A Re-Analysis: Assessing the Impact of Methodological Choices on Landscape Restoration Effectiveness

Andrea Ingrosso, Nick Schmitthenner, Niko Kron

February 28, 2025

Abstract

Ruijsch et al., 2023 quantifies the effectiveness of landscape restoration projects by analyzing greening trends derived by MODIS NDVI and EVI. After a thorough re-analysis of their data, we found slight variations in our results, though they remain within an acceptable range. We quantified uncertainties by systematically modifying the analysis and demonstrate that each modification — subregional analysis in Africa, changing the vegetation index, or using updated MODIS data — substantially impacts the results. We conclude that methodological choices play a crucial role and must be carefully considered when interpreting the findings.

The greening analyzed in Ruijsch et al., 2023 underwent several preprocessing steps, resulting in what is termed *local greening*. First, the neighborhood greening trend was subtracted from each pixel, yielding a background-corrected greening trend. This corrected trend was then further filtered. Finally, the fraction of local greening around landscape restoration projects, or simply projects, was compared to control regions using a t-test. However, the original study relied on a single t-test statistic with an arbitrary α level, limiting result reliability and uncertainty assessment. To address this and other limitations, we conducted a systematic reanalysis.

We modified Ruijsch et al., 2023's analysis in four ways to assess both result uncertainties and the sensitivity of findings to methodological choices. Specifically, we examined how changes in the analysis influenced p-values from the t-test (project vs. control local greening) and the mean and standard error of the fraction of local greening, aggregated over all projects or specific project types (see Ruijsch et al., 2023 Fig. 3B). Our focus on revegetation, natural regeneration and the aggregation of all project types ensures comparability with the original study (Ruijsch et al., 2023). To isolate the effect of each modification, we tested one change at a time, ranking them by impact size from strongest to negligible.

(1) To address potential biases arising from aggregated data analysis (Steel and Holt, 1996), we split Africa into four quadrants (see fig.A4). We show that results in Ruijsch et al., 2023 are dominantly driven by projects in East Africa with 339 of 516 projects (see fig.1). In contrast, North and South Africa are underrepresented with respectively 31 and 54 projects. West Africa includes 92 projects. Further, we demonstrate that projects show no increase of greening compared to the control region for South and West Africa, which is roughly half of the study area. We conclude that the dominance of East Africa should be considered when interpreting the results.

(2) We replaced NDVI and EVI (Didan, 2021) with NIRv and kNDVI, which better capture vegetation productivity and minimize background noise (Badgley et al., 2017; Wang et al., 2023). Using these improved indices, we conducted nine runs to quantify t-test uncertainty and assessed the project's impact range. The aggregation of all project types exhibited inconsistent significance levels, indicating an undefined impact range, whereas revegetation and natural regeneration projects showed stable significance for specific ranges, suggesting a well-defined impact range, which could be used for further analysis (see fig. 2). By inferring significance from multiple runs, we increased robustness in the results. For example, while Ruijsch et al., 2023 found significant revegetation greening only within a radius of 1 km (Fig. 3B), our approach suggests low uncertainty in P values, revealing significant greening even at 0.5 km (see fig. 2).

(3) To assess the impact of satellite data choice on our findings, we re-ran the analysis using an updated MODIS product (Didan, 2021) instead of the previously used version (Didan, 2015). This revealed notable differences in results for revegetation projects while confirming the robustness of aggregated project outcomes. For revegetation and natural regeneration projects, the fraction of local greening showed no

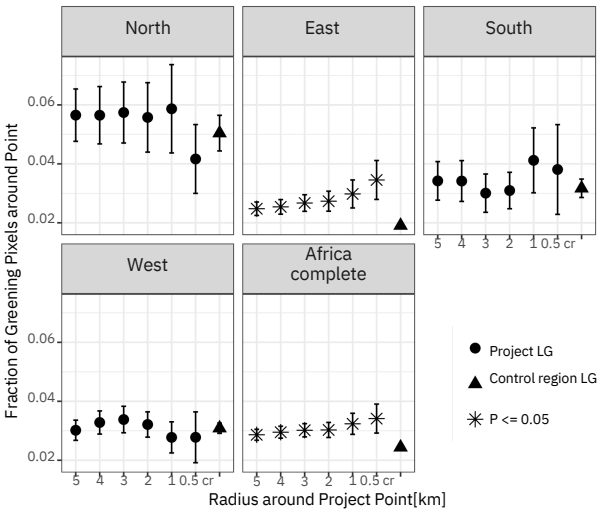


Figure 1: Mean and standard error of the fraction of local greening around all projects shown per buffer size around the projects. Stars indicate a significant increase in local greening around a project compared to a control region, based on a one-sided t-test ($\alpha = 0.05$). Each subplot represents a subregion of Africa, while the final subplot presents results for the entire continent. Results are shown only for aggregation of all project types, as absence of specific project in subregions was too high for a t-test. Projects with a median $NDVI > 0.15$ were filtered out. Local greening is based on EVI and NDVI data from MOD13Q1.006.

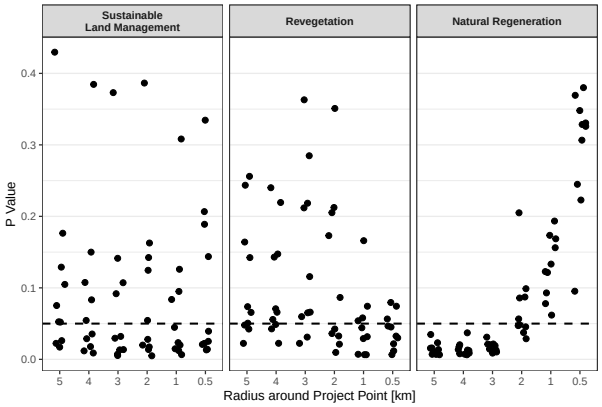


Figure 2: P-values indicating whether the fractions of local greening differ significantly from those in the same control regions for aggregation of all project types (sustainable land management) and specific project types (revegetation, natural regeneration). Each vegetation index (NDVI, EVI, NIRv, and kNDVI), as well as each combination of two indices, has one point per radius around the project locations. Projects with a median $NDVI < 0.15$ were filtered out. Local greening is based on data from MOD13Q1.006.

clear trend along the buffer sizes, contrary to the original analysis (see fig. A9b). Also, t-test results diverged. In contrast, results for the aggregation of all project types did not alter substantially. We conclude that the aggregated results of all projects are definitively more robust than project-type-specific results. We advise interpreting project-type-specific results with caution due to their uncertainty.

(4) We masked out radii of 0.5, 1, 2, 3, 4, and 5 km around each project when calculating local greening in the control regions. This ensures that the comparison of local greening fractions considers only the control region, excluding the greening effect of the project itself. We observed virtually no difference from the original results (see Tables A6 and A7), probably because the control region is too large relative to the masked radii to influence the analysis.

Our reanalysis of Ruijsch et al., 2023 highlighted the sensitivity of the results to methodological choices, affecting the mean fraction of local greening and t-test outcomes. Aditionally, splitting Africa into four subregions showed that there is no greening effect in South and West Africa and that East African projects mainly drive the results. Further, variations in vegetation indices and the use of updated MODIS data changed the results. While results remained stable for the aggregation of all projects, project-type-specific outcomes are more uncertain. We emphasize the need for cautious interpretation, especially regarding statistical significance, regional biases and specific project types.

Appendix: Reproducible analysis with comments and conclusions for each step

Contents

A. Introduction

A.1 Summarization of the study

This study used a spatial context approach within Google Earth Engine (Gorelick et al., 2017) to separate small-scale greening caused by land restoration and different project types from background trends driven by natural climate variability. The study examines local greening trends across Africa from 2001 to 2021, with a particular focus on semi-arid regions. The approach used neighborhood-based analysis to detect sudden increases in vegetation cover, which were then classified as local greening. This methodology helped differentiating greening caused by land management from broader climatic trends.

Within the framework of several conducted land management projects, the study examined the effectiveness of different land restoration methods. It found that active revegetation strategies, such as tree planting, were more effective in achieving rapid greening, while natural regeneration, although slower, had a greater spatial impact and was more cost effective. This implies that while active restoration techniques may be more efficient for smaller areas, natural regeneration could be a more sustainable and less costly approach for larger regions, assuming enough time is allowed for the processes to unfold.

Sustainable Land Management Projects (SLM) are merely point coordinates and therefore do not represent an area of influence where the effects of the projects would be visible. To overcome this limitation, radial buffers were generated in a prior step within Google Earth Engine (GEE) to approximate the areas impacted by the SLM projects. The buffers analyzed in this script correspond to those used in Ruijsch et al. (2023), with sizes of 0.5 km, 1 km, 2 km, 3 km, 4 km, and 5 km. The control regions (CR) are defined as the intersections of an aridity map, a land use map, and country boundaries.

Local greening (LG) is detected using the Breaks For Additive Seasonal and Trend (BFAST) algorithm in GEE (Verbesselt et al., 2010) in order to identify changes in the time series of the selected vegetation indices. These time series were created using the spatial context approach by subtracting the mean of each pixel's neighborhood for different radii whereby a circle with a radius of 1 km around the certain pixel was not included, in order to exclude climate variability and remain with trends driven by local processes and land management causes compared to the surrounding areas (Ruijsch et al., 2023). Several conditions have to be met to assign LG to an analyzed pixel time series according to (Ruijsch et al., 2023): LG refers to pixels where BFAST identifies a significant slope change after a detected breakpoint ($p = 0.05$), followed by positive post-break trends that are stronger than before, with the pixel time series also showing a positive trend. Greening hotspots are areas where the trends detected by BFAST for both vegetation indices align, providing greater reliability and reducing noise (Ruijsch et al., 2023). For a consistent comparison, the fraction of LG around an SLM project is only compared to the fraction of LG within the control region in which the SLM project is located. To find significant differences, a t-test with a p-value threshold of $p = 0.05$ was applied.

As a reminder, the "fraction of LG (LG) around a SLM project" (SLM_LG) is defined as

$$\text{SLM\_LG} = \frac{\text{Number of pixels around the SLM showing LG}}{\text{Total number of pixels around the SLM}} \tag{1}$$

The "fraction of LG within the control region" (CR_LG) is defined as

$$\text{CR\_LG} = \frac{\text{Number of pixels within the CR showing LG}}{\text{Total number of pixels within a CR}} \tag{2}$$

For a spatial correction, using a 25 km radius, 2.1% of the observed area exhibited LG over the analyzed time period. Most breakpoints occured between 2015 and 2018 and were unevenly distributed across Africa. Many local greneing events were found in the Sahel region, Kenya, Tanzania, and parts of

southern Africa. The spatial distribution showed for all radii nearly the same pattern, but fewer areas showed LG (10km with 1.9% and 5 km with 1.8%).

Semi-arid regions, which include only 26% of the study area, represented 39. 1% of local greening detected, highlighting their disproportional contribution. In terms of land cover, the main drivers of LG were shrublands, grasslands and savannas. Other areas such as, for example, Lake Chad also yielded greening results, where declining water levels with low NDVI/EVI values result in increasing land area and therefore the vegetation index. Similar findings were observed in agricultural areas.

Therefore, in semi-arid regions, SLM projects showed a significant increase in LG compared to the detected control region greening. Revegetation projects had a substantially higher percentage of LG for lower radii (0.5 km, 1 km) whereas larger distances did not result in significant differences. The biggest effects were achieved due to planting trees and fruit trees. Natural regeneration projects showed weaker effects across all spatial scales. Still, significant differences were found for 3 - 5 km distances around the project.

The findings matched the author's expectation, that semi-arid regions had a very large increase in LG due to sparsely vegetated areas where SLM projects lead to an increase in vegetation cover compared to more humid regions with already existing vegetation. Over small areas, revegetation and tree planting seem to be more efficient than natural regeneration. When it comes to larger areas, natural regeneration shows significant increases in LG compared to revegetation. Differences in the two compared methods lie in the temporal and monetary aspects, where natural regeneration takes much longer with lower costs and revegetation yields results in less time but is more work and therefore monetary resources. Not all SLM projects aim for local greening; some focus on biodiversity conservation, income generation, policy implementation, or cultural objectives (Ruijsch et al., 2023). For this reason, some SLM projects did not result in revegetation because of a different intention of this project.

Some uncertainties exist in the spatial context approach. For example, local greening may be detected even when it is unrelated to SLM projects. Therefore, further differentiation in the spatial context approach would be necessary in order to focus on land restoration only. Also, accurate boundaries would improve results. MODIS resolution (250 m) limits the detection of small projects, with 7% being smaller than this resolution. The spatial-context method cannot detect greening over areas larger than $\approx 2500\,\mathrm{km}^2$, though only 4% of projects exceed this size and 38% lack size data. Seasonal greenness changes could aid policymakers, but improper land restoration risks biodiversity loss, water issues, and ecosystem disruption. More research and a comprehensive, open-access database of land restoration projects are needed to prevent unintended harm.

B. Reproducing the original study

B.1 Data

In order to reproduce Ruijsch et al., 2023's analysis, multiple datasets were included in the analysis as shown in table A1.

Table A1: Description of spatial and temporal resolution of datasets used in the study

Name	Source	Spatial Resolution	Temporal Resolution	Time Period
NDVI	Terra MODIS Vegetation Indices (MOD13Q1.006)	250 m	16-day composites	2001-01-01 to 2022-01-01
EVI	USGS Landsat-7 ETM+ Level 2, Collection 2, Tier 1	30 m	16-day land surface reflectance images	2001-01-01 to 2022-01-01
Land cover	MODIS Land Cover Type (MCD12Q1)	250 m	16-day composites	2001-01-01 to 2022-01-01
Aridity index	CRU TS4.04	0.5°	Yearly	1901 to 2029
SLM projects	WOCAT SLM technologies	Point coordinates	—	—

The data of the SLM projects are available in the "WOCAT Global sustainable land management database", n.d. On this website, it is necessary to click through all the projects and copy the necessary information manually into a spreadsheet. As this is a lot of work for 434 SLM projects, and as this has already been done by the authors of Ruijsch et al., 2023, we decided to use their datasheet. It is available here.

MODIS data (MOD13Q1.006) including global NDVI and EVI data (Didan, 2015) with a resolution of 250 m are directly available inside the Google Earth Engine.

Land Cover types (Sulla-Menashe et al., 2019) are available in the MODIS Land Cover Type collection (MCD12Q1) and are also directly available in the Google Earth Engine.

Aridity Index classes are available in the Harris and Osborn, 2020 database.

B.2 Analysis in Google Earth Engine

Table A2: List of the Google Earth Engine Scripts by us Ruijsch et al. (2023)

Script ID	Description with link
1	Calculate NDVI LG pixels
2	Calculate EVI LG pixels
3	Combine NDVI and EVI LG pixels
5	Compute statistics and study LG at sustainable land management projects

We want to reproduce the analysis in the Google Earth Engine. As in this analysis, a lot of assumptions are made and a pipeline including a lot of functions is used, it would not be realistic to reproduce these scripts from scratch. Instead, the existing scripts by Ruijsch et al. (2023) (see table A2) were used. For each of the improvements and corrections, we created new scripts based on the original ones. All scripts that we adapted, are listed separately (see table A3).

Table A3: List of the Google Earth Engine Scripts adapted by us

Based on ID	Description with link
	Analyze Regional Differences
-	Separating Africa into 4 parts & Export
	Choice of Vegetation Indices
1	Calculate LG pixels with kNDVI instead of NDVI
1	Calculate LG pixels with NIRv instead of NDVI
5	Combine kNDVI and EVI, compute statistics and study LG at SLMs
5	Combine kNDVI and NIRv, compute statistics and study LG at SLMs
5	Combine kNDVI and NDVI, compute statistics and study LG at SLMs
5	Combine NDVI and NIRv, compute statistics and study LG at SLMs
5	Combine EVI and NIRv, compute statistics and study LG at SLMs
5	Compute statistics and study LG at SLMs with NIRv only
5	Compute statistics and study LG at SLMs with NDVI only
5	Compute statistics and study LG at SLMs with kNDVI only
5	Compute statistics and study LG at SLMs with EVI only
-	Calculate correlation between the different indices
	Updated MODIS Data
1	Calculate NDVI LG pixels with updated MODIS-061 data
2	Calculate EVI LG pixels with updated MODIS-061 data
3	Combine NDVI and EVI LG pixels with updated MODIS-061 data
5	Compute statistics and study LG at SLMs with updated MODIS-061 data
	Masking Other Projects
5	Compute statistics and study LG at SLMs including SLM-masking

B.3 Calculating P-values, Mean, and SE in R

In this section we first load in the data generated with Google Earth Engine in R, then reproduce the mean and standard error of the LG around the projects (SLM_LG), and LG in the control region (CR_LG). Also, we do the t-test and reproduce the p-values. Finally, fig. 3B from Ruijsch et al., 2023 is reproduced to visualize and compare the mentioned metrics for all SLMs and the subsets revegetation and natural regeneration.

The following packages in R are required:

```
library(dplyr)
library(ggplot2)
library(tidyr)
library(tibble)
library(patchwork)
library(reticulate)
library(purrr)
library(xtable)
```

B.3.1 Load Data

Now we load in the LG data:

```
unfiltered_SLM <- read.csv("evi_ndvi/unfiltered_technologies.csv") %>% arrange(SAMPLE_NDV)
filtered_SLM <- read.csv("evi_ndvi/projectSamples_v12_0500mBuffer.csv") %>% arrange(SAMPLE_NDV)
# in GEE: filter out SAMPLE_NDV < 1500

output_text <- paste0("unfiltered SLMs row number: ",
```

```
    length(unfiltered_SLM$SAMPLE_NDV))
output_text <- paste0(output_text,"\n")
summary(unfiltered_SLM$SAMPLE_NDV)
head(unfiltered_SLM, 10)[,7:8]
output_text <- paste0(output_text,"there are ",
    length(which(unfiltered_SLM$NDVImask == 0 & unfiltered_SLM$SAMPLE_NDV > 1500)),
    " cases, where NDVImask == 0 & SAMPLE_NDV > 1500")


output_text <- paste0(output_text,"\n_____________________________________________________\n\n")
output_text <- paste0(output_text,"filtered SLMs row number: ",
    length(filtered_SLM$SAMPLE_NDV))
output_text <- paste0(output_text,"\n")
summary(filtered_SLM$SAMPLE_NDV)
head(filtered_SLM, 10)[,7:8]
output_text <- paste0(output_text,"there are ",
    length(which(filtered_SLM$NDVImask == 0 & filtered_SLM$SAMPLE_NDV > 1500)),
    " cases, where NDVImask == 0 & SAMPLE_NDV > 1500")
cat(output_text)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  299.6  2127.4  4114.0  4175.4  6114.5  8655.0       8
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1511    2649    4666    4577    6299    8655
unfiltered SLMs row number: 628
there are 28 cases, where NDVImask == 0 & SAMPLE_NDV > 1500
------------------------------------------------------------

filtered SLMs row number: 550
there are 28 cases, where NDVImask == 0 & SAMPLE_NDV > 1500
```

The authors filtered out rows, where $NDVImask == 0$ but. Note that NAs were not filtered out. We can't comprehend, what the variable NDVImask means and why they filter again, as the $NDVI < 1500$ filtering happened already in the GEE. The GEE NDVI filtering was done on the variable, as the following code shows. Also, presupposing that NDVImask codes for $NDVI > 1500$ or < 1500, it doesn't align with SAMPLE-NDV. Here is the code of the authors:

```
# Define file paths and buffer sizes
buffer_sizes = [5000, 4000, 3000, 2000, 1000, 500]
file_paths = [f'evi_ndvi/projectSamples_v12_{size}mBuffer.csv' for size in buffer_sizes]
file_paths[-1] = "evi_ndvi/projectSamples_v12_0500mBuffer.csv"
randomStatisticsMean = (
    pd.read_csv('evi_ndvi/mean_AI_country_LU_region_slmMask0.csv', sep=',')
    .sort_values(by='index', ascending=True)
    .groupby(by='index')
    .mean(numeric_only=True)
)

project_statistics = {}
for size, file_path in zip(buffer_sizes, file_paths):
```

```
    df = (
        pd.read_csv(file_path, sep=',')
        .sort_values(by='index', ascending=True)
        .set_index('index')
    )
    df['mean_areaMean'] = randomStatisticsMean['mean']
    project_statistics[size] = df
```

Additionally, the authors drop duplicated values for *mean* and *definition*. This seems to contradict their statement: "If a single SLM project contained multiple locations, we considered it as multiple projects, resulting in 628 project locations" (Ruijsch et al., 2023, Chapter 2.1). Another indication that dropping rows is inappropriate is that it results in a varying number of data points across differently sized SLM buffers. This occurs because the differing buffer sizes around the SLM yield distinct mean values (i.e., LG within the buffer), causing the rows dropped to vary with each buffer size. The next chunks demonstrate this: without dropping each buffer has 550 data points.

```
for size in buffer_sizes:
  print(f"Buffer {size}m: {len(project_statistics[size])} rows")
```

```
Buffer 5000m: 550 rows
Buffer 4000m: 550 rows
Buffer 3000m: 550 rows
Buffer 2000m: 550 rows
Buffer 1000m: 550 rows
Buffer 500m: 550 rows
```

After dropping, the data points vary between the buffer sizes.

```
for size in buffer_sizes:
    project_statistics[size] = project_statistics[size].drop_duplicates(
      subset=['mean', 'definition'])
    print(f"Buffer {size}m: {len(project_statistics[size])} rows")
```

```
Buffer 5000m: 461 rows
Buffer 4000m: 457 rows
Buffer 3000m: 456 rows
Buffer 2000m: 451 rows
Buffer 1000m: 430 rows
Buffer 500m: 413 rows
```

We don't see a reason to drop data points and think it is wrong to do so. Therefore, in our analysis, we don't drop any data points. However, we do additionally filter out $NDVImask == NA$.

Regardless of which variable is more appropriate for filtering, we think that NAs should be excluded when present in $NDVImask$ or $SAMPLE_NDV$. Yet, we don't know how NAs are produced in the raw data and what they actually mean.

B.3.2 Load in SLM and CR and subset CR

Here we load in the SLM project data and the CR data. We take the mean value of the cr_list because each SLM pojects can be assigned to multiple control regions, as the latter can be spatially separated. Also, we create a subset of control regions by project type.

```r
# SLM
setwd("evi_ndvi/")
files <- list.files(pattern = "project", full.names = F)
x <- sub(".*v12_(\\d{4}).*", "\\1", files)
var_name_p <- paste0("slm_", x)
slm_list <- lapply(seq_along(files), function(x){
  res <- read.csv(files[x])
  return(res)
})
names(slm_list) <- var_name_p #give df names for clarification
index_ndvi_zero <- slm_list$slm_0500$index[slm_list$slm_0500$NDVImask == 0 |
                                           is.na(slm_list$slm_0500$NDVImask)]

slm_list <- lapply(slm_list, function(df) {
  df <- df %>%
    mutate(mean = round(mean, 10)) %>%
    filter(NDVImask != 0 & !is.na(NDVImask)) %>%        # filtering out 0 and NA
    # distinct(mean, definition, .keep_all = TRUE) %>%  # drop duplicates
    mutate(slm_lg = mean) %>%
    select(definition, index, slm_lg, type) %>%
    arrange(index)

  return(df)
})

# control regions
cr <- read.csv("mean_AI_country_LU_region_slmMask0.csv") %>%
  filter(!index %in% index_ndvi_zero) %>%
  group_by(index) %>%
  summarise(cr_lg = mean(mean)) %>%
  select(cr_lg,index)

# combine SLM and CR
slm_cr <- lapply(slm_list, function(x){
  x <- x %>% inner_join(y = cr, by = "index")
})

# CR subset by project type
types_s <- unique(slm_cr$slm_0500$type)
subnames <- paste0("type_",gsub("\\s+", "", types_s)) # deleting white spaces

sub_ls <- list()

for (i in types_s) {
  types <- gsub("\\s+", "", i)
  sub_ls[[types]] <- lapply(slm_cr, function(x){
    x %>% filter(type == i)
  })
}
```

B.3.3 Explanation of the Variable Names

For an easier comprehension of our analysis, we explain the meaning of each variable in data frames (see table A4).

The list slm_list consists of 6 data frames, all with a unique numeric ending. This ending stands for the radial buffer around the SLM project [m]. Inside this buffer, the SLM_LG is calculated.

Table A4: Description of the columns of each data frame in slm_list

Column Name	Description
definition	Detailed explanation of the project.
index	Index of the SLM.
slm_lg	Fraction of LG (LG) within the selected buffer.
technology	Unique code of the SLM.
type	One of the following SLM types: erosion prevention, agriculture management, social measures, natural regeneration, animals, water management, water harvesting, river and coastal restoration, agroforestry, revegetation, and fire management. Subsets are created based on this variable.

The list cr_list consists of 7 data frames. In the first "no mask" no masking was applied inside the control region, like in Ruijsch et al. (2023). For the other 6 data frames, the CR_LG was calculated with an SLM-Mask inside the control region. The numbers at the end of the data frame name indicate the size [km] of the radial buffer around the SLM project that was used for masking. Masking buffers were the same as buffers for calculating SLM_LG.

Table A5: Description of the columns of each data frame in cr_list

Column Name	Description
cr_lg	fraction of LG within the selected CR.
index	Index of the SLM.

B.3.4 T-Test

We defined a t-test function for a one-sided t-test, to test if the fraction of LG pixels around the SLM projects significantly differs from their control regions:

```
T_test_custom_simp <- function(slm_cr, alternative, print_results) {
  # Initialize vectors for results
  pvals <- numeric()
  mean_x <- numeric()

  # Populate pvals and mean_x
  for (i in names(slm_cr)) {
      t_result <- t.test(
        slm_cr[[i]]$slm_lg,
        slm_cr[[i]]$cr_lg,
        paired = FALSE,
        alternative = alternative,
        var.equal = FALSE,
        mu = 0
      )
```

```
    pvals[i] <- t_result$p.value
    mean_x[i] <- t_result$estimate["mean of x"]
  }

  # Print results if print_results is TRUE
  if (print_results) {
    cat("P-values:\n")
    print(pvals)
    cat("\nMeans of x:\n")
    print(mean_x)
  }

  # Prepare results
  ttest_results <- data.frame(
    pvals = pvals,
    mean_x = mean_x
  )
  return(ttest_results)
}
```

If we apply this function to all SLM projects, we get

```
all_prj_1sided <- T_test_custom_simp(slm_cr = slm_cr,
                                     alternative = "greater",
                                     print_results = T)
```

```
P-values:
   slm_0500    slm_1000    slm_2000    slm_3000    slm_4000    slm_5000
0.025220505 0.014799962 0.013749603 0.007188633 0.008848033 0.017112777

Means of x:
  slm_0500   slm_1000   slm_2000   slm_3000   slm_4000   slm_5000
0.03413497 0.03236138 0.03027815 0.03017526 0.02951626 0.02863921
```

We can also do that for a project-type-specific SLM subset by

```
sub_ttest <- lapply(sub_ls, function(x){
  T_test_custom_simp(slm_cr = x,
                     alternative = "greater",
                     print_results = F)
})
```

Next, we calculate the metrics mean and standard error (se) for LG around the SLM projects (SLM_LG) and LG in the control region (CR_LG) and join the p-values from before.

```
# take mean lg_slm and lg_cr for all
all_slm_sum <- sapply(slm_cr, function(x){
  x <- x %>%  summarise(mean_slm_lg = mean(slm_lg),
                        se_slm_lg = sd(slm_lg) / sqrt(n()),
                        mean_cr_lg = mean(cr_lg),
                        se_cr_lg = sd(cr_lg) / sqrt(n()))
}, simplify = TRUE) %>%
```

```
  t() %>%
  as.data.frame() %>%
  rownames_to_column(var = "buffer") %>%
  pivot_longer(- buffer, values_to = "vals", names_to = "vars")

# bring the pvals in a similar format
all_p_slm <- all_prj_1sided %>%
  rownames_to_column(var = "buffer") %>%
  inner_join(y = all_slm_sum, by = "buffer") %>%
  mutate(type = "all") %>%
  select(type, everything(), -mean_x)


# take mean lg_slm and lg_cr for subsets
sub_slm_ls_sum <- lapply(sub_ls, function(x){
  lapply(x, function(y){
    y <- y %>%  summarise(mean_slm_lg = mean(slm_lg),
                          se_slm_lg = sd(slm_lg) / sqrt(n()),
                          mean_cr_lg = mean(cr_lg),
                          se_cr_lg = sd(cr_lg) / sqrt(n()))
  })
})

# unwrap the lists and join them
sub_slm_sum <- map_dfr(sub_slm_ls_sum,
                       ~ map_dfr(.x, ~ .x, .id = "buffer"), .id = "type") %>%
  pivot_longer(-c(type, buffer), names_to = "vars", values_to = "vals")
sub_p <- map_dfr(sub_ttest,
                 ~ .x %>% rownames_to_column("buffer"), .id = "type") %>%
  select(-mean_x)

sub_p_slm <- sub_p %>%
  inner_join(sub_slm_sum, by = c("buffer", "type"))
```

B.3.5 Plot the data

Now we have calculated all the fractions of LG and significances and create a plot similar to fig. 3B from Ruijsch et al., 2023.

```
fig3_pre1 <- rbind(sub_p_slm, all_p_slm) %>%
  mutate(buffer = case_when(vars == "mean_cr_lg" ~ "cr",
                            vars == "se_cr_lg" ~ "cr",
                            TRUE ~ buffer))

fig3_pre2 <- fig3_pre1 %>%
  distinct(type, buffer, vars, vals, .keep_all = T) %>%
  pivot_wider(id_cols = c(buffer, type, pvals),
              names_from = vars,
              values_from = vals)  %>%
  mutate(mean_ = as.numeric(coalesce(mean_slm_lg, mean_cr_lg)),
```

```
        se_ = as.numeric(coalesce(se_slm_lg, se_cr_lg))) %>%
  select(mean_, se_, type, buffer, pvals) %>%
  mutate(shape = case_when(buffer == "cr" ~ buffer,
                           TRUE ~ "slm")) %>%
  mutate(key = case_when(pvals <= 0.05 & shape == "slm" ~ "s",
                         pvals > 0.05 & shape == "slm" ~ "r",
                         shape == "cr" ~ "t"))

# filter out all, revegetation, naturalregeneration
fig3 <- fig3_pre2 %>%
  filter(type %in% c("all", "revegetation", "naturalregeneration")) %>%
  mutate(type = factor(type,
                       levels = c("all",
                                  "revegetation",
                                  "naturalregeneration")),
         buffer = factor(buffer, levels = c("slm_5000",
                                            "slm_4000",
                                            "slm_3000",
                                            "slm_2000",
                                            "slm_1000",
                                            "slm_0500",
                                            "cr")))
ggplot(data = fig3,
       aes(y = mean_, x = buffer)) +
  geom_point(aes(shape = key),
             position = position_dodge(width = 0.5), size = 3) +
  geom_errorbar(aes(ymin = mean_ - se_, ymax = mean_ + se_),
                position = position_dodge(width = 0.5),
                width = 0.2) +
  scale_shape_manual(values = c(16, 8, 17),
                     labels = c("SLM_LG\nmean", "P <= 0.05", "CR_LG\nmean")) +
  scale_x_discrete(labels = c("5", "4", "3", "2", "1", "0.5", "cr")) +
  labs(
    title = "",
    x = "Radius around Project Point[km]",
    y = "Fraction of Greening Pixels around Point",
    cex = 2,
    shape = NULL
  ) +
  theme_bw() +
  theme(
    legend.position = c(0.5, 0.35),
    legend.justification = c(2, -1),# Position legend at the top
    legend.direction = "horizontal",  # Arrange items horizontally
    legend.box = "horizontal",
    strip.text = element_text(size = 12),
    legend.background = element_rect(fill = "white", color = "white"), # White box
    legend.box.margin = margin(0, 0, 0, 0)# Ensure the legend is a single box
  ) +
  guides(shape = guide_legend(nrow = 3, ncol = 1)) +  # Set legend layout
```

```
facet_wrap(~type, ncol = 3, labeller = labeller(type = c(
"naturalregeneration" = "Natural Regeneration",
"revegetation" = "Revegetation",
"all" = "Sustainable Land\nManagement"
)))
```

With that code we create the plot shown in fig.A3.

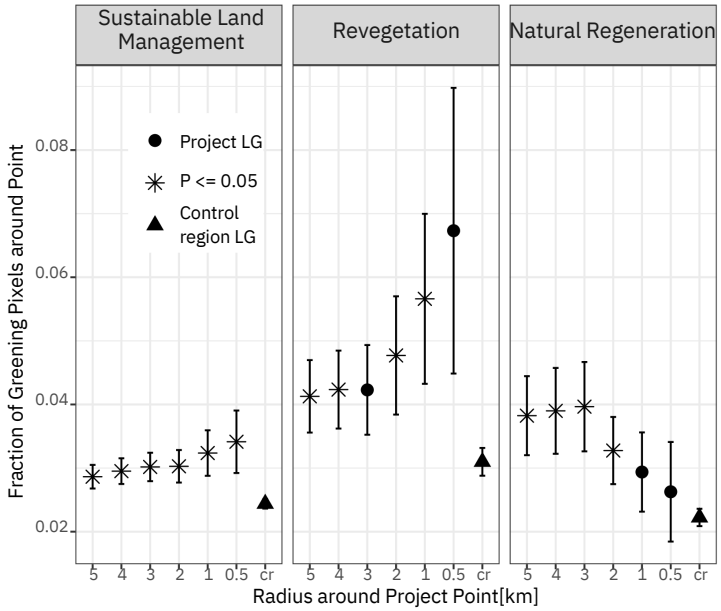


Figure A3: Mean and standard error (SE) of the fraction of LG in all SLM (left) and SLM subsets by project type (mid, right). Mean and SE are given for each buffer size around the SLM. Stars show significant differences between SLM_LG and CR_LG from a one-sided t-test. Projects with a median NDVI > 0.15 were filtered out. LG based on NIRv and NDVI data from MOD13Q1.006

All SLM projects combined show significant differences in greening pixels and total pixels per given radius as indicated with the star (see fig.A3). As the radii increase, the relative amount of LG pixels decreases. However, these still show larger fractions than the LG ratio for the control regions, which are marked with triangles. By separating the SLMs by project type, revegetation shows a similar relationship with radius size. Smaller radii yield larger LG fractions, but two buffer sizes did not show significant differences. Additionally, the standard error increases visibly as the radius size decreases. For natural regeneration, a different pattern emerges: As the radii decrease, the relative sum of greening pixels also decreases, while the standard deviations remain nearly constant across all sizes. The lowest two radii didn't show significant differences. It is also noticeable that the control region's LG fractions were the lowest across all three categories, with the smallest standard deviations.

Comparing fig.A3 with fig.3B in Ruijsch et al., 2023, we observe very similar patterns and absolute values. However, the differences in our analysis (one-sided t-test instead of two-sided and slightly different data filtering as described in section B.3.1) led to slightly different results. In particular, the p-values in our analysis tend to be lower. For example, four radii showed significant differences in our analysis for revegetation projects, whereas only one radius showed significant differences in Ruijsch et al., 2023's analysis. This difference is likely due to our analysis having more data points and using a one-sided t-test, whereas the original study used a two-sided t-test. Since a two-sided test splits the significance level between both tails of the distribution, it is more conservative and results in higher p-values compared to a one-sided test focused on a specific directional effect Ruxton and Neuhäuser, 2010.

C. *Improvements and Corrections*

To the best of our knowledge, the original study was conducted thoughtfully and without major flaws. In our reanalysis, we primarily extend the GEE scripts to contribute additional insights to the study. By comparing our results with the original findings, we aim to evaluate the robustness of these findings. Next, each paragraph describes a modification we added to the GEE analysis.

C.1 Considering regional differences in Africa

The partitioning of Africa is based on the SLM project distribution across the continent, with each quadrant containing multiple projects. Given Africa's geographical extent in both latitude and longitude, we expect the results to vary accordingly Shi et al., 2023. The analysis is then geographically divided into these quadrants, and the differences between them are considered. Fig.A4 shows the SLM project locations and their control regions differentiated by each quarter.

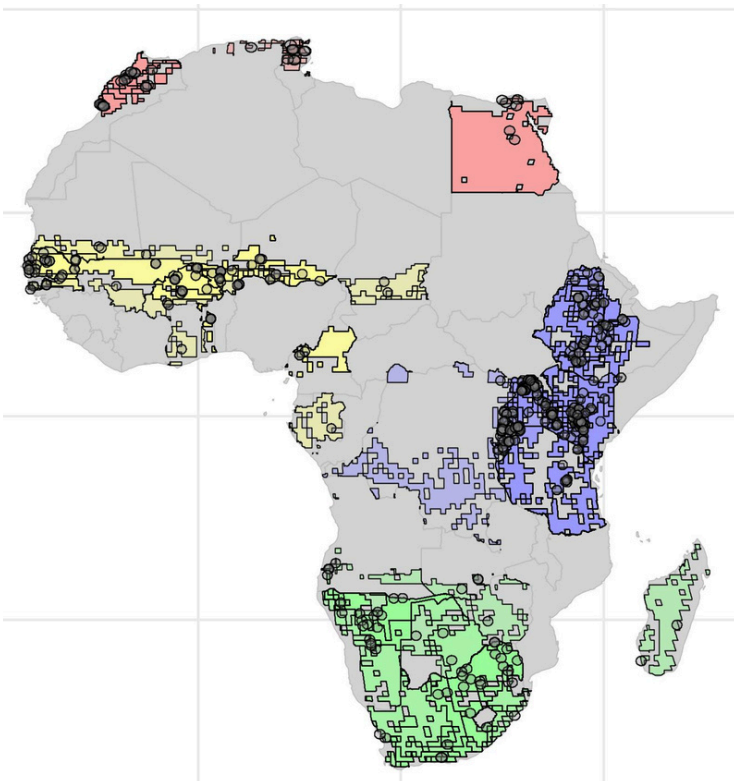


Figure A4: The grey dots represent landscape restoration projects. Projects are rendered with low opacity, meaning that darker spots show high project density. A buffer is added around the project point coordinate to make it visible. The highlighted areas are the control regions against which the LG around the projects are compared. Control regions share the same aridity index, land use, and country boundary with the project. Colors indicate different quadrants.

Control regions are differentiated based on three factors: country, aridity index, and land cover. A unique ID was derived for each pixel by combining the three categories. Pixels sharing the same ID, which represent equal aridity, country, and land cover type, are grouped within a specific control region. This provides distinguishable analyses where only similar regions are compared/used for detecting greening processes. Spatial differentiation through latitudinal and longitudinal splits reveals that a significant number of SLM projects were conducted in eastern Africa (see fig.A4). Blue parts represent the east (339 out of 516 SLM projects), yellow parts the west (with 92 SLM projects), read north (with 31 SLM projects), and green the south (with 54 SLM projects) of Africa.

Loading Data

We created a GeoTIFF file for each of the North, East, West, and South regions of Africa in Google Earth Engine (using the script listed in table A3), with the outputs shown in fig. A4. Then we loaded the created data into R. We provide an example for northern Africa here:

```
tif_n <- rast("quarter_africa/NESW_tif/north.tif")
extent_n <- ext(tif_n)
```

We used the output from Script 5 by Ruijsch et al., 2023, including LG data, and control region data, and loaded it into R as well:

```
# SLM
setwd("evi_ndvi/")
files <- list.files(pattern = "project", full.names = F)
x <- sub(".*v12_(\\d{4}).*", "\\1", files)
var_name_p <- paste0("slm_", x)
slm_list <- lapply(seq_along(files), function(x){
  res <- read.csv(files[x])
  return(res)
})
names(slm_list) <- var_name_p #give df names for clarification
index_ndvi_zero <- slm_list$slm_0500$index[slm_list$slm_0500$NDVImask == 0 |
                                           is.na(slm_list$slm_0500$NDVImask)]

slm_list <- lapply(slm_list, function(df) {
  df <- df %>%
    mutate(mean = round(mean, 10)) %>%
    filter(NDVImask != 0 & !is.na(NDVImask)) %>%        # filtering out 0 and NA
    # distinct(mean, definition, .keep_all = TRUE) %>%  # drop duplicates
    mutate(slm_lg = mean) %>%
    select(definition, index, slm_lg, type, X, Y) %>%
    arrange(index)

  return(df)
})

# combined regions
cr <- read.csv("mean_AI_country_LU_region_slmMask0.csv") %>%
  filter(!index %in% index_ndvi_zero) %>%
  group_by(index) %>%
  summarise(cr_lg = mean(mean)) %>%
  select(cr_lg,index)

# combine SLM and CR
slm_cr <- lapply(slm_list, function(x){
  x <- x %>% inner_join(y = cr, by = "index")
})

# CR subset by project type
types_s <- unique(slm_cr$slm_0500$type)
subnames <- paste0("type_",gsub("\\s+", "", types_s)) # deleting white spaces
```

```
sub_ls <- list()

for (i in types_s) {
  types <- gsub("\\s+", "", i)
  sub_ls[[types]] <- lapply(slm_cr, function(x){
    x %>% filter(type == i)
  })
}
```

Split the data into the quartes

Now we only extract the data for the quarter of Africa that we want to analyze (here, again, the northern quarter):

```
# Convert dataframe to spatial points
slm_cr_vec <- lapply(slm_cr, function(x){
   vect(x, geom = c("X", "Y"), crs = crs(tif_n))
   }
)

# north
slm_cr_n <- lapply(slm_cr_vec, function(x){
   x[extent_n, ] %>%
   as.data.frame(.)
   }
)
```

Calculate the results

With the list slm_cr_n that contains data frames with LG data for each of the radii around the projects, we apply the t-test function as shown in section B.3.4. We repeated this procedure for each quarter and generated the plot shown in fig.1 shown on the first page with similar code as shown in the section B.3.5.

Results

Only East Africa displayed continuous significant differences between greening around the SLM projects and greening in the control region. For every other part of Africa, there was no significant difference to be found. This result is not entirely surprising, given the lower number of data points available for the other regions. Interestingly, in some cases, control regions exhibited larger fractions of LG than the radii around the SLM projects. The largest fractions of LG were detected in northern Africa, where the control regions showed a substantially greater mean compared to all the other control region means. For the whole African continent, such as for the eastern part only, all radii resulted in significant differences (see fig.A3), as the results show a very similar pattern as in East Africa since the data are dominated by SLM projects in East Africa.

C.2 Choice of Vegetation Index

Introducing NIRv and kNDVI as alternative Vegetation Indices

The original study by Ruijsch et al. (2023) employed NDVI and EVI for measuring LG, as these indices are both well-established in vegetation monitoring and readily accessible in the widely used MOD13Q1.006 product Son et al., 2014; Xue and Su, 2017.

For our comparative analysis, we tried near-infrared reflectance of vegetation (NIRv) as an alternative, calculating it according to the methodology outlined in Zhang et al., 2022. Given the similarity between median EVI and NIRv values, we anticipated that substituting would not significantly affect our results (see fig.A5). Moreover, NIRv's stronger correlation with gross primary productivity (GPP) suggested it might provide enhanced measurement precision Badgley et al., 2017.

The NIRv is, according to Zhang et al., 2022, defined as

$$\text{NIRv} = (\text{NDVI} \times C) \times \rho_2 \tag{3}$$

where the parameter C is set to 0.08, as recommended by Badgley et al., 2017. rho_2 is the MODIS band 2 which is radiation with wavelengths between 841 and 876 nm.

Using kNDVI brings a number of benefits, which also do yield some disadvantages as stated in the following: kNDVI captures higher-order differences between the NIR and red reflectance bands, allowing it to model nonlinear relationships more effectively than traditional indices such as NDVI.. Unlike NDVI and other vegetation indices, which suffer from saturation in densely vegetated areas, kNDVI can handle higher vegetation cover and peak seasonal values due to its kernel-based formulation. By adjusting the kernel parameter, kNDVI can increase sensitivity to specific vegetation conditions, making it adaptable to different biomes or regions, including sparsely and densely vegetated areas. On the other hand, kNDVI relies on kernel parameters, particularly the lengthscale Sigma, which need to be carefully tuned or estimated. Misestimation can lead to suboptimal results. Also, the performance of kNDVI depends heavily on the quality of the input data. Sensor noise, clouds, and non-vegetated regions (e.g., water or snow) can introduce errors if not properly handled Wang et al., 2023.

The kNDVI is, according to Camps-Valls et al., 2021, defined as

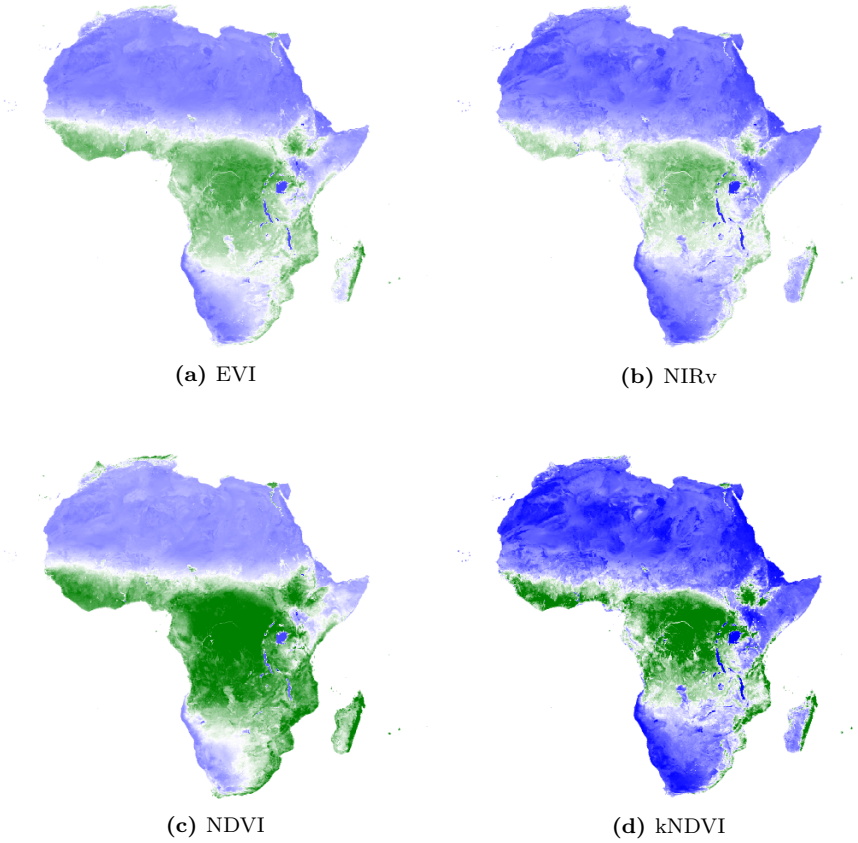
$$\text{kNDVI} = \tanh\left(\left(\frac{\text{NIR} - \text{red}}{2\sigma}\right)^2\right) \tag{4}$$

where NIR stands for near-infrared radiation and the σ parameter controls the notion of distance between the NIR and red bands and was set to 1 by us.

The pipeline of getting results for the new vegetation indices is very similar to the above-shown pipeline to reproduce the study. In the Google Earth Engine, NIRv and kNDVI were calculated and LG based on them was calculated. After that, each combination with single vegetation indices or combining two vegetation indices was tried to calculate LG around the SLM projects and in the control regions. For that the scripts 1 and 5 from Ruijsch et al., 2023 were adapted as shown in table A3.

Afterward, we applied the same pipeline to calculate fractions of LG and significances as described before when reproducing the study. We collected fractions of greening and p-values from all combinations of vegetation indices and visualized the combined results of p-values in 2, and fig.A8, and the combined results of fractions of greening in fig.A6, and fig.A7.

Combined project types for multiple combinations of VIs show certain differences in variation and mean per radius. With decreasing radii, the variation and mean between the VI combinations increase. Similar results for the revegetation projects, but with far higher means and variances (see fig.A6). Natural regeneration on the other hand displays a different pattern, where the means in general decrease with



(a) EVI **(b)** NIRv

(c) NDVI **(d)** kNDVI

Figure A5: Illustration of EVI, NIRv, NDVI und kNDVI for the African continent. Green colors represent larger values and blue colors lower values.

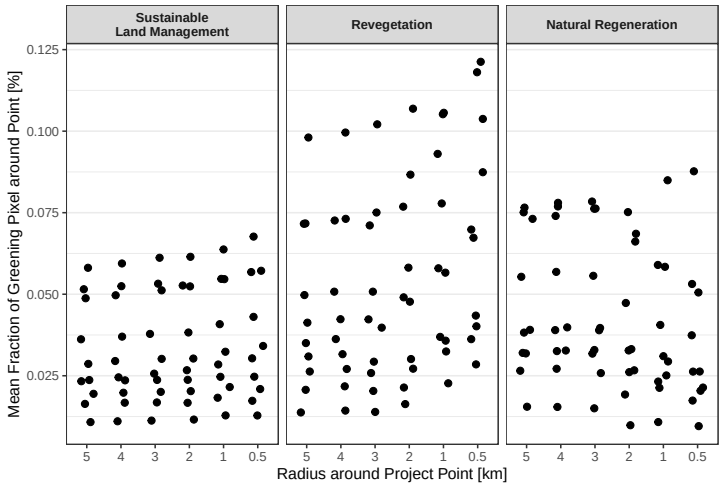


Figure A6: Mean of the fraction of LG in all SLM (left) and SLM subsets by project type (mid, right). For each vegetation index (NDVI, EVI, NIRv, and kNDVI) and each combination of two of them, there is one point for each radius around the SLM points. Projects with a median NDVI < 0.15 were filtered out. LG based on data from MOD13Q1.006

decreasing radii, but due to two large fraction means from a VI combination the variation remains roughly the same. Without these points, the variation would decrease drastically with decreasing radius.

The comparison of LG by radii and control region yields interesting patterns. The shared projects show reducing variances with decreasing radii, whereby many differences were found to be significantly different. Similarly seen for revegetation, where lower radii around projects result in nearly all differences

being significant with very low variance per radius. Natural regeneration on the other hand, shows the clearest relationship between p-values and radius size. Higher radii result in all compared vegetation indices to be significant from the control region, but by lowering the radius below 3 km, the variance increases visibly and all tests showed non-significance for radii 1 and 0.5 km, displaying an exponential relationship between radius and significant differences (see fig.2).

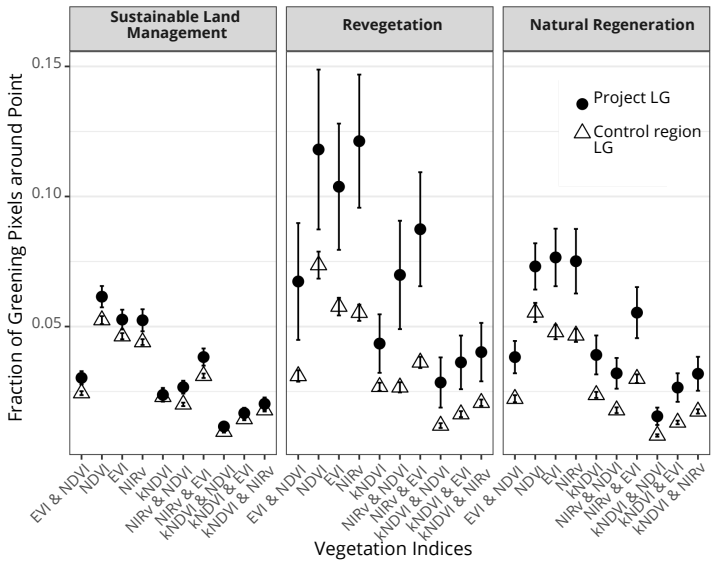


Figure A7: Mean and standard error (se) of the fraction of LG in all SLM (left) and SLM subsets by project type (mid, right). For each vegetation index (NDVI, EVI, NIRv, and kNDVI) and each combination of two of them, there is one column. The mean and standard error of fractions of LG for corresponding control regions are shown as triangle. For all SLMs a radius of 2 km was chosen, for Natural Regeneration a radius of 5 km, and for Revegetation a radius of 0.5 km. Projects with a median NDVI < 0.15 were filtered out. LG based on data from MOD13Q1.006

Inspecting Combined Projects revealed similar trends in LG fractions for pixels within a 2 km radius for both SLM and control region greening (see fig.A7). The highest mean values per VI group were achieved for NDVI, EVI, and NIRv individually. Multiple indices did not provide that large averages. When accounting for revegetation projects only, there is much more variance within the indice fractions as seen in the error bars. Also, SLM-LG and CR-LG differ visibly more, whereas for all VIs the SLM-LG means were consistently larger than CR-LG. Similarly, for natural regeneration, NDVI, EVI, and NIRv displayed the highest proportions of LG. Every combination of the other indices on the other hand showed to be lower, where the combination of NIRv&NDVI, kNDVI&EVI, and kNDVI&NIRv were the lowest. In general, the fraction distribution across the indices followed the same trend across all three project-type differentiations, for both SLM and CR LG, but differed in their absolute values.

SLMs for all project types with a selected radius of 2 km, which yielded the best p-values, showed consistently larger relative amounts of LG for SLM greening for all VIs compared to the CR-LG (see fig.2). The single VI greening fractions always yielded the largest values (see fig.A7). This was the case for all 3 types. Revegetation, with 0.5 km, had substantially differing values within each VI comparison of SLM-LG and CR-LG, also much larger variances for SLM-LG. Combined VIs were showing lower values, but in general the same trend across all project types. Natural regeneration, with 5 km radius, had visibly lower variances for the VI combinations. In general, revegetation showed for SLM-LG the largest values. CR-LG was roughly the same regarding the values of the fractions.

For comparison of LG around SLM projects (SLM-LG) and in the control regions (CR-LG), nearly all vegetation index combinations in all three project types yielded significant differences (see fig.A8).

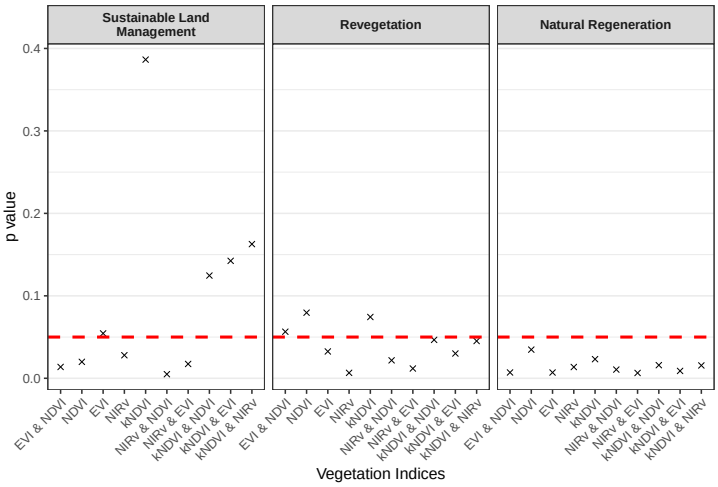


Figure A8: P values that show if the fractions of LG are significantly different from the fractions of LG in the same control regions in all SLM (left) and SLM subsets by project type (mid, right). For each vegetation index (NDVI, EVI, NIRv, and kNDVI) and each combination of two of them, there is one column. For all SLMs a radius of 2 km was chosen, for Natural Regeneration a radius of 5 km, and for Revegetation a radius of 0.5 km. Projects with a median $NDVI < 0.15$ were filtered out. LG based on data from MOD13Q1.006

Natural regeneration had no not significant results. Revegetation on the other hand did not result in significant differences for EVI&NDVI, NDVI, and kNDVI. In SLM-shared projects, 5 combinations did not yield significant differences.

C.3 Using updated MODIS Data

Ruijsch et al., 2023 used MODIS data including NDVI and EVI for 20 years between 2001-01-01 and 2022-01-01. They stated many breakpoints in the last year when breakpoints were allowed (in 2019) and concluded from this that a particularly large number of projects have had an impact this year and that this is where the LG trend begins. We wonder whether this claim is really valid. Given the short duration of data, two years, after the breakpoint, we question whether the identified breakpoints are truly indicative of significant changes in greening trends. After all, it could simply be that in the last two years, the growing conditions for plants happened to be slightly better than before because of natural variations. To look at this, we repeat the analysis and use more recent MOD13Q1.061 data up to 2024 instead of the MOD13Q1.006 data up to 2022. We also set the minSpacing parameter to 0.2 instead of 0.15 so that breakpoints are allowed to be less at the beginning or end of the time series. It should be noted that the newer MODIS data has been improved by newer quality management methods, consisting of new calibration and polarization correction, and is therefore not exactly comparable with the older data set (Didan et al., 2015).

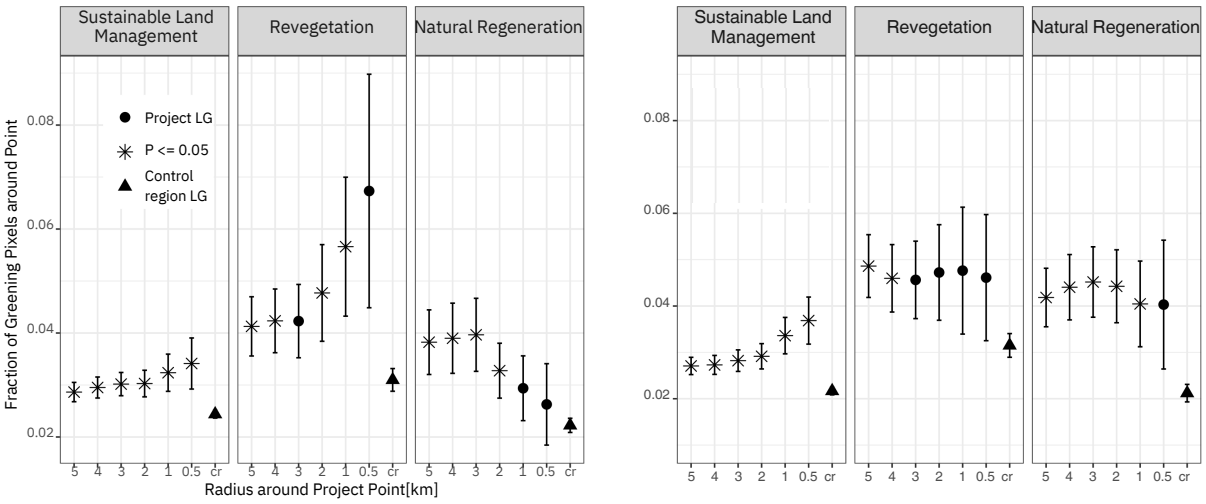
(2) Instead of using MOD13Q1.006 from January 2001 until January 2022, we used the updated product MOD13Q1.061 from January 2001 until November 2024. Particularly, we were interested in how this changed the breakpoint analysis. Ruijsch et al., 2023 extracted NDVI and EVI from the, by now deprecated satellite product MOD13Q1.006. The breakpoint analysis in the original paper ran with data from 2001-01-01 to 2022-01-01. As suggested by the authors, it is striking, that pixels with breakpoints and LG spiked from 2017 to 2018 (see fig.A10). In contrast, the increase in previous years was more moderate. The authors concluded that in 2018 a huge area became greener compared to its surroundings. Additionally, we checked the trend of all pixels with breakpoints (with and without LG) revealing an even more pronounced increase, with their numbers doubling from 2017 to 2018. As the breakpoints peaked at the end of the time series, we wondered if this was caused by the low number of years (3) after the breakpoints.

... we checked this by using new MODIS data ....

questioned the validity of the peaking breakpoints at the end of the time series could be influenced and questioned. After all, it could simply be that in the last two years, the growing conditions for plants happened to be slightly better than before. To look at this, we repeat the analysis and use more recent MOD13Q1.061 data up to 2024 instead of the MOD13Q1.006 data up to 2022. We also set the minSpacing parameter to 0.2 instead of 0.15 so that breakpoints are allowed to be less at the beginning or end of the time series. It should be noted that newer quality management methods have improved the newer MODIS data and are therefore not exactly comparable with the older data set.

When analyzing the data by project type, distinct outcomes emerged regarding the proportion of greening pixels at various distances from project sites (see Fig. A9(a)). For combined project types, significant differences were observed across all radii when comparing SLM-LG and combined region-LG (CR-LG). The CR-LG fractions were always displayed as the right tick at each project-type plot. Revegetation displayed significant results for radii larger than 1 km, for lower radii the fraction results were too similar to the control regions. In natural regeneration were more radii values with significant differences, only 0.5 km showed no notiecable difference. When only accounting for SLM without subsetting, the largest effect of greening is detected with a 0.5 km radius. When considering subsetting, revegetation yields the most differences when using 5 km radius and natural regeneration for 3 km.

The older dataset (see fig.A9(a)) yields the same content as the newer one (see fig.A3). The updated dataset (MOD13Q1.061) led to noticeable changes, as shown in panel (b). While combined project types still exhibit similar trends and significance, the uncertainty and variability in the fraction values have increased. The same can be said about revegetation and natural regeneration, where visibly increase



(a) Results based on MOD13Q1.006

(b) Results based on MOD13Q1.061

Figure A9: Mean and standard error of the fraction of LG in all project types (left) and specific project types (middle, right). Mean and standard error are given per buffer size around the SLM. Stars indicate a significant increase in LG around a project compared to a control region, based on a one-sided t-test. Projects with a median NDVI < 0.15 were filtered out. LG is based on EVI and NDVI data from MOD13Q1.006 (left) and MOD13Q1.061 (right).

of variability can be derived from the error bars. Also noticeable, with the updated time series seems natural regeneration exhibits more significance towards smaller radii, shifting from less than 2km to 1km significant differences. Revegetation also changed in significance, where now radii smaller than 2km dont result in differences compared to the radii 3 and 0.5km as in (a). Another trend that arose from updated MODIS data is the lowering fraction values with decreasing radius values for revegetation. Besides that, the trends from (a) and (b) are nearly similar except the absolute values.

The shorter time series from MOD13Q1.006 displays a general trend of increasing detected breakpoint frequency per year, where the last year, 2018, contained the most breakpoints. Due to the filtering process, many detected breakpoints were excluded which resulted in roughly 10 times fewer pixels (see fig.A10). The fraction of pixels with EVI or NDVI where LG was found display a different pattern for detected breakpoints. Before, NDVI & EVI showed nearly similar relationship before and after filtering. After accounting for breakpoints in pixels with LG only, the last year had a decrease in detections. In general, NDVI had troughout higher detection rated compared to EVI across all three comparisons. Similar outcomes are shown in the row below, for MOD13Q1.061, except for the frequency intensity. Here, they exceeded the detected breakpoints from the previous time series but still had the same trend during the overlapping time. The general trend for filtered breakpoints also seems quite similar, also besides the further increase of the updated time series. The share image of LG with detected breakpoints per pixel yields an increasing frequency towards the last year, while still moving in similar frequency intensities as the older time series.

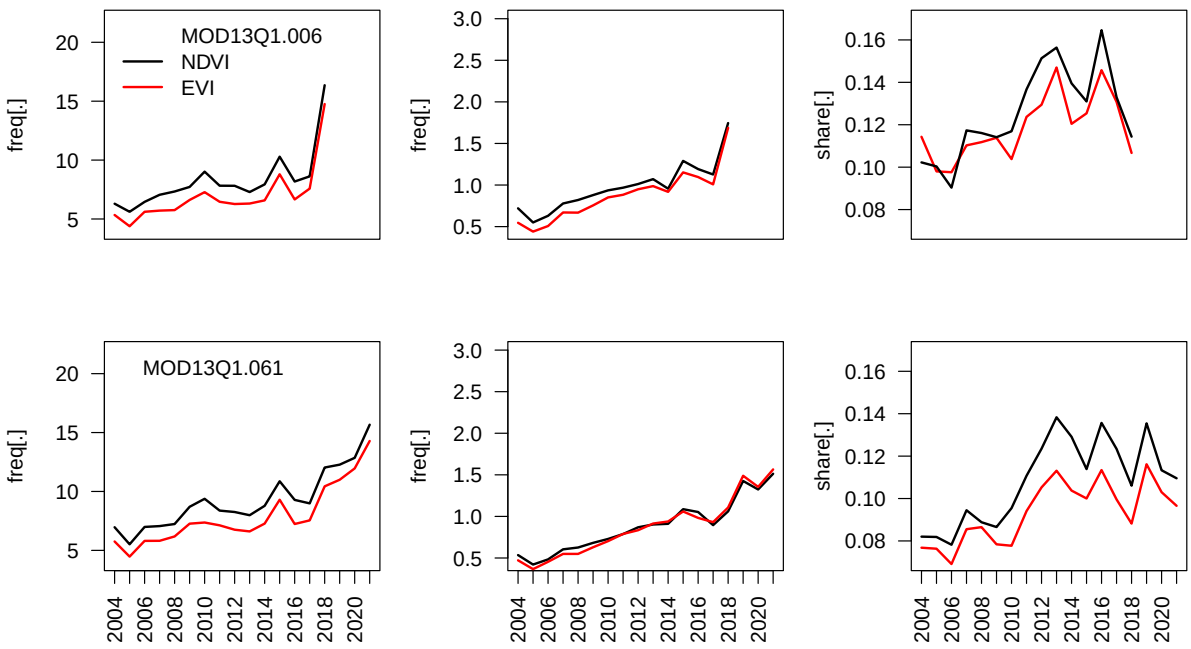


Figure A10: For the upper row MOD13Q1.006 background adjusted (spatial context approach) NDVI or EVI data from 2001-01-01 to 2022-01-01 was used with the function argument minSpacing = 0.15 (as Ruijsch et al., 2023). For the lower row MOD13Q1.061 from 2001-01-01 to 2024-11-01 with minSpacing = 0.13. Left: frequency of all pixels with identified breakpoints. Mid: frequency of pixels with identified breakpoints that passed the filtering and were classified as EVI or NDVI LG. Right: fraction of pixels with EVI or NDVI LG in pixels with identified breakpoints. A Breakpoint could only be detected once per pixel by the BFAST algorithm (ee.Algorithms.TemporalSegmentation.StructuralChangeBreakpoints()).

C.4 Masking other projects

Besides comparing SLM_LG with CR_LG in a t-test, we assess, whether the t-test results differ, if masking the SLM projects during the calculation of the control region's LG CR_LG. We introduce this additional step because the SLM-LG should be compared with non-SLM-LG to assess whether SLMs are contributing to greening, as their effect is best evaluated by excluding their own influence from the control region. Masking SLM projects within the control region ensures that only non-SLM areas are used for comparison, isolating the control region's greening effect from any SLM-induced influence. The appended code does this exemplarly for masking a buffer of 0.5 km.

```
# SLM
rm(list = ls())
setwd("evi_ndvi/")
files <- list.files(pattern = "project", full.names = F)
x <- sub(".*v12_(\\d{4}).*", "\\1", files)
var_name_p <- paste0("slm_", x)
slm_list <- lapply(seq_along(files), function(x){
  res <- read.csv(files[x])
  return(res)
})
names(slm_list) <- var_name_p #give df names for clarification

index_ndvi_zero <- slm_list$slm_0500$index[slm_list$slm_0500$NDVImask == 0 |
                                           is.na(slm_list$slm_0500$NDVImask)]
slm_list <- lapply(slm_list, function(df) {
```

```
  df <- df %>%
    mutate(mean = round(mean, 10)) %>%
    filter(NDVImask != 0 & !is.na(NDVImask)) %>%        # filtering out 0 and NA
    # distinct(mean, definition, .keep_all = TRUE) %>%  # drop duplicates
    mutate(slm_lg = mean) %>%
    select(definition, index, slm_lg, type) %>%
    arrange(index)

  return(df)
})

# control regions
files <- list.files(pattern = "mean_AI_country_LU_region", full.names = F)
x <- sub(".*Mask(\\d+).*", "\\1", files)
var_name_cr <- paste0("km_", x)
var_name_cr[1] <- "nomask"
cr_list <- lapply(seq_along(files), function(x){
  res <- read.csv(files[x])
  return(res)
})
names(cr_list) <- var_name_cr

cr_list <- lapply(cr_list, function(df) {
  df <- df %>%
    filter(!index %in% index_ndvi_zero) %>%
    mutate(cr_lg = mean) %>%
    select(cr_lg,sumsum,index)
  return(df)
})

# combine SLM and CR
lg_cr <- lapply(cr_list, function(x){
  x %>%
    group_by(index) %>%
    summarize(new = mean(cr_lg, na.rm = TRUE))
}) %>% bind_cols() %>% select(1,2,4,6,8,10,12,14)

names(lg_cr) <- c("index_cr", var_name_cr)

slm_cr <- lapply(slm_list, function(x) bind_cols(x, lg_cr))
```

Next, we created an adapted t-test function:

```
T_test_custom <- function(slm_cr, var_name_cr, alternative) {
  # Initialize data frames for results
  pvals <- data.frame(matrix(nrow = length(names(slm_cr)), ncol = length(var_name_cr)))
  colnames(pvals) <- var_name_cr
  row.names(pvals) <- names(slm_cr)

  mean_x <- data.frame(matrix(nrow = length(names(slm_cr)), ncol = length(var_name_cr)))
  colnames(mean_x) <- var_name_cr
```

```
  row.names(mean_x) <- names(slm_cr)

  mean_y <- data.frame(matrix(nrow = length(names(slm_cr)), ncol = length(var_name_cr)))
  colnames(mean_y) <- var_name_cr
  row.names(mean_y) <- names(slm_cr)

  # Populate pvals, mean_x, and mean_y
  for (i in names(slm_cr)) {
    for (j in var_name_cr) {
      t_result <- t.test(
        slm_cr[[i]]$slm_lg,
        slm_cr[[i]][[j]],
        paired = FALSE,
        alternative = alternative,
        var.equal = FALSE,
        mu = 0
      )

      pvals[i, j] <- t_result$p.value
      mean_x[i, j] <- t_result$estimate["mean of x"]
      mean_y[i, j] <- t_result$estimate["mean of y"]
    }
  }

  # Combine results into a single list
  results <- list(
    P_Values = pvals,
    Mean_X = mean_x,
    Mean_Y = mean_y
  )

  return(results)
}
```

Now, we can call the function and get our output:

```
# Apply function and display tables in LaTeX format
all_prj_1sided <- T_test_custom(slm_cr = slm_cr,
                                var_name_cr = var_name_cr,
                                alternative = "greater")
all_prj_2sided <- T_test_custom(slm_cr = slm_cr,
                                var_name_cr = var_name_cr,
                                alternative = "two.sided")
```

We show the output in the tables A6 and A7.

Table A6: P-values for all possible control regions after a One-Sided T-Test

	nomask	km_05	km_1	km_2	km_3	km_4	km_5
slm_0500	**0.02522**	**0.02520**	0.02518	0.02516	0.02514	0.02524	0.02539
slm_1000	**0.01480**		**0.01476**	0.01475	0.01473	0.01482	0.01495
slm_2000	**0.01375**			**0.01369**	0.01366	0.01377	0.01395
slm_3000	**0.00719**				**0.00713**	0.00720	0.00732
slm_4000	**0.00885**					**0.00887**	0.00902
slm_5000	**0.01711**						**0.01744**

Table A7: Means of fraction of LG for all possible control regions after a One-Sided T-Test

	nomask	km_05	km_1	km_2	km_3	km_4	km_5
slm_0500	**0.02439**	**0.02438**	0.02438	0.02438	0.02438	0.02439	0.02440
slm_1000	**0.02439**		**0.02438**	0.02438	0.02438	0.02439	0.02440
slm_2000	**0.02439**			**0.02438**	0.02438	0.02439	0.02440
slm_3000	**0.02439**				**0.02438**	0.02439	0.02440
slm_4000	**0.02439**					**0.02439**	0.02440
slm_5000	**0.02439**						**0.02440**

Table A6 presents the p-values from t-tests comparing local SLM greening (SLM-LG) with control region greening (CR-LG) under various masking buffer sizes. The masking buffers exclude pixels within a certain radius around other SLM projects within the CR of the analyzed SLM. For example, km_05 excludes all pixels within a radius of 0.5 km around any SLM project within the CR of the analyzed SLM before performing the comparison.

We compared the results of each SLM radius without masking (nomask) to those from control regions with larger masking radii than the SLM radius, to evaluate if excluding regions with potential SLM influence improves the results. With this approach, we wanted to check if the results get better when we invent a buffer range in which pixels are not considered as part of the SLM region nor as the control region and therefore excluded from the t-test.

For most SLM radii, the km_03 masking produced the lowest p-value. However, for slm_4000 and slm_5000, the lowest p-values were observed with masking radii of 4 or 5 km. However, these differences are minimal and can be considered negligible for the purpose of our analysis.

Table A7 shows the fractions of LG of the control regions depending on whether they are masked or not shows the same result. Masking SLM projects with a larger radius of 5,km results in a marginal increase in the fraction of LG, but this effect is negligible for our analysis.

This suggests that the masking buffer has minimal impact on the comparison between SLM_LG and CR_LG, indicating that the masking step does not significantly contribute to the analysis. This result is not entirely surprising, as control regions are generally large, and only small portions of them are influenced by SLM projects.

D. Conclusion

Splitting Africa revealed significant regional differences, with the eastern part of the continent hosting the majority of the conducted projects. This was reflected in the analysis, where only the eastern region showed significant differences when compared to the control region (see fig.1). Based on this approach, this study and its findings only find impactable evidence for the eastern part of Africa where most of the projects have taken place (see fig.A4). Regarding applications and realizations of further projects, accounting for the location within Africa is crucial due to the unproven effects of projects in western, northern, and southern Africa. In East Africa, both NDVI and EVI confirmed that greening strategies had a relevant and consistent effect, regardless of the radius size used for comparison. For the other three regions, more projects would be needed to secure reliable recommendations and instructions regarding greening effects for land management projects.

Replacing the initially used vegetation indices NDVI and EVI with kNDVI and NIRv did not yield significant changes. Across all three partitionings by project type, each index or combination of 2 followed a certain trend which was roughly the same for all (see fig.A6). Additionaly, each combination of 2 VIs yielded less greening within the projects radius per control region. This arises trough additional filtering of both indices where more potential greening pixels are excluded by only accounting for a single index (see fig.A7). Some indices and certain combinations did not result in significant differences, mostly when kNDVI was included (see fig.A8). Due to the fact, that besides 5 combinations for SLM and 3 in revegetation are insignificant, no improvement in replacing the already used indices was found. Therefore, the replacement did not yield new results. Still, the chosen radius for the project impacts the uncertainty noticeably (see fig.2). Revegetation, across nearly all included combinations, showed significant differences for radii smaller than 2 km whereas the lowest uncertainty comes with 0.5 km. For natural regeneration, the relationship with radius size was the opposite: radii larger than 2,km showed significant differences, indicating lower uncertainty with larger radii. Combined project types yielded more variance, with a small trend of decreasing uncertainty with decreasing radius size. Still, this relationship comes from all combinations of the included VIs which show that the trend would be nearly similar unregarding of index choice which is why NDVI and EVI have no need to be replaced.

The updated MODIS data revealed different results compared to the original paper Ruijsch et al., 2023. The trend of breakpoints, which was previously proposed, is now more pronounced with the updated data (see fig.A10). As well as for all detected breakpoints as for breakpoints that passed the filtering, this trend is visible. With a nearly doubled amount of detected breakpoints, the proposed result regarding the efficiency of landscape restoration can be underlined. Still, there remains doubt that this is only caused by greening trends or due to other factors that were not included. More investigation and research are advised in order to securely assign the whole development to greening caused by the projects.

Masking other projects within the same control region while analyzing the greening effect of SLMs across different radii did not result in significant differences. The p-values showed minimal variation (see table A6). This suggests that including or excluding SLMs from the control region does not lead to substantial differences, as shown consistently across all SLM radii. This is shown for all possible SLM radii consistently. Meaning, there is no influence of SLM on these scales impacting the greening effect. Another explanation would be, that the greening patterns are already consistent, no matter what buffer size is chosen. The same can be derived from table A7, where differing radii for SLMs were compared to every mask buffer. The same result, equal proportions across all radii for each buffer possibility, was found. In conclusion, masking does not provide any significant benefits or improvements for the analysis, and therefore, does not enhance analysis performance

References

Badgley, G., Field, C. B., & Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances*, *3*(3), e1602244. https://doi.org/10.1126/sciadv.1602244
C parameter in NIRv.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., Jung, M., Gamon, J. A., Reichstein, M., & Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, *7*(9), eabc7447. https://doi.org/10.1126/sciadv.abc7447
kndvi.

Didan, K. (2015). MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. https://doi.org/10.5067/MODIS/MOD13Q1.006
MODIS Database including NDVI and EVI.

Didan, K. (2021). MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061. https://doi.org/10.5067/MODIS/MOD13Q1.061

Didan, K., Munoz, A. B., Solano, R., & Huete, A. (2015). MODIS vegetation index user's guide (MOD13 series). *University of Arizona: Vegetation Index and Phenology Lab*, *35*, 2–33. Retrieved February 14, 2025, from https://modis-land.gsfc.nasa.gov/pdf/MOD13_User_Guide_V61.pdf

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Harris, I., & Osborn, T. (2020). CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901-Dec. 2019). https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9/
Aridity Index Data base.

Ruijsch, J., Teuling, A. J., Verbesselt, J., & Hutjes, R. W. A. (2023). Landscape restoration and greening in Africa. *Environmental Research Letters*, *18*(6), 064020. https://doi.org/10.1088/1748-9326/acd395

Ruxton, G. D., & Neuhäuser, M. (2010). When should we use one-tailed hypothesis testing? *Methods in Ecology and Evolution*, *1*(2), 114–117. https://doi.org/10.1111/j.2041-210X.2010.00014.x

Shi, S., Yang, P., & Van Der Tol, C. (2023). Spatial-temporal dynamics of land surface phenology over Africa for the period of 1982–2015. *Heliyon*, *9*(6), e16413. https://doi.org/10.1016/j.heliyon.2023.e16413

Son, N., Chen, C., Chen, C., Minh, V., & Trung, N. (2014). A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agricultural and Forest Meteorology*, *197*, 52–64. https://doi.org/10.1016/j.agrformet.2014.06.007
NDVI EVI usage.

Steel, D. G., & Holt, D. (1996). Analysing and Adjusting Aggregation Effects: The Ecological Fallacy Revisited. *International Statistical Review / Revue Internationale de Statistique*, *64*(1), 39. https://doi.org/10.2307/1403423

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., & Friedl, M. A. (2019). Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, *222*, 183–194. https://doi.org/10.1016/j.rse.2018.12.013
Land Cover Classes.

Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, *114*(12), 2970–2980. https://doi.org/10.1016/j.rse.2010.08.003

Wang, Q., Moreno-Martínez, Á., Muñoz-Marí, J., Campos-Taberner, M., & Camps-Valls, G. (2023). Estimation of vegetation traits with kernel NDVI. *ISPRS Journal of Photogrammetry and Remote Sensing*, *195*, 408–417. https://doi.org/10.1016/j.isprsjprs.2022.12.019

WOCAT Global sustainable land management database. (n.d.). https://qcat.wocat.net/en/wocat/

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, *2017*, 1–17. https://doi.org/10.1155/2017/1353691
Often used NDVI source.

Zhang, J., Xiao, J., Tong, X., Zhang, J., Meng, P., Li, J., Liu, P., & Yu, P. (2022). NIRv and SIF better estimate phenology than NDVI and EVI: Effects of spring and autumn phenology on ecosystem production of planted forests. *Agricultural and Forest Meteorology*, *315*, 108819. https://doi.org/10.1016/j.agrformet.2022.108819
NIRv.