Comparing ‘Woody’ and ‘Non-Woody’ soil carbon distributions

Ruan de Wet

ruan@meatnaturallyafrica.com

2025-01-03

Table of contents

1 Introduction

During the VCS validation audit for the VM0042 project titled Grassland Restoration and Stewardship in South Africa (GRASS), a Forward Action Request (FAR) was raised regarding the sampling applied to monitor soil organic carbon stocks between rangelands of different tree densities (greater and less than 10% canopy cover). The wording of the FAR is provided below:

> VVB in the first verification will assess the stratified sampling applied by PP for the parameter SOCbsl,i,t (Areal-average soil organic carbon stocks in the baseline scenario for sample unit i in year t) to check the sampling approach and its conservativeness. PP has proposed to apply stratified random sampling in selecting sampling points to distinguish between rangelands of different tree densities (greater and less than 10% canopy cover).

Additional soil samples were collected in the 2023 sampling campaign to address this FAR. Soil samples are grouped into strata from rangelands with greater and with less than 10% canopy cover - labelled 'woody' and 'non-woody' strata, respectively.

1.1 Sampling approach

1.1.a Initial sampling strategy

Following the requirements of Section 8.6 of the VM0042 Methodology, soil sampling sites were randomly assigned across the project accounting area using `st_sample()` from the `{sf}` package (Pebesma 2018; Pebesma and Bivand 2023) in R (R Core Team 2024). This generates a simple random sample of points within the defined spatial features - the eligible grazing land within the active associations under the GRASS project. Variation of the site location was allowed based on field conditions within approximately 200 meters. This is to account for unforeseen edge effects from foot paths, watering sites, rocky outcrops, or any other factors that could skew the representativeness of the samples collected from that site relative to the surrounding landscape.

Initially, no distinction was made between grazing land with or without tree cover and the only strata of interest were the implementation landscapes (e.g. uMkhomazi Catchment) to ensure that sufficient sampling density was obtained within each landscape. Given that the statistical power of the soil assessments are inversely correlated with the number of strata, no further stratification was originally made within the implementation landscapes.

The minimum target sampling density per landscape was one sample site per 500 hectares. The field sampling team collected additional samples beyond this minimum requirement at their discretion based on their assessment of the variability of the topology, land use history, and ecological context of the grazing lands.

1.1.b Defining eligible grazing lands

An additional stratum was introduced with the raising of this FAR, solely for the purposes of assessing whether any meaningful distinctions exist across grazing lands of different tree densities (greater and less than 10% canopy cover).

This was necessitated because the Verified Carbon Standard (VCS) Program Definitions defines "Grasslands" as also including savannah ecosystems that have a mosaic of grass and tree covers, which aligns with the latest ecological science on open ecosystems (Bond 2019). The following description of a grassland is provided:

> Areas dominated by grasses with a density of trees too low to meet an internationally accepted definition of forest, including savannas (i.e., grasslands with scattered trees). Grasslands also include managed rangeland and pastureland that is not considered cropland where the primary land use is grazing, and which may also include grass-dominated systems managed for conservation or recreational purposes.

Within the initial validation audit, a Corrective Action Request was made by the reviewers with the expectation that a maximum woody cover density threshold of 10% should be applied as this is included in the FAO definition of a forest. This definition, however, would exclude almost all definitions of a savannah, which contradicts the VCS Program Definitions that explicitly incorporate savannahs in the definition of a grassland. Savannahs include a wide range of tree densities that have been long described in southern Africa to exceed 10% canopy cover density while still supporting grazing practices (Edwards 1983).

The complexity of contradicting forest definitions has been addressed in South Africa through legislation as well as in the South African court. This is supported by a technical definition that includes a largely contiguous canopy with a cover threshold of 75%, the dominance of trees, and the absence of grasses and fire. This definition is consistent with the objective of the VCS eligibility criteria, as it distinguishes forested land that is not grazed (due to a lack of grass dominance) from treed savannahs that *are* grazed.

To assess whether or not additional stratification is required to distinguish land that falls within the stricter definition of a savannah (>10% woody cover) from land that falls within a stricter definition of a non-treed grassland (<10% woody cover), soil carbon measurements were collected across these strata. If statistically significant difference in soil carbon concentrations can be detected between sampling sites in the more woody cover relative to the less woody land, this would suggest that further interrogation of the monitoring plan may be justified and, perhaps, additional stratification of the project area would be required.

The 2018 South African National Land Cover (SANLC) dataset was used to distinguish between the strata. This dataset is developed, maintained and distributed by the South African Department of Forestry, Fisheries, and the Environment (DFFE).

SANLC Classes 3, 4, and 42 are the only eligible land cover classes that have >10% woody canopy cover. Additional sampling sites were randomly assigned within these land cover classes. The total extent of these eligible woody classes account for approximately 8%-10% of the total area, depending on the landscape.

1.2 Analysis approach

A two-sample t-test is used in this assessment to compare measured soil carbon results across the two strata. The null hypothesis is tested that the true difference in means between the two groups is equal to 0.

To determine the required sample size in the `Woody` stratum, a power analysis has been run using the `pwr.t2n.test()` from the `{pwr}` package (Champely 2020). This test takes as input the existing sample size of the `Non-Woody` stratum, the standard deviation of the same sample as the effect size, and calculates the minimum required sample size of the `Woody` stratum to detect a difference between the samples with a power of at least 0.5.

```
     t test power calculation

             n1 = 198
             n2 = 40.93928
```

```
          d = 0.337854
  sig.level = 0.05
      power = 0.5
alternative = two.sided
```

To maximize the power of this analysis, more than 41 sites needed to be sampled from the `Woody` stratum. Given that 64 sites were sampled from the `Woody` stratum, it can be concluded that the sample size was more than sufficient.

2 Exploratory Data Analysis

2.1 Soil carbon data

An initial visual exploration of the soil data will be helpful to understand what the distributions look like.

Soil carbon (%)

Summary statistics across woody (>10% cover) and non-woody (<10% cover) strata

Group	count	mean	sd
Non-Woody	198	1.8520	1.4154
Woody	64	1.8679	1.2418

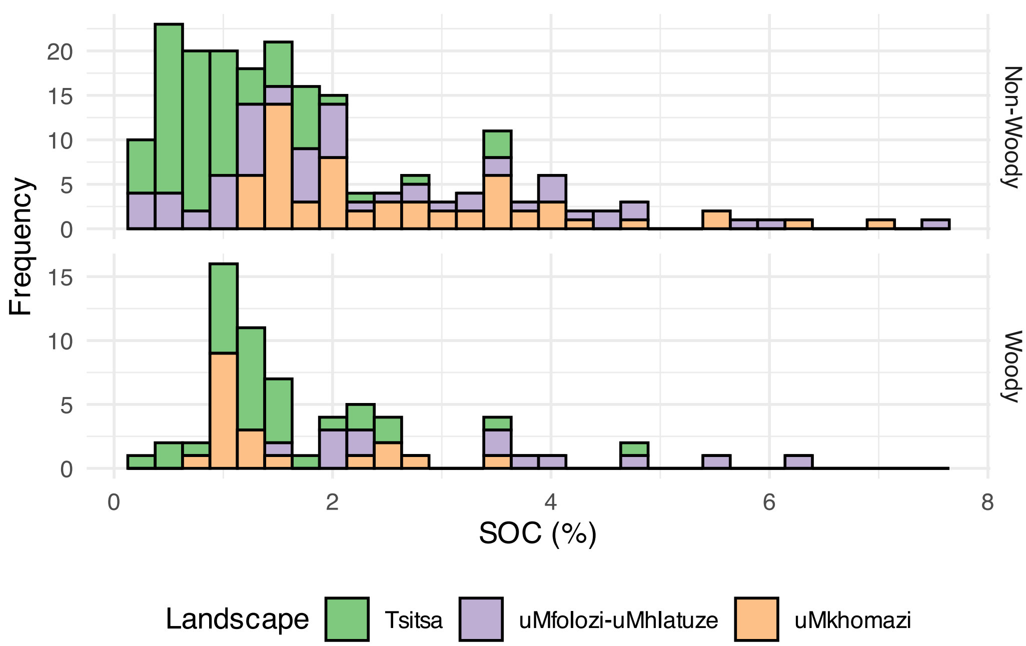

Figure 1: Soil carbon (%) distribution across target landscapes and `woody` (>10% cover) and `non-woody` (<10% cover) strata.

There is a considerable skew to the data, so a log10 transformation might better represent the SOC data as a Gaussian response variable.

Log10-transformed soil carbon (%)

Summary statistics across woody (>10% cover) and non-woody (<10% cover) strata

Group	count	mean	sd
Non-Woody	198	0.1452	0.3379
Woody	64	0.1934	0.2561

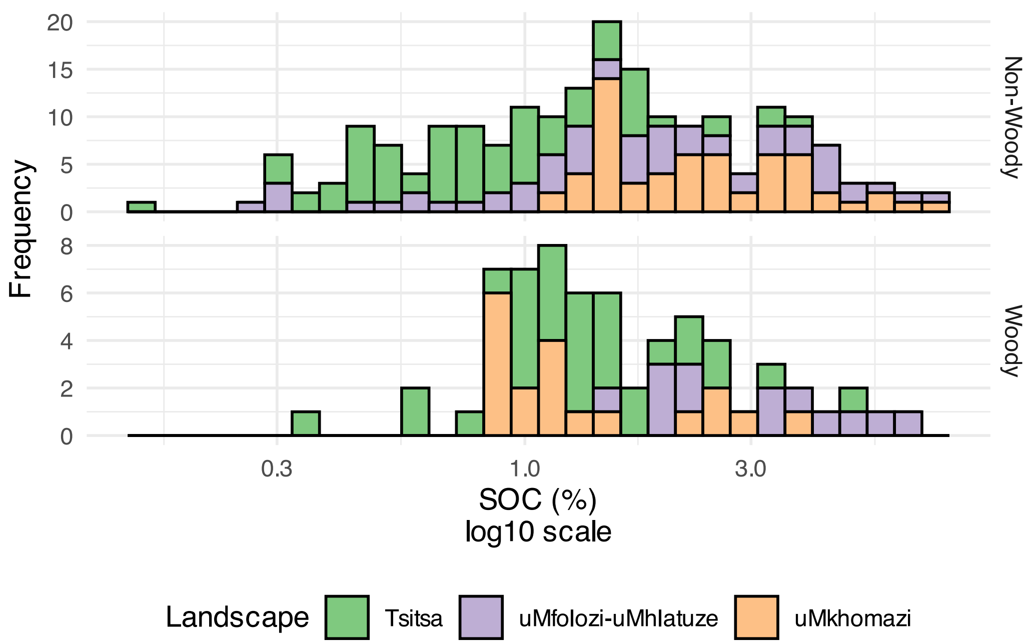

Figure 2: Log10-transformed soil carbon (%) distribution across target landscapes and `woody` (>10% cover) and `non-woody` (<10% cover) strata.

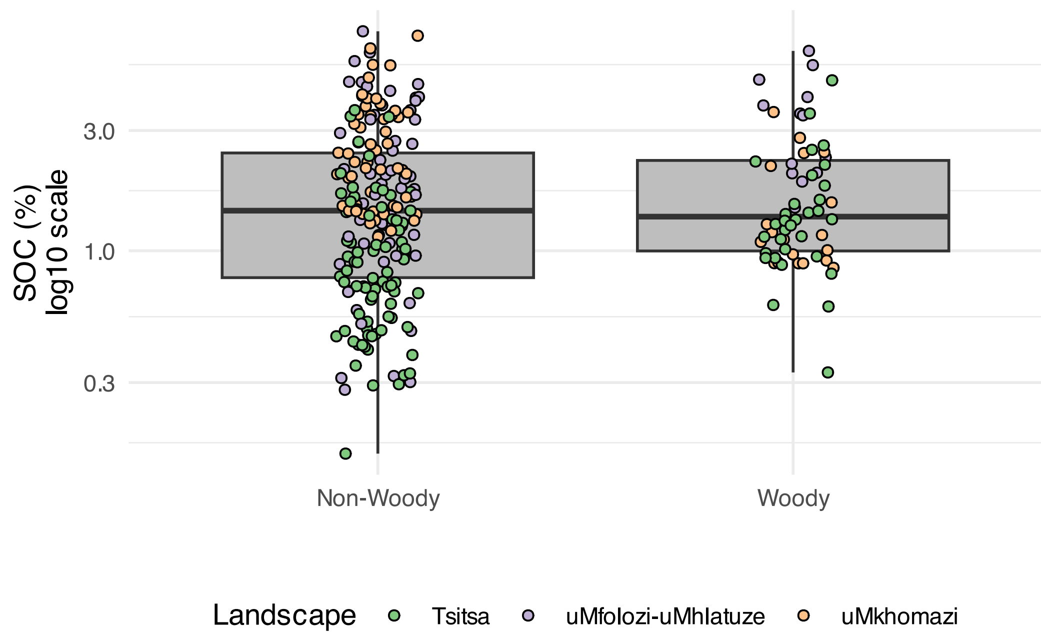

Figure 3: Log10-transformed soil carbon (%) boxplots across `woody` (>10% cover) and `non-woody` (<10% cover) strata.

2.2 GRASS VCS Project Boundary

The first VCS monitoring period includes associations that joined the GRASS project in 2021 and 2022. The boundary of the total area under management for each association was mapped, but this includes a variety of land cover classes that are ineligible for inclusion in the VCS project. Details of the eligibility criteria and further exclusions that were made go beyond the scope of this report and can be reviewed in the GRASS validation audit documentation on the public Verra registry.

The `Woody` and `Non-Woody` strata are represented within the eligible boundary areas.

Total and eligible area (ha)

Across all landscapes represented in the first monitoring period

Landscape Name	Total area (ha)	Eligible area (ha)
uMfolozi-uMhlatuze	33,301.93	27,346.82
uMkhomazi	26,092.19	18,081.98
Tsitsa	58,433.76	37,937.28

2.3 Land cover data

The SANLC2018 land cover classifications for all associations boundaries in the first monitoring period have been extracted and summarized per landscape.

After excluding ineligible land cover, the remaining eligible classes can be categorised as either `Woody` (classes 3, 4, and 42) or `Non-Woody` (all other classes). The eligible `Woody` classes account for 8%-10% of the total land cover, depending on the landscape.

South African National Land Cover (SANLC)

Eligible classes occuring across GRASS associations and their assignment to Woody and Non-Woody strata

	Class	Class Name
Woody	3	Dense Forest & Woodland (35 - 75% cc)
	4	Open Woodland (10 - 35% cc)
	42	Fallow Land & Old Fields (Trees)
Non-Woody	8	Low Shrubland (other regions)
	13	Natural Grassland
	27	Eroded Lands
	31	Other Bare
	43	Fallow Land & Old Fields (Bush)
	44	Fallow Land & Old Fields (Grass)
	45	Fallow Land & Old Fields (Bare)
	46	Fallow Land & Old Fields (Low Shrub)

The simple random site allocation for sampling within the `Woody` stratum ensures that the samples are representative. Given that the `Non-Woody` stratum accounts for a much larger extent of the eligible land than the `Woody` stratum (~90/10), equal sampling efforts across both strata result in an unequal sample size. Additional sites were, therefore, sampled to increase the `Woody` stratum sample size and improve the sample ratio (~70/30).

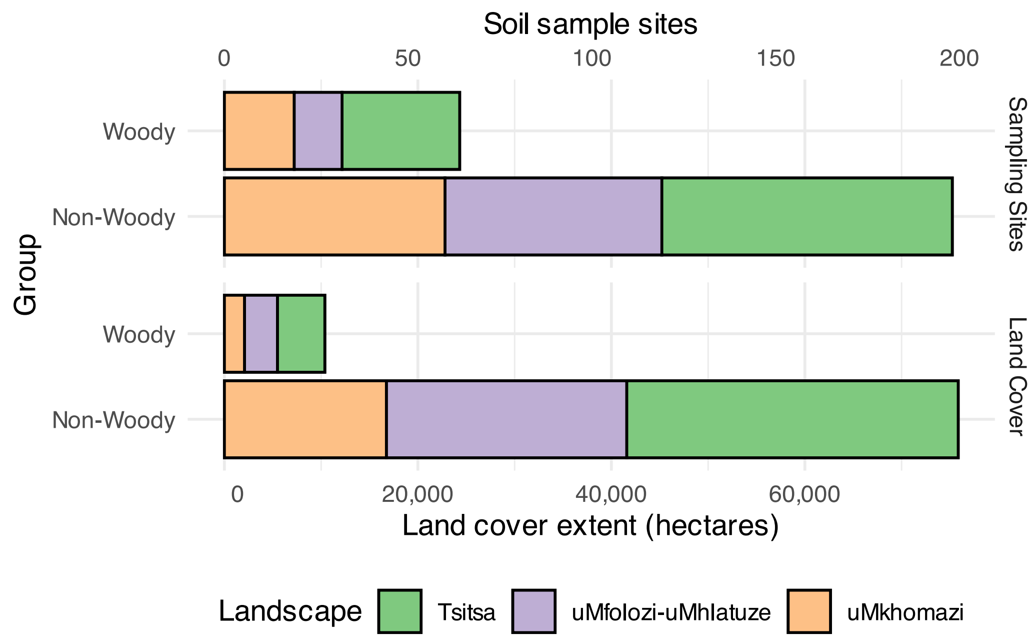

Figure 4: Eligible land cover extent (ha) and number of soil sampling sites between `woody` (>10% cover) and `non-woody` (<10% cover) strata across the relevant landscapes.

3 Parametric Test Assumptions

In order to apply a parametric t-test, the data should be Gaussian (normally distributed) and homoscedastic (equal variance). If these assumptions are violated, then an alternative test, such as the Welch t-test, may be required.

This STHDA article provides more information on the steps followed in this analysis.

3.1 Normal distribution

A visual inspection of the log10-transformed SOC data indicates that the data are normally distributed. The Shapiro-Wilk normality test can be used to test this more quantitatively. The null hypothesis is that the data are normal. If p-value > 0.05, then the data are not significantly different from normal.

```

	Shapiro-Wilk normality test

data:  SOC_log10[Group == "Non-Woody"]
W = 0.9884, p-value = 0.1072
```

```

	Shapiro-Wilk normality test
```

```
data:  SOC_log10[Group == "Woody"]
W = 0.96643, p-value = 0.07873
```

For both "Woody" and "Non-Woody" data, the p-values are > 0.05. We can therefore assume normality for both datasets.

3.2 Equal variance

The homogeneity in variances of the two groups can be tested with an F-test. The null hypothesis is that the ratio of variances is equal to 1.

```

	F test to compare two variances

data:  SOC_log10 by Group
F = 1.7401, num df = 197, denom df = 63, p-value = 0.01134
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.137268 2.551662
sample estimates:
ratio of variances
          1.740069
```

The p-value is < 0.05. We can therefore reject the null hypothesis and infer that the ratio of variances are not equal to 1. The data have therefore failed the test of equal variance and a non-parametric two-sample t-test (Welch t-test) needs to be applied.

3.3 Welch two-sample t-test

The null hypothesis of the t-test is that the true difference in means between the two groups is equal to 0. There's no prior expectation that one group would be greater or less than the other, so a two-sided hypothesis is applied by default.

```

	Welch Two Sample t-test

data:  SOC_log10 by Group
t = -1.2051, df = 139.67, p-value = 0.2302
alternative hypothesis: true difference in means between group Non-Woody and
group Woody is not equal to 0
95 percent confidence interval:
 -0.12734721  0.03089267
sample estimates:
mean in group Non-Woody     mean in group Woody
              0.1452139               0.1934412
```

4 Conclusion

The p-value is > 0.05, so the null hypothesis cannot be rejected, and it can be concluded that **there is no difference in SOC between the `Woody` and `Non-Woody` strata**.

Bibliography

Bond, William J. 2019. "Open Ecosystems: Ecology and Evolution Beyond the Forest Edge". https://doi.org/10.1093/oso/9780198812456.001.0001

Champely, Stephane. 2020. "Pwr: Basic Functions for Power Analysis". https://cran.r-project.org/package=pwr

Edwards, E. 1983. "A Broad-Scale Structural Classification of Vegetation for Practical Purposes". *Bothalia* 14 (3/4): 705–12. https://doi.org/10.4102/abc.v14i3/4.1231

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data". *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009

Pebesma, Edzer, and Roger Bivand. 2023. *Spatial Data Science: With applications in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

R Core Team. 2024. "R: A Language and Environment for Statistical Computing". https://www.r-project.org/