Data and Codes used for the Publication – "Modelling the Impact of Ecosystem Fragmentation on Ecosystem Services in the Degraded Ethiopian Highlands"

Tegegne Molla Sitotaw, Louise Willemen, Derege Tsegaye Meshesha, Martha Weldemichael, and Andrew Nelson

These datasets and codes in the DANS repository were compiled as part of the study *"Modelling the Impact of Ecosystem Fragmentation on Ecosystem Services in the Degraded Ethiopian Highlands,"* focusing on the Lake Tana basin. The study explored variations in four ecosystem services - grass biomass supply, heat stress regulation, crop pollination, and nature-based tourism - caused by ecosystem fragmentation between 2000 and 2020.

The provision of the four ES varies significantly across different ecosystem types and is shaped by distinct ecosystem processes inherent to each service type. This variation highlights the unique roles that specific ecosystems, such as forests, wetlands, and grasslands, play in supporting and regulating ES flow. Additionally, the four ES are influenced differently by fragmentation indices, such as area, perimeter-area ratio, and proximity, which capture different aspects of landscape structure and spatial configuration. For example, while the area might directly determine the capacity of an ecosystem to support biodiversity or biomass production, the perimeter-area ratio often relates to edge effects that influence ecological interactions. Proximity, as a fragmentation index, further emphasises the spatial arrangement of patches relative to one another, which is particularly relevant for services that depend on connectivity or spillover effects, such as pollination. Importantly, the calculation of proximity and other fragmentation indices is tailored to reflect specific buffer distances for each ES, ensuring that each service's unique spatial scale and flow dynamics are appropriately captured and modeled. This nuanced approach allows for a more precise understanding of how fragmentation impacts the spatial provision of multiple ES over time.

Four zipped folders containing raw ES data and spatial predictors:

grass_biomass_data.zip – contains raw data on grass biomass and spatial predictor variables.

heat_stress_regulation_data.zip – contains raw data on air temperature data, cooling intensity, and spatial predictor variables.

crop_pollination_data.zip– contains raw data on pollinator visitation rate and spatial predictor variables.

nature_tourism_data.zip - contains raw data on geotagged photographs from Flickr and Twitter platforms as well as spatial predictor variables.

Table S1. *Description of all data types, predictor variables, and sources used to explore the variations of ES*

Ecosystem services	Field data and spatial predictors	Descriptions	Data sources
Grass biomass	plot_grassagb_raw.csv	This file contains fresh and dried grass biomass supply data collected in 2019 and 2020 used as a response variable and also contains the geographic coordinates of sampling field plots.	Field measurement
	ndvi_biomass_2000.tif ndvi_biomass_2020.tif	These files represent the normalized difference vegetation index (NDVI) and indicate vegetation health and productivity.	NDVI values for 2000 and 2020 were calculated from Landsat images downloaded within the periods of the grass biomass data collection period, ensuring accurate temporal alignment with the fieldwork.
	area_biomass_2000.tif area_biomass_2020.tif	Area represents the wetland fragment size in hectares for 2000 and 2020, extracted from the land cover classifications of those years to capture historical size variations.	Wetland land cover class derived from Landsat images downloaded in October and November as the biomass data collection to ensure temporal alignment.
	peri-area_biomass_2000.tif peri-area_biomass_2020.tif	The perimeter-area ratio, which captures the ratio of wetland perimeter to area for 2000 and 2020, provides insights into wetlands' shape and edge effects.	Wetland land cover of 2000 and 2020
	prox_biomass_2000.tif prox_biomass_2020.tif	Proximity measures the distance from wetland fragments to the nearest wetland patch, which can influence biomass by affecting connectivity and ecological interactions.	Wetland land cover classes of 2000 and 2020
Microclimate regulation	airtemp_data_raw_2020.csv	This CSV file contains raw air temperature measurements recorded across the distance from green patches (forest and wetland patches) during the hot months of April and May 2020, essential for analysing spatial and temporal variations in heat stress regulation.	Field recorded air temperature using Heat Stress Tracker instrument.
	fvc_cooling_2000.tif fvc_cooling_2020.tif	A raster file representing the Fractional Vegetation Cover (FVC) in 2020 at 30 m resolution, providing insights into vegetation density and land cover characteristics.	Landsat images downloaded for April and May of 2000 and 2020.
	Impersurf_cooling_2000.tif Impersurf_cooling_2020.tif	These raster files represent impervious surface areas of 2000 and 2020, indicating built-up areas that can affect the cooling intensity of green patches in benefiting areas.	Impervious surfaces were extracted from the GMIS) Dataset from Landsat https://cmr.earthdata.nasa.gov/search/concepts/C1517102461-SEDAC.html.
	area_cooling_2000.tif area_cooling_2020.tif	These raster files show the spatial extent of cooling areas (forest and wetland green patches) in hectares for 2000 and 2020, useful for assessing areas providing ecosystem cooling services.	Landsat-based forest and wetland land cover classes
	peri-area_cooling_2000.tif peri-area_cooling_2020.tif	These raster files represent the perimeter-to-area ratio of cooling patches in 2000 and 2020, providing metrics to assess the shape and spatial complexity of these areas, which influence the heat stress regulation of green patches.	Landsat-based forest and wetland land cover classes

	prox_cooling_2000.tif prox_cooling_2020.tif	These raster files represent the proximity of grid cells of benefiting areas to surrounding cooling patches within an Euclidean distance of 300 m (threshold cooling decay) for 2000 and 2020.	Locations of temperature records and Landsat-based forest and wetland covers
Crop pollination	pollinator_visitation_data_raw.csv	This CSV file contains raw data on pollinator visitation rates, including counts of pollinator visits to crop flowers recorded at four buffer distances (200, 500, 1,000, and 1,500 m) from church forest habitats on crop fields during the summer crop growing of 2019 and 2020.	Field recorded pollinator visitation rate
	fvc_pollin_2000.tif fvc_pollin_2020.tif	These raster files showing FVC for 2000 and 2020 provide spatial information on variations in vegetation cover that influence pollinator visitation rates on nearby crop fields.	Landsat satellite images for 2000 and 2020
	area_pollin_2000.tif area_pollin_2020.tif	Raster files represent the spatial extent of church habitat areas around crop fields for 2000 and 2020, availability and distribution of habitats supporting pollinators.	Church forest habitats extracted from land covers using church locations.
	Peri-area_pollin_2000.tif Peri-area_pollin_2020.tif	These raster files indicate the perimeter-to-area ratio of pollinator habitats for 2000 and 2020, providing insights into habitat fragmentation.	Church forest patches extracted from land covers.
	prox_pollin_2000.tif prox_pollin_2020.tif	This raster file shows the spatial proximity of each grid cell of crop plots to surrounding pollinator habitats within a distance of pollinator foraging ranges (1,500 m) calculated for 2000 and 2020.	Location of visitation plots and forest habitat patches
Nature-based tourism	pud_tud_geotagged_data_2020.csv	This CSV file contains geotagged photo data from Flickr and Twitter for 2020. The mean annual number of PUDs and TUDs from 2016 - 2020 within a 1 × 1km grid cell	Photographs uploaded to the photo-sharing websites Flickr and Twitter https://www.flickr.com and https://developer.twitter.com
	area_pud_2000.tif area_pud_2020.tif	Raster files represent the spatial extent of natural and cultural landscapes (wetlands and sacred church forests) for 2000 and 2020, used to analyse the areas valuable for nature tourism. Area of these land covers intersects with geotagged photo locations within a grid of 1 × 1 km.	Natural and cultural covers derived from Landsat images.
	peri-area_pud_2000.tif peri-area_pud_2020.tif	Raster files indicate the perimeter-to-area ratio of natural and cultural land covers for 2000 and 2020, useful for exploring landscape fragmentation on nature visitation.	Natural and cultural landscapes derived from Landsat images.
	prox_pud_2000.tif prox_pud_2020.tif	These raster files encompass information on the proximity of grid cells of pud_tud data to natural and cultural landscape features for 2020, providing insights into the accessibility of these areas for recreation and tourism within 500 m tourist walking preference.	pud_tud grid center locations and Landsat-based natural and cultural land cover
	nat_cult_counts_pud_2000.tif nat_cult_counts_pud_2020.tif	Raster files presenting the count of cultural sites within 1 km grid cells for 2000 and 2020, cultural heritage distribution and density over time, and point count within a grid of 1 × 1km.	Obtained from Amhara Region Culture and Tourism Bureau and https://www.openstreetmap.org
	facilities_counts_pud_2000.tif facilities_counts_pud_2020.tif	These raster files represent the count of tourism-related facilities within 1 × 1 km grid cells for 2000 and 2020, service availability for visitors.	Obtained from Amhara Region Culture and Tourism Bureau and https://www.openstreetmap.org
	road_prox_pud_2000.tif road_prox_pud_2020.tif	Raster files indicating the proximity of grid cells of pud_tud to roads, accessibility to natural and cultural landscapes in 2000 and 2020—road proximity calculated based on a walking distance of 500 m from roads computed using distance accumulation.	Main roads obtained from Amhara Region Transport Bureau and https://www.openstreetmap.org

	line_of_sight_pud_2000.tif line_of_sight_pud_2020.tif	This line-of-sight raster file shows the visibility of natural and cultural landscapes up to 6 km for photographing from roads for 2000 and 2020, to assess the visual aesthetics of the Ethiopian highland rocky landscapes. Main roads and ALOS PALSAR 12.5 m resampled to 30 m Raster surface locations visibility along main roads using geodesic viewshed methods (Spatial Analyst tool). The observer's offset was 1.7 m (the average tourist height above the ground to capture photographs).	ALOS PALSAR 12.5 m from the Alaska Satellite Facility https://asf.alaska.edu

Table S2. *Description of the four R codes used to model and plot the variations of each ES due to fragmentation indices.*

R codes	Descriptions
1. GAMs_biomass_analysis_prediction_plotting.R 2. GAMs_cooling_Intensity_analysis_prediction_plotting.R 3. GAMs_pollination_analysis_prediction_plotting.R 3. GAMs_nature-tourism_analysis_prediction_plotting.R	These R files contain a detailed analysis of the relationship between ES response data and different spatial predictors (Table 1) using GAMs. Four main parts including: **Part I: extraction of raster predictors:** This section extracts spatial predictor variables from raster datasets, such as area, perimeter-to-area ratio, and proximity indices, which will be used in subsequent GAMs model building. **Part II: GAM model building:** The file constructs GAMs using the response variable of each ES with the extracted spatial predictor variables. **Part III: prediction and plotting:** After fitting the GAMs, the script generates new predictions using the model and visualises the relationship between fragmentation indices (such as area, perimeter-to-area ratio, proximity) and ES response data. This part provides visualisations of how different fragmentation indices influence ES supply. **Part IV: contour Plot for biomass supply:** The final part uses the *vis.gam* function to generate contour plots that display the predicted spatial distribution of ES supply, visually representing the model's predictions across the study area.

plot_grassAGB_raw.csv

A CSV file containing raw data on fresh and dried grass biomass g measurements collected from field plots across wetland meadows. Field-based grass biomass data was collected in October and November 2019 and 2020 when the wetland pasture had peak biomass productivity. The comprehensive sampling method recommended by Hankins et al. (2004) and Liu et al. (2017) was used to design the field survey to acquire accurate estimates of grass biomass for grazing lands in Lake Tana wetlands. Therefore, 25 sampling sites were randomly selected across the wetland

meadow. At each field sampling site, a circle with a size 15 m radius (705.5 m^2) was established, the recommended area for wetland vegetation assessment (Hankins et al., 2004; Liu et al., 2017). At each circle, three radial transects were laid out, and four quadrats with a dimension of 0.5 m × 0.5 m (0.25 m^2) were placed along the transects with equally spaced inside each circle. Grass biomass was collected at each quadrant using the clip-harvest method at the soil or water surface (Raynolds et al., 2006). The sizes of the sampling circles were comparable to the pixel size of the Landsat Satellite images, which is recommended to overcome the errors related to geometric correction (Edirisinghe et al., 2012). The freshly clipped vegetation biomass was weighed, and the dried biomass was also measured using a digital weighing balance after drying at 60 °C to a constant weight.

---

airtemp_data_raw_2020.csv

This CSV file containing raw air temperature measurements recorded at various field plots in 2020, essential for analyzing spatial and temporal variations in microclimatic conditions. What these datasets describe in general: this file includes air temperature data collected in and around 34 sites of forest and wetland patches (24 forests and 10 wetland patches), each larger than 2 ha. At each patch, transect temperature monitoring was conducted from 1:00 to 3:00 pm in the afternoon when daytime temperatures were high. The weather was generally hot, with clear and sunny days and low wind speed. Air temperature was recorded by the Kestrel 5400 Heat Stress Tracker, which has a stated accuracy of ±0.5 °C within a measurement range of -29.0 to 70.0 °C (https://kestrelmeters.com). The tracker was set to measure air temperature every 10 seconds while holding the tracker 1.5 m above the ground. The transect surveys started from the centre of natural ecosystems to the surrounding benefiting areas. Temperature records were made in each transect, up to a maximum distance of 300 m from vegetative patches. Concurrently, the corresponding geographic coordinates and cooling distances were recorded with each temperature measurement using a Garmin 64s handheld GPS along the transects.

---

pud_tud_geotagged_data_2020.csv

This file includes the average annual number of PUD from 2016 - 2020 within 1 × 1 km grid cell calculated based on photographs uploaded to the photo-sharing website Flickr https://www.flickr.com and Twitter https://developer.twitter.com. Areas within a buffer distance of 10 km (Balmford et al., 2015) from the Lake Tana basin boundary were considered in order to avoid location biases related to geotagged photos (Crampton et al., 2013). The location of the geotagged photos captured using mobile devices is accurate to at least 10 m (Zandbergen and Barbeau, 2011) and satisfies the requirement for assessing cultural ecosystem services. The details of the method and codes to extract information about the geotagged photos are available in the GitHub link: https://github.com/MarthaWD/Socialmedia_analytics. The distances were calculated using the Euclidean distance tool with a 1 km cell size for the predictor variable.

The following seven major spatial predictor variables are considered to affect tourists' preferences for nature-based tourism in the Lake Tana basin. These variables were site-specific, i.e., biodiversity of church forests and wetlands (areas attracting endemic fauna and flora, church forest with indigenous and rare tree species, global bird areas); cultural sites (cultural resources including churches, monasteries, and artifacts in the Lake Tana, natural and man-made heritages); scenic landscape (lakeshore, waterfalls, volcanic landscape); road accessibility; topography and tourism facilities. From Martha

---

Accuracy assessment of the land cover classes of 2000 and 2020

We collected reference polygons from GPS-based ground truth in 2020 and the acquisition of high-resolution satellite imagery from Planet satellite images and Google Earth for 2000 and 2020. 224 polygons representing each land cover class were obtained from this data for each year's land cover classification and accuracy assessment. The polygons were randomly divided into training (75%) and validation (25%), ensuring comprehensive representation across the various land cover classes. We used a support vector machine (SVM) approach to classify land cover, considering seven distinct categories: forest, shrub, wetland, grassland, cultivated land, water body, and built-up. The classification accuracy of 2000 and 2020 are presented in Tables S3 and S4.

Table S3. *Classification accuracy assessment of 2000 land cover map*

	Reference data for classified land cover							Total reference points	User's accuracy (%)
Classified map	Grassland	Built-up	Waterbody	Shrub	Cultivated	Wetland	Forest		
Grassland	**698**	0	0	85	121	96	Forest	1000	69.8
Built-up	2	**41**	4	0	1	3	0	51	80.4
Waterbody	0	5	**1200**	3	0	24	0	1232	97.0
Shrub	84	0	0	**1133**	381	41	2	1641	69.0
Cultivated	66	2	0	55	**5232**	9	228	5592	93.6
Wetland	32	7	100	91	132	**753**	11	1126	66.9
Forest	7	0	0	111	34	3	**844**	999	84.5
Total classified points	889	55	1304	1478	5901	929	1085	**11641**	
Producer's accuracy (%)	78.5	74.6	92.0	76.7	88.7	81.1	77.8	**Overall accuracy (%):**	**85.0**
Kappa coefficient =			**78.78**						

Table S4. *Classification accuracy assessment of 2020 land cover map*

	Reference data for classified land cover							Total reference points	User's accuracy (%)
Classified map	Grassland	Built-up	Waterbody	Shrub	Cultivated	Wetland	Forest		
Grassland	**719**	0	0	55	245	35	0	1054	68.2
Built-up	2	**41**	3	0	2	3	0	51	80.4
Waterbody	0	6	**980**	3	0	35	1	1025	95.6
Shrub	12	0	0	**1262**	221	32	121	1648	76.6
Cultivated land	87	1	0	65	**8693**	12	11	8869	98.0
Wetland	30	2	143	221	105	**552**	21	1074	51.4
Forest	5	0	0	232	45	4	**982**	1268	77.4
Total classified points	855	50	1126	1838	9311	673	1136	**14989**	
Producer's accuracy (%)	84.1	82.0	87.0	68.7	93.4	82.0	86.4	**Overall accuracy (%):**	**88.3**
Kappa coefficient =			**75.6**						

Land cover dynamics from 2000 to 2020

Table S5. *Land cover change matrix from 2000 – 2020 (values in %).*

		Land cover classes 2000						
		Grassland	Built-up	Waterbody	Shrub	Cultivated	Wetland	Forest
Land cover classes 2020	Grassland	**13.2**	0.3	0	2.3	0.9	2.2	0.6
	Built-up	4.1	**87.5**	0	6.7	6.3	2	1.9
	Waterbody	0.2	0	**97.2**	0.3	0.1	2	0
	Shrub	10.8	1.9	0	**18.3**	3.1	4.4	22.1
	Cultivated	66.7	9	1.7	66.7	**87.5**	39.6	42
	Wetland	3.1	0.3	1	1.8	0.4	**47**	2.1

	Forest	1.6	0.7	0	3.7	1.6	2.9	**31.2**

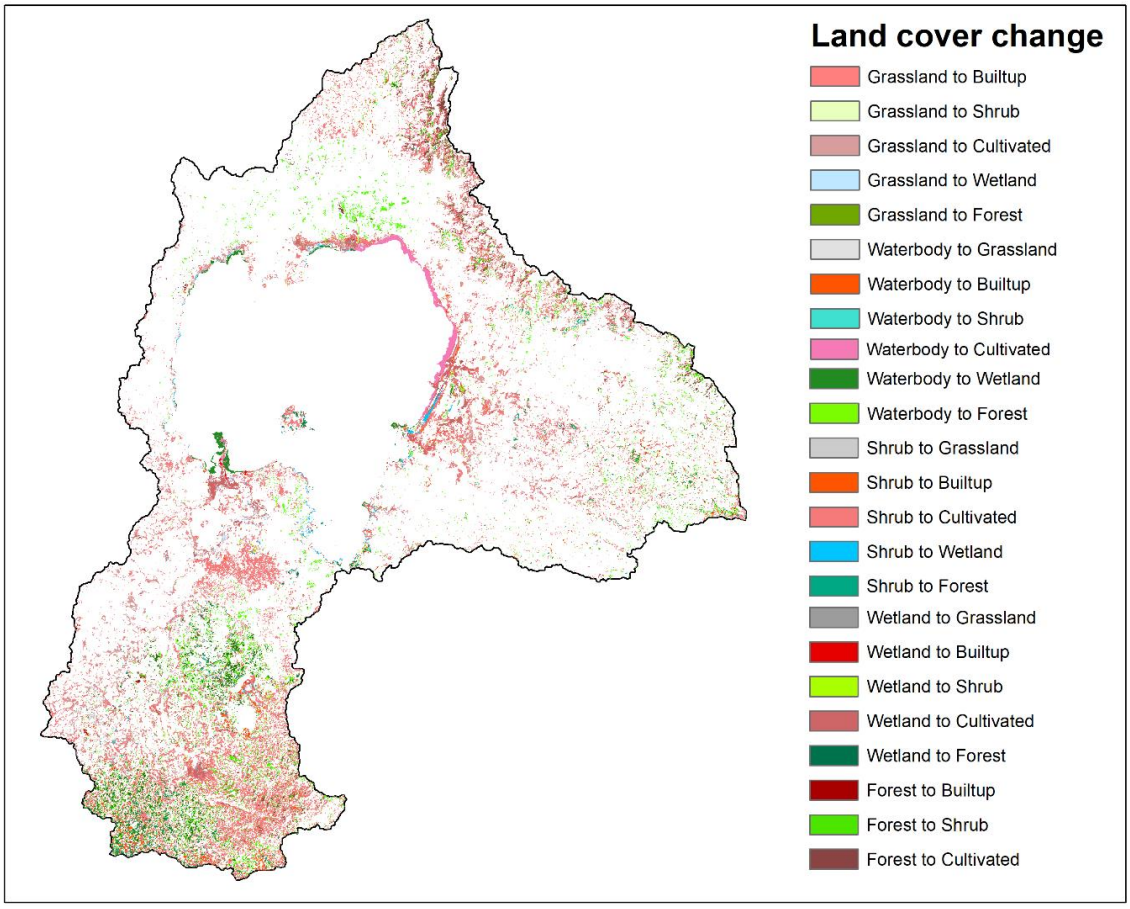

Figure S1. Transformation of of targeted land cover type to another type of land cover from 2000 to 2020

References

Balmford, A., Green, J.M.H., Anderson, M., Beresford, J., Huang, C., Naidoo, R., Walpole, M., Manica, A., 2015. Walk on the Wild Side: Estimating the Global Magnitude of Visits to Protected Areas. PLoS Biol. https://doi.org/10.1371/journal.pbio.1002074

Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W., Zook, M., 2013. Beyond the geotag: Situating "big data" and leveraging the potential of the geoweb. Cartogr. Geogr. Inf. Sci. https://doi.org/10.1080/15230406.2013.777137

Edirisinghe, A., Clark, D., Waugh, D., 2012. Spatio-temporal modelling of biomass of intensively grazed perennial dairy pastures using multispectral remote sensing. Int. J. Appl. Earth Obs. Geoinf.

https://doi.org/10.1016/j.jag.2011.11.006

Hankins, J., Launchbaugh, K., Hyde, G., 2004. Rangeland Inventory as a Tool for Science Education. Rangelands 26. https://doi.org/10.2458/azu_rangelands_v26i1_hankins

Liu, S., Cheng, F., Dong, S., Zhao, H., Hou, X., Wu, X., 2017. Spatiotemporal dynamics of grassland aboveground biomass on the Qinghai-Tibet Plateau based on validated MODIS NDVI. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-04038-4

Raynolds, M.K., Walker, D.A., Maier, H.A., 2006. NDVI patterns and phytomass distribution in the circumpolar Arctic. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2006.02.016

Zandbergen, P.A., Barbeau, S.J., 2011. Positional accuracy of assisted GPS data from high-sensitivity GPS-enabled mobile phones. J. Navig. https://doi.org/10.1017/S0373463311000051