series a temporal, geographic and thematic scope needs to be determined. The temporal scope is determined by a specific time span which can be subdivided in defined intervals. The geographical scope depends on the desired study area which can be global, regional, or local.

To develop meaningful time series three components are of high importance (Kuenzer et al., 2015). Firstly, long term directional upward or downward trends, secondly, seasonal variations and finally short-term fluctuations. Each of these three components can be investigated using time series, simultaneously or separately. Long-term directional trends would be the rise of sea-level, vegetation loss or drought. Short-term fluctuations are the occurrence of hazards, fires, or plant diseases. To observe trends or fluctuations, land surface attributes are artificially classified in land cover categories (Lambin & Linderman, 2006). This categorisation always represents a simplification to a more complex reality. However, by representing the land surface in distinct categories, the detection of changes such as seasonal patterns, vegetation loss or urbanisation can better be observed, quantified and interpreted. To classify the land cover into thematic categories, representing desired classes, mainly two methods are applied: the object-based and pixel-based classification (Weih & Riggan, 2010). Thus, the supervised and unsupervised pixel-based classifications are the most common methods, whereas a third, the object-based approach, has gained popularity in recent times. Unsupervised pixel-based classification methods cluster the land surface into groups without sample data sets, one such method is the K-means algorithm. The K-means algorithm clusters the pixels of an

image into a number of groups based on their similarities, the number of groups is represented by k (Likas et al., 2003). Supervised pixel-based classification methods such as the Support Vector Machine (SVM) or the Random Forest (RF) classification are machine learning algorithms which cluster the land surface on the basis of training data which has to be initialised as a manual input. The SVM segregates the pixels of an image into desired classes by searching for a boundary between two classes with the biggest existing margin (Hsu et al., 2003). The more scattered the classes are, the more difficult it is to find appropriate boundaries. The RF algorithm is based on decision trees which split the information stored in a pixel, and as a result assign each pixel a land cover class. The more decision trees are used the more accurate are the results (Gislason et al., 2006). According to Maxwell et al., (2018) it is "…impossible make a universal statement of what classification algorithm is most suitable for remote sensing classifications". Hence, the choice of a suitable classification algorithm depends on the scale and complexity of the study area and commonly requires the trial of different algorithms, to achieve the best accuracies possible.

Long-term time series can also be created based on geophysical or index variables (Kuenzer et al., 2015). Geophysical variables are e.g., the top-of-atmosphere reflectance (TOA), the Land Surface Temperature, and the Leaf Area Index. All space-based sensors record imagery containing TOA. When examining the land surface on the basis of TOA imagery, atmospheric distortion needs to be corrected first. Dimensionless index variables are e.g., the Normalised Difference Vegetation Index (NDVI), Normalised Difference Built-up Index (NDBI), and the Modified Normalised Difference Water Index (MNDWI). The NDVI, NDBI and MNDWI are all indices describing the difference of particular spectral bands and support the detection of certain land surface properties such as vegetation, built-up areas or water bodies. NDVI is also a commonly used variable to detect environmental degradation in areas of mining production (Liu et al., 2019).

3. Data and Methods

The concept of time series was applied in this research to monitor change resulting from extractive activities at the source of energy transition supply chains. As concluded in 2.1 creating time series is a widely used and tested approach to do so. Particularly interesting in this context is the use of open access satellite data and cloud computing. By applying free-to-use open access satellite images and cloud computing the applicability and accessibility of this model is enhanced. To test the model two case studies were conducted.

3.1. Case study design.

3.1.1. Case study selection.

The two case studies are located in the province of Cabo Delgado in northern Mozambique. Mozambique is situated in Sub-Saharan Africa. Sub-Saharan Africa constitutes an underrepresented area in the geographical case study-based research on energy justice issues (Jenkins et al., 2021). Moreover, while aiming to focus on the extraction of resources at the source of energy transition supply chains, Cabo Delgado in Mozambique is particularly interesting as it is home to large high-quality graphite reserves and natural gas reserves (Brown et al., 2021; Macuane et al., 2018; Salimo et al., 2020).

The province of Cabo Delgado is situated in a subequatorial zone and affected by the tropical rain belt which causes monsoon rains in the region from January to March. Heavy rainfalls and cloud cover are abundant during this time. Contrary to that, the region experiences a hot, dry season from July to October. To investigate land surface dynamics cloud cover and dry seasons constitute a barrier as spectral information on satellite imagery is less distinct, which hampers the classification process of land covers. Cabo Delgado is also endowed with abundant natural resources and home to numerous

locations, extracting ruby, graphite, marble and natural gas (Els & Chelin, 2021). In my research I focus on two extraction locations which can directly be linked to resource extraction for the energy transition, namely graphite and natural gas. Namely, one graphite mine and one onshore area which supports the offshore extraction of natural gas are considered. Tab. 1 provides an overview of the study areas and relevant key data for constructing the time series.

Tab. 1 Overview of the study areas Balama Graphite Project and the Mozambique LNG Park.

	Balama Graphite Project	Mozambique LNG Project
District	Balama	Palma
Lead	Syrah Resources Ltd.	TotalEnergies
Concession		
- Granted	2013	2012
- Expires	2038	n.a.
Start of production	2018	2024
Area [ha]	11,031	7,078

The extraction of graphite in Cabo Delgado is currently taking place in three mines, the Balama, Ancuabe and Montepuez graphite project. The Balama Graphite Project (BGP), situated over the largest high-grade graphite deposit in the world and is managed by Syrah Resources. Together, the BGM and the Ancuabe Graphite Project have increased their production from 802 tonnes in 2017 to 113,803 tonnes in 2019 (Brown et al., 2021). Additionally, there are 8 ongoing graphite exploration projects in Cabo Delgado (Mitchell & Deady, 2021). To provide an overview of the current situation the concession areas of the respective extraction locations are depicted in Fig. 1.

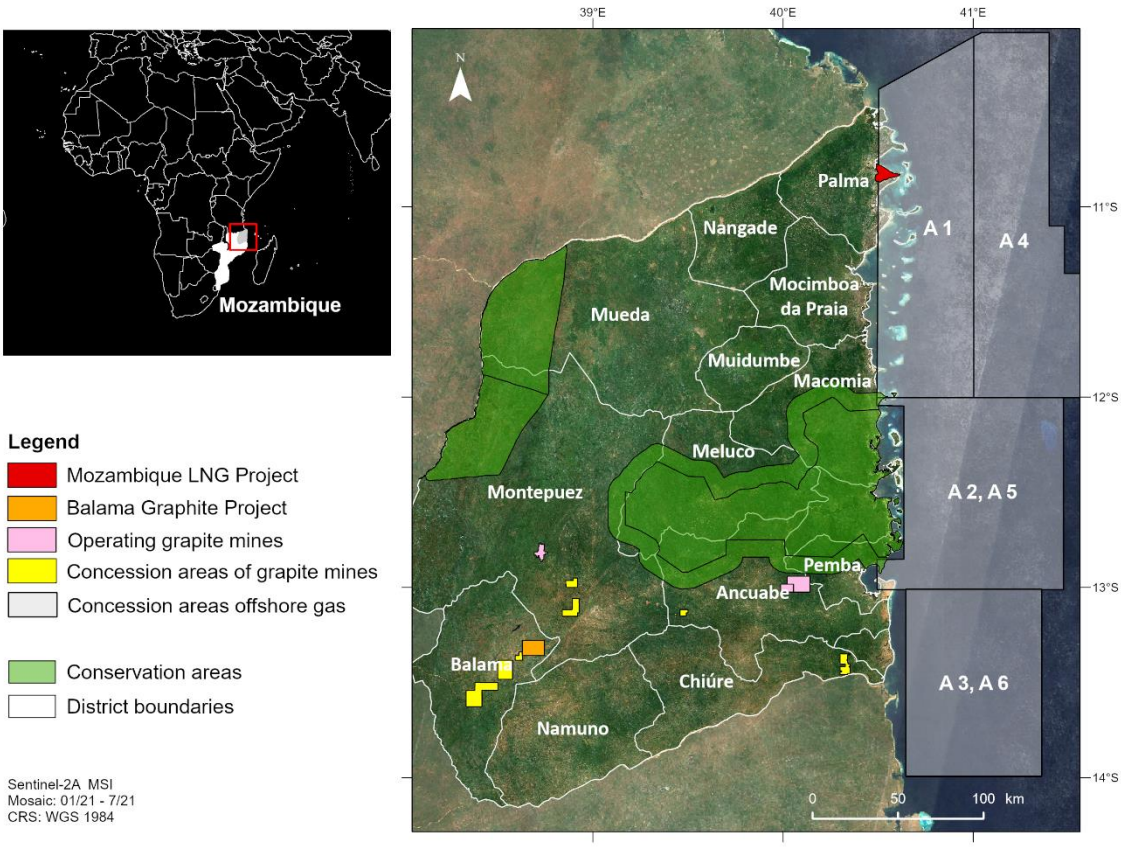


Fig. 1 Overview map of Cabo Delgado with the Balama Graphite Project (orange), Mozambique LNG Project (LNGP), current graphite mine concession areas (yellow), district boundaries and conservation areas.

Large gas fields in the Rovuma Basin just off the coast from Cabo Delgado were discovered in 2006. Currently, concessions are given for two extraction areas, in total covering an area of around 7,500 square kilometres (INP, 2014). Area 1 has been operated by the multinational Total since 2006. Area 4 was awarded to the companies Eni and ENH in 2006. The licensing of five additional areas off the coast of Cabo Delgado is currently taking place, the results of this 6th licensing round will be available in November 2022. To process and distribute the gas from the offshore fields Total is building onshore support facilities and an airport, together forming the Mozambique LNG Project (LNGP) (Kirsch et al., 2021). The onshore area of the LNGP is subject to the third case study and as well depicted in Fig. 1.

3.1.2. Study area boundaries.

Identifying distinct boundaries of the study areas for the case studies and by that limiting the geographical scope to a meaningful area is important to limit the computing power required for the classification processes. Generally, the boundaries of extractive activities could be determined based on the granted concession areas. However, this would neglect the possibility that the prospection, exploration and exhaustion of the respective resources, supporting infrastructure and the development of mining areas might exceed or be smaller than the concession area itself (Song et al., 2020). Particularly, this relates to the construction of roads or harbours, demographic effects on surrounding conurbations or resettlement measures. These effects might cause as well significant land cover change and environmental degradation. Generally, the study area boundaries have to be set in line with the aim of the research. For my researcg, I did consider the location of Project-Affected Communities (PACs) to draw a polygon depicting the study area. The PACs are listed in the resettlement action plans of the EIAs of both extraction projects (ENI, 2014b; EOH, 2014). By determining the study area boundary on the basis of the PACs, I do have a broader geographical scope than considering solely the concession areas. I do expect by including the PACs, that major changes in the surrounding of the case studies resulting from e.g. demographic effects and expanding supporting infrastructure can be detected. The final study areas and all considered PACs are depicted in Fig. 2. The study area for the BGP case study is 44,362 ha and the LNGP case study is 40,195 ha.

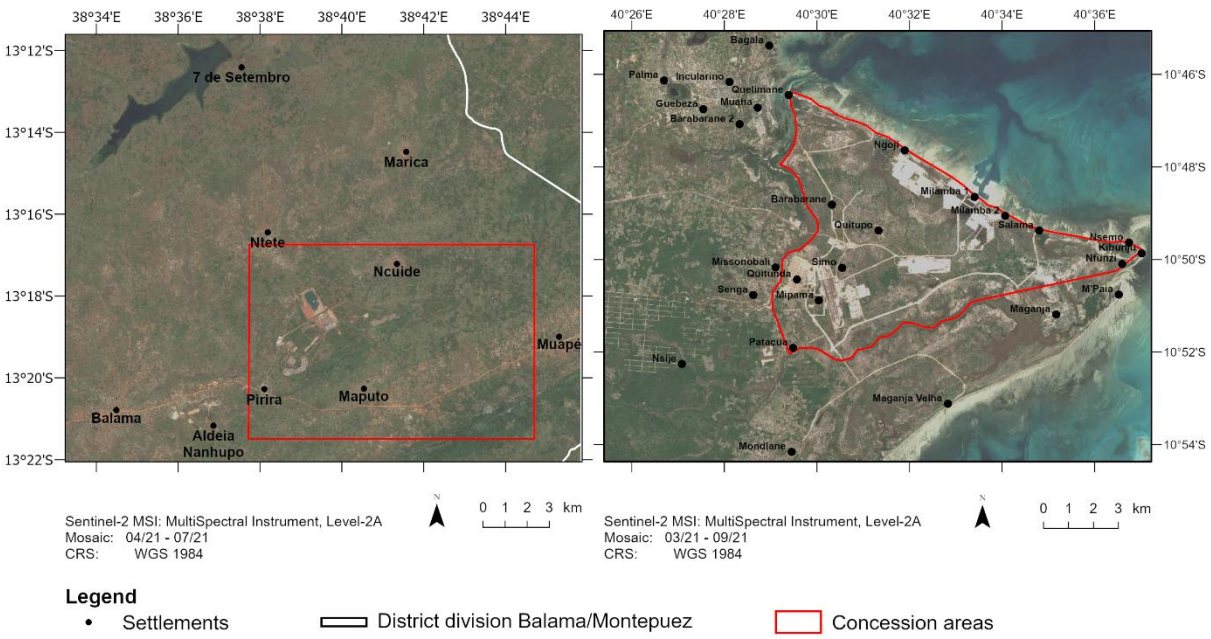


Fig. 2 The Balama Graphite Project (right) and the Mozambique LNG Project (left) study areas. Indicated are the concession areas and Project-Affected Communities.

3.2. A model to map change in areas of resource extraction.

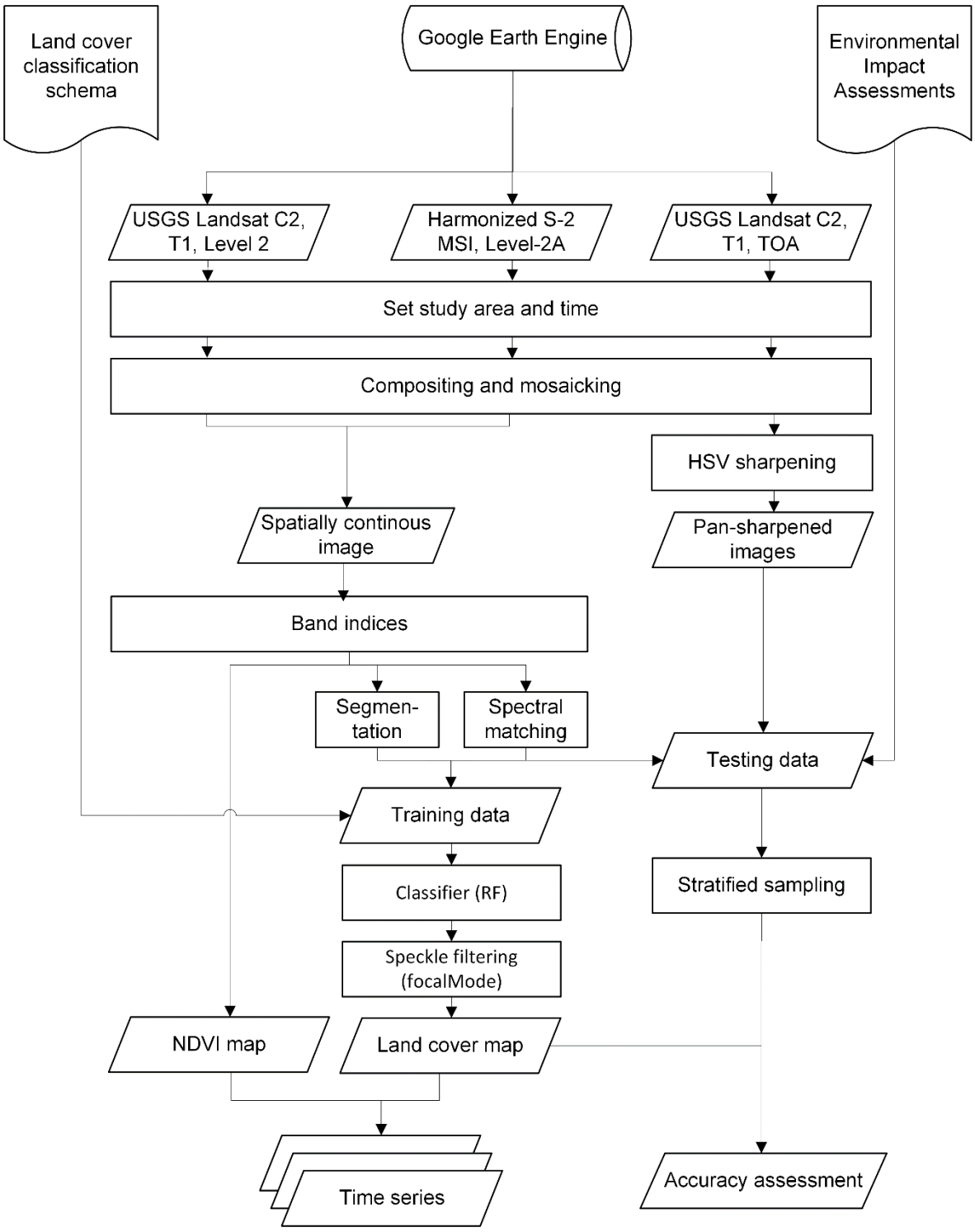


Fig. 3 The model to create time series and map change in areas of resource extraction.

Creating remote sensing time series requires a variety of working steps and variables as introduced in 2.2. The main aim of my thesis was to develop a scalable, replicable workflow to create time series to detect land cover change and environmental degradation due to extractive activities. The workflow can be conducted and altered by anyone who brings the technical interest or know-how but does not require any exalted financial or technical means as it is based on GEE and open access satellite data. By creating multiple land cover maps of each study area, time series over a defined time span and within a determined geographical area can be created. Fig. 3 shows the final working process for the operational framework to create the time series and test their reliability.

The land cover classification to construct the time series was conducted with a supervised pixel-based classification. A supervised pixel-based classification method requires the use of training data. If the training data is collected a machine learning algorithm is used to assign all pixels of the image to a class label. By creating the operational framework and testing its applicability in the study areas sub-question one was answered.

3.2.1. Data selection.

The selection of appropriate satellite imagery is essential to successfully conduct land cover classification over a distinct temporal and geographical scope. The temporal scope is determined by the lifetime of the extraction locations as shown in Tab. 1. The geographic scope of these study areas is well below province level. Considering these constraints and the aim to use open access satellite imagery, there are two satellite systems providing suitable imagery, Landsat 4 – 8 and Sentinel 2.

Landsat 4 - 8 provide imagery since 1970 with a spatial resolution of 30 m in the multispectral ranges and 15 m in the panchromatic band. Landsat imagery is provided in a variety of datasets ranging from raw scenes to high-quality imagery, composed and pre-processed by USGS. The highest-quality imagery are the Level 2, Collection 2, Tier 1 datasets. Thereby, Tier 1 indicates that it is the highest data quality currently available, Level 2 stands for atmospherically corrected data and Collection 2 indicates improved geometric accuracy (Landsat Missions, n.d.) The Level 2, Collection2, Tier 1 datasets are well suited for the land cover classification process as they provide a homogenous database over a long period of time. However, considering the geographic scope of the study areas, imagery with a higher resolution is favourable, so distinct objects such as roads or pit water can accurately be represented as well. Sentinel 2 does provide a spatial resolution of 10 m in the visible and near-infrared (NIR) spectrum and images are recorded since 2017. Therefore, classification maps from 2017 onwards are preferably be constructed on the basis of Sentinel 2 data.

Tab. 2 Satellite data used to conduct the land cover classification.

Dataset	Data availability	Sensor	Pixel size [m]	Bands
USGS Landsat 8, Level 2, Collection 2, Tier 1	2013/03/18 – up to date	Landsat 8 OLI/TIRS	30	coastal aerosol, blue, green, red, NIR, SWIR (1,2)
USGS Landsat 5, Level 2, Collection 2, Tier 1	1984/03/16 – 2012/05/05	Landsat TM	30	blue, green, red, NIR, SWIR (1,2)
Harmonized Sentinel-2 MSI, Level 2A	2017/03/28 – up to date	Sentinel-2 MSI	10	blue, green, red, NIR
			20	red edge (1,2,3,4), SWIR (1,2)
			60	aerosols, water vapor, cloud mask

Furthermore, to support the land cover classification with manual interpretations Landsat TOA, Collection 2, Tier 1 data was combined with the Level 2, Collection 2, Tier 1 datasets to construct pan-sharpened images. The TOA dataset contains the panchromatic band which is needed for the pan-

sharpening process. Pan-sharpening is the improvement of the resolution of multispectral imagery by integrating the geometric details of high-resolution pan-chromatic images (Du et al., 2005). The pan-sharpening was conducted with GEE and the Hue, Saturation, and Value (HSV) method in GEE. Tab. 2 summarises the chosen datasets, their characteristics and temporal availability. The desired data can directly be obtained and altered in GEE.

Tab. 3 Auxiliary data used to support the land cover classification process with manual interpretations.

Dataset	Data availability	Sensor	Pixel size [m]	Bands
USGS Landsat 8, TOA, Collection 2, Tier 1	2013/03/18 – up to date	Landsat 8 OLI/TIRS	15	panchromatic, cirrus
			30	coastal aerosol, blue, green, red, NIR, SWIR (1,2), TIR (1,2)
Google Earth	3D representation of the earth based on multiple satellite images at multiple points of time.			

3.2.2. Data pre-processing.

To construct meaningful land cover maps for the time series, the imagery must be cloud cover free and provide distinct spectral information. Considering the location of the study areas as described in 3.1. cloud cover and drought are major challenges. The former makes it impossible to obtain spectral information from the multispectral imagery and the latter makes it difficult to distinguish between different spectral profiles and identify distinct objects. Due to seasonal variations, a time window must be chosen were little cloud cover and sufficient spectral information to distinguish vegetation, rocks, and other land cover types can be expected.

For Cabo Delgado, this time window was found to be from May to July, between the raining season and the dry summer season. These three months outline the time span to construct quality mosaics for the classification. For most years in the study areas this time span proved to be suitable to obtain cloud free mosaics.

To further support the classification process and better distinguish between different land cover classes, the images are provided with additional band indices. Band indices such as NDVI or NDBI provide additional data for the classification algorithm to distinguish certain objects. The band indices added to the images before conducting the land cover classification are listed in Tab. 4. Additionally, the classification is based on the bands from the visible spectrum (Blue, Green, Red) and the NIR and SWIR bands.

Tab. 4 Band indices included in the classification process.

Band	Range	Description	Calculation	
NDVI	(-1) to (+1)	Water and barren land ((-1) – 0), Dry, sparse veg. (0.1 – 0.5), Healthy, dense veg. (0 – (+1)).	$\frac{NIR - Red}{NIR + Red}$	(Eq. 1)
NDBI		Water and vegetation ((-1) – 0), Built-up (0 – (+1)).	$\frac{\text{SWIR 1} - \text{NIR}}{\text{SWIR 1} + \text{NIR}}$	(Eq. 2)
MNDWI		Built-up, soil, veg. ((-1) – 0), Water features (0 – (+1)).	$\frac{\text{Green} - \text{SWIR 1}}{\text{Green} + \text{SWIR 1}}$	(Eq. 3)

3.2.3. Land cover classes and training data.

Eventually, the land cover classes must represent the spatial dimension of the extractive activities, the characteristics of its surrounding and the desired aim of the research. Thereby, land cover classifications are always a simplified representation of the actual circumstances. If the classification process does not include elaborate manual alterations, an optimal representation of the region is restricted by technical constraints, such as the spectral information which can be obtained from the imagery, or the inaccuracies introduced by the classification algorithm. Hence, the choice of classes depends on the characteristics of the respective study area and the complexity of the classification process. The land cover classification schema depicted in Tab. 5 was developed in an iterative process with the classification process. The iterative process contained the trial of different land cover classes while observing the resulting classification accuracies.

Tab. 5 Land cover classification schema and information for the training data collection.

Class	Definition	Collection
Mining (BGP)	Mining pit, mining waste, tailings.	Identification by maps from EIA, shape and location and high-resolution imagery (Sentinel-2A).
Wetland (LNGP)	Wetland, saltmarshes, flooded vegetation.	Identification by spectral signature and image segmentation.
Dense vegetation	Trees, woodland, dense shrub/scrub.	Identification by spectral signature and image segmentation.
Sparse vegetation	Crops, grass, herbaceous areas, sparse shrub/scrub, deforested areas.	Identification by spectral signature and image segmentation.
Unvegetated	Built-up areas, infrastructure, bare ground.	Identification by shape and location, spectral signature and high-resolution imagery (Sentinel-2A).
Water	Lakes, rivers, sea, pit water.	Identification by shape and location, spectral signature and image segmentation.

The land cover classes are a prerequisite to identify and collect the training data for the classification process. The training data consists of a pixel or a set of pixels in an image, these represent a particular theme or object which in turn represent the land cover class. On the basis of the training data the remaining parts of the image were allocated to a specified land cover class by the classification algorithm. In this study, training data was compiled by collecting a set of pixels representing a class. Depending on the similarity of one class to another more or less pixels need to be collected, to avoid under- and overfitting. Additionally, to ensure the comparability of the classification results between the years, the training data was collected from locations which did provide the same land cover throughout the whole time series.

Tab. 5 states the collection process of the training data per class. Each class is represented in both case studies, except from Mining which is solely investigated in the BGP case study and Wetland which is solely investigated in the LNGP study area. As a first step, the training data was identified due to obviously interpretable elements, such as location, shape, tone, and size. To support these interpretations the auxiliary data from Tab. 3 was used. Furthermore, the EIA of the BGP site was used to manually identify and verify elements of the extraction site itself (CES, 2014). Where the manual interpretation wasn’t leading to explicit results the training data collection was enhanced by adding image segmentation and the identification of spectral signatures to the process. The former is indicated in Fig. 3 as *Segmentation* and the latter as *Spectral matching*. In the case of a low resolution (scene elements are smaller than the image pixels) the additional methods can provide more accurate training data. Image segmentation is the process of grouping a set of pixels based on their spectral and

spatial similarities and thereby partitioning the image in different segments. The identification of spectral signatures was conducted by inspecting the spectral curve of the selected training data and compare it to idealised spectral reflectance curves. The idealised reflectance curves, provided by institutions such as the John Hopkins University and the USGS, are compounded in the ECOSTRESS spectral library from the NASA Jet Propulsion Laboratory (NASA/JPL, n.d.). The ECOSTRESS library does provide spectral profiles from a variety of different minerals, vegetation, water types etc. In this research the level of detail for the land cover classes was not as specific and did not differentiate between various minerals or vegetation types. To validate the self-collected spectral profiles, I did compare them with profiles from species from the same class, such as shrub, grass, sea water and road. The classes are indicated in the ECOSTRESS library. Subsequently, the selection of training data was adapted to resemble the idealised reflectance curves.

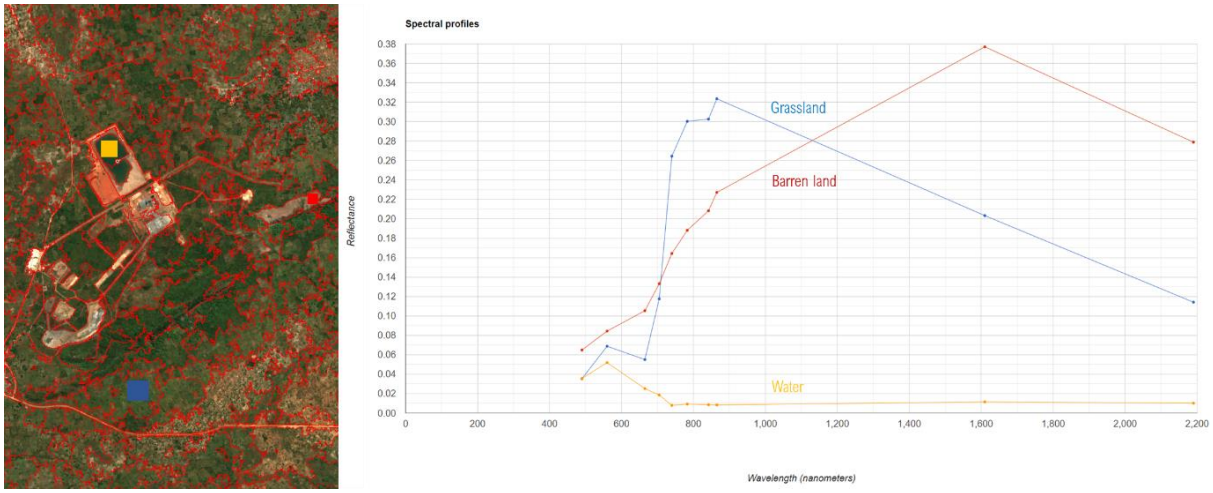


Fig. 4 Example for the training data collection. On the left, image segmentation used to identify suitable regions of interest for pixel collection. On the right, spectral profiles from the GEE console of aggregated pixels representing the training data of one class. The spectral profiles, obtained from the aggregated pixels, were used to compare their similarity to idealised reflectance curves of the respective land cover type.

3.2.4. Land Cover Classification.

To classify the remaining image on the basis of the collected training data the RF machine learning algorithm was used in GEE. In the first place, different supervised classification algorithms were tested to achieve the highest accuracies possible. Supervised classification algorithms available in GEE are the Classification and Regression Trees (CART), the SVM, the NaïveBayes and the RF. For this study the SVM and RF algorithm were tested, as they constitute the most popular and accurate algorithms. The testing was part of preliminary research in which the RF achieved the highest accuracies and was subsequently implemented in the workflow as the default classifier. Other than that, the RF classifier was tested on a different number of decision trees to improve the accuracy of the results. In the course of classification procedures in remote sensing a variety of settings on the number of decision trees of the RF classifier were already tried and suggested, whereas a number of 50, 100 or 500 are most commonly used (Maxwell et al., 2018). Eventually, I did set 100 trees as a default in the workflow. The accuracies achieved with 100 trees were slightly higher than with a setting of 50. There was no difference in the accuracies observed between 100 and 500 trees. Yet, the classification process was notably faster with 100 trees.

After the classification process, the workflow was added a post-classification step as depicted in Fig. 3. Misclassified pixels or mixed pixels can cause a speckled salt and pepper effect in the classification results. An example would be a pixel classified as road in the middle of the ocean. The speckle or salt and pepper effect is a common result of pixel-based classifications, particularly on

high-spatial resolution imagery with smaller pixel sizes. The aim of the post-classification step is to filter the speckle of the classified image. To filter the speckle the focalMode operation was used in GEE. The focalMode is a reducer operation which uses a squared kernel to aggregate and cluster pixels, based on their majority. As a result, single pixels are integrated into their environment to harmonise the image.

3.2.5. Time series and change detection

As depicted in Fig. 3, time series was constructed for land cover maps and the NDVI index.

The time series for the land cover maps is based on parameters such as the award of the concession and the start of production of the extraction locations. These parameters differ per extraction location as shown in Tab. 1. Eventually, these years constitute the intermittent time intervals. As both extraction locations are still active, the end year of the time series is determined by the most recent year, which is for this study 2021. The start year represents the pre-extraction time of the study areas, showing the areas without the influence of any potential extraction-induced changes. The pre-extraction year is determined by taking the same time span between granted concession and most recent year and subtracting it from the granted concession year. To support the land cover time series land cover change maps were constructed. Land cover change maps compare two particular years and visualise what land cover was substituted by the other. In doing so, it can be observed what land cover classes had the highest impact on change in the study area. The land cover change maps compare the year of granted concession with 2021 for both study areas.

The time series for the NDVI index was applied on a yearly basis from the start of production until the most recent year. For the time series the mean NDVI was calculated for each year, in the time window with little cloud cover and sufficient spectral information. The mean NDVI was used to observe the directional trend of the variable in the study area. Additionally, image differencing was applied for the NDVI. With the differencing method major changes in the study areas can be detected. By comparing the NDVI values per pixel for a chosen start and end year, vegetation loss can be observed and located. To detect the differences, the differencing method subtracts the values of two particular years as follows in Eq. 4, whereas X stands for NDVI. The method can be applied with any other index or

$$X_i = \mathrm{X}_{\mathrm{year\ n}} - \mathrm{X}_{\mathrm{year\ m}} \qquad \text{(Eq. 4)}$$

The land cover maps are used in this study to quantify the land cover change and observe it over time. The NDVI index is used to particularly observe the environmental degradation in the study areas.

3.3. Accuracy assessment.

Accuracy assessments are a crucial part of any research containing land cover classifications. The value of a map is highly dependent on its accuracy, as wrong land cover classifications lead to wrong conclusions. The most common way to test the accuracy of land cover maps is to compare ground truth data of the case study with the classification results. This is done by collecting a number of testing points in the field, assign a land cover class to each of them and compare them to the land cover class which was assigned to the exact same location in the classification process. For this approach ground truth data needs to be collected or an expert, familiar with the local conditions, needs to be consulted. The former and the latter was not available for my research. However, the absence of ground truth data or an local expert is a common issue in remote sensing research projects. As a consequence, it is a common approach to take the same imagery which was used for the classification or high-resolution imagery as a surrogate and collect testing data on it (Foody, 2002).

The accuracy assessment in this research was conducted by compiling a number of 50 test points for each land cover class, each year of the time series and each case study. A minimum of 50 control points per class is suggested by Congalton & Green, (2019) for study areas of less than 1 million acres and with fewer than 12 land cover classes. The study areas are described in 3.1.2 and are much smaller than 1 million acres (BGP: 109,620, LNGP: 99,324 acre). Additionally, the number of land cover classes is limited to 5 classes per case study. To distribute the testing points I did define testing polygons for each land cover class. The testing polygons are set to avoid spatial autocorrelation between test and training data. Spatial autocorrelation describes the effect which occurs when two observations have a high spatial proximity, resulting in a higher probability to share similar characteristics (Ploton et al., 2020). The occurrence of spatial autocorrelation is a common issue when conducting accuracy assessments for land cover classifications leads to biased results and a tendency for too high accuracies (Congalton, 1991). To avoid spatial autocorrelation between the test data and the train data, the testing polygons were collected with as much distance as possible to the training polygons. This distance had to be kept dynamic as it depends on the areal extent of the study area and the prevalence of each land cover class. The allocation of the testing polygons was conducted in the same manner as the training data collection, supported by information retrieved from land cover maps provided in the EIAs (CES, 2014; ENI, 2014a). Subsequently, to distribute the test points throughout the testing polygons, a stratified sampling approach was chosen. By using the stratified sampling approach, I follow the suggested sampling design from Olofsson et al., (2014).

According to Foody, (2008) there is a widely used minimum level as target accuracy in current remote sensing research of 85 %. The author criticises the 85 % as a too ambitious target accuracy for most current mapping applications but states it as appropriate for mapping broad land cover classes at a small cartographic scale. The cartographic scale of the study areas in this research and the amount of land cover classes let me conclude that the target accuracy of 85 % is suitable for my research.

By adding an accuracy assessment which tests the reliability of the land cover classifications, sub-question three was answered.

4. Model results

The results of my research methods displayed in detail in Fig. 3 are presented in this section. These methods were applied on two case studies. By applying the workflow to both case studies, I created time series that visualise and measure the land cover and environmental change in the case study over time. Additionally, the reliability of these results was tested by conducting an accuracy assessment. The time series data analysis was entirely conducted using the cloud-computing platform Google Earth Engine, results were exported and the layouts for the maps and the diagrams shown in this chapter were compiled with ArcGIS Pro and Excel.

4.1. Time series and change detection.

In section 3 the geographical was determined. The study areas introduced in section 3.1.2 represent the overall geographical scope, for which the analysis was conducted. To visualise and present the results, I decided to curtail the areal extent of the study areas to the concession areas of the project LNGP and BGP. The reason for this is laid out in the following.

Both case studies represent extraction projects in an early-stage development. The BGP site started production in 2018 and the LNGP is still under construction and aims at starting production from 2024 onwards (Tab. 1). Due to the recency of both projects, neither explicit upward or downward directional trends of land cover and environmental change could be identified in the overall study areas. In the course of the research, it was increasingly apparent, that the measurable change that the

case studies cause in their surroundings is limited yet. For this reason, the results section of my thesis focusses on the geographical scope of the concession areas, where change was more distinct and measurable. By limiting the geographical scope in accordance with the present state of the extraction projects, the visualisation and interpretability of the created maps is enhanced as areas subject to change are better visible.

The land cover classification for each year of the time series is conducted with the respectively best available image composite (based on spatial resolution, available bands, and cloud cover). The process to obtain the best available image composite is laid out for each case study.

4.1.1. Mozambique LNG Project.

To monitor change in the LNGP study area a cloud-free image of the licensed area for each year of the time series was required. Thereby, the regular presence of clouds was a limiting factor. Due to the high frequency of clouds, I adapted the initially planned time horizon of the NDVI, and land cover time series, and the time span used to construct quality mosaics for each of the years. By following the approach for the time series intervals, as it is laid out in section 3.2.5, the time series intervals 2003, 2012, 2021 were determined. The year 2012, in which the concession was granted, showed a particularly dense cloud cover. To bypass the cloud issues, the land cover time series uses 2005 as a pre-extraction year, 2013 as the year of granted concession and 2021 as the most recent year. Furthermore, due to the high frequency of clouds, the initially determined time span (May to June) to create the cloud-free image for each year was adjusted as well. Eventually, the time span from June until September proved to be more suitable. The time span from June until September was affected by less clouds and provided sufficient spectral information to differentiate the respective land cover classes. Other than expected, the spectral information during the summer months proved to be sufficient for this study area as the presence of wetlands cause the area to be much soggier.

Land cover time series

The mosaic for 2005 was created based on images recorded from the Landsat 5 satellite. The mosaic for 2013 is based on Landsat 8 images, and 2021 is based on Sentinel-2 images. Sentinel-2 provides a much finer spatial resolution than the Landsat images and is always used preferably, when aiming for a land cover classification. The land cover time series is represented in Fig. 5, showing the change over time of the land cover classes wetland, dense vegetation, sparse vegetation, unvegetated and water.

The year 2005 depicts the area of interest 8 years before the LNGP was initiated. Therefore, no construction sites or facilities of the project were identified. Even though, several settlements do exist in the area, these settlements are mostly so small and scattered so that they are not identified as a coherent patch of unvegetated areas. Though, Quitupo, the largest settlement inside the concession area, is recognised and indicated in the map. The rest of the unvegetated areas in 2005 are identified as barren ground. The areas classified as sparse vegetation do majorly consist of sparse shrub/scrub, whereas dense vegetation is identified as dense shrub/scrub.

In 2013, one year after the license for the project was granted, the area showed first alterations in its land cover composition, with a slight increase of unvegetated areas. As 2013 represents the year of granted concession, no major construction activities have taken place in the area yet. In between 2013 – 2021, a huge step in the development of the project has taken place. Almost all unvegetated areas in the region can be identified as facilities or supporting infrastructure. The outlines of buildings are clearly visible. At the same time dense vegetation did slightly increase as well.

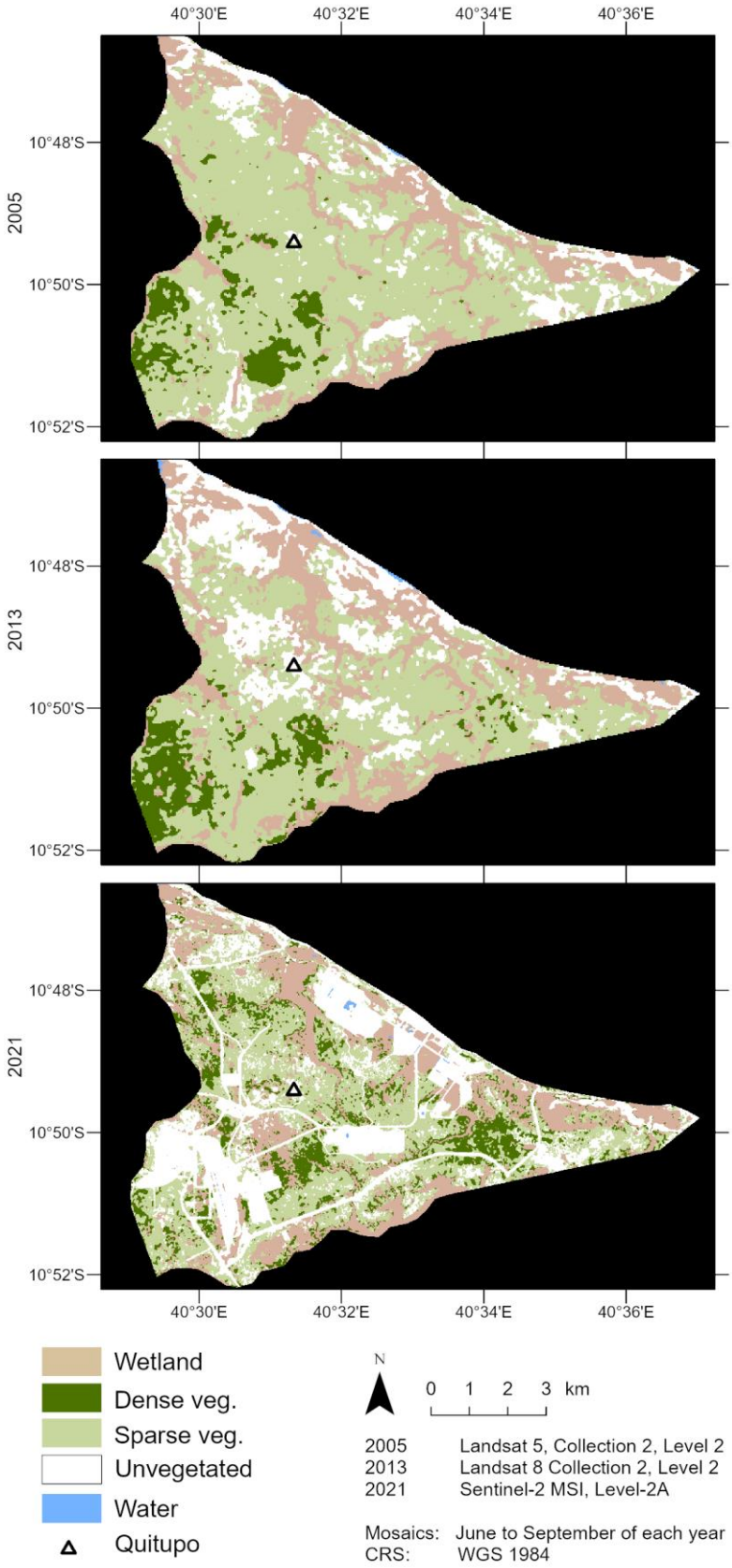
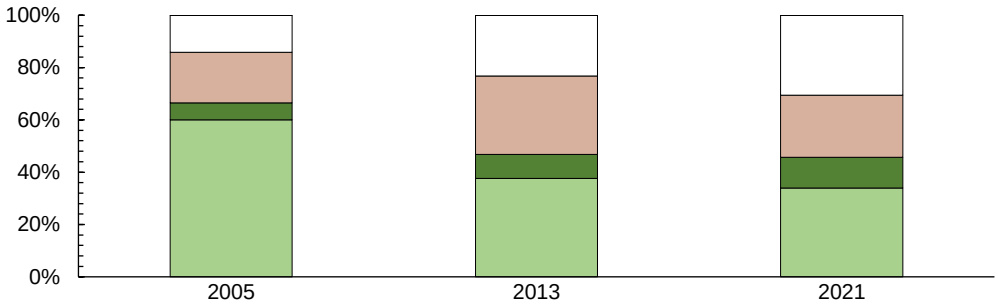


Fig. 5 LNGP case study: Time series of the land cover change in the concession area before the granted concession (2005), the year of granted concession (2013) and most recent (2021). Over time, the land cover is increasingly dominated by unvegetated areas due to the construction of facilities and infrastructure for the project which can be identified in 2021. The area's largest settlement is indicated in the map by the black triangle.

The data on the different land cover classes obtained from the time series is provided in Fig. 6. For the concession area the share of the land cover classes is depicted in the stacked bar chart and underneath the bar chart, a table with the precise areas in hectare for each class and year is provided. The total oarea (7,078 ha) refers to the size of the overall concession area, as indicated in Tab. 1. From the land cover maps it can be seen that the development of the LNGP caused a fairly uniform increase in unvegetated areas. When looking at the data for the concession area in Fig. 6, and comparing the numbers of unvegetated areas in 2005 and 2021 an increase of 116 % of the land cover class can be observed. This outlines the sharpest increase of all the land cover classes being studied. Furthermore, dense vegetation and wetlands also gained area when comparing the numbers of 2005 and 2021. The former increases by 81 % and the latter by 19 %. However, for the landcover class wetland, significant inconsistencies can be observed. The wetlands constitute a highly dynamic land cover in the region of interest, undergoing major changes with tides and seasonal differences. These inconsistencies can be deduced from Fig. 6, where the land cover class wetland shows highly fluctuating values. This dynamic also impacts the classification results of the land cover classes sparse and dense vegetation. Depending on the aridity of the year or the turn of the tides and the correlating changes in moisture, a particular piece of land might rather be classified as dense vegetation than as wetland, due to the absence of water. Or dense vegetation might be classified as sparse vegetation, due to less water content in the plants, making the spectral reflectance curve less distinct.



Class area [ha]	2005	2013	2021
Water bodies	6	5	11
Unvegetated	998	1,642	2,155
Wetland	1,364	2,116	1,681
Dense vegetation	460	649	832
Sparse vegetation	4,250	2,666	2,399
Total	7,078	7,078	7,078

Fig. 6 LNGP case study: Share [%] and areal extent [ha] of the land cover classes in the concession area for the years of the granted concession (2005), the year of granted concession (2013) and most recent (2021). The data is obtained from the land cover maps shown in Fig. 5.

Due to these seasonal inconsistencies, it is favourable to investigate on the natural areas collectively, when aiming to evaluate the impact of the LNGP on the environment. Taking wetland, sparse and dense vegetation together, it can be observed that they diminished by 19 % from 2005 to 2021. Therefore, the landcover classes dense and sparse vegetation and wetland respond to the increase of unvegetated areas with an overall decrease in their areal extent. This decrease constitutes a loss of natural and vegetated areas. The landcover water does play a minor role when observing land cover change in the concession area, as it makes up a share of max. 0.23 % (in 2013) throughout the years.

To better understand the introduced land cover change the following Fig. 7 provides a land cover change map. The map compares the year of granted concession (2013) with the most recent year (2021), to observe what land cover was replaced by one another. To reduce the number of classes

and enhance the readability of the map, the dense and sparse vegetation class are combined in one class. Consequently, changes from sparse to dense vegetation are not depicted. Overall, 2,808 ha of the total area has undergone change, whereas the remaining 4,270 ha did not change or solely altered from dense to sparsely vegetated areas. A total of 1304 ha vegetation did change to unvegetated area of which most are unambiguously identified as facilities and infrastructure of the LNGP. This outlines the most popular land cover change in the area and time investigated. Furthermore, 657 ha of vegetation turned to wetland and a total of 304 ha of unvegetated area became vegetated again. Finally, the wetland areas changed to an extent of 195 ha to unvegetated areas and 115 ha of it became vegetation.

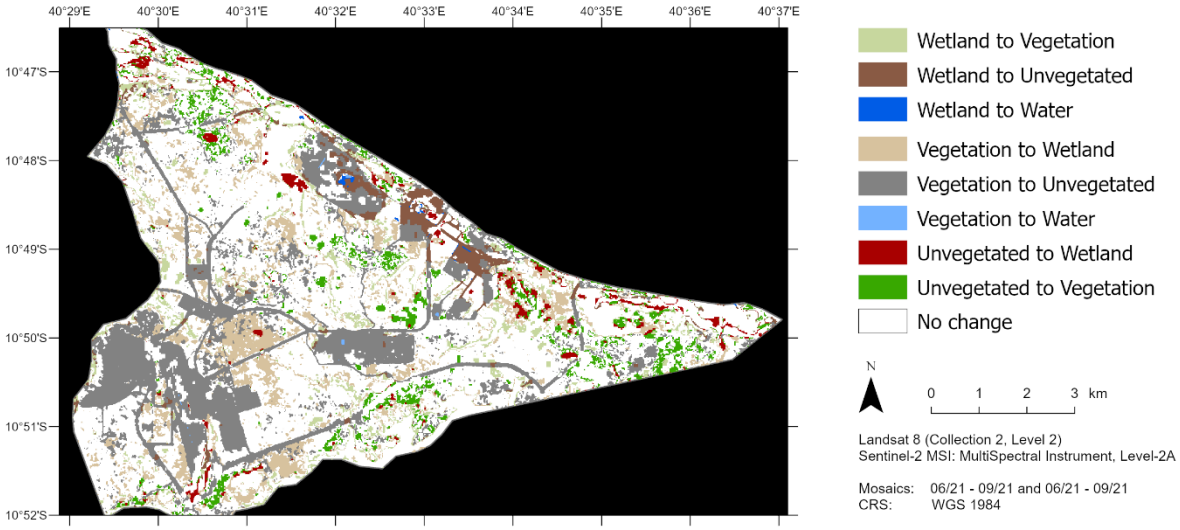


Fig. 7 LNGP case study: Land cover change map of the concession area comparing the years 2013 and 2021. The land cover change map shows which land cover class was substituted by one another between 2013 and 2021, indicating a large share of vegetated areas being dispelled by unvegetated areas (grey areas).

NDVI time series

To further inspect the effect of the LNGP development on the local vegetation, a NDVI time series was added to the analysis. Fig. 8 depicts this time series from 2013 to 2021 with yearly time intervals and the corresponding NDVI maps for the starting year 2013 and end year 2021. The NDVI is an index used to measure the presence and healthiness of vegetation, as introduced in Tab. 4. The colour scale of the NDVI maps indicates healthy, dense vegetation with dark green, unvegetated areas with white and the presence of water in red and can help the map reader quickly identify particularities in vegetation covers. The index is used in this research to identify and measure the loss of vegetation over time in the case study areas by observing the expansion of project related facilities and infrastructure. The NDVI map for 2021 does distinctively indicate the facilities and infrastructure which have been constructed for the LNGP. On the other hand, the map for 2013 shows the absence of any facilities and shows a majorly vegetated land surface. In the course of the research, an NDVI map for each year of the NDVI time series was created. From the analysis it was deduced that major construction activities started in 2018 whereas before just minor alterations on the land surface could be observed. Nevertheless, the time series for the mean NDVI shows no directional downward trend. However, the years 2020 and 2021 show the lowest values of 0.55 and 0.52, indicating the construction and expansion of the LNGP. The values of the min. NDVI are fluctuating disparately. Albeit the highly fluctuating values the min. NDVI shows a downward trend over time from (- 0.24) in 2013 to (- 0.62) in 2021. Nevertheless, to see if more distinct trend patterns develop over time both trendlines would need to be continued.

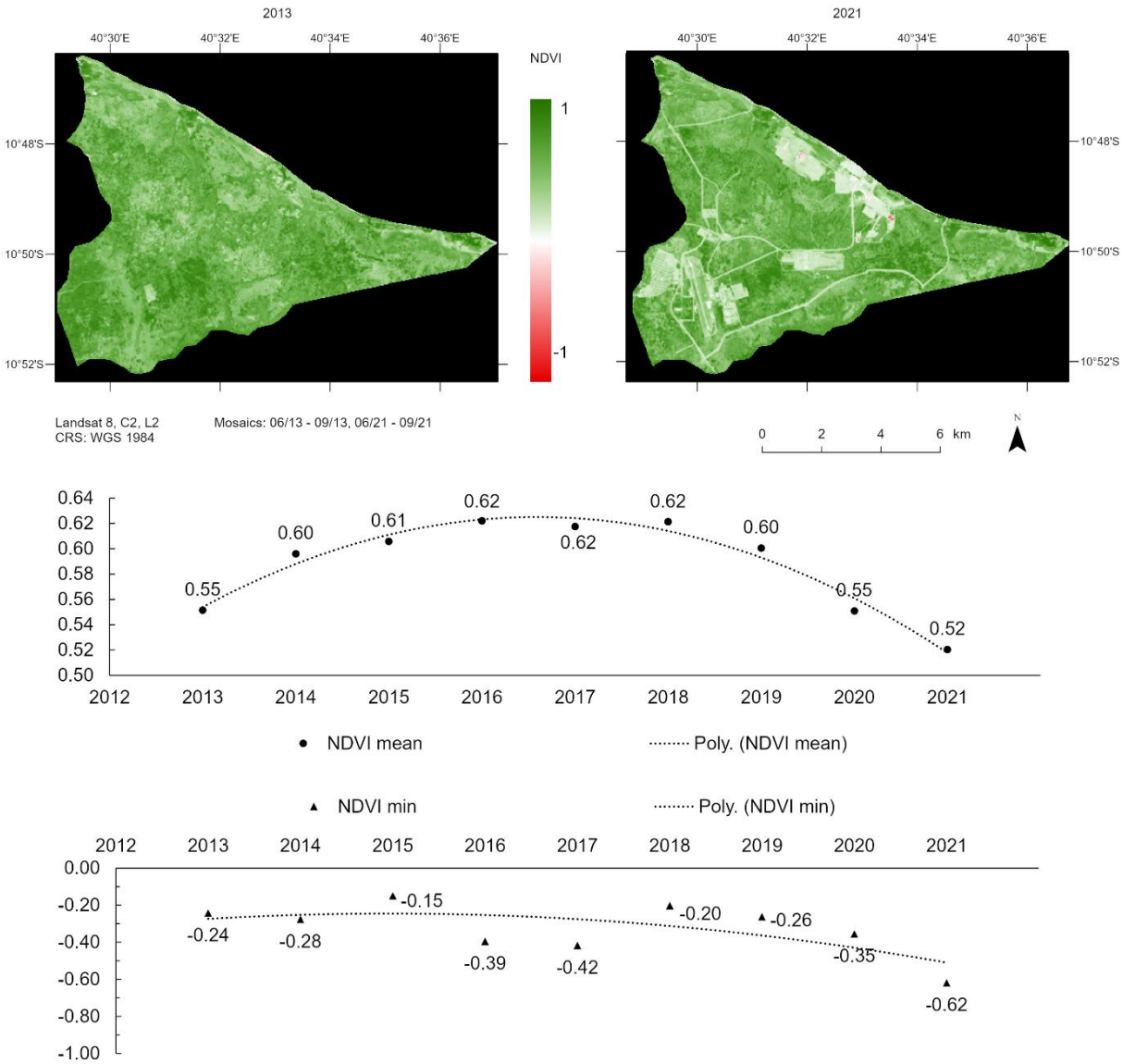


Fig. 8 LNGP case study: NDVI maps of the concession area for the years 2013 and 2021, indicating the intensity of present vegetation covers. Deduced from the maps and plotted in the graphs below are the mean and min. NDVI in yearly intervals from 2013 – 2021, indicating a slight decrease of vegetation from 2018 – 2021.

To enhance the interpretation of the NDVI time series a NDVI differencing map was created, see Fig. 9. The NDVI differencing map compares the start and end year of the preceding NDVI time series. For the legend of the differencing map the values are grouped based on a natural breaks classification. As a result, two vegetation loss categories are introduced, indicating severe losses. On the contrary, the gain of vegetation was significantly lower. The map shows distinctively that the vegetation loss follows the contours of the constructed facilities and infrastructure, proving a high interrelation between the loss of vegetation and the development of the LNGP.

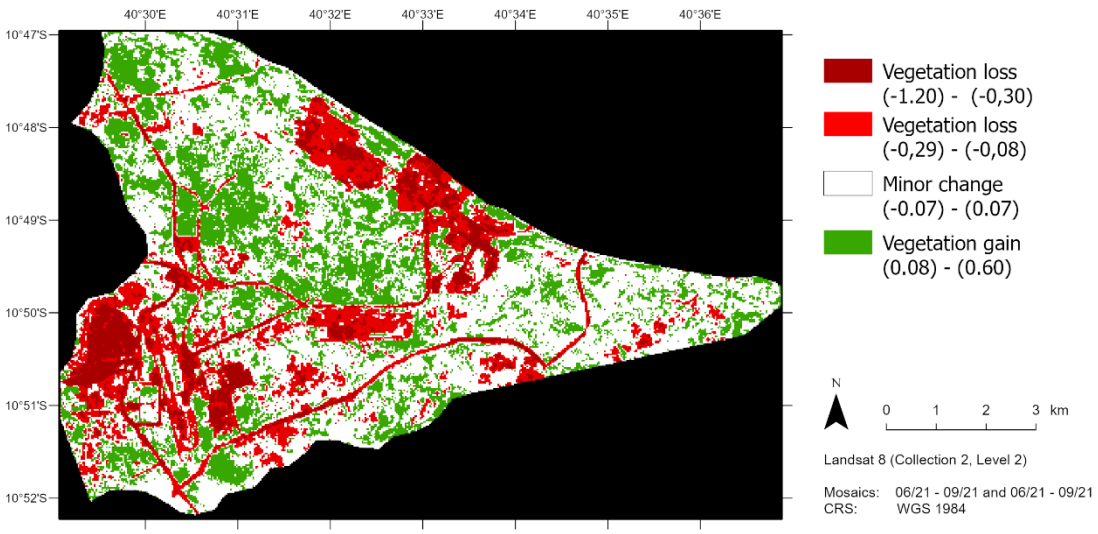


Fig. 9 LNGP case study: NDVI differencing map of the concession area deducting the NDVI values of 2013 and 2021, indicating the highest vegetation loss in the areas where facilities and infrastructure for the LNGP were constructed.

As a result of the analysis, it can be concluded, that the implementation of the LNGP, introduced significant land cover changes and vegetation losses in the area of interest. The construction of the facilities and infrastructure for the project caused the vegetated landcover to decrease substantially. The healthiness of the vegetated areas was measured with an NDVI time series of the mean and min. NDVI which did not show an explicit directional trend. Yet, the min. NDVI did decrease notably over the last three years.

4.1.2. Balama Graphite Project.

The BGP is subject to the second case study. Obtaining quality mosaics for the area of the BGP site was much easier than for the LNGP site. The frequency of clouds was much less and did not limit the creation of mosaics. Therefore, the time series for this case studies was conducted with the initially planned temporal parameters. The year 2005 constitutes the time before the extraction project, in 2013 the concession was granted, 2018 represents the year were production started and 2021 constitutes the most recent year. For each year a mosaic was created for the time span of May to July.

Land cover time series

The mosaic for 2005 is based on Landsat 5 images, for 2013 Landsat 8 images were used and for 2018 and 2021 Sentinel-2 images were used. Other than in the region of the LNGP, aridity in the summer months does have a significant effect on this study area. The aridity was the major problematic factor while composing the mosaics for this region but did not cause any adaptations. The results of the land cover classification are shown in Fig. 10.

The first year of the time series (2005) does show the concession area without any impacts of the later planned project. The year 2005 did prove to be difficult in detecting any distinct shapes, representing particular objects. As in the LNGP area, settlements are very small and scattered making it difficult to map them as coherent patches of unvegetated areas. Yet, the settlement Ncuide in the north of the map is identified unambiguously as a patch of unvegetated area and also indicated in the map. The rest of the map is majorly covered with sparse and dense vegetation. Sparse vegetation was

majorly identified as areas with grassland and crop but also sparse scrub/shrub. Densely vegetated areas contain trees, woodland and dense scrub/shrub.

In the year of granted concession 2013 the settlement Maputo and the street in the south of the concession area are much more distinct and can be identified on the map unambiguously. Particularly, the road N14 leading through the concession area is detected very well. Sparse and dense vegetation show and overall similar pattern in 2013 compared to 2005.

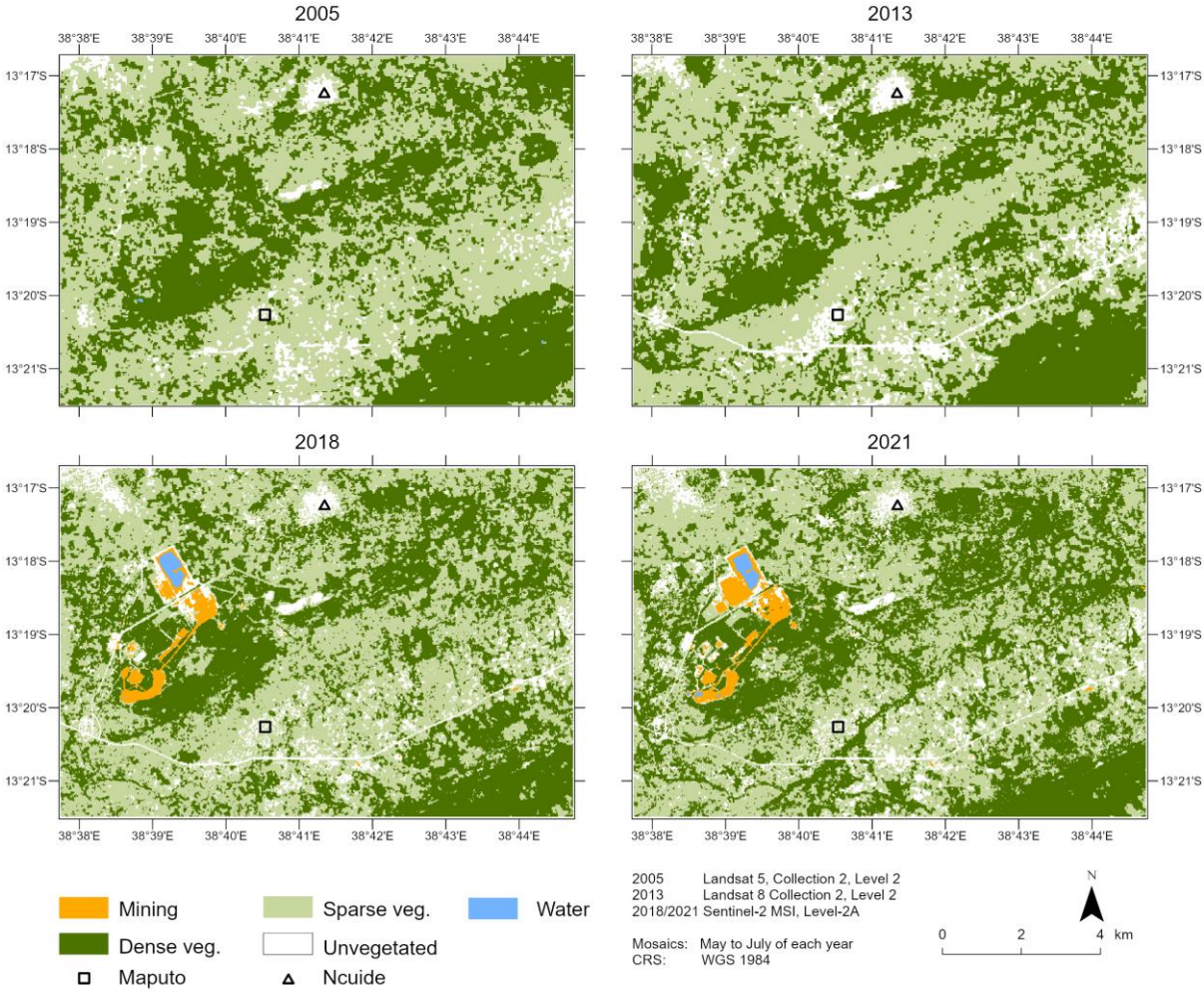
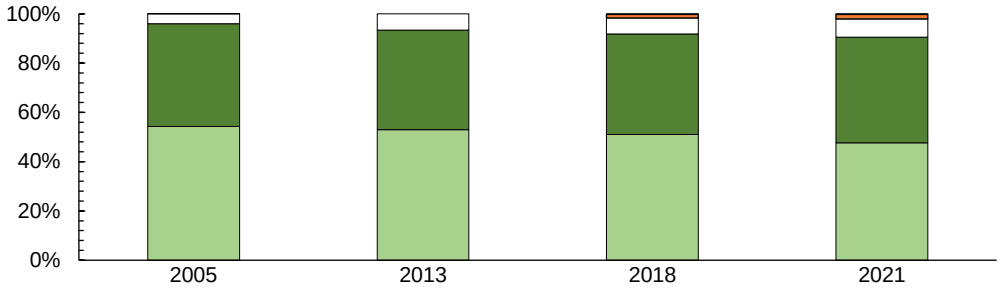


Fig. 10 BGP case study: Time series of land cover change in the concession area, before the granted concession (2005), the year of granted concession (2013), start of production (2018) and most recent (2021). 2018 and 2021 show the contours of the established mining areas. 2021 shows an increase of the tailing area next to the pit water. The settlements Maputo and Ncuide are indicated in the map with a black rectangle and triangle.

From the year 2013 to 2018, a huge expansion of the extraction project has taken place. The land cover map shows the contours of the mining pits in the south of the facility distinctively. In the north, we see the tailing dam and next to it, the tailing storage. The street leading from the mining pit towards the northeast located processing facilities, and a large part of the processing facilities, are identified as mining by the supervised classification, rather than unvegetated areas. This can be explained with the fact, that the transportation and processing of the mining ore causes dust of the mining ore to spread and eventually settle on the road and surrounding objects.

The most recent year 2021 does not reveal many changes in the concession area when comparing it to 2018. However, the tailing storage area next to the tailing dam expanded considerably. We also see that in the mining pit, pit water was identified. Furthermore, the settlement Maputo does seem to diminish from 2013 to 2021. The distribution of dense and sparsely vegetated areas remains the same. 2021 does have a higher share of dense vegetation which is attributed to seasonal differences between 2018 and 2021. The share of the different land cover classes on the total study area is shown in Fig. 11. Apparently, the construction of the extraction site did have a low areal impact on the observed area, so that a change in the proportional shares of each land cover class is barely visible, as it is depicted in bar chart in Fig. 11.



Class area [ha]	2005	2013	2018	2021
Water bodies	2	0	33	33
Mining	0	0	156	193
Unvegetated	444	723	711	822
Dense vegetation	4,600	4,464	4,501	4,724
Sparse vegetation	5,986	5,844	5,631	5,259
Total	11,031	11,031	11,031	11,031

Fig. 11 BGP case study: Share [%] and areal extent [ha] of the land cover classes in the concession area over time. The numbers are obtained from the land cover maps shown in Fig. 10 and show the increase of the mining area in the concession area.

Nevertheless, Fig. 10 does show an increasing trend towards more mining and unvegetated areas in the concession area. Adding up both land cover classes their share increased from 4 % in 2005 to 9 % in 2021 of the total area of 11,031 ha. Nevertheless, from 2005 to 2021, vegetated areas decreased by only 0.06 % in the concession area. The time series would have to be continued to observe more dramatic changes in terms of land cover change.

To provide a better insight into the actual land cover change, I created a change map, depicted in the following Fig. 12. To enhance the interpretability and readability of the change map, sparse and dense vegetation were merged into one vegetation class. At the first glance, the change map does provide the impression that the largest share of change accounts to the change from vegetation to mining. When examining the data on which the map is based, we can conclude that this is a wrong impression. The change from vegetation to mining accounts to an area of only 186 ha. A total area of 537 ha changed from being vegetation to unvegetated and 431 ha underwent the opposite change. In total, 1,174 ha of the 11,031 ha big concession area did undergo change, whereas the remaining 9,857 didn’t change or solely changed from sparse to dense vegetation.

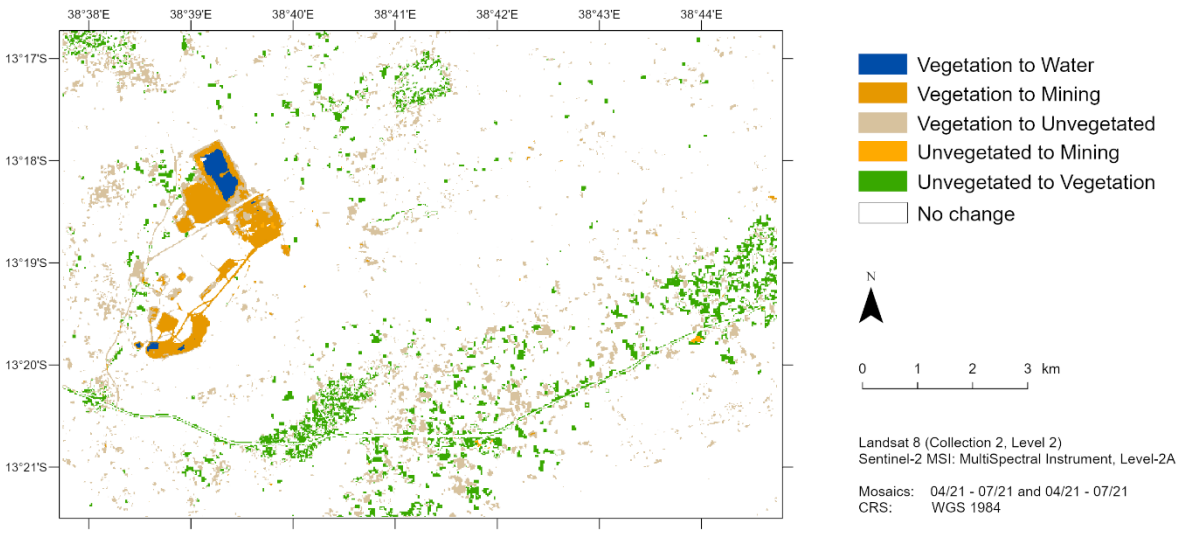


Fig. 12 BGP case study: Land cover change map of the concession area, comparing the years of granted concession (2013) and most recent (2021). The map shows what land cover class was substituted by one another, indicating the change from vegetated areas to mining areas.

NDVI time series

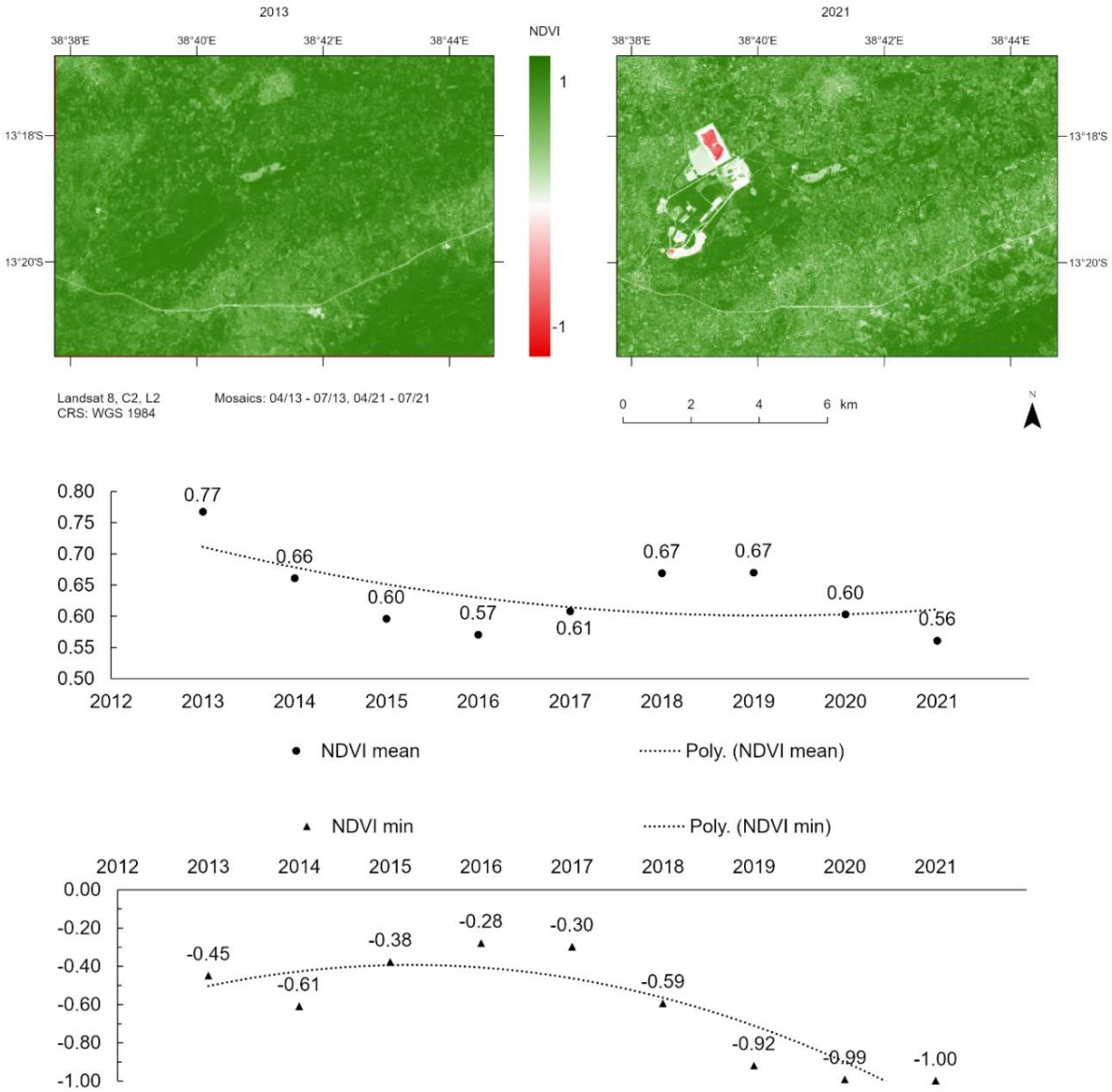


Fig. 13 BGP case study: NDVI maps of the concession area for the years 2013 and 2021, indicating the intensity of present vegetation covers. Deduced from the maps and plotted in the graphs below are the mean and min. NDVI in yearly intervals from 2013 – 2021, indicating a decreasing trend of vegetation from over time.

The NDVI time series of the concession area for the BGP is depicted in Fig. 13 and covers a time span from 2013 to 2021 in yearly intervals. The start and end year of the time series are added as NDVI maps to Fig. 13. The NDVI map of 2021 does outline distinctively the contours of the mining area and the street with corresponding NDVI values of (+/-) 0. The tailing dam is clearly recognisable as the red patch in the NDVI map of 2021, with values of approximately (- 1). The trendlines for both, the mean and min. NDVI value, for the area shows an overall decrease from 2013 - 2021. The decrease indicates a correlation between the development of the mining site and the degradation of the local vegetation. The mean NDVI does fluctuate over the years but decreases from 0.77 in 2013 to the

lowest value of the whole time series, 0.56 in 2021. The min. NDVI fluctuates as well, but latest with the presence of the tailing dam from 2019 balances out on negative values smaller than (- 0.90).

To support the interpretation of the NDVI time series the NDVI differencing map was produced, comparing the values of 2013 with the values of 2021. The differencing map is depicted in Fig. 14. The differencing map shows a distinct contour of the extraction site which is attributed to the loss of vegetation. Small spots of vegetation loss can also be observed in other locations but are randomly spread. The vegetation gain is fairly equally spread over the concession area.

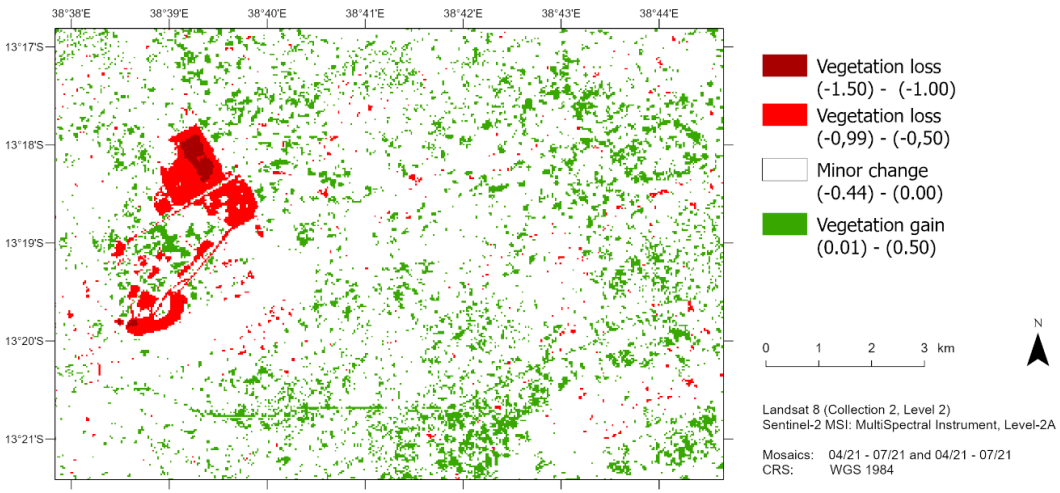


Fig. 14 BGP case study: NDVI differencing map of the concession area, deducting the NDVI values of 2013 and 2021. The map shows that the highest vegetation loss occurs in the areas of resource extraction.

4.2. Accuracy assessment.

The accuracy assessment was conducted for the land cover classification in GEE and is embedded as a default in the model workflow. The results are presented in this chapter. Overall, the assessment of the land cover classification for the LNGP and BGP case study provides high accuracies as listed in Tab. 6. For each case study and each year of the time series, accuracies higher than 90 % were achieved.

Tab. 6 Overall accuracies of the LNGP and BGP case studies, in percentage for the land cover classifications of 2005, 2013, 2018 and 2021.

Overall accuracy		2005	2013	2018	2021
LNGP	%	96	94	-	94
BGP	%	94.5	93.5	92.8	92.8

To better understand the overall accuracies, an error matrix of the land cover classification for 2021 for each of the case studies is provided in the following. The results section contains one error matrix per case study, depicting only the year 2021, being exemplary for the other years of the respective time series. For each year of the time series, an error matrix exists, as they are automatically created as part of the workflow model. The matrices for each year of the time series are not all presented in this section as the confusion between the classes is comparable throughout the years. To keep the main part of this research concise, the matrices of the other years can be found in Appendix 9.2. The error matrices show to what extent the test points of each class were confused with different land covers during the classification process. The error matrix for the LNGP and BGP case

studies are depicted in Tab. 7 and Tab. 8, respectively. Based on the error matrix, the users and producers' accuracy were calculated. The user's accuracy is calculated by dividing the number of correctly classified testing points of a class by the total of testing points that were eventually attributed to that class. A test point was correctly classified if the land cover class assigned to it matches the land cover class assigned by the classification algorithm. Thereby, the user's accuracy indicates the reliability of the testing point being correctly classified on the ground. The producer's accuracy is calculated by dividing the number of correctly classified test points by the total number of test points which were collected for each class respectively. With the producer's accuracy we can verify the accuracy of the land cover classification process. The producer's accuracy is of higher importance in this research, as it provides us an estimate on how well the classification algorithm in the workflow model performs. For each land cover class, a number of 50 test points was collected.

The LNGP study area was divided into the classes' wetland, dense vegetation, sparse vegetation, unvegetated and water. Tab. 7 shows the error matrix of the land cover classification for the year 2021. Each class did achieve high producer's accuracies, in most cases exceeding 90 %. This indicates high accuracy and that there was limited confusion between the land cover classes. In Tab. 7, we see that the class wetland was confused the most with densely vegetated areas and, to a lower extent, with sparse vegetation. This confusion is very likely and depends highly on the current state of the wetland areas. If vegetation is temporarily flooded, it is most likely classified as wetland, whereas at times, the region is drier; it will be classified as dense vegetation or even sparse vegetation. This error can be explained by the dynamic character of the land cover wetland as it was described in section 4.1.1. Depending on the aridity and water level, the spectral signature of a particular piece of land will alter. Land containing sparse vegetation might be confused with unvegetated or dense vegetation as the spectral signatures will resemble one another with changing aridity.

Tab. 7 LNGP case study: Error matrix of the testing data, showing the confusion between the land cover classes as a result of the classification process in 2021.

Class/Class	Wetland	Dense veg.	Sparse veg.	Unveg.	Water	Total	Users' accuracy
Wetland	43	0	0	0	0	43	100%
Dense veg.	5	48	0	0	0	53	91%
Sparse veg.	2	2	47	3	0	54	87%
Unvegetated	0	0	3	47	0	50	94%
Water	0	0	0	0	50	50	100%
Total	50	50	50	50	50	250	
Producers' accuracy	86%	96%	94%	94%	100%		
Overall accuracy	= (43+48+47+50)/250 = 94.0%						

Moreover, there is a regular confusion between dense and sparse vegetation and sparse vegetation and unvegetated. The confusion between sparse vegetation and unvegetated can be traced back to the scattered structure of settlements but also to the healthiness of the plants. The latter also has a significant impact on the confusion of dense and sparsely vegetated areas. The more arid a certain patch of vegetation is, the drier and less healthy it will be the vegetation. For vegetation, this has a significant effect on their spectral response curve, making it very likely to be confused within the vegetation classes dense and sparse or even with the class unvegetated. This is because the lower the water content, the less distinct the spectral response curve of vegetation. Consequently, macroclimatic conditions or weather trends in the respective year have had a considerable impact on the classification of the earth's surface in the study areas. The confusion between sparse and dense vegetation and unvegetated areas account for both case studies. In both study areas, the class water reaches the highest accuracies. As shown in Fig. 4, the spectral reflectance curve of water is very distinct, which makes it easy to distinguish from the other classes. Nevertheless, with the occurrence

of pit water the accuracy of water decreased slightly. The pit water contains much of the mining ore, causing confusion between the classes water and mining as it can be observed in the error matrix for the BGP in Tab. 8. Instead of the wetland class, the BGP area contains a mining landcover class. The class wetland was disregarded entirely in the BGP case study as it does not represent a typical land cover of the area. Overall, the error matrix for the BGP site shows very high producers' accuracies for each class. Whereas the mining pits, tailings and waste were to a large extent allocated to the mining class some confusion with the unvegetated areas is observed. The unvegetated areas do represent bare ground containing gravel, sand and soil which is also contained to larger or smaller extent in the mining ore in the pit areas. Depending on the purity of the ore, a mining area might also be identified as bare ground and be classified as unvegetated.

Tab. 8 BGP case study: Error matrix of the testing data, showing the confusion between the land cover classes as a result of the classification process in 2021.

Class/Class	Mining	Dense veg.	Sparse veg.	Unveg.	Water	Total	Users' accuracy
Mining	45	0	0	0	1	46	98%
Dense veg.	0	49	5	1	0	55	89%
Sparse veg.	0	1	44	4	0	49	90%
Unvegetated	5	0	1	45	0	51	88%
Water	0	0	0	0	49	49	100%
Total	50	50	50	50	50	250	
Producers' accuracy	90%	98%	88%	90%	98%		
Overall accuracy	= (45+49+45+49)/250 = 92.8%						

Overall, the accuracy assessment provided results which are more than satisfactory. However, the results need to be treated cautiously. This is discussed in the next chapter.

5. Discussion

This research offers a replicable proof of concept approach to measure land use change over time using open access imagery and data processing. This was done by creating a model to map and identify change in areas of resource extraction and testing that model on two case studies. In doing so, I aimed to map the areal extent of land cover and environmental change resulting from extractive activities at the source of energy transition supply chains. The method will be useful for the follow up stage of EIAs to monitor and ensure project initiators, in fact hold the promises made through the EIA process. Furthermore, NGOs and concerned parties can use the proposed workflow to study the effects of extractive activities in vulnerable regions and raise awareness. To enhance the accessibility and cost-efficiency of the model, it was created by using solely free of charge satellite data and cloud computing. The reproducibility and scalability of the model were tested by applying the model to two case studies. Additionally, I tested the model's reliability by conducting an accuracy assessment for the land cover classifications. Therefore, the key aspects of this research were:

1. Building a reproducible and scalable model,
2. Testing it on two case studies and
3. Conduct an accuracy assessment for the obtained results.

Each aspect was addressed by one of my research questions and is discussed in a separate sub-chapter in the following.

5.1. The accessibility, reproducibility, and scalability of the created workflow model.

The first sub-question stated: ”What is an example of a reproducible, scalable model to map change in areas of resource extraction using open access satellite data and cloud computing?”. This question identifies the first objective of my research, as I aimed at developing an example of an applicable workflow model to map change in areas of resource extraction. In this sub-chapter, I discuss the accessibility of the workflow model, the reproducibility and scalability, and lastly, potential enhancements and recommendations.

I createed the example model using GEE as a cloud computing platform and open access satellite data. The model uses common methods to create time series and conduct change analysis, as introduced in section 3.2. The GEE code editor was used to create the time series and conduct change analysis. The GEE code editor works with the programming languages JavaScript and Python. In this case several scripts were written using JavaScript. The scripts are based on commands and algorithms from GEE Guides and JavaScript libraries. To create the workflow model, the algorithms and commands were adapted in the code editor by the researcher and compiled in the GEE. GEE and the satellite data are free to use, meaning even those with limited resources can reuse the scripts and use the platform and data presented here. Moreover, using GEE as a cloud computing platform prevents the need for an end device with high computational power and the storage of large amounts of data. Even though GEE has existed since 2010, a study by Wagemann et al., (2021), showed that the prevailing mode of data handling in Earth data science is still to download the required data onto local machines and process it locally with a combination of programming and desktop-based software. I argue that GEE could be a solution for the future to avoid the need for these working steps and high performance hardware if accessible workflows are provided. However, despite the increased accessibility of the workflow technical know-how, basic equipment and a stable internet connection to apply the model is still required. Required technical know-how involves a basic understanding of the applied EO data analyses techniques and programming language. To conduct analysis with the workflow in a region of interest, the following adaptions must be made in the scripts:

1. Determine the region of interest,
2. Define the land cover classes and subsequently collect training and testing data, respectively,
3. Determine temporal parameters such as the time series year and the time span to define what satellite imagery to classify, based on the seasonality of the region and location specific decisions.

The scripts indicate how and where these adjustments must be made. If desired, more exhaustive adjustments can be made to the scripts, such as using a different classification algorithm. The selection of classification algorithms available in GEE is limited however, a skilled programmer could write their own algorithms to meet their needs. In conclusion, the workflow proved to be reproducible and scalable in different locations and for different study areas.

To conduct analysis with the proposed workflow model, all required scripts are made available on GitHub. GitHub is a webpage that provides developers a space to store, manage and share their scripts with interested parties. Information on how to use and access the scripts on GitHub is shared in Appendix 9.1. By sharing the workflow and required scripts, my research combines cloud based services and earth data science and provides an alternative to desktop-based solutions and the downloading of enormous amounts of data. Thus, investigating change in areas of resource extraction and identifying possible implications arising from extractive activities is made more accessible. Stakeholders such as public and non-profit organisations or concerned parties can use the workflow model to analyse and use the results for EIA follow-ups. This can provide valuable insights into the actual effects extractive activities have on their surroundings, as the EIAs provided by the extractive companies are often weak. Moreover, even though extractive companies are obliged to conduct a

follow-up, it is often being neglected as it requires substantial resources in terms of money, time, and expertise (Arts & Morrison-Saunders, 2012; Marshall et al., 2005). The developed model provides an opportunity to counteract the neglect of monitoring the actual effects of mines on their surroundings.

To further prove the reproducibility and scalability of the workflow, it was tested by applying it to two different case studies. Both case studies are located in Cabo Delgado in Mozambique but differ in their spatial extent, location and thematic scope. The study area of the LNGP is smaller, located next to the seaside and required the evaluation of different land cover classes than in the BGP study area. The location next to the seaside required the adjustment of the initially planned time series intervals as described in 4.1.1 due to cloud cover issues.

In summary, the compiled model provided reliable results in the two different case studies by altering a few parameters in the scripts. Therefore, the workflow can be scaled and reproduced in any study area subject to land cover change and environmental degradation. Applying the workflow does not require any financial means; the storage of vast data amounts or high computational power which makes it more accessible and open for diverse stakeholders. While the model provides an opportunity to analyse land covers and environmental change, it is limited in providing sophisticated layouts and diagrams for the results, that is why ArcGIS or QGIS are recommended to use for map making. Future research could focus on providing required scripts and working steps which enable the output of layouts and diagrams within GEE. This would outline a valuable extension of current the workflow.

Providing the workflow model was initially motivated by the anticipated threat through an expanding extraction frontier caused by the material demands for the global energy transition (IEA, 2021b; Hund et al., 2020). While developing and applying the model, it was increasingly apparent that it is also suitable to analyse the change in study areas not subject to extractive activities but to similar challenges such as land cover change and environmental degradation. The model could also be used to map the cumulative impacts of extraction activities in one particular region or of one particular commodity by accumulating the results of different and multiple case studies. The neglect of cumulative impacts of mining activities in one particular region or of one particular commodity in current research was criticised by Werner et al. (2019). Cabo Delgado might be an interesting subject for such a cumulative analysis, as there are 8 ongoing exploration projects for graphite extraction (Brown et al., 2021). Investigating all these locations together and studying their cumulative impact is an interesting possibility for future research.

5.2. The extent of change monitored at the source of energy transition supply chains.

The second objective of my research was to apply and test the workflow and monitor and reveal change resulting from extractive activities at the source of energy transition supply chains. Subsequently, the second sub-question asked, “What extent of change can be monitored in the case study areas over time by applying the workflow?”. With this objective, I aimed to emphasise the potential implications resulting from the resource extraction for a sustainable energy transition. The case studies BGP and LNGP represent two extraction projects, extracting graphite and natural gas, two resources central to the energy transition.

The results for the LNGP showed a strong impact of the project on the local natural areas such as wetlands and dense and sparse vegetation. While natural areas diminished, the share of unvegetated areas increased significantly by 116 %, indicating a significant loss of vegetation and natural areas. In the LNGP case study, a share of 49,6 % of all the land cover change from 2005 to 2021 was attributed to the alteration from natural to unvegetated areas. This outlines a severe impact of the development of the LNGP on its environment. While investigating the environmental impact by building yearly time series for the mean and min. NDVI from 2013 to 2021, a severe deterioration of vegetation could not be observed. However, the two lowest mean NDVI values of 0.55 and 0.52 appear in 2020 and

2021. In 2020 and 2021, major construction activities took place in the area. It is likely that the low mean NDVI values of these two years are a result of the intensification of the construction activities. The LNGP has not even started production yet, and it is planned to do so in 2024. Due to attacks and political instability in Cabo Delgado the company TotalEnergies was already forced to stop its activities on the project site and declare force majeure (TotalEnergies, n.d.). The company announced to resume the project in 2022, but the situation remains unclear due to remerging attacks and political unrest in the region. However, if the project is continued, further changes in the study area can likely be monitored and identified. In this case, it is of high importance and value to continue the time series in the upcoming years to continue monitoring the project's impact on its surroundings.

Investigating change in the concession area of the BGP could determine the impact on land cover change and the environment. The implementation of the BGP instituted the construction of 193 ha of mining area between 2013 and 2021. Formerly, there was no mining area, and 96 % of the area classified as mining area in 2021 was previously sparse or densely vegetated. Additionally, a substantial part of the project-associated infrastructure and facilities have been classified as unvegetated and need to be attributed to the development of the project. Before the implementation of the BGP, 444 ha of the concession area were classified as unvegetated. This value increased to 822 ha in 2021. This indicates that the development of the BGP did have an impact on the environment causing land cover change by setting up supportive infrastructure and facilities.

The results for the NDVI time series showed a clearer trend than in the area of the LNGP. For both the mean and min. NDVI, a clear downward trend was observed as depicted in Fig. 13. The mean NDVI in 2013 was 0.77 and decreased to 0.56 in 2021. In the same time period, the min. NDVI decreased from – 0.45 to -1. The development of the BGP site involved the construction of a tailing dam. The NDVI values for water are very low, close to -1. Therefore, the trend observed for the NDVI time series is heavily affected by the tailing dam's presence, which must be considered when interpreting the results. To support the NDVI time series, the NDVI differencing map depicted in Fig. 14 was constructed. This map explicitly indicates the loss of vegetation in the areas of resource extraction and tailings. The BGP is a graphite extraction project and part of a vastly developing graphite extraction frontier in the district of Cabo Delgado. Continuing the time series for the project together or adding other graphite extraction projects in the region, such as the Ancuabe Graphite Project from Graphit Kropfmuehl GmbH, would provide valuable insight into the cumulative impacts of these projects on their surroundings. Investigating on these cumulative effects of extracting graphite in Cabo Delgado outlines an exciting opportunity for future research.

This research provides an approach to investigate the land cover change and environmental implications of extractive activities at the source of energy transition supply chains. The energy transition aims at lowering greenhouse gas emissions in the final consumption of sectors such as households, mobility, industry to combat global warming. While technologies such as storage systems and solar panels do contribute significantly to limit the emission of carbon to the atmosphere, the extraction of the required resources has, under current practices, a severe impact on the environment, biodiversity and land cover. By conducting this research, I aimed to emphasise these trade-offs at the source of energy transition supply chains, posing a contradiction to the efforts for a sustainable energy transition (Lèbre et al., 2020).

While investigating the land cover change in the areas of interest, it was recognised that societal implications such as demographic effects and resettlements could be examined by applying the workflow. For example, the concession areas of both extraction projects are adjacent to conurbations, as shown in Fig. 2. The conurbations are the city of Balama next to the BGP and Palma next to the LNGP. While analysing the overall study areas, it was observed that these cities grew considerably with the development of the projects. Furthermore, Fig. 10 of the BGP site shows an increase in vegetation over time in the region of Maputo village, which is indicated on the map. In the EIA from

the BGP, Maputo is listed as a settlement subject to resettlement measures (EOH, 2014). This likely indicates the abandonment of the village and a return of vegetation. This conclusion would need to be validated by local citizens or regional experts. Also, the settlements within the LNGP concession have changed, such as Quitupo, which is indicated on the land cover maps in Fig. 7. According to the resettlement plan, the resettlement measures for Quitupo took place 20019/2020 (ENI, 2014b). Focussing on societal implications such as the resettlement measures from a spatial analysis perspective provides an opportunity for future research. Notably, current research about extractive activities in Cabo Delgado majorly focuses on societal impacts such as resettlement measures and the loss of farming and fishing livelihood and could be supplemented with the spatial analysis perspective. Additionally, Cabo Delgado is subject to violent attacks provoked by jihadist extremists in the region. Some publications have already made a connection between the occurrence of armed conflicts and extractive activities in the region (Macuane et al., 2018; Mate, 2021). Approaching this topic with spatial analysis and mapping the locations of armed conflicts together with the developments of mining projects over time could provide further insights into the effects associated with extractive activities.

5.3. The reliability of the model's land cover classification.

After the first two sub-questions have been treated, the third sub-question aimed at assessing the land cover classification's accuracy, to test the reliability of the results. The research question asked "How reliable are the results obtained from a combination of satellite data and cloud computing in monitoring the change?". The results of the accuracy assessment proved to be reliable, as it is laid out in 3.3.

According to Foody, (2008) there is a widely used minimum level as target accuracy in current remote sensing research of 85 %. To obtain comparable accuracies, I had to adapt the land cover classes several times during the research. Initially, I aimed to distinguish between 8 land cover classes, including the differentiation of mining pit and tailings or grassland and crop. Without comprehensive manual alterations of the resulting land cover maps, a thematic resolution of 8 land cover classes would have fallen far below accuracy values of 85 %. Particularly problematic proved to be a differentiation between built-up areas, bare ground, grassland and agriculture. The differentiation between built-up and bare ground was challenging due to the structure of the settlements. The settlements are made up off scattered tiny houses connected by pathways made of permeable surfaces, with bushes and trees between them. The differentiation between grassland and crop was challenging due to the spectral resemblance of grass and crops. Without adding further spectral unmixing techniques or manual alterations as a post-classification step to the workflow, no reliable classification results could have been obtained. Obtaining acceptable accuracies with the developed workflow and the given time frame in my thesis was solely possible by reducing the thematic resolution to five classes per case study. Eventually, the accuracies shown in Tab. 6 are all above the threshold value of 85 % and mostly above 90 %. However, even though the accuracies are high, they need to be interpreted cautiously. The accuracy assessment process was conducted using the same imagery used for the classification process. Moreover, the accuracy assessment did not involve the expertise of local parties or individuals, who could confirm the analysis results unbiased. Conducting accuracy assessments for time series does pose major limitations when aiming for the involvement of local experts. In most cases, historical land cover maps and auxiliary data on the past land cover would be required. Digitalised land cover maps for Mozambique were available for this research. The respective maps were made available to this research as it feeds into the inFront research project at the University of Utrecht. I did aim to include these maps in the accuracy assessment, but the maps provided the land cover classification results for Cabo Delgado on a much broader scale. Objects such as extraction locations or villages were not captured, making the results of my maps incomparable to the maps provided as I investigated a much higher level of detail. After all, the accuracy assessment

for my study was a key challenge and was conducted on a basic level as ground truth data was not accessible, and imagery with the highest resolution was already used for the classification itself.

Collecting ground truth data is requisite to improve the accuracy assessment and the classification process. During my research, the testing data was gathered based on a similar selection process to the training data. Even though spatial correlation was avoided, this approach does pose the risk of overoptimistic accuracy results as testing data is not collected independently but with a similar approach as the training data. This is also the reason for the high accuracies obtained during this research. Moreover, the selection of the testing data and training data and the accuracy results are highly dependent on the person collecting the respective data. The workflow model does provide support for collecting training and testing data by plotting the spectral reflectance curves of the collected data and by segmenting the image in spectrally similar areas, as introduced in 3.2.3 and Fig. 4. However, the collection of the data is dependent on a person who does introduce a high degree of potentially biased results. To conduct the accuracy assessment in a less biased way, it is best to use and collect ground-truth data and use it as a surrogate to the training points applied in this study.

It is hoped that this workflow be implemented by those who live near mining activities so ground truthing will be much easier to those communities.

6. Conclusion

This research outlines a proof of concept with the aim to demonstrate that open access satellite data and cloud computing can make an important contribution to engage on sustainability issues emerging at the source of energy transition supply chains. Therefore, the main research question was, “How may open access satellite data and cloud computing be used to monitor change in areas of resource extraction at the source of energy transition supply chains?”. The approach was to develop a reproducible, scalable and accessible model to construct remote sensing time series and prove its applicability in two case studies. The model represents a workflow using EO observation data analyses to map land cover and environmental change in areas of resource extraction over time. This workflow will be helpful to anyone who aims to monitor change over time. This workflow will be especially useful for those aiming to conduct follow up assessments set forth in the EIA process.

One of the conditions of this research was to ensure that the workflow would be reproducible, scalable and accessible. To ensure accessibility, the model is based solely on open access satellite data and cloud computing. GEE is used as a cloud computing platform; it is free of charge and avoids the need for high performance hardware. The combination of both makes the workflow model applicable in circumstances of low financial and technical means. To conduct analysis, technical know-how and a working internet connection are needed as GEE is cloud based and requires a basic understanding of the programming language JavaScript. All the required scripts to follow the workflow are open accessibly and free to use. By applying the developed model to two different case studies, it proved to be reproducible and scalable. A subsequent accuracy assessment of the land cover classifications was used to test the model’s reliability. The assessment results showed high accuracies for each investigated year of the constructed time series. However, the accuracy assessment was conducted without ground truth data or the expertise of local parties. This must be mentioned as a critical drawback of this research. The collection of the testing data was based on a similar approach as the training data, introducing the risk of biased, overly optimistic results. Using testing data based on field observations or consulting a local expert to verify the land cover classification maps is recommended. In conclusion, the workflow model proved to be reproducible, scalable, reliable and as accessible as possible, making it applicable for investigations with similar challenges, particularly in vulnerable regions with low financial and technical means.

Another condition of this research was to emphasise sustainability issues at the source of energy transition supply chains. In light of the pace and scale of the global energy transition, the associated resource demand, and the intensifying energy crisis, the energy transition was identified as a significant threat to the local environments of resource rich countries. This constitutes a contradiction in the global efforts for a sustainable energy transition. Therefore, the case study design aimed specifically at selecting study areas directly linked to the extraction of energy transition resources. The results obtained by applying the model to the case studies revealed land cover change and environmental degradation, which could be directly linked to the implementation of the extraction projects. By identifying and quantifying these changes, private and non-profit decision makers can develop and enhance plans to preserve, manage and restore adjacent land and nature and protect local communities. In this study, it was found that the implementation of the LNGP and the BGP caused local vegetation covers to decrease and introduced an increase of unvegetated and mining areas. In the locations of the case studies, vegetation health decreased and was monitored using a NDVI time series. However, both investigated extraction projects are in an early stage of development, and the areal extent of investigated impacts of the extractive activities have on their surrounding was small compared to the areal extent of the entire study areas. Consequently, it is recommended to continue the created time series for the chosen case studies. Additionally, the province of Cabo Delgado is a hot spot for extracting natural gas and graphite. The time series can be complemented by investigating the cumulative impacts of one or both of these commodities in the entire province.

By conducting this research, an example was provided for a reproducible, scalable and accessible workflow. Testing the workflow in two case studies, emphasised the environmental impacts of an expanding energy transition-induced extraction frontier and the workflow was found to be reliable. Hence, the model has the potential to contribute to a more sustainable energy transition by providing a practical workflow for public and non-profit stakeholders and concerned parties, to do regular EIA follow ups and conduct analysis to raise awareness with the results.

7. Acknowledgements

Writing this thesis was a lot of work in which I gratefully wasn't alone. For this reason, I would like to express my gratitude to everybody who did support and help me to finish this project.

Major thanks go to Britta Ricker (PhD) and Emilinah Namaganda (PhD candidate). Thank you both for supervising my thesis and helping me to progress and improve continuously. Major thanks go also to the 2nd reader Dr. ir. C.G.M. Kees Klein Goldewijk, who took the time to read my proposal and final thesis and provided crucial feedback to it.

I also want to thank the GEE team for providing helpful guides[1] which make the use of the platform very approachable. Furthermore, the availability of question-and-answer platforms such as gis.stackexchange[2] and the continuous and voluntary contributions there, were substantial for building all the required scripts in this research. Therefore, every user who keeps such networks of mutual support alive with answering questions and providing feedback has my deepest gratitude. I want to particularly thank the following users: Noel Gorelick[3], Justin Braaten[4], Daniel Wiell[5], Kevin Reid[6] and Rodrigo E. Principe[7].

[1] https://developers.google.com/earth-engine/guides
[2] https://gis.stackexchange.com/
[3] https://gis.stackexchange.com/users/90192/noel-gorelick
[4] https://gis.stackexchange.com/users/68792/justin-braaten
[5] https://gis.stackexchange.com/users/154371/daniel-wiell
[6] https://gis.stackexchange.com/users/39720/kevin-reid
[7] https://gis.stackexchange.com/users/23470/rodrigo-e-principe

8. Bibliography

Abbas, I. I., & Ukoje, J. A. (2009). Application of Remote Sensing (Rs) and Geographic Information Systems (Gis) to Environmental Impact Assessment (Eia) for Sustainable Development. *Research Journal of Environmental and Earth Sciences*, *1*(1), 11–15.

Agusdinata, D. B., Liu, W., Eakin, H., & Romero, H. (2018). Socio-environmental impacts of lithium mineral extraction: towards a research agenda. *Environ. Res. Lett*, *13*, 123001. https://doi.org/10.1088/1748-9326/aae9b1

Arts, J., & Morrison-Saunders, A. (2012). *Assessing Impact : Handbook of EIA and SEA Follow-up*. Routledge.

Banza Lubaba Nkulu, C., Casas, L., Haufroid, V., de Putter, T., Saenen, N. D., Kayembe-Kitenge, T., Musa Obadia, P., Kyanika Wa Mukoma, D., Lunda Ilunga, J. M., Nawrot, T. S., Luboya Numbi, O., Smolders, E., & Nemery, B. (2018). Sustainability of artisanal mining of cobalt in DR Congo. *Nature Sustainability 2018 1:9*, *1*(9), 495–504. https://doi.org/10.1038/s41893-018-0139-4

Belward, A. S., & Skøien, J. O. (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 115–128. https://doi.org/10.1016/J.ISPRSJPRS.2014.03.009

Brown, T. J., Idoine, N. E., Wrighton, C. E., Raycraft, E. R., Hobbs, S. F., Shaw, R. A., Everett, P., Deady, E. A., & Kresse, C. (2021). *World Mineral Production 2015 - 2019*. http://nora.nerc.ac.uk/id/eprint/529923/1/WMP_2015-2019_COMPLETE.pdf

CES. (2014). *ENVIRONMENTAL, SOCIAL AND HEALTH IMPACT ASSESSMENT Balama Graphite Mine*. http://www.cesnet.co.za/pubdocs/Syrah%20Balama%20Graphite%20Mine%20Eng%20CB290814/Syrah%20Draft%20ESHIAR%20Full.pdf

Charou, E., Stefouli, M., & Vasileiou, E. (2010). Using Remote Sensing to Assess Impact of Mining Activities on Land and Water Resources Environmental Hydrogeology View project Hydrogeology of mines View project. *Article in International Journal of Mine Water*. https://doi.org/10.1007/s10230-010-0098-0

Church, C., & Crawford, A. (2018). *Green Conflict Minerals: The fuels of conflict in the transition to a low-carbon economy IISD REPORT*.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Congalton, R. G., & Green, K. (2019). Assessing the Accuracy of Remotely Sensed Data : Principles and Practices, Third Edition. *Assessing the Accuracy of Remotely Sensed Data*. https://doi.org/10.1201/9780429052729

Du, Q., Gungor, O., & Shan, J. (2005). Performance evaluation for pan-sharpening techniques. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *6*, 4264–4266. https://doi.org/10.1109/IGARSS.2005.1525860

Edwards, D. P., Sloan, S., Weng, L., Dirks, P., Sayer, J., & Laurance, W. F. (2014). Mining and the African environment. *Conservation Letters*, *7*(3), 302–311. https://doi.org/10.1111/CONL.12076

Els, W., & Chelin, R. (2021). MOZAMBIQUE, CABO DELGADO INSURGENCY: EXTRAORDINARY MINERAL RESOURCES AND LIQUID NATURAL GAS, A BLESSING THAT MAY BE A CURSE OR ARE WE MISSING THE POINT? In G. Dr. Segell, S. Dr. Kostelyanets, & H. Pro. Solomon (Eds.), *Terrorism in Africa: New Trends and Frontiers.* (pp. 106–116). Institute for African Studies.

ENI. (2014a). *ENVIRONMENTAL IMPACT ASSESSMENT REPORT FOR THE LIQUEFIED NATURAL GAS PROJECT IN CABO DELGADO*. https://www.afdb.org/pt/documents/mozambique-mozambique-lng-p-mz-fd0-003-full-eia

ENI. (2014b). *Initial Resettlement Plan*. https://mzlng.totalenergies.co.mz/sites/g/files/wompnd2311/f/atoms/files/annex_i_-_lng_final_eia_sept_2014_eng.pdf

EOH. (2014). *FINAL RESETTLEMENT ACTION PLAN; PART 6 OF THE ENVIRONMENTAL AND SOCIAL HEALTH IMPACT ASSESSMENT PROCESS*. http://www.cesnet.co.za/pubdocs/Syrah%20Balama%20Graphite%20Mine%20Eng%20CB290814/Resettlement%20Action%20Plan_Part1.pdf

ESRI. (2018). *The Geographic Advantage ™ GIS Solutions for Mining*. www.esri.com/mining

Firozjaei, M. K., Sedighi, A., Firozjaei, H. K., Kiavarz, M., Homaee, M., Arsanjani, J. J., Makki, M., Naimi, B., & Alavipanah, S. K. (2021). A historical and future impact assessment of mining activities on surface biophysical characteristics change: A remote sensing-based approach. *Ecological Indicators*, *122*, 107264. https://doi.org/10.1016/J.ECOLIND.2020.107264

Foody, G. (2008). Harshness in image classification accuracy assessment. *Http://Dx.Doi.Org/10.1080/01431160701442120*, *29*(11), 3137–3158. https://doi.org/10.1080/01431160701442120

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*(1), 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Gallo, A. B., Simões-Moreira, J. R., Costa, H. K. M., Santos, M. M., & Moutinho dos Santos, E. (2016). Energy storage in the energy transition context: A technology review. *Renewable and Sustainable Energy Reviews*, *65*, 800–822. https://doi.org/10.1016/J.RSER.2016.07.028

Gharehbaghi, K., & Scott-Young, C. (2018). GIS as a vital tool for Environmental Impact Assessment and Mitigation. *IOP Conference Series: Earth and Environmental Science*, *127*(1). https://doi.org/10.1088/1755-1315/127/1/012009

Gielen, D. (2021). *Critical materials for the energy transition*. www.irena.org

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/J.PATREC.2005.08.011

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55–72. https://doi.org/10.1016/J.ISPRSJPRS.2016.03.008

Hird, J. N., DeLancey, E. R., McDermid, G. J., & Kariyeva, J. (2017). Google earth engine, open-access satellite data, and machine learning in support of large-area probabilisticwetland mapping. *Remote Sensing*, *9*(12). https://doi.org/10.3390/RS9121315

Höhne, N., Lui, S., Skribbe, R., Moisio, M., Hare, B., Ramalope, D., Ancygier, A., & Heck, S. (2022). *Climate Action Tracker Global reaction to energy crisis risks zero carbon transition Analysis of government responses to Russia's invasion of Ukraine Climate Action Tracker*. https://climateactiontracker.org/publications/global-reaction-to-energy-crisis-risks-zero-carbon-transition/

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). *A Practical Guide to Support Vector Classification*. http://www.csie.ntu.edu.tw/~cjlin

Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., Gong, P., Yang, J., Bai, Y., Zheng, Y., & Zhu, Z. (2017). Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sensing of Environment*, *202*, 166–176. https://doi.org/10.1016/J.RSE.2017.02.021

Huber, S. T., & Steininger, K. (2022). Critical sustainability issues in the production of wind and solar electricity generation as well as storage facilities and possible solutions. *Journal of Cleaner Production*, *339*, 130720. https://doi.org/10.1016/J.JCLEPRO.2022.130720

Hund, K., Porta, D. la, Fabregas, T. P., Laing, T., & Drexhage, J. (2020). *Minerals for Climate Action: The Mineral Intensity of the Clean Energy Transition*. www.worldbank.org

IEA. (2021a). *Net Zero by 2050 - A Roadmap for the Global Energy Sector*. https://www.iea.org/reports/net-zero-by-2050

IEA. (2021b). *The Role of Critical Minerals in Clean Energy Transitions*. https://www.iea.org/reports/the-role-of-critical-minerals-in-clean-energy-transitions

IEA. (2021c). *World Energy Outlook 2021*. https://www.iea.org/reports/world-energy-outlook-2021

INP. (2014). *Current Areas of Exploration and Production / Exploration & Production / Home - INP*. http://www.inp.gov.mz/en/Exploration-Production/Current-Areas-of-Exploration-and-Production

Jenkins, K., Sovacool, B. K., Mouter, N., Hacking, N., Burns, M. K., & McCauley, D. (2021). The methodologies, geographies, and technologies of energy justice: A systematic and comprehensive review. *Environmental Research Letters*, *16*(4). https://doi.org/10.1088/1748-9326/ABD78C

Khassab, B. (2021). *The paradox of sustainability in the extraction of "green minerals" from highly volatile contexts: A case study of graphite mining in Cabo Delgado Province, Mozambique*. University of Utrecht.

Kirsch, A., Opena Disterhoft, J., Marr, G., McCully, P., & Breech, R. DilworthT. (2021). *Banking on Climate Chaos 2021 - A Fossil Fuel Financing Report 2021*. https://www.ran.org/wp-content/uploads/2021/03/Banking-on-Climate-Chaos-2021.pdf

Kuenzer, C., Dech, S., & Wagner, W. (2015). Remote Sensing Time Series Revealing Land Surface Dynamics: Status Quo and the Pathway Ahead. In *Remote Sensing and Digital Image Processing* (Vol. 22, pp. 1–24). Springer, Cham. https://doi.org/10.1007/978-3-319-15967-6_1

Lamb, W. F., Wiedmann, T., Pongratz, J., Andrew, R., Crippa, M., J Olivier, J. G., Wiedenhofer, D., Mattioli, G., al Khourdajie, A., House, J., Pachauri, S., Figueroa, M., Saheb, Y., Slade, R.,

Hubacek, K., Sun, L., Kahn Ribeiro, S., Khennas, S., de la Rue du Can, S., … Minx, J. (2021). A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. *Environ. Res. Lett*, *16*, 73005. https://doi.org/10.1088/1748-9326/abee4e

Lambin, E. F., & Linderman, M. (2006). Time series of remote sensing data for land change science. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1926–1928. https://doi.org/10.1109/TGRS.2006.872932

Landsat Missions. (n.d.). *Landsat Collection 2 | U.S. Geological Survey*. Retrieved March 14, 2022, from https://www.usgs.gov/landsat-missions/landsat-collection-2

Lèbre, É., Stringer, M., Svobodova, K., Owen, J. R., Kemp, D., Côte, C., Arratia-Solar, A., & Valenta, R. K. (2020). The social and environmental complexities of extracting energy transition metals. *Nature Communications*, *11*(1). https://doi.org/10.1038/S41467-020-18661-9

Lechner, A. M., Owen, J., Li Ern Ang, M., Edraki, M., Aklima Che Awang, N., & Kemp, D. (2019). *Historical socio-environmental assessment of resource development footprints using remote sensing*. https://doi.org/10.1016/j.rsase.2019.100236

Legg, C. (1994). *Remote sensing and geographic information systems : geological mapping, mineral exploration, and mining*. Wiley.

Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, *36*(2), 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2

Liu, W., Agusdinata, D. B., & Myint, S. W. (2019). Spatiotemporal patterns of lithium mining and environmental degradation in the Atacama Salt Flat, Chile. *International Journal of Applied Earth Observation and Geoinformation*, *80*, 145–156. https://doi.org/10.1016/J.JAG.2019.04.016

Macuane, J. J., Buur, L., & Monjane, C. M. (2018). Power, conflict and natural resources: The Mozambican crisis revisited. *African Affairs*, *117*(468), 415–438. https://doi.org/10.1093/AFRAF/ADX029

Marshall, R., Arts, J., & Morrison-Saunders, A. (2005). International principles for best practice EIA follow-up. *Impact Assessment and Project Appraisal*, *23*(3), 175–181. https://doi.org/10.3152/147154605781765490

Mate, R. (2021). Mining Concession requests increase as armed conflict in Cabo Delgado intensifies - Who are the lords of the mining licenses in Cabo Delgado? *Centro de Integridade Pública - Extractive Industry*, *06*. https://eiti.org/files/documents/eiti_-_validacao_de_mocambique_-_relatorio_sobre_a_recolha_de_dados_inicial_e_consulta_as_partes_envolvi-

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). *International Journal of Remote Sensing Implementation of machine-learning classification in remote sensing: an applied review Implementation of machine-learning classification in remote sensing: an applied review*. https://doi.org/10.1080/01431161.2018.1433343

Mitchell, C., & Deady, E. (2021). *Graphite resources, and their potential to support battery supply chains, in Africa*. http://nora.nerc.ac.uk/id/eprint/531119/1/Graphite%20supply%20chains%20in%20Africa_Report.pdf

Mustard, J. F., & Sunshine, J. M. (1999). Spectral analysis for earth science: investigations using remote sensing data. In *Manual of Remote Sensing, Remote Sensing for the Earth Sciences* (3rd ed., Vol. 3, pp. 251–307). John Wiley & Sons Inc,.

NASA/JPL. (n.d.). *ECOSTRESS Spectral library*. Retrieved July 9, 2022, from https://speclib.jpl.nasa.gov/library

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. v., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, *148*, 42–57. https://doi.org/10.1016/J.RSE.2014.02.015

Paull, D., Banks, G., Ballard, C., & Gillieson, D. (2008). Monitoring the Environmental Impact of Mining in Remote Locations through Remotely Sensed Data. *Http://Dx.Doi.Org/10.1080/10106040608542372*, *21*(1), 33–42. https://doi.org/10.1080/10106040608542372

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications 2020 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-18321-y

Salimo, P., Buur, L., & Macuane, J. J. (2020). The politics of domestic gas: The Sasol natural gas deals in Mozambique. *Extractive Industries and Society*, *7*(4), 1219–1229. https://doi.org/10.1016/J.EXIS.2020.05.017

Song, W., Song, W., Gu, H., & Li, F. (2020). Progress in the remote sensing monitoring of the ecological environment in mining areas. In *International Journal of Environmental Research and Public Health* (Vol. 17, Issue 6). MDPI AG. https://doi.org/10.3390/ijerph17061846

TotalEnergies. (n.d.). *Total Mozambique LNG Project*. Retrieved July 6, 2022, from https://mzlng.totalenergies.co.mz/

UNDP, & UN Environment. (2018). *Managing mining for sustainable development - A Sourcebook*.

Wagemann, J., Siemen, S., Seeger, B., & Bendix, J. (2021). Users of open Big Earth data – An analysis of the current state. *Computers & Geosciences*, *157*, 104916. https://doi.org/10.1016/J.CAGEO.2021.104916

Weih, R. C., & Riggan, N. D. (2010). OBJECT-BASED CLASSIFICATION VS. PIXEL-BASED CLASSIFICATION: COMPARITIVE IMPORTANCE OF MULTI-RESOLUTION IMAGERY . *Remote Sensing and Spatial Information Sciences*, *XXXVIII–4*(C7). https://www.isprs.org/proceedings/xxxviii/4-c7/pdf/Weih_81.pdf

Werner, T. T., Bebbington, A., & Gregory, G. (2019). Assessing impacts of mining: Recent contributions from GIS and remote sensing. *Extractive Industries and Society*, *6*(3), 993–1012. https://doi.org/10.1016/J.EXIS.2019.06.011

Wiegink, N., & Kronenburg García, A. (2022). Surplus to extraction: Resettlement as a “make live” intervention in Mozambique. *The Extractive Industries and Society*, *9*, 101036. https://doi.org/10.1016/J.EXIS.2021.101036

9. Appendix

9.1. GitHub repository.

The proposed workflow model used in this research was developed using common JavaScript commands and GEE machine learning algorithms. The commands and algorithms were composed in the GEE code editor to provide a basis for earth observation analyses. Each script requires a few input parameters to run successfully. How to use each script, what to consider and where to input required data is described in a README file respectively. All scripts which were created during this research can be accessed via the following GitHub repository:

- "Land-change-monitoring-using-JavaScript-and-GEE" https://github.com/steichtmann/Land-change-monitoring-using-JavaScript-and-GEE-.git (will be active from the 23rd of August 2022 on).

The repository contains all scripts required to conduct analysis as depicted in Fig. 3 and obtain results as they are presented in this research. The following scripts are available and can be used and adapted in the GEE code editor (in order of application):

1. Define Region of Interest (ROI).
2. Land cover classification and accuracy assessment (Landsat 4,5,7).
   Land cover classification and accuracy assessment (Landsat 8).
   Land cover classification and accuracy assessment (Sentinel 2).
3. Land cover change analysis.
4. NDVI Differencing.

Auxiliary scripts are provided to support the analysis processes above:

- Indices NDVI, NDBI, NDMI, MNDWI (Landsat 4,5,7).
- Indices NDVI, NDBI, NDMI, MNDWI (Landsat 8).
- Indices NDVI, NDBI, NDMI, MNDWI (Sentinel 2).

9.2. Error Matrices of the land cover maps.

In the following the error matrices of the LNGP and BGP case study for the remaining years of the time series, respectively.

Appendix Tab. 1 LNGP case study: 2005 error matrix of the testing data of the land cover classification.

	Wetland	Dense veg.	Sparse veg.	Unveg.	Water	Total	Users accuracy
Wetland	49	0	0	0	0	49	100%
Dense vegetation	0	46	1	0	0	47	98%
Sparse vegetation	1	4	49	4	0	58	84%
Unvegetated	0	0	0	46	0	46	100%
Water	0	0	0	0	50	50	100%
Total	50	50	50	50	50	250	
Producers accuracy	98%	92%	98%	92%	100%		
Overall accuracy	= (49+46+49+46+50)/250 = 96 %						

Appendix Tab. 2 LNGP case study: 2013 error matrix of the testing data of the land cover classification.

	Wetland	Dense veg.	Sparse veg.	Unveg.	Water	Total	Users accuracy
Wetland	46	0	0	0	0	46	100%
Dense vegetation	1	48	1	0	0	50	96%
Sparse vegetation	2	2	44	3	0	51	86%
Unvegetated	0	0	5	47	0	52	90%
Water	1	0	0	0	50	51	98%
Total	50	50	50	50	50	250	
Producers accuracy	92%	96%	88%	94%	100%		
Overall accuracy	= (46+48+44+47+50)/250 = 94 %						

Appendix Tab. 3 BGP case study: 2005 error matrix of the testing data of the land cover classification.

	Mining	Dense veg.	Sparse veg.	Unveg.	Water	Total	Users accuracy
Mining	-	0	0	0	0	0	-
Dense vegetation	-	48	4	0	0	52	92%
Sparse vegetation	-	2	46	5	0	53	87%
Unvegetated	-	0	0	45	0	45	100%
Water	-	0	0	0	50	50	100%
Total	-	50	50	50	50	200	
Producers accuracy	-	96%	92%	90%	100%		
Overall accuracy	= (48+46+45+50)/200 = 94,5 %						

Appendix Tab. 4 BGP case study: 2013 error matrix of the testing data of the land cover classification.

	Mining	Dense veg.	Sparse veg.	Unveg.	Water	Total	Users accuracy
Mining	-	0	0	0	0	0	-
Dense vegetation	-	47	3	0	0	50	94%
Sparse vegetation	-	2	47	1	0	50	94%
Unvegetated	-	0	7	43	0	50	86%
Water	-	0	0	0	50	50	100%
Total	-	49	57	44	50	200	
Producers accuracy	-	96%	82%	98%	100%		
Overall accuracy	= (47+47+43+50)/200 = 92,8 %						

Appendix Tab. 5 BGP case study: 2018 error matrix of the testing data of the land cover classification.

	Mining	Dense veg.	Sparse veg.	Unveg.	Water	Total	Users accuracy
Mining	47	0	0	0	1	48	98%
Dense vegetation	0	47	3	0	0	50	94%
Sparse vegetation	1	3	44	5	0	53	83%
Unvegetated	2	0	3	45	0	50	90%
Water	0	0	0	0	49	49	100%
Total	50	50	50	50	50	250	
Producers accuracy	94%	94%	88%	90%	98%		
Overall accuracy	= (47+47+44+45+49)/200 = 92,8 %						