Remote Sensing Improved Forages

Preview Results

November 20, 2021

Abstract

Key words:
JEL Classification:

Contents

List of Figures

List of Tables

1 Land Use Classification

1.1 Descriptive Statistics

- There are 8 Kebeles (subdivisions)
- In these Kebeles 6% (16) of the survey plots (249) are improved foraged plots .
- We kept in the data set polygons that are classified either as: Improved Forage, Crop, Grazing, and Tree. In the survey of the area some polygons were classified as Tree/Crop, Bare, but only in the Dil-betegel subdivision, and is less than 2% of the area. To maintain consistency I also drop those polygons classified as other, since it is missing for some of the plots.[1]. Table 2 shows land use by Kebele and percentage area. Improved forage accounts for 4.17%, whereas Crop 29.84%, Grazing 40.75%, and Tree 25.23% of the total area.

Table 1

kebele	Crop	Grazing	Improved.forage	Tree
Afesa	3	6	1	10
Bachema	16	8	2	9
Dil-betegel	12	8	4	13
Enashenefalen	15	4	3	35
Guiete	8	8	1	3
Gulet Abeshekan	11	7	2	4
Legaba	21	10	2	10
Wufeta Dati	7	1	1	4
total	93	52	16	88

Table 2

kebele	Polygons	mean_area	Crop	Grazing	Improved.forage	Tree
Afesa	20	37,657.14	22.86	17.01	7.07	53.05
Bachema	35	113,392.40	32.79	36.90	4.89	25.43
Dil-betegel	37	117,456.30	46.11	25.01	3.00	25.88
Enashenefalen	57	297,524.70	22.10	3.96	1.33	72.61
Guiete	20	76,724.22	39.07	37.55	2.53	20.84
Gulet Abeshekan	24	75,565.21	72.56	10.34	4.74	12.36
Legaba	43	629,447.60	12.96	79.36	0.67	7.01
Wufeta Dati	13	132,115.20	90.62	2.02	1.50	5.86

1.2 Remote sensing

We use Sentinel 2 products Bottom-Of-Atmosphere scenes at different spatial resolution levels. For land use classification, we use 76 scenes from Sentinel-2 beginning on 2019-09-09 and ending 2021-10-08, each

[1] we can potentially drop the Afesa Kebele that is missing the "Other" class

scene has a 10 day difference. We construct a time series of NDVI using the near infrared (NIR) and red bands at the 10 m spatial resolution:

$$NDVI = (NIR - Red)/(NIR + Red) \tag{1}$$

As a result we have 19,033 time series, the approximate distribution land use of these pixels are

Table 3. Pixel in sample aproximate coverage

kebele	Crop	Grazing	Improved.forage	Tree
Afesa	124	113	35	291
Bachema	560	567	86	434
Dil-betegel	731	392	64	464
Enashenefalen	932	170	67	2,742
Guiete	438	409	35	218
Gulet Abeshekan	702	149	60	160
Legaba	1,148	5,413	67	614
Wufeta Dati	1,408	45	35	120
All	6,043	7,258	449	5,043

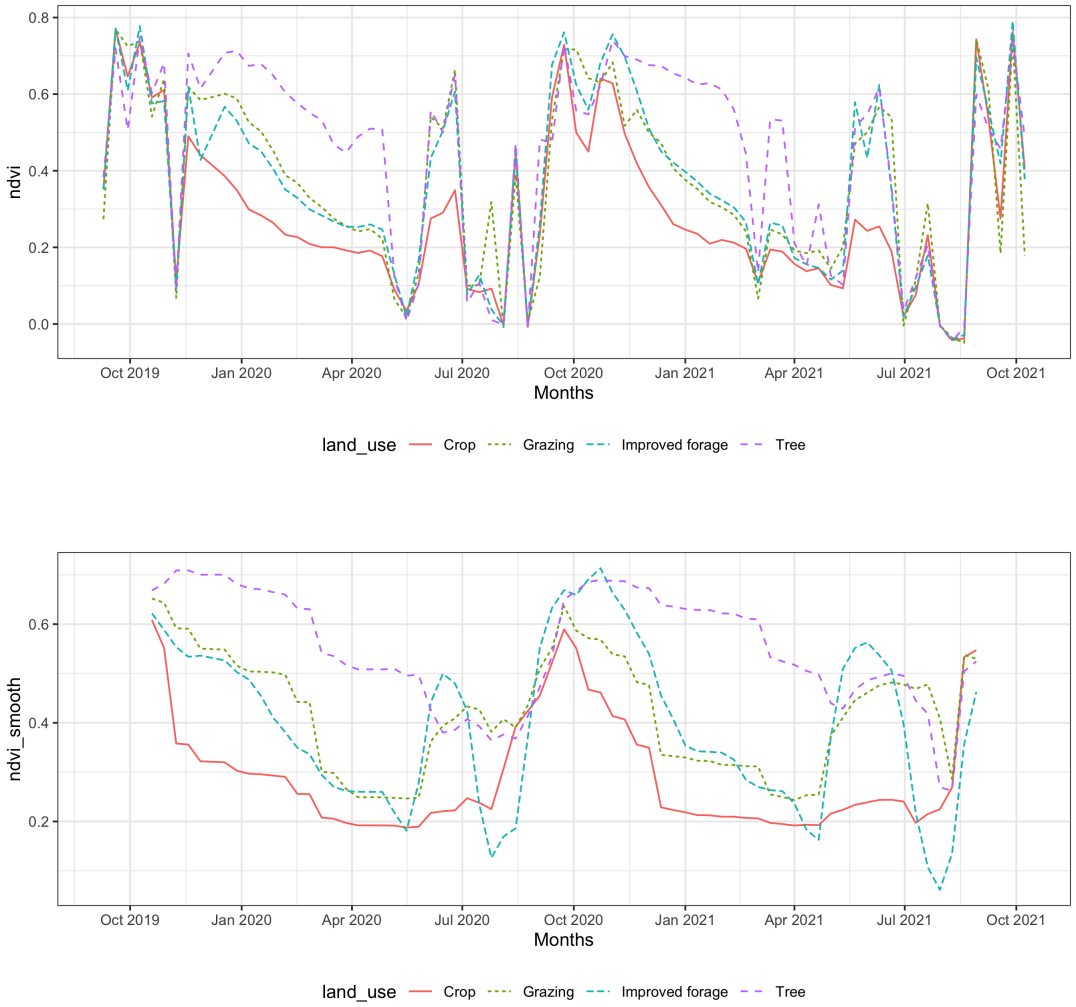
there are multiple pixels that have multiple land use, so we keep 11,052 pixels that have 100% coverage to train our model

Table 4. Pixel 100% coverage

kebele	Crop	Grazing	Improved.forage	Tree
Afesa	56	20	15	116
Bachema	201	283	29	167
Dil-betegel	369	202	14	150
Enashenefalen	443	79	16	1,511
Guiete	188	185	11	100
Gulet Abeshekan	385	32	14	43
Legaba	520	4,569	25	288
Wufeta Dati	949	21	11	40
All	3,111	5,391	135	2,415

Since there is cloud coverage in these scenes I smooth them out with a centered 90 day moving average and weight each scene by the inverse of cloud coverage, this way clear scence get more weight.

This allows us to capture the time trend of the series without loosing data or interpolating. Cloud coverage attenuates the reflectance, but since we are not interested in between dates comparison but within dates, and the overall trend, the strategy works well.[2]



1.3 Model

We used the k-shape clustering algorithm introduced by Paparrizos and Grabvano (2015). This is a clustering algorithm that proposes as a distance measure, a normalized version of the cross-correlation measure to consider the shapes of time series while comparing them. The idea is that, given the set of time series search for a "centroid" and assign the series to the closest centroid.

I only use NDVI and a smoothed version of NDVI (which is the one that performs the best), and an undersampling approach.

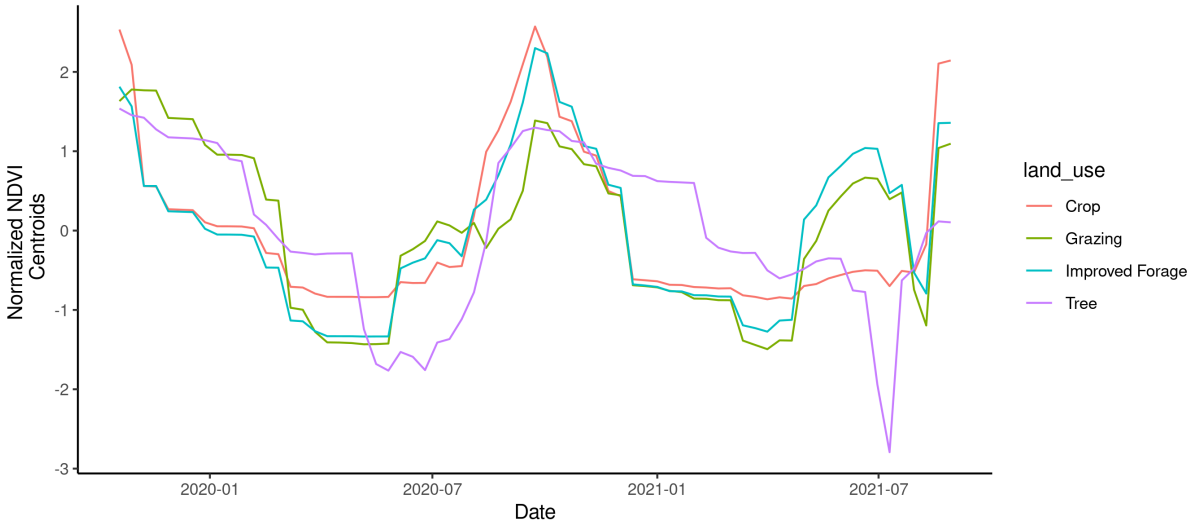
[2]Some preliminary robustness for 30,40,50,60,70, and 90 days moving averages; 90 day MA works best

1.4 k-shapes

Given the small number of pixels with improved forage we use a bootstrap unbalanced approach. In each of 1,000 bootstrap iterations we sample pixels from each kebele (subdivision) in a way that the number of pixels with different land use evens out. We then average out the centroids given across this bootstrap samples, in this way we smooth out the randomness introduced by the undersampling.

The resulting centroids with their closest land use can be seen in the following figure:

Figure 1. Centroids by Land Use



The model perfors quite satisfactory given the small number of classes for the improved forrage, table 5 shows the confusion matrix for 100% coverage pixels. Cluster 2 identifies Crop with above 91% precision, while Centroid 4 for Tree, with similar precision. Cluster 1 and Cluster 3 identify Grazing and Improved Forage, which identifies correctly about 74% of the series. The main challenge here is separating Grazing from Improved Forages. 24% of the Improved Forage series is assigned to Cluster 1 that identifies Grazing, whereas 17% of the grazing series are assigned to the Improved Forage cluster.

Table 5. Confusion Matrix

	Cluster				
Land use	1	2	3	4	# Series
Crop	2.83	91.67	3.05	2.44	3,111
Grazing	74.09	6.44	17.46	2.02	5,391
Improved forage	24.44	0.74	74.81	0	135
Tree	5.34	2.09	0.70	91.87	2,436

1.5 Where I'm missing

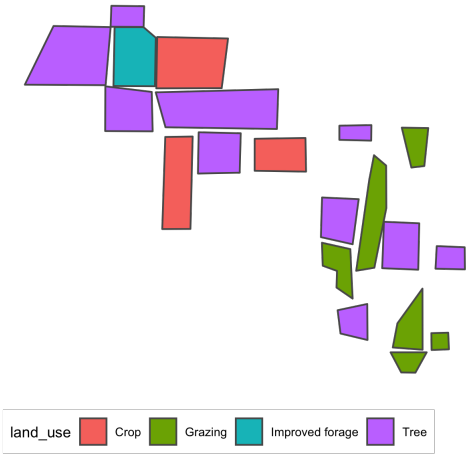
2 Soil Erosion

3 Figures

3.1 Land Use by Kebeles

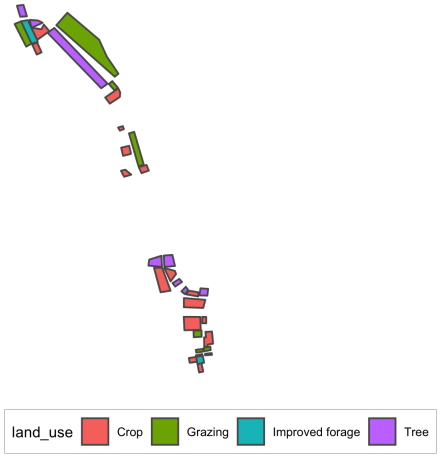
3.1.1 Afesa

Figure 2. Land Use



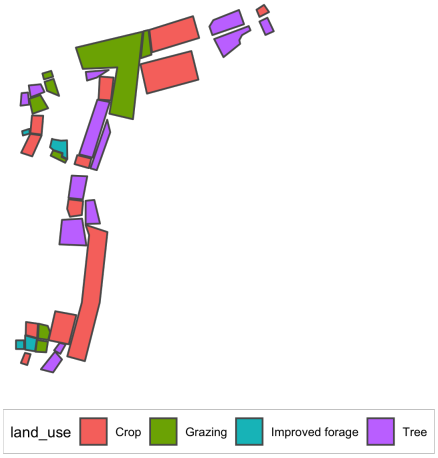
3.1.2 Bachema

Figure 3. Land Use



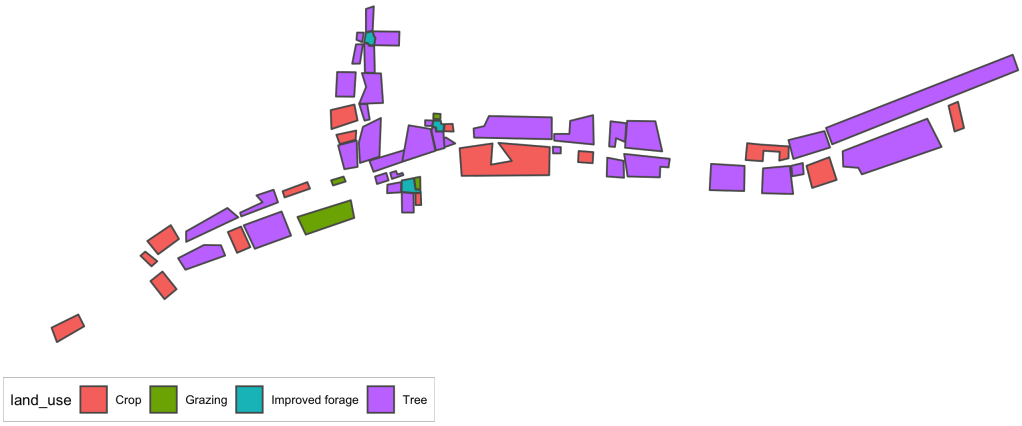
3.1.3 Dil-betegel

Figure 4. Land Use



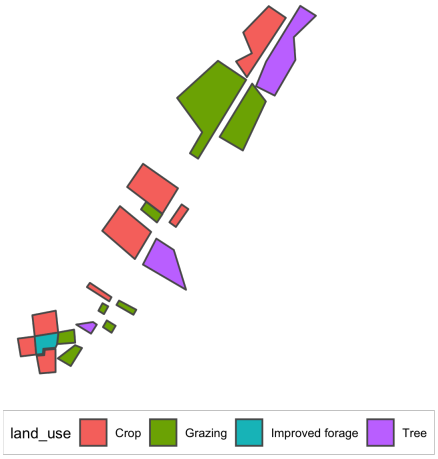
3.1.4 Enashenefalen

Figure 5. Land Use



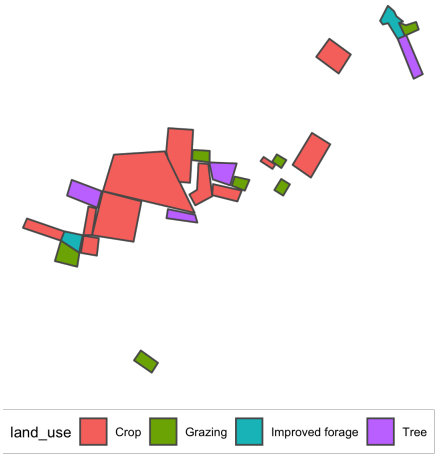
3.1.5 Guiete

Figure 6. Land Use



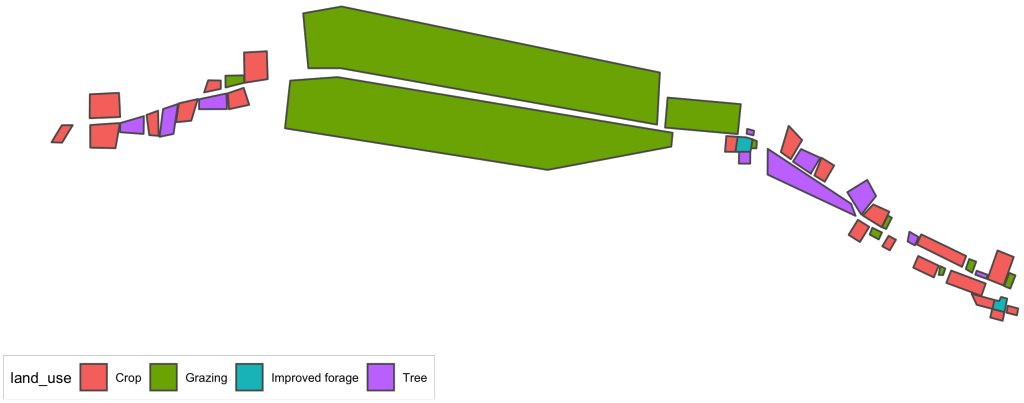
3.1.6 Gulet Abeshekan

Figure 7. Land Use



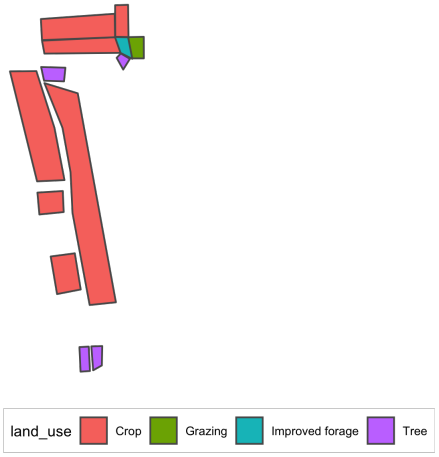
3.1.7 Legaba

Figure 8. Land Use



3.1.8 Wufeta Dati

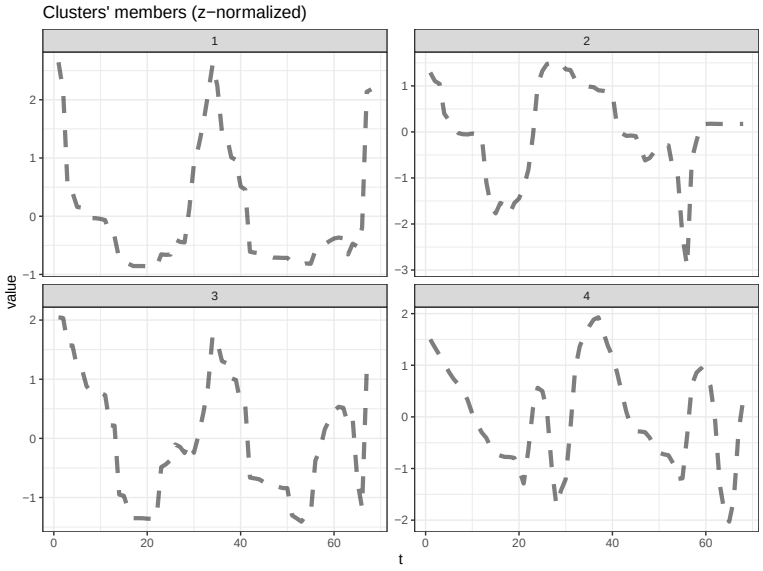
Figure 9. Land Use

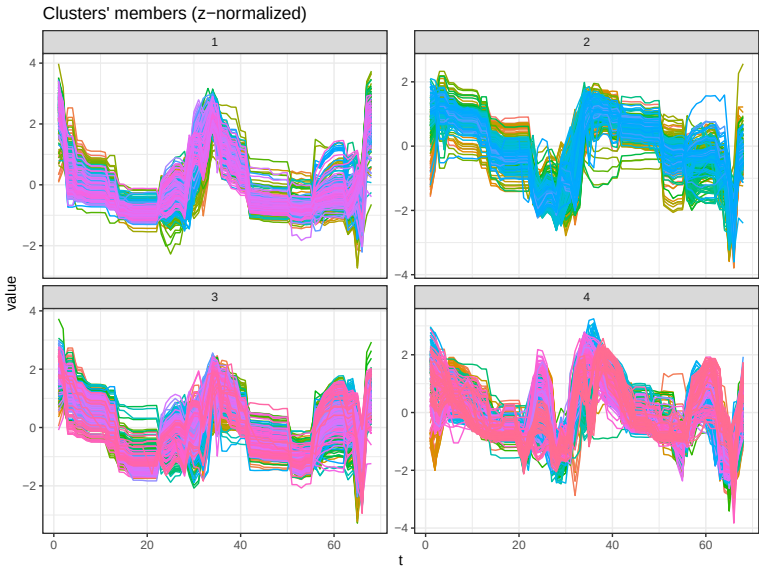


A Appendix: K-shape algorithm

A.0.1 One iteration using smoothed NDVI

	1	2	3	4
Crop	287	9	35	29
Grazing	55	8	256	41
Improved forage	0	4	45	311
Tree	11	256	57	36





B Appendix: Tables and Figures

Table A.1

regions	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_10	k_11	k_12	k_13	k_14	k_15	k_16
Afesa	0.46	0.64	0.61	0.49	0.52	0.51	0.46	0.47	0.42	0.52	0.48	0.42	0.46	0.40	0.42
Bachema	0.29	0.40	0.53	0.58	0.50	0.51	0.47	0.54	0.52	0.55	0.52	0.53	0.55	0.40	0.54
Dil-betegel	0.35	0.47	0.39	0.39	0.41	0.41	0.38	0.35	0.38	0.34	0.33	0.37	0.30	0.35	0.30
Enashenefalen	0.43	0.53	0.67	0.68	0.60	0.46	0.57	0.53	0.46	0.39	0.49	0.42	0.36	0.34	0.35
Guiete	0.47	0.63	0.57	0.55	0.50	0.38	0.43	0.52	0.40	0.45	0.38	0.40	0.46	0.48	0.52
Gulet Abeshekan	0.53	0.57	0.57	0.56	0.50	0.58	0.47	0.52	0.35	0.40	0.53	0.43	0.47	0.40	0.42
Legaba	0.43	0.42	0.58	0.51	0.48	0.51	0.49	0.43	0.39	0.36	0.29	0.33	0.36	0.38	0.38
Wufeta Dati	0.41	0.61	0.66	0.60	0.59	0.59	0.49	0.52	0.47	0.42	0.44	0.36	0.47	0.44	0.42
Results_all	0.31	0.38	0.40	0.37	0.32	0.31	0.33	0.32	0.32	0.29	0.23	0.29	0.24	0.27	0.32

Table A.2

regions	clusters	sil	max	ind
Afesa	k_3	0.64	0.64	1
Bachema	k_5	0.58	0.58	1
Dil-betegel	k_3	0.47	0.47	1
Enashenefalen	k_5	0.68	0.68	1
Guiete	k_3	0.63	0.63	1
Gulet Abeshekan	k_7	0.58	0.58	1
Legaba	k_4	0.58	0.58	1
Wufeta Dati	k_4	0.66	0.66	1
Results_all	k_4	0.40	0.40	1

Characteristic	**Bare**, N = 1,140[1]	**Crop**, N = 353,172[1]	**Grazing**, N = 482,372[1]	**Improved forage**, N = 49,400[1]	**Other**, N = 52,516[1]
Blue	2,165 (3,237)	2,318 (3,252)	2,196 (3,255)	2,422 (3,380)	2,457 (3,130)
Geen	2,291 (2,917)	2,476 (2,900)	2,363 (2,900)	2,570 (3,025)	2,576 (2,786)
Red	2,468 (2,616)	2,569 (2,679)	2,349 (2,724)	2,553 (2,850)	2,602 (2,579)
NIR	3,187 (2,301)	3,714 (2,166)	3,776 (2,105)	4,085 (2,142)	3,596 (2,127)
NDVI	0.22 (0.14)	0.29 (0.24)	0.37 (0.26)	0.37 (0.28)	0.25 (0.19)
cloudcov	33 (35)	33 (35)	33 (35)	33 (35)	33 (35)

[1]Mean (SD)

kebele	Polygons	mean_area (m^2)	Crop	Grazing	Improved forage	Tree	Tree/Crop	Other	Bare	NA
Afesa	22	41804.45	0.21	0.15	0.13	0.48	0.04	NA	NA	NA
Bachema	44	126129.41	0.29	0.33	0.08	0.23	NA	0.07	NA	NA
Dil-betegel	40	121036.30	0.45	0.24	0.04	0.25	0.01	NA	0.01	NA
Enashenefalen	63	301719.35	0.22	0.04	0.02	0.72	NA	0.01	NA	NA
Guiete	29	93043.98	0.32	0.31	0.04	0.25	NA	0.07	NA	NA
Gulet Abeshekan	32	121280.96	0.45	0.06	0.06	0.08	NA	0.35	NA	NA
Legaba	46	633953.41	0.13	0.79	0.01	0.07	NA	0.00	NA	NA
Wufeta Dati	19	145149.11	0.82	0.02	0.03	0.11	NA	0.02	NA	NA
NA	4	5742.78	NA	NA	NA	NA	NA	NA	NA	5742.78

Figure A.1. Average Smoothed NDVI by Land Use

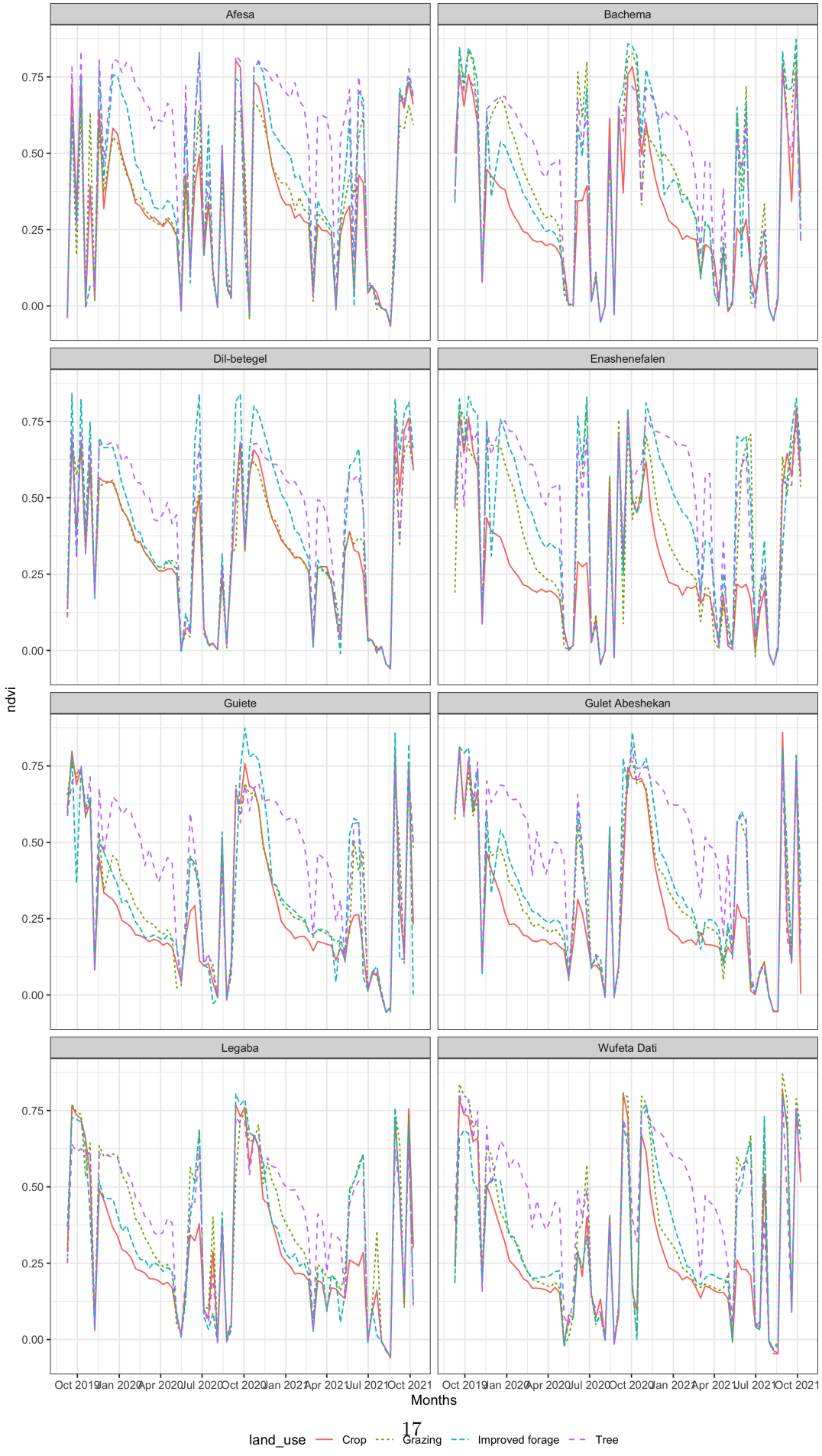
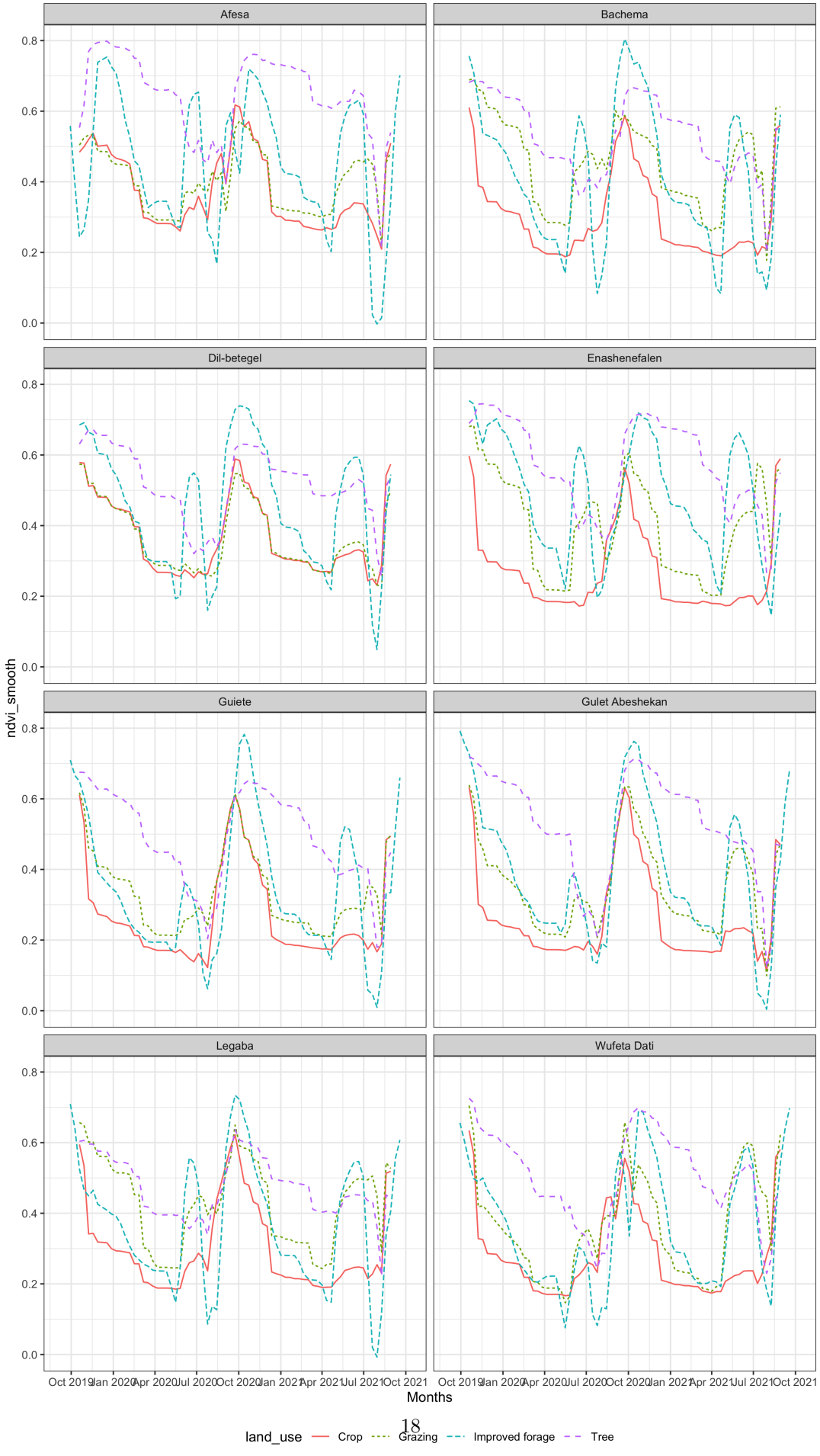


Figure A.2. Average Smoothed by Land Use



C Appendix: Original Land Use and NDVI by Kebele

Figure A.3. NDVI: Average time series (smoothed)

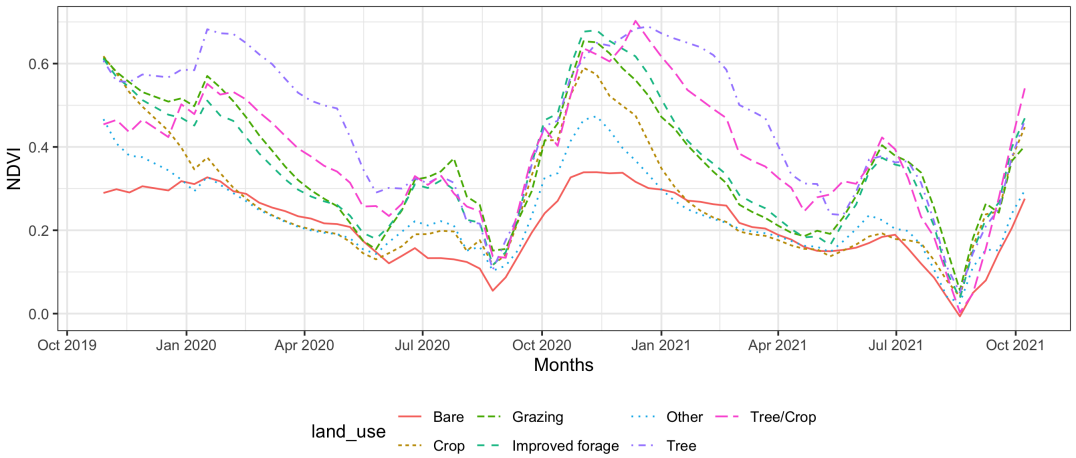


Figure A.4. NDVI: Average time series

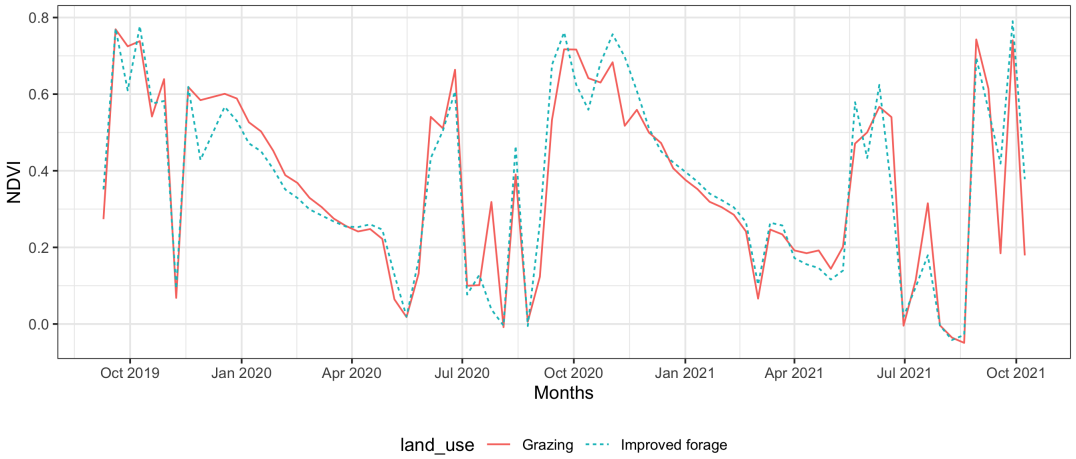
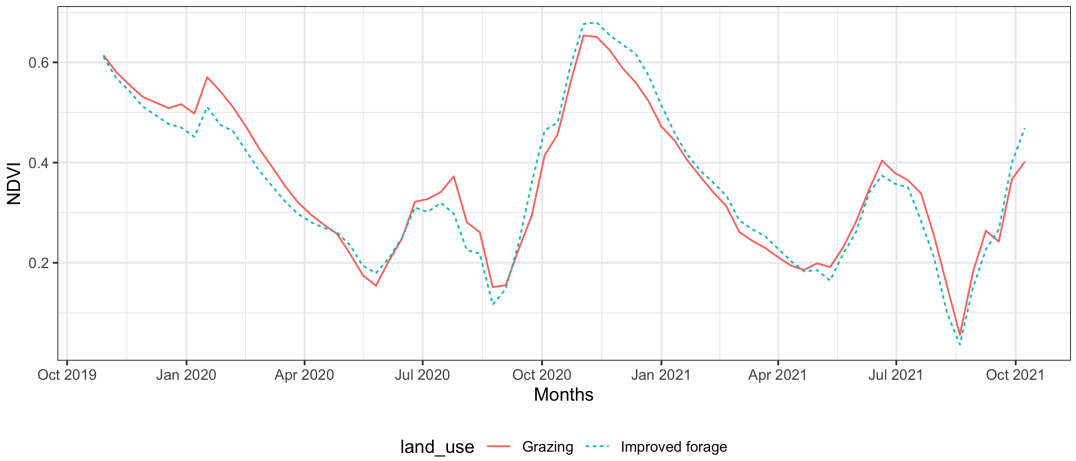


Figure A.5. NDVI: Average time series (smoothed)



C.1 By Kebeles

C.1.1 Afesa

Figure A.6. NDVI: Average time series

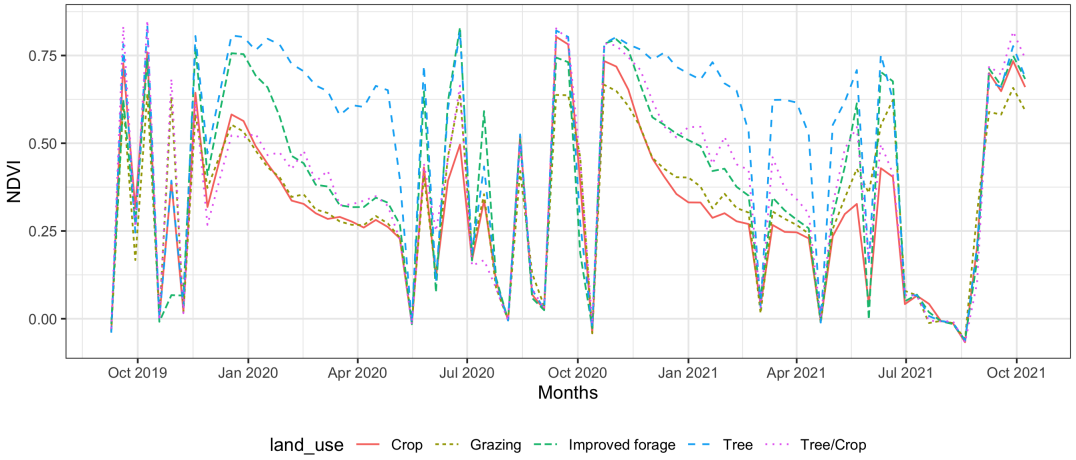
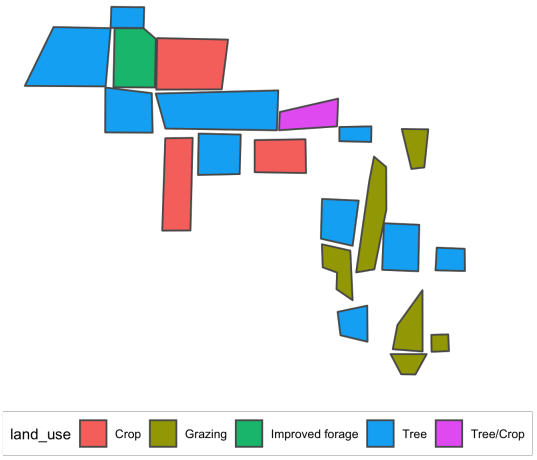


Figure A.7. Land Use



C.1.2 Bachema

Figure A.8. NDVI: Average time series

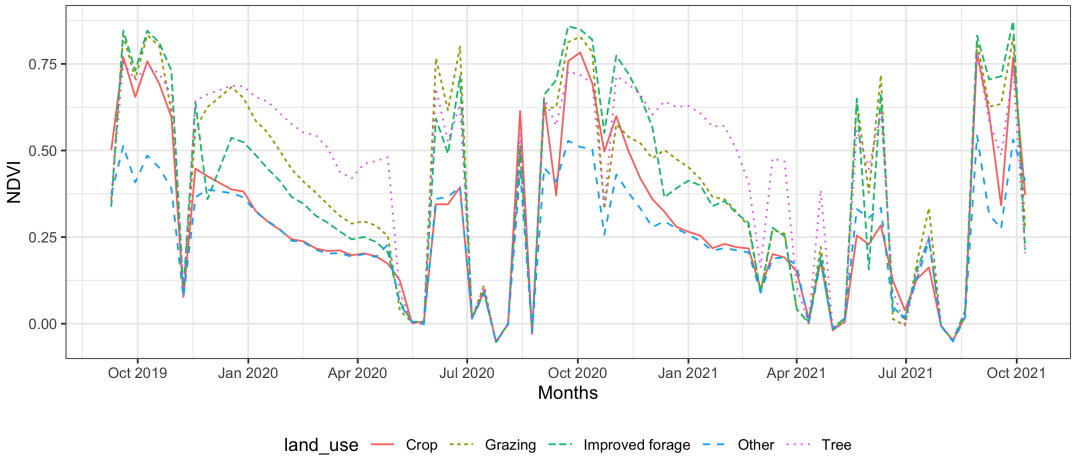
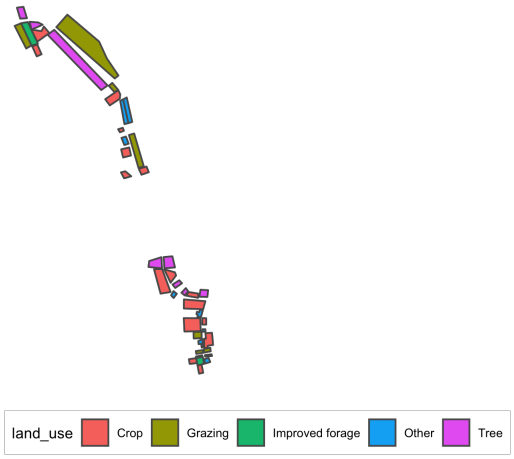


Figure A.9. Land Use



C.1.3 Dil-betegel

Figure A.10. NDVI: Average time series

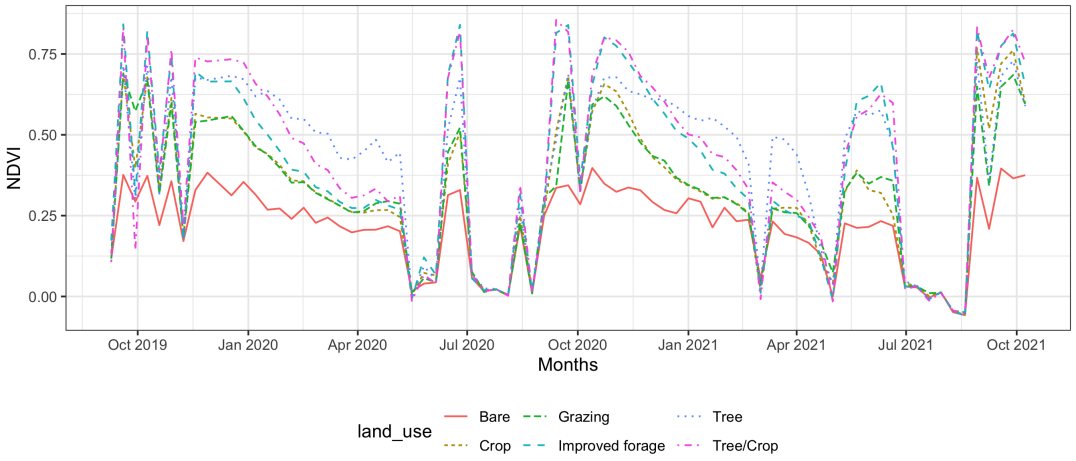
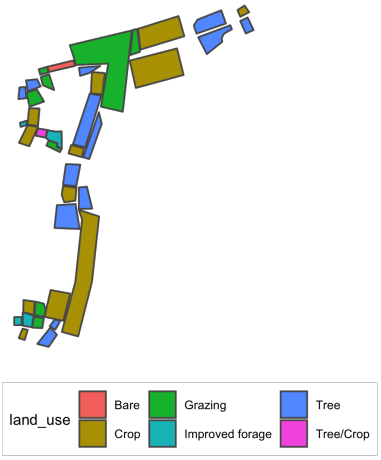


Figure A.11. Land Use



C.1.4 Enashenefalen

Figure A.12. NDVI: Average time series

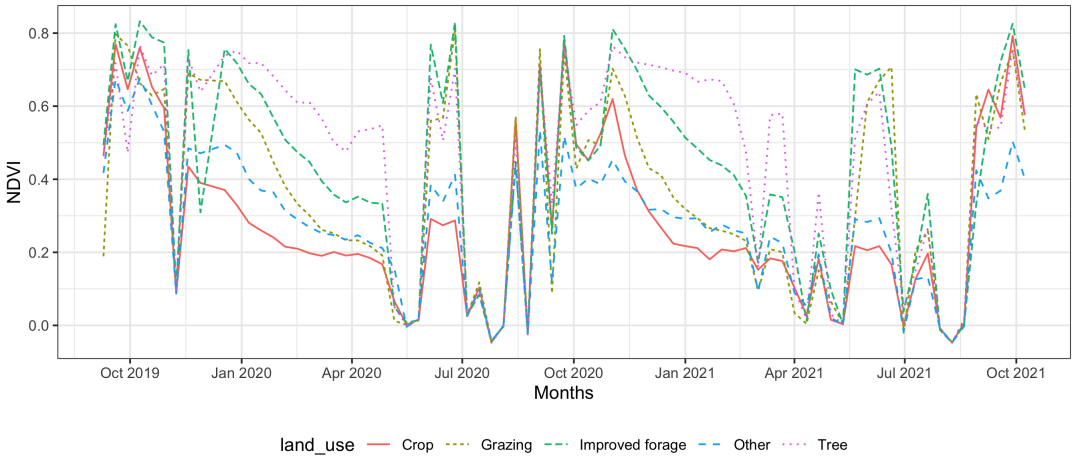


Figure A.13. Land Use

C.1.5 Guiete

Figure A.14. NDVI: Average time series

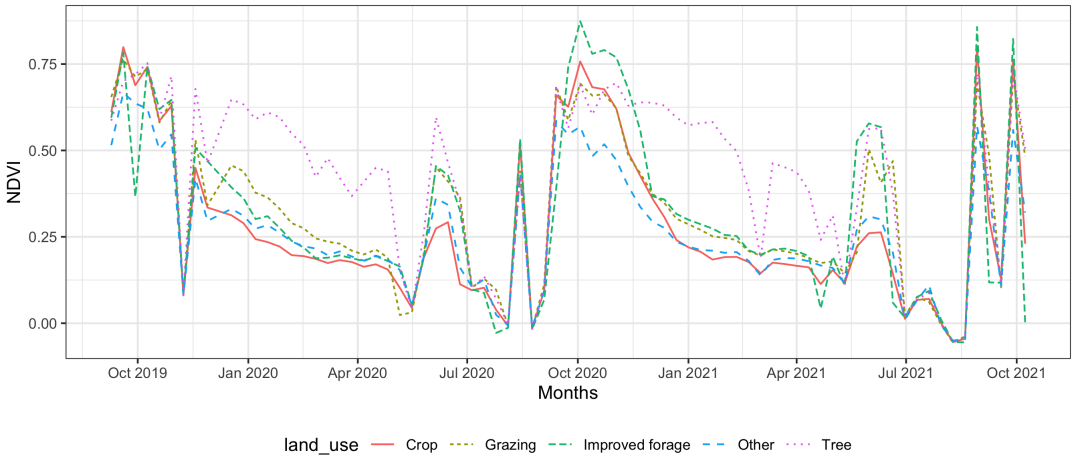
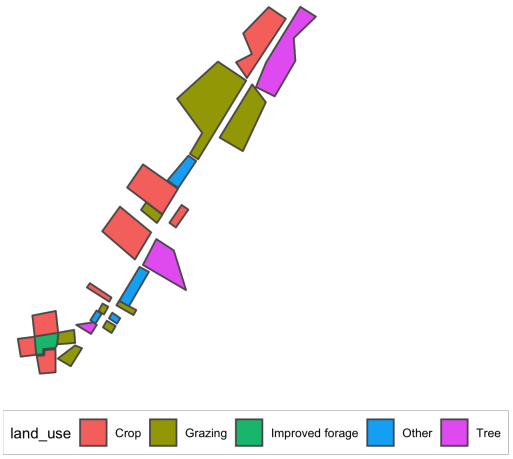


Figure A.15. Land Use



C.1.6 Gulet Abeshekan

Figure A.16. NDVI: Average time series

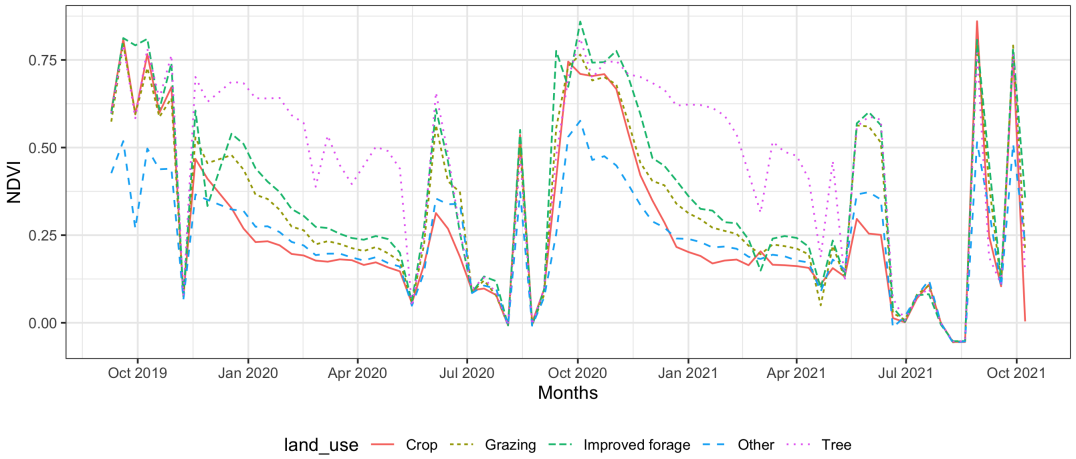
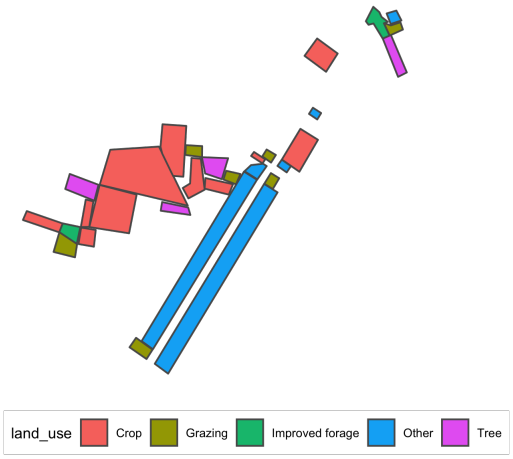


Figure A.17. Land Use



C.1.7 Legaba

Figure A.18. NDVI: Average time series

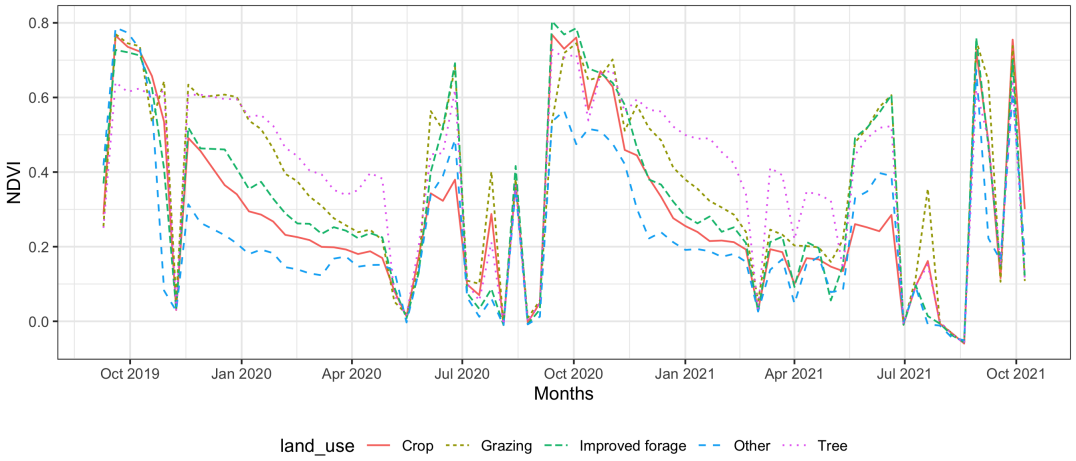
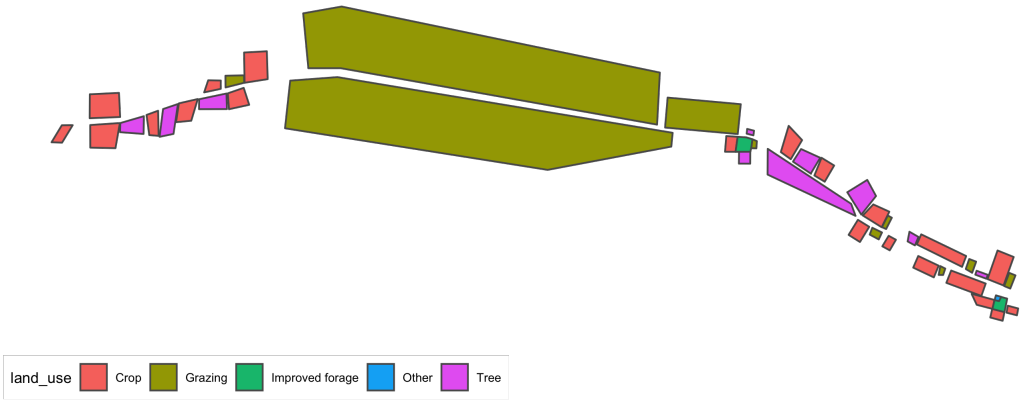


Figure A.19. Land Use



C.1.8 Wufeta Dati

Figure A.20. NDVI: Average time series

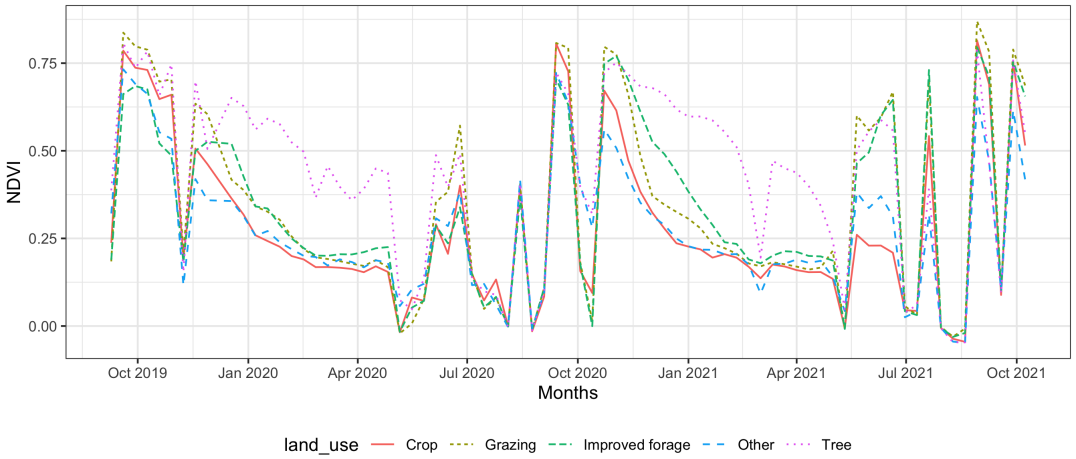
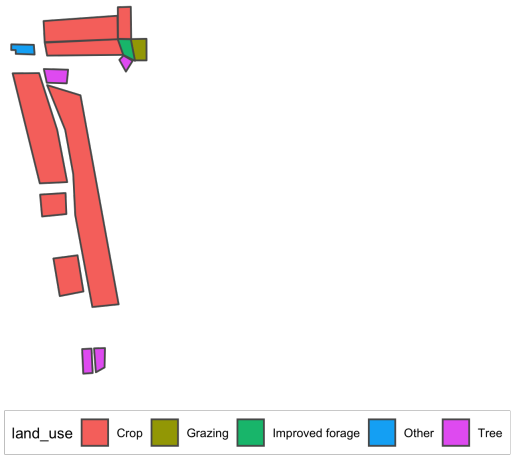


Figure A.21. Land Use



D Other unsuccessful approaches

D.1 Cross Sectional Variation

I modeled land use as a function of the 4 bands and the NDVI:

$$LandUse_{ij} = f(Blue, Green, Red, NIR, NDVI) + u_{ij}$$

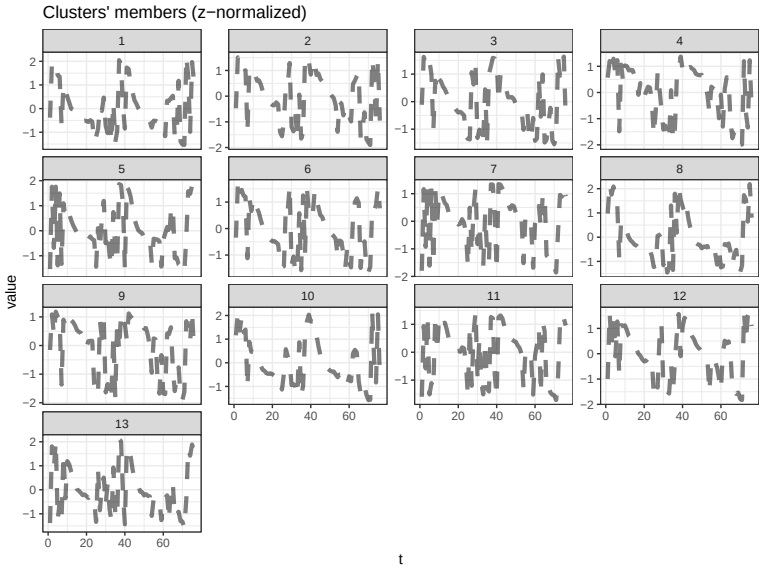
where $LandUse_{ij}$ is the land use in pixel i of polygon j. f is the ML function that we use to predict land use. We explore 3 classes of models:

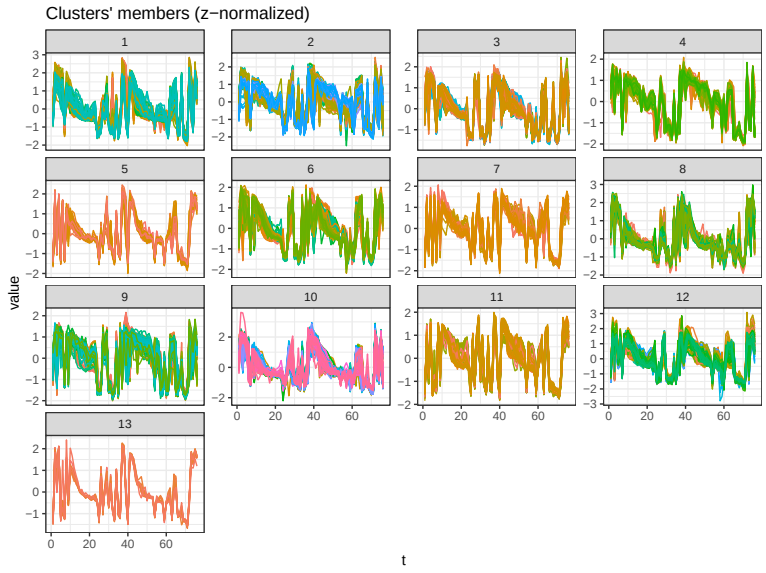
- Binary classifications: In this case $LandUse_{ij} = I(LandUse_{ij} == Improved\,Forage)$ and 0 are the other classes
- Multiclass classification where $LandUse_{ij}$ takes 4 classes: Improved Forage, Crop, Grazing, and Tree

The challenge was incorporating the time dimension, one approach was collapse to a single cross section and in different variables consider multiple moments of the distribution of predictors: mean, sd, iqr, max, min. I did these for within pixel and between pixels of the same polygon for the entire time series, and quarterly (approximating the seasons, although there are no seasons in Ethiopia, but the idea was trying to capture the cycles seen in the time series). None of these were successful so far...

D.2 Time Series Variation: K-shapes, Raw NDVI time series

	1	2	3	4	5	6	7	8	9	10	11	12	13
Crop	55	9	9	0	13	6	6	102	5	97	1	43	14
Grazing	29	55	65	1	15	54	19	0	3	76	0	43	0
Improved forage	29	45	56	8	0	48	0	0	18	79	34	43	0
Tree	14	73	17	70	0	1	13	0	104	4	21	43	0





regions	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_10	k_11	k_12	k_13	k_14	k_15	k_16
Afesa	0.40	0.45	0.42	0.43	0.36	0.45	0.42	0.35	0.38	0.44	0.34	0.35	0.42	0.38	0.45
Bachema	0.41	0.38	0.39	0.43	0.44	0.45	0.52	0.48	0.42	0.47	0.50	0.45	0.41	0.40	0.51
Dil-betegel	0.40	0.44	0.40	0.39	0.40	0.32	0.28	0.37	0.38	0.35	0.27	0.35	0.35	0.24	0.40
Enashenefalen	0.41	0.30	0.47	0.44	0.42	0.30	0.45	0.40	0.37	0.39	0.39	0.40	0.42	0.38	0.39
Guiete	0.58	0.61	0.58	0.63	0.41	0.46	0.31	0.43	0.46	0.29	0.40	0.35	0.34	0.36	0.25
Gulet Abeshekan	0.52	0.47	0.57	0.51	0.59	0.40	0.45	0.53	0.48	0.49	0.52	0.43	0.41	0.41	0.40
Legaba	0.30	0.31	0.36	0.35	0.40	0.45	0.38	0.45	0.46	0.44	0.42	0.44	0.46	0.42	0.45
Wufeta Dati	0.51	0.59	0.43	0.53	0.55	0.45	0.45	0.50	0.42	0.47	0.34	0.42	0.45	0.45	0.32
Results_all	0.26	0.31	0.32	0.34	0.32	0.37	0.34	0.33	0.34	0.31	0.38	0.40	0.35	0.34	0.38

regions	clusters	sil
Afesa	k_7	0.45
Bachema	k_8	0.52
Dil-betegel	k_3	0.44
Enashenefalen	k_4	0.47
Guiete	k_5	0.63
Gulet Abeshekan	k_6	0.59
Legaba	k_10	0.46
Wufeta Dati	k_3	0.59
Results_all	k_13	0.40