



Summer 2024

South Africa Ecological Conservation
Identifying and Mapping Riparian Areas in South Africa with Earth
Observations

DEVELOP Technical Report

August 9th, 2024

Mina Nada (Project Lead)
Andrew Saah
Madison Elowitt
Katya Beener

Advisors:

Dr. Arnaud Cerbelaud (NASA Jet Propulsion Laboratory, California Institute of Technology)
Dr. Kerry Cawse-Nicholson (NASA Jet Propulsion Laboratory, California Institute of Technology)
Benjamin M. Holt (NASA Jet Propulsion Laboratory, California Institute of Technology)

Lead:

Michael Pazmino (California – JPL)
Jane A. Zugarek (California – JPL)

1. Abstract

Riparian zones are thin strips of wetland that occur along the edges of rivers, lakes, and other water bodies. They provide many ecosystem and societal benefits such as stream bank stabilization, flood control, and habitat stability, making these zones very important areas to preserve and protect. Recognizing the need to accurately map these ecosystems in a cost-effective way, we partnered with the South African National Biodiversity Institute (SANBI) and the Biodiversity Survey of the Cape (BioSCape) to develop a GIS methodology using Esri's ArcGIS Pro and data from Landsat 9 Operational Land Imager-2, Sentinel-2 Multispectral Instrument, and the Shuttle Radar Topography Mission to identify potential riparian zones (PRZ) and observed riparian vegetation (ORV) in the Southern Cape and the North West Province of South Africa to estimate the actual riparian vegetation (ARV) within those areas. Potential riparian ecosystems were calculated by utilizing a Topographic Wetness Index (TWI) and a water feature layer, while observed riparian vegetation were calculated by overlaying a Normalized Difference Vegetation Index (NDVI) and TWI. Actual Riparian Vegetation is determined by the overlap of ORV with PRZ and is done with conditional statements; ORV within PRZ is ARV. This method significantly improved a current riparian land cover classification since it combined high resolution optical imagery and topographic data. We anticipate such maps will be used by conservationists and practitioners interested in riparian monitoring and management.

Key Terms

remote sensing, riparian ecosystems, Landsat 9, Sentinel-2, South Africa, SRTM

2. Introduction

Freshwater is one of, if not the most important natural resources for humanity, yet it only makes up a very small amount of the total water supply. Freshwater and the ecosystems that come along with it have been at the center of human development for most of history and are relevant to many cultures. Beyond human-based needs for this limited resource, this marginal amount of water is home to nearly 6% of all known species (Dudgeon et al., 2006). These freshwater ecosystems range from simple tributaries in headlands to landscape-scale reveries that are hundreds of square kilometers, and so are not only highly biodiverse but also physically diverse, which leads to invaluable ecological services and societal benefits: benefits range from sediment deposition stabilizing stream banks to natural filtration of water from natural and man-made pollutants (Naiman et al., 2005). The interfaces between these freshwater ecosystems and terrestrial upland areas are unique in their characteristics, borrowing traits from both surrounding ecozones. These highly dynamic zones are called riparian zones, which are characterized by their high spatially and temporally dynamic nature. Many definitions of these zones exist for specific reasons; however, for this study, we classify riparian zones as regularly inundated transitional regions extending between waterbodies and the terrestrial lands that surround them (Naiman et al., 2005).

Because of the importance to humans, riparian ecosystems are often cultivated, and in highly developed areas such as Europe and North America, up to 90% of riparian zones are functionally extinct (Tockner & Stanford, 2002). Tockner & Stanford (2002) and Dudgeon et al. (2006) found that the main anthropogenic pressures on these ecosystems are habitat destruction, flow and flood pattern alteration, water pollution, and invasive exotic species, with the loss rate only accelerating. Because of their rapid loss and ecological importance, understanding and monitoring the spatial and temporal patterns of riparian zones is the first step for responsible and effective management. Earth observations (EO) are a highly advantageous monitoring technology because of the extent of coverage and the ability to collect data at regular return intervals.

There are numerous ways to categorize riparian zones. Fernández et al. (2012) argued that floodplain extent should be used as a primary riparian delineator by using remotely sensed Digital Elevation Models (DEMs). Their reasoning stems from the fact that hydrological regimes can influence both vegetation and geological structures. In addition to elevation data, remotely sensed multispectral data integrated with GIS has been used

to effectively delineate vegetation cover estimation from Yang (2007) integrated remotely sensed multi-spectral EO data with Geographic Information Systems (GIS) to effectively delineate and map riparian vegetation community patterns using their unique spectral signature. With the rapidly increasing availability of remotely sensed EO data and GIS software, EO data products are becoming standard in the management of riparian ecosystems. However, this management method is only available to the few with access to the data and software. Often, the most threatened riparian ecosystems are also in areas with both low spatial data availability and field measurements, rendering this powerful decision support tool inaccessible in many cases (Stutter et al., 2021). As the human population in these regions increases, the impacts on the riparian ecosystems will also rise, with some of the most threatened Riparian ecosystems being in East Asia, North America, and Africa (Tockner & Stanford, 2002). South Africa provides an example of a region with a high impact on riparian ecosystems but low access to high-resolution EO data, which would help support mitigation efforts for some of the impacts.

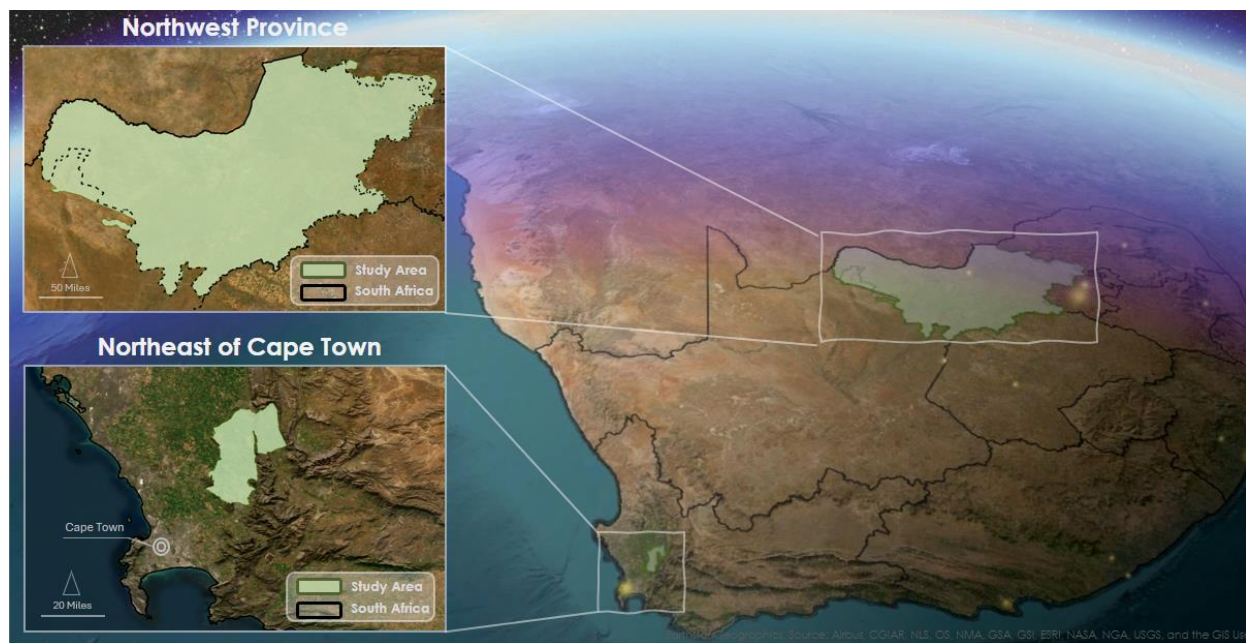


Figure 1. Study site map of the Western Cape and North West Province in South Africa

In partnership with the South African National Biodiversity Institute (SANBI) and the Biodiversity Survey of the Cape (BioSCape), our project created a methodology that helps to identify riparian areas not highlighted in prior map products. SANBI provides public access to biodiversity data, promotes sharing information and knowledge, advancing policy, and conserving biodiversity. Currently, SANBI utilizes a combination of methods for mapping riparian ecosystems. This includes field observations, soil mapping, and heads-up digitization of Google Earth and historic maps. This process is labor intensive and relies on datasets that may not be available for the entire country, making riparian ecosystems difficult to identify. BioSCape is a research collaboration between the US and South Africa that combines remote sensing and field data to better understand biodiversity on land and water in the Greater Cape Floristic Region.

Historically, land cover classification efforts have not accurately captured riparian zones due to the highly dynamic nature that riparian zones inhabit. We determined 2 sites to carry out this methodology due to the differences in the environmental conditions and in riparian data availability (Figure 1). The study site near Cape Town is wetter, more populated, and has higher quality data available such as Light detection and ranging (LiDAR); The North West (NW) province has a far more arid environment, with a lower population and high-quality data is scarce. Due to the dramatic differences in climate, population, and data quality, a

methodology that maps riparian ecosystems throughout South Africa—not just in places with favorable environments or data availability—must be developed. This methodology should combine the geospatial data that SANBI possesses with NASA EOs. While a second project team will work on mapping the entirety of South Africa, our team concentrated on refining the methodology for mapping the selected sites. We modeled our methodology after Weissteiner et al. (2016) and combined "potential" riparian zones (PRZ; locations where riparian zones could exist) with "observable" riparian zones (ORZ; locations where riparian zones are observed to exist) to produce an "actual" riparian zone for the year of 2023 (locations where PRZ and ORZ overlap; Vegetation Data Came from 2023 as we wanted the most up to date riparian maps possible; however Landcover data comes 2022 and our DEM uses data from 2000). This involved using river networks from open-source datasets and ground elevation and calculating the Normalized Difference Vegetation Index (NDVI), and the Topographic Wetness Index (TWI). Our end products will help partners understand where their riparian zones are most effective and where to target future conservation efforts. Overall, this project provides repeatable GIS-based methods which use freely available EO data to delineate riparian zones that can be used to guide future monitoring and decision-making.

3. Methodology

3.1 Data Acquisition

We selected two time periods for each study site: peak rainy season (April-October) for the North West Province, and the peak rainy season (August-September) for the Western Cape to identify the annual maximum NDVI and TWI. As the most recent land cover classification map of South Africa was published in 2022, we focused on post-2022 date ranges for our study sites. We used Landsat 9 Operational Land Imager 2 (OLI-2) and Sentinel-2 Multispectral Instrument (MSI) imagery to analyze vegetation health from January 2023- December 2023 (Table 1). For the Landsat data, we used Collection 2 Tier 1 Surface Reflectance because it is the highest quality data available that was preprocessed and atmospherically corrected by the USGS. We acquired the 2022 South African National Land Cover map (SANLC), which covered the entire country, and a geodatabase which included rivers, mapped wetlands and potential riparian area shapefiles from SANBI to analyze how accurately riparian zones were mapped (Table 2).

To acquire the necessary atmospherically corrected Sentinel-2 MSI (Level 2A) and Landsat 9 OLI-2 data, we used the Google Earth Engine JavaScript API (GEE) platform. We filtered the data for the year of 2023 with a five percent cloud cover threshold and filtered the images for the highest mean NDVI values (section 3.2.6) to best represent vegetation greenness patterns in the study area. Since a Digital Terrain Model (DTM) was not available, our team chose a DEM of 1 arc-second or approximately 30 meters derived from SRTM. To account for any edge-effects in future calculations, our team created a buffer around the study area. To match the resolution of the Sentinel-2 images, we resampled the DEM to a 10 meter resolution.

Table 1

Earth observations used in this project

Platform & Sensor	Processing Level	Date Range Used	Use Cases
Landsat 9 OLI-2	Level-2 SR Collection 2 Tier 1	January 2023 – December 2023	NDVI
Sentinel-2 MSI	Level-2A	January 2023 – December 2023	NDVI
NASA SRTM	N/A	2000	DEM, TWI

Table 2

Ancillary data used in this project

Name	Product Type	Data Type	Data Provider	Use Cases
The South African National Land Cover map (SANLC)	Land Use Land Cover map	Raster, 20 meters	Republic of South Africa: Department of Forestry, Fisheries, and the Environment	Water feature extraction and agriculture mask
River Polylines	Centerline Shapefile	Vector	SANBI	Water feature extraction
National Wetland Map 5 (2018)	Polygon	Vector	SANBI	Delineated riparian zone accuracy assessment
Potential Riparian Areas	Polygon	Vector	SANBI	Delineated riparian zone accuracy assessment

3.2 Data Processing

3.2.1 Water Feature Data

The water feature surface used in this study comprised two authoritative datasets. The 2022 geographic South African National Land Cover map (SANLC) had water body pixels extracted to create a water mask. We extracted pixel values that represented Natural Rivers, Natural Lakes, Natural Pans (flooded @ observation times), Artificial Dams and Canals, Herbaceous Wetlands, and Bare Riverbed Material. The mask excluded all other classes and non-freshwater water sources like lagoons and estuaries. These pixels were extracted to a new layer with the exact resolution of the parent dataset of 20 meters. Because of the highly variable climate in South Africa, many of the rivers contained little to no water during the SANLC classification campaign. Hence, the resulting dataset underrepresented the water bodies in South Africa. To counter this effect, SANBI provided a shapefile containing all rivers in South Africa. Rivers are highly dynamic and disturbance-prone; they dance, change course, dry up, and flood. Because of this, and for ease of analysis, a 30-meter buffer was applied to the rivers to approximate their location better and rasterize them in the same step. The shapefile was rasterized to the same spatial resolution as the SANLC and assigned a 30-meter pixel width. The extracted water feature raster was merged with this river raster, creating a raster composite of water bodies and features that more accurately represent river networks in more arid environments in ArcGIS Pro.

3.2.2 Agriculture Mask

To create a more accurate ORZ, areas that were classified as any type of agriculture or cultivated lands in the 2022 geographic South African National Land Cover map (SANLC) were removed. Agriculture is often adjacent to riparian vegetation and can be confused with ORV when using only NDVI, so it was important to mask out these pixels to ensure they were not included in any final riparian layers.

3.2.3 Calculating Total Wetness Index

Because riparian vegetation is water dependent and usually river adjacent (Huylenbroeck et al., 2020), we calculated topographic wetness index (TWI) to determine where water could potentially accumulate (Eq. 1, Winzeler, 2022). In equation 1, A represents the catchment portion of the land (area upstream of a land unit from which water flows) while B represents the slope of the land. The geoprocessing tools fill (to fill sinks), flow direction, flow accumulation, and slope were run on the SRTM DEM. We then used the Raster Calculator function in ArcGIS Pro to calculate TWI. To ensure accurate calculations and avoid null values, a small constant of 0.00001 was added to all slope values. A subjective threshold of 0.3 was chosen based on visual inspection.

$$TWI = \ln \left[\frac{A}{\tan B} \right] \quad (1)$$

3.2.4 NDVI Vegetation Threshold

An extensively used metric for assessing the density and health of vegetation is NDVI. To exclude all non-vegetated areas in the analysis, our team calculated NDVI on the Sentinel-2 and Landsat 9 imagery using the near-infrared and red bands (Eq. 2; Przyborski, 2000). A low NDVI threshold of 0.3 was selected to minimize the risk of inadvertently excluding vegetation.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2)$$

3.3 Data Analysis

3.3.1 Delineation of Groundwater-dependent Ecosystems (GWDE)

Raster cells that met both threshold criteria for NDVI and TWI, were likely to be Ground-Water Dependent Ecosystems (GWDE). The structure, composition, and overall functioning of groundwater-dependent ecosystems are primarily maintained by groundwater, which provides all or part of the water needed for them (Huylenbroeck et al., 2020). This identification assumed that areas that satisfied both predefined threshold criteria for NDVI and TWI values, are more likely to be influenced by groundwater. These areas were extracted for further analysis.

3.3.2 Classifying Riparian Vegetation

Initially, sample points to use for classification were selected from iNaturalist observations. Our team only used “research-grade” observations (observations that were agreed upon by the iNaturalist community) as there was a consensus on the species identification. All vegetation points within the study sites in 2023, both native and invasive. However, points often overlapped or were in close proximity to each other. Since our team had limited knowledge about South African vegetation, we did not feel comfortable doing a supervised classification. Instead, an unsupervised classification was executed on the areas identified as GWDEs. The "Iso Cluster Unsupervised Classification" in ArcGIS Pro was utilized as the main classifier for this process. We used spectral reflectance, standard deviation, rectangularity, and compactness as segment attributes.

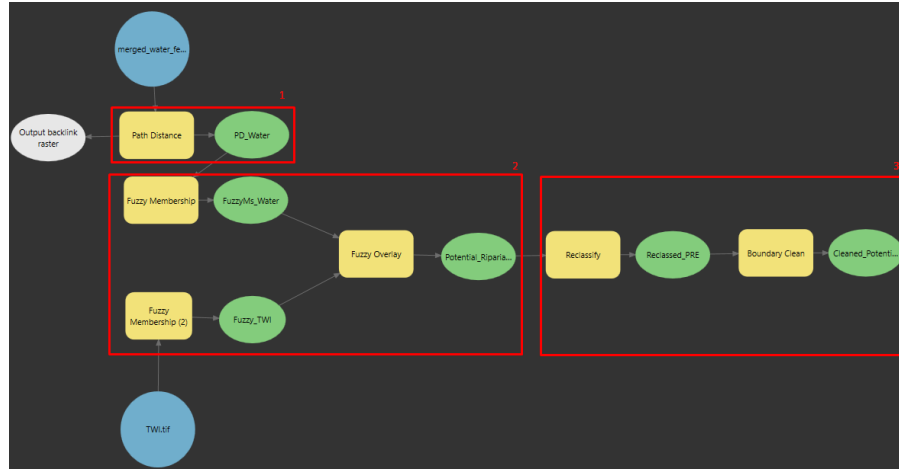


Figure 2. Shows the simple method for calculating PRZ including performing a path distance analysis (Section 1), combining this analysis with the TWI dataset (Section 2), and finally reclassifying the data as discrete values to differentiate areas of overlap (Section 3). The output shows the PRZ.

3.3.3 Potential Riparian Zone (PRZ) Calculation

In essence, the PRZ looks at the interactions of water features within a physical landscape and has two inputs: water features and a TWI. Superficial characteristics like distance from water features and slope can provide enough context to rule out areas for PRZ. This is important since riparian zones are distinct from upland vegetation communities, often separated by a dramatic change in slope (Naiman et al., 2005).

Section 1 in figure 2 shows a path distance analysis, which searches for areas close to water features until a fixed distance is reached. Path distance is similar to Euclidian distance (straight line), but it differs when considering elevation surface straight line distance. ArcGIS Pro's Path distance tool allows users to input a cost surface, reducing or blocking the tool from searching further. In this scenario, the slope positions itself perfectly as a cost surface and limits the search distance when a significant slope value is encountered. The combination of the two properties of the tool had the unintended effect of mapping dry or extinct riverbeds where no water features were located while ignoring ridgelines and bluffs. This had a similar effect as a floodplain mapping model, another critical indicator of riparian zones (Fernández et al., 2012).

Section 2 combines the two datasets, the TWI created previously and the path distance to water features created just above. Because riparian zones have nondescript boundary lines, it is fitting to use a nondescript method for calculating them. Fuzzy Logic is a nondescript way to help categorize data between two different datasets. It has two major parts: Membership and Overlay. Fuzzy membership allows pixel values of all sizes to be remapped to values between 0 and 1, where favorable data is assigned numbers closer to 1, and unfavorable data is assigned numbers closer to 0. This nondescript way of assigning numbers accounts for the gradual transitions within the PRZ and helps provide a nuanced understanding of the landscape's likelihood of being a PRZ. For the path distance to water features, smaller numbers signified less distance from water features and were used to assign membership. Specifically, in ArcGIS Pro's Fuzzy membership tool, it was assigned MSSmall and membership mean of 0.1, with a standard deviation of 1. Regarding the TWI, more extensive features represented more flow accumulation and larger catchment areas, and so had larger values signify membership. TWI was assigned an MSLarge and membership mean and standard deviation of 1.

Fuzzy Overlay combines fuzzy membership layers to make another layer whose values represent a composite of the input membership layers, with values closer to 1 more likely to be a PRZ. The two resulting fuzzy membership datasets were combined using ArcGIS Pro's fuzzy overlay tool with an overlay type of sum. The sum is not to be confused with the additive sum. This overlay type of sum heavily favors areas with high pixel value overlap and is based on a linear combination of the input features unique data distributions. This analysis favored areas closer to water with a large flow accumulation-drainage basin.

Section 3 aimed at refining the PRZ so that areas of little or no overlap between fuzzy membership layers were removed. These values will vary from analysis to analysis because of the nondescript nature of fuzzy logic. While values ranging from 0 –1 are intuitive, they still include areas of little or no agreement between fuzzy membership layers. Removing these non-overlapping areas can be done by reclassifying pixel value ranges into NODATA and integers. The range from 0 to 0.8 was given NODATA, 0.8 to 0.9 0, and 0.9 to 1 in the analysis. This ensures that only the areas with the highest degrees of overlap could be PRZ. This method of discrete boundaries on the data produces holes and islands and was hard to use in further analysis, so a continuous surface was needed. Cleaning the boundary of the raster removed voids in the data and helped smooth out jagged edges. This was achieved using the boundary clean tool in ArcGIS Pro.

3.3.4 *Actual Riparian Vegetation (ARV) Calculation*

Actual Riparian Vegetation is determined by the overlap of ORZ with PRZ and is done with conditional statements; ORZ within PRZ is ARV. The ARV represents vegetation by considering slope, proximity to water features, flow accumulation, drainage area size, and NDVI. The resulting vegetation resembles the biophysical characteristics of true riparian vegetation.

3.3.5 *Accuracy Assessment & Confusion Matrix*

In GEE, we performed an unsupervised classification on areas that only were identified as ARV that could potentially contain riparian landcover. As the 2022 geographic South African National Land Cover map (SANLC) identified 5 types of Landcover which could contain riparian vegetation (including contiguous low forest & thicket, dense forest & woodlands, open woodland, herbaceous woodland, and lowland shrubland), we chose to use 5 classes for the plant communities could contain riparian vegetation. As this was an unsupervised classification and we had no on the ground data to compare to, we named these classifications 'Riparian 1', 'Riparian 2', 'Riparian 3', 'Riparian 4', & 'Riparian 5' in lieu of more specific identification (Figure 4). We performed an accuracy assessment of the potential riparian vegetation map using the potential riparian vegetation map to showcase how well the unsupervised classifier delineated the 5 classes. The assessment used stratified random sampling, with 2000 points in the sample, making sure each of the 5 classes was represented equally and robustly in the sample equally (Table 1).

4. Results and Discussions

4.1 *Potential Riparian Zone*

4.1.1 *Analysis of Results*

The resulting methodology produced just over 6 million hectares of potential riparian zones across both study sites. Area in hectares was achieved by equation 3 below, with pixels mapped in meters:

$$Area (Ha) = \frac{(Pixel Side Length^2)}{10,000} \quad (3)$$

The total PRZs classified encompass nearly half the 12,579,304-hectare study area, making up 48% of the total area between both study sites. Of those 12.5 million hectares of the study area, 9.79% were considered high confidence PRZ. Of the high confidence PRZ, 6.09% were above the 0.9 threshold, and 3.70 were below 0.9 and above 0.8. The PRZ > 0.9 was 12.7% of the total PRZ area classified, and the PRZ > 0.8 but < 0.9 made up 7.71 percent of the total PRZ area classified. Statistics were not calculated for any PRZ values below 0.8, which comprised 79.59 percent of the remaining PRZ classified (Table 3).

Table 3

Simple data analysis ran on the classified PRZ

Zone Type	Pixel Count (px)	Area (Ha)	Percent of Study area (12,579,304 Ha)	Percent of Total PRZ area
Total PRZ classified	67,113,997	6,040,259.73	48.01	
Highest Confidence PRZ > 0.9	8,525,459	767,291.31	6.09	12.7
High Confidence PRZ > 0.8	5,178,279	466,045.11	3.70	7.71

SANBI had conducted a preliminary assessment of PRZ using an aggregation of soil data, hydrological extents, hand-drawn boundaries from aerial images, and more. No campaign has validated these, so they cannot be used to validate our model. Instead, they were used to assess over/underfitting. In order to compare the spatial patterns of our PRZs with those of SANBIs, we used a simple proportion index to demonstrate if our model was under or overfit when compared to a reference PRZ. The equations for the Proportion Index are below where Equation 4 details how the proportion inside was calculated and Equation 5 details how the proportion outside was calculated:

$$\text{Proportion Inside} = \frac{\text{Count Inside}}{\text{Count Inside} + \text{Count Outside}} \quad (4)$$

$$\text{Proportion Outside} = \frac{\text{Count Outside}}{\text{Count Inside} + \text{Count Outside}} \quad (5)$$

The total area of PRZ, which agreed with SANBI's reference dataset, spanned 354,166.38 hectares. The majority of both >0.9 and > 0.8 confidence PRZs were located outside the reference dataset, 66.6 percent and 33.3 percent, respectively. Conversely, 33.3 percent of >0.9 PRZ and 21.1 percent of >0.8 PRZ were within the reference dataset (Table 4).

Table 4

Simple data analysis ran on the classified PRZ against reference PRZ data

Zone Type	Pixel Count	Area (Ha)	Proportion of PRZ inside Reference area (%/Ha)	Proportion of PRZ outside Reference area (%/Ha)
Highest Confidence PRZ > 0.9	8,525,459	767,291	33.3/255,815	66.6/511,476
High Confidence PRZ > 0.8	5,178,279	466,045	21.1/98,351	78.9/367,694

4.1.2 Discussion

In total, the applied method identified a little over 6 million hectares of land in which riparian zones could theoretically exist. When comparing SANBI's PRZ zones, our model appeared to overestimate PRZ. Also note that the reference dataset has never been validated. The proportion index for each PRZ class demonstrated that most PRZ pixels are not within the reference dataset for digital numbers > 0.8 and >0.9. This can be primarily attributed to the water features being classified as riparian. This issue relates to how the water feature layer was made. Large reservoirs and manmade lakes, along with natural lakes, are mistakenly classified as PRZ because of their high proximity to water and high flow accumulation. Removal of these zones should significantly reduce the outside reference area. However, this analysis was never run since the

reference data could not be verified. However, an agreement can be seen visually between the two, further supporting the hypothesis that outside pixel counts are obscured by prominent water features misclassified as PRZ.

Further analysis needs to ensure extensive water features (e.g., lakes, reservoirs, and artificial water bodies) are not classified as PRZ within themselves but instead around their edges. Secondly, further studies should be far more liberal with the thresholds to characterize PRZ and use a lower threshold like 0.5 or none. PRZ zones were critical in determining ARV, so they should be as broad as possible while still relevant to the water features they relate to. Our PRZ zones showed great promise in defining the theoretical boundaries of riparian ecosystems, but further analysis should incorporate flood zones, inundation levels, and historical flood heights to understand the hydrological regime better. Finally, and most importantly, future studies should not include TWI as a member of PRZ. TWI seemed to serve little purpose in calculating PRZ; it rarely connected PRZ zones and caused gaps in the final data product. Instead, we propose further studies that utilize wetness coefficients, as Weissteiner et al. (2016) have used. Several wetness coefficients have been derived using data from mainstream sensors, including MSI on Sentinel 2 (Nedkov, 2017); Such methods have been used to derive highly accurate vegetation characteristics in arid and semi-arid environments (Lamqadem et al., 2018). Even with the statistics pointing to our model overfitting PRZ, our model intended to do just that. Combined with other analyses, the proposed method provides a significant first step for narrowing in on riparian ecosystems within a landscape when given only water features and a DEM, which makes this method especially attractive option for aiding wetland land management & conservation planning.

4.2 Actual Riparian Vegetation

4.2.1 Analysis of Results

Table 5
ARV Statistics by Study Area

Study Area	Area (Ha)	Percent of Site-Specific GWDE	Percent of Total ARV
North West ARV	140967.18	20.80	96.84
Cape Town ARV	4597.92	45.86	3.16

Within the over 6 million hectares of identified PRZ across both study sites, 145,565 hectares were classified as ARV. Area in hectares was achieved by using equation 1, with pixels mapped in meters. It is important to note that this ARV is based on a GWDE figure that had urban development and agriculture removed. The North West province study area has 20.80% of its GWDE classified as ARV, which comprised most of the ARV classified, at 96.84%. Cape Town had 45.86% of its GWDE classified as ARV, which made up just 3.16% of the total ARV classified. Because the SANBI-provided PRZ zones had no relation to vegetation, they were not used to assess the ARV results.

Five classes were created for the unsupervised classification (Figure 4). Since, we are not familiar with South African vegetation, input was needed to correctly identify these riparian communities. Overall accuracy compared to reference data was high, at 97.3%. And while there was little confusion between classes (Table 6) and the accuracy (i.e., level of agreement with reference data) was very high, there are drawbacks to using this type of accuracy assessment. For example, there is no ground truth/validation data to compare the classification against, so it is only testing how well the classifier works, not necessarily if it is right.

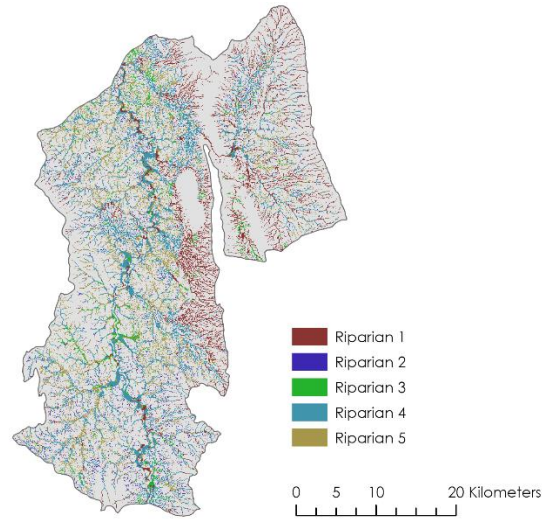


Figure 4. Unsupervised classification of riparian vegetation in the Study Site Northeast of Cape Town

Table 6

Confusion Matrix for unsupervised classification of potential riparian vegetation compared to available reference data

	Riparian 1	Riparian 2	Riparian 3	Riparian 4	Riparian 5
Riparian 1	1,006	0	13	3	8
Riparian 2	0	419	0	3	0
Riparian 3	19	0	946	7	13
Riparian 4	3	7	26	991	0
Riparian 5	14	0	8	0	974

4.3.2 Discussion

Due to the lack of validation data, the summary statistics provided little information about the accuracy of the ARV. When visually inspecting the data, many interesting spatial patterns were observed in the resulting ARV, indicating some major takeaways. Primarily, and most importantly, the proposed methods suggest that large-scale riparian zone classification can be conducted with freely available EO data and produce usable results. Figure 3 shows how the process can distinguish riparian vegetation from cultivated vegetation and how well it can navigate through a heavily cultivated area full of disturbances.



Figure 3. This figure shows ARV in blue and GWDE in yellow in the North West study site.

Figure 5 again demonstrates how well the method can distinguish between cultivated vegetation and vegetation likely to be riparian. However, this method was imperfect, and some drawbacks mainly related to the TWI. Visually examining the spatial patterns of the ARV showed that the TWI was far more effective at predicting ARV in the more mountainous Cape Town study site, as seen in Figure 5. TWI also played a dramatic role in determining where vegetation was. The first example is that water in perennial rivers with elevated NDVI values was classified as ARV, while the surrounding vegetation was not. Figure 6 is an example of this anomaly occurring. While this was the case for many downstream waterways, headwaters, and waterways before and after artificial water features seemed to be exempt from this phenomenon, as seen in Figure 6. Even though the aerial imagery was not taken at the same time the multispectral imagery was collected, these initial conclusions suggest that slower-moving waters might support more photosynthetic organisms, which could be included in the final GWDE. The second example of TWI incorrectly influencing where ARV was classified can be seen at the bottom left of Figure 6, where forested vegetation communities were not classified as ARV. There could be a handful of reasons, ranging from DEM's spatial resolution not capturing the slope accurately to the NDVI values of the vegetation not being high enough to meet the ARV threshold. We speculate, however, that it was related to the TWI in that area not breaching the threshold used to classify GWDE, pointing to the sporadic distributions of GWDE and, thus, ARV as evidence for the arbitrary nature of the incorporation of TWI. However, these issues seemed unique to the flatter areas and epically evident in North West province.

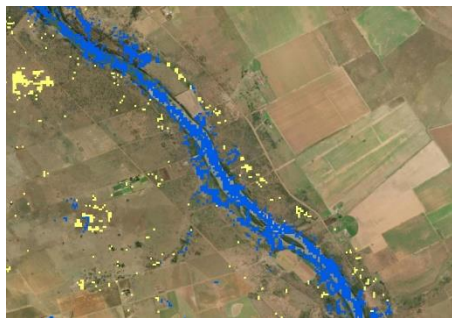


Figure 5. This shows ARV in blue and GWDE in yellow in the Cape Town Study Site.



Figure 6. This shows perennial waterways being classified as ARV in the NW Study Site.



Figure 7. This shows that faster-moving waterways avoided being classified as ARV in NW

While being sporadic, the method for classifying ARV narrows down the search of riparian vegetation within a landscape significantly, and initial visual comparison shows significant promise for the refinement of this method to produce accurate results. Specifically, improvements with the physical predictors of ARV mainly relate to the TWI index used. TWI shines in areas with more topographic relief. TWI produced GWDE with matched vegetation on the ground; however, when using a PRZ zone to search for vegetation, our analysis showed that TWI seemed to limit our ability to capture vegetation accurately as it narrowed vegetation communities too much. Further analysis should avoid using TWI because the dynamic nature of riparian zones doesn't tend to match the ridged predictions of flow accumulation that TWI predicts.

5. Conclusions

Our team used remotely sensed imagery from Landsat 9, Sentinel-2, and NASA SRTM to classify and detect riparian zones in South Africa. We developed a methodology using a two-pronged approach that utilized both physical and biological features characteristic of riparian areas. The parameters considered in our model are NDVI, TWI, and elevation. We produced maps of riparian zones in our two study sites using data from 2023, to provide the partners with information about the current extent of these ecosystems. Using these methods, we detected 145,565 hectares of actual riparian vegetation occurring on our two study sites.

We found it feasible to use remote sensing methods for our project's ecological conservation application. Our results suggest that the methods successfully detect riparian zones using data that was freely available to the public. This benefits our partners as it is much less expensive and labor-intensive than previously used methods of detecting these wetland areas. Further projects may consider using ground validation data to improve result accuracy and incorporate a temporal aspect to this analysis, as these ecosystems are disturbance prone.

The products delivered to our partners can be used to help assess current riparian ecosystem health and assist in planning and protecting these important ecological areas. The methodologies established by our project can be applied to other areas in South Africa to help create a cohesive map across the nation and increase accessibility to riparian extent data. Our products will also assist SANBI in contributing to compilation of the South African National Land Cover Map.

6. Acknowledgements

We would like to thank our partners Anisha Dayaram, Nancy Job, Dr. Andrew Skowno, and Dr. Cherie Forbes, for providing guidance with this project. Our sincere gratitude extends to Michael Pazmino, Jane Zugarek, Benjamin Holt, Dr. Arnaud Cerbelaud, and Dr. Kerry Cawse-Nicholson. We appreciate their insight, constructive feedback, time, and support.

This material contains modified Copernicus Sentinel data (2023), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract 80LARC23FA024.

7. Glossary

ARV – Actual Riparian Vegetation

DEM – Digital Elevation Model

DTM – Digital Terrain Model

Earth Observations – Satellites and sensors that collect information about the Earth’s physical, chemical, and biological systems over space and time.

Fuzzy membership – Scales numerical values from zero to one.

GEE – Google Earth Engine.

GIS – Geographic Information Systems

GWDE – Groundwater Dependent Ecosystems

MSI - Multispectral Instrument.

NASA SRTM – NASA Shuttle Radar Topography Mission

NDVI - Normalized Difference Vegetation Index

NDWI - Normalized Difference Vegetation Water Index

OLI - Operational Land Imager.

PRZ – Potential Riparian Zones

Riparian Zone - thin strips of land that occur along the edges of rivers, lakes, and other water bodies

TIRS - Thermal Infrared Sensor.

TWI - Topographic Wetness Index

8. References

- Copernicus Sentinel-2 (processed by ESA), 2021, MSI Level-2H/F Harmonized/Fused Reflectance Product. Collection 1. European Space Agency. <https://doi.org/10.57780/esa-4862e7e>
- Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81(2), 163–182. <https://doi.org/10.1017/S1464793105006950>
- Earth Resources Observation and Science (EROS) Center. (2013). *Collection-2 Landsat 8-9 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) Level-2 Science Products* [Data set]. U.S. Geological Survey. <https://doi.org/10.5066/P9OGBGM6>
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D.E. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45(2), RG2004. <https://doi.org/10.1029/2005RG000183>
- Fernández, D., Barquín, J., Álvarez-Cabria, M., & Peñas, F. J. (2012). Quantifying the performance of automated GIS-based geomorphological approaches for riparian zone delineation using digital elevation models. *Hydrology and Earth System Sciences*, 16(10), 3851–3862. <https://doi.org/10.5194/hess-16-3851-2012>
- Huylenbroeck, L., Laslier, M., Dufour, S., Georges, B., Lejeune, P., & Michez, A. (2020). Using remote sensing to characterize riparian vegetation: A review of available tools and perspectives for managers. *Journal of Environmental Management*, 267(110652), 1–18. <https://doi.org/10.1016/j.jenvman.2020.110652>
- Lamqadem, A. A., Saber, H., & Pradhan, B. (2018). Quantitative Assessment of Desertification in an Arid Oasis Using Remote Sensing Data and Spectral Index Techniques. *Remote Sensing*, 10(12), Article 12. <https://doi.org/10.3390/rs10121862>
- Naiman, R. J., Décamps, H., McClain, M. E., & Likens, G. E. (2005). 1—Introduction. *Riparia*, 1–18. <https://doi.org/10.1016/B978-012663315-3/50002-2>
- Nedkov, R. (2017). ORTHOGONAL TRANSFORMATION OF SEGMENTED IMAGES FROM THE SATELLITE SENTINEL-2. *Comptes Rendus de l'Académie Bulgare Des Sciences: Sciences Mathématiques et Naturelles*, 70, 687–692.
- Przyborski, P. (2000). Measuring Vegetation (NDVI & EVI). NASA Earth Observatory, 30, https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php
- Republic of South Africa Department of Forestry, Fisheries, and the Environment. (2022). South African National Land Cover (SANLC), 2022. [dataset]. <https://egis.environment.gov.za/sa-national-land-cover-datasets>
- Stutter, M., Baggaley, N., Ó hUallacháin, D., & Wang, C. (2021). The utility of spatial data to delineate river riparian functions and management zones: A review. *Science of The Total Environment*, 757(143982), 1–14. <https://doi.org/10.1016/j.scitotenv.2020.143982>
- Tockner, K., & Stanford, J. A. (2002). Riverine flood plains: Present state and future trends. *Environmental Conservation*, 29(3), 308–330. <https://doi.org/10.1017/S037689290200022X>

- Weissteiner, C. J., Ickerott, M., Ott, H., Probeck, M., Ramminger, G., Clerici, N., Dufourmont, H., & De Sousa, A. M. R. (2016). Europe's Green Arteries—A Continental Dataset of Riparian Zones. *Remote Sensing*, 8(11), 925. <https://doi.org/10.3390/rs8110925>
- Winzeler, H.E., Owens, P.R., Read, Q.D., Libohova, Z., Ashworth, A., & Sauer, T. (2022). Topographic Wetness Index as a Proxy for Soil Moisture in a Hillslope Catena: Flow Algorithms and Map Generalization. *Land*, 11(11), 2018. <https://doi.org/10.3390/land11112018>
- Yang, X. (2007). Integrated use of remote sensing and geographic information systems in riparian vegetation delineation and mapping. *International Journal of Remote Sensing*, 28(2), 353–370. <https://doi.org/10.1080/01431160600726763>