



Fall 2024

South Africa Ecological Conservation II

Identifying and Mapping Riparian Areas in South Africa with Earth Observations

DEVELOP Technical Report

November 22nd, 2024

Authors: Joseph Gibbs (Project Lead), Sera Tolgay, Anjan Rana Magar, Daniella Asturias (Analytical Mechanics Associates)

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)

Previous Contributors:

Mina Nada
Andrew Saah
Madison Elowitt
Katya Beener

Lead:

Michael Pazmino (California – JPL)

1. Abstract

This study, in collaboration with the South African National Biodiversity Institute (SANBI) and the Western Cape Government, used NASA Earth observations to develop a geospatial tool for estimating the extent of woody riparian vegetation across South Africa. SANBI currently lacks comprehensive GIS datasets on these wetlands, hindering effective conservation planning and management of these critical ecosystems. Our approach involved delineating potential riparian zones using path distance analysis and classifying woody vegetation within those zones using NDVI thresholds for each watershed area. This methodology was performed across the entire nation to obtain a national woody riparian vegetation map. Additionally, a more complex methodology centered around a Random Forest classification was performed on a few regions of interest. The results of this classification were very similar to the simpler NDVI threshold method. These results provide SANBI with an estimate of woody riparian vegetation distribution, supporting more informed conservation decisions. While limitations exist, the project enhances efforts to protect South Africa's riparian ecosystems and mitigate climate change impacts on biodiversity. The simplified methodology enabled timely completion of a nationwide dataset, though further refinements are possible with additional computational resources.

Key Terms

remote sensing, woody riparian vegetation, GIS, Earth observation, Random Forest classification, riparian zones, SANBI, land use change

2. Introduction

Riparian zones, the transitional areas between riverine and terrestrial ecosystems, are critical for global biodiversity. This is particularly true in water-scarce but biodiverse regions like South Africa. South Africa relies heavily on riparian ecosystems to maintain ecological balance and support both aquatic and terrestrial life. However, these ecosystems face mounting pressures from urbanization, deforestation, water diversion, and invasive species (O'Connor, 2010, Tockner & Stanford, 2002;). Their degradation threatens biodiversity and reduces vital ecosystem services such as flood regulation, water purification, and habitat provisioning, which are economically and ecologically vital (Riis et al., 2020). Addressing these challenges requires accurate mapping and a thorough understanding of riparian ecosystem distribution (González et al., 2017). Unfortunately, the lack of geospatial data on riparian zones in South Africa hinders effective conservation planning. Developing a precise geospatial inventory is essential to protect these ecosystems and their functions amid climate change and land-use transformations.

Due to the variability of riparian zones across different landscapes, it is often challenging to outline and classify them. Their highly dynamic nature is shaped by hydrological regimes, soil properties, and local topography, making it difficult to establish an exact and all-encompassing definition of riparian. Riparian zones are also often conflated with other ecosystems, such as wetlands, due to their proximity to water and similar vegetation types. As Naiman & Décamps (1997) observed, riparian zones can be defined specifically as ecotones—transitional areas where water, soil, and plant communities interact to form complex habitats. This transitional nature has led to a lack of consensus on their boundaries, which complicates efforts to conserve and manage these critical areas (Ward et al., 1999). The variability and intricate boundaries of riparian zones pose challenges for geospatial tools, making it difficult to accurately delineate and classify these areas (Sweeney et al., 2002).

In South Africa, the lack of comprehensive GIS-based datasets for riparian areas presents significant challenges for organizations like the South African National Biodiversity Institute (SANBI). SANBI plays a critical role in protecting the country's biodiversity by conducting research, maintaining biodiversity databases, and supporting conservation planning. However, the absence of detailed riparian mapping has limited SANBI's ability to identify, classify, and prioritize these ecosystems for protection. Currently, SANBI relies on limited field data and expert judgment for decision-making, which can be inconsistent and less effective without a comprehensive spatial framework. As a result, a more detailed and geospatially informed

understanding of riparian zones is essential for SANBI's work, particularly as climate and land use changes continue to threaten these ecosystems (Capon et al., 2013).

To address these challenges, our team partnered with SANBI to conduct a feasibility study aimed at improving the methods used to delineate and classify riparian zones. We also partnered with the Western Cape Government Department of Environmental Affairs and Developmental Planning, who provided us with feedback on our methodology and local expertise. Our primary objective was to determine if we could create a robust, scalable geospatial tool to estimate the extent of woody riparian vegetation across the entire nation using open-source Earth observation data. The use of open-source data is crucial in this context because it allows for widespread access, transparency, and the ability to build and refine the tool collaboratively, all while minimizing costs for SANBI and other stakeholders. These products will inform SANBI's conservation strategies, enabling them to better target conservation priorities and guide future land management decisions. This partnership will ultimately support SANBI's efforts to protect these critical ecosystems and help mitigate the impacts of climate change on South Africa's biodiversity.

We built upon the work of the summer 2024 NASA DEVELOP South Africa Ecological Conservation project, which developed a GIS-based protocol in partnership with SANBI to estimate the extent of riparian areas in two study sites: the Northwest Province and a smaller region northeast of Cape Town. Their work laid a critical foundation for delineating riparian areas, but the complexity and variability of South Africa's landscapes required a more comprehensive and scalable approach. Our team aimed to expand and refine their efforts by creating a model capable of delineating woody riparian vegetation across the entire country, providing SANBI with a comprehensive estimate of the distribution and location of these ecosystems nationwide.

Our methodology, informed by Weissteiner et al. (2016), involved generating a potential riparian zone layer based on proximity to river systems. We then clipped Harmonized Landsat and Sentinel-2 imagery to this riparian zone layer and performed a Random Forest classification to identify woody riparian vegetation. The results were further refined using a land cover map from SANBI to assess the likelihood of woody vegetation and the Normalized Difference Vegetation Index (NDVI) to filter out sparse vegetation. We applied this analysis to each of South Africa's 22 catchments individually, using imagery from April 2023 as a representative month for average annual precipitation (Figure 1). To reduce computational demand, we split the nation into these catchments, and the results were later combined into a national dataset. This new approach builds on previous work, supporting SANBI's goal of better identifying and conserving riparian ecosystems across South Africa.

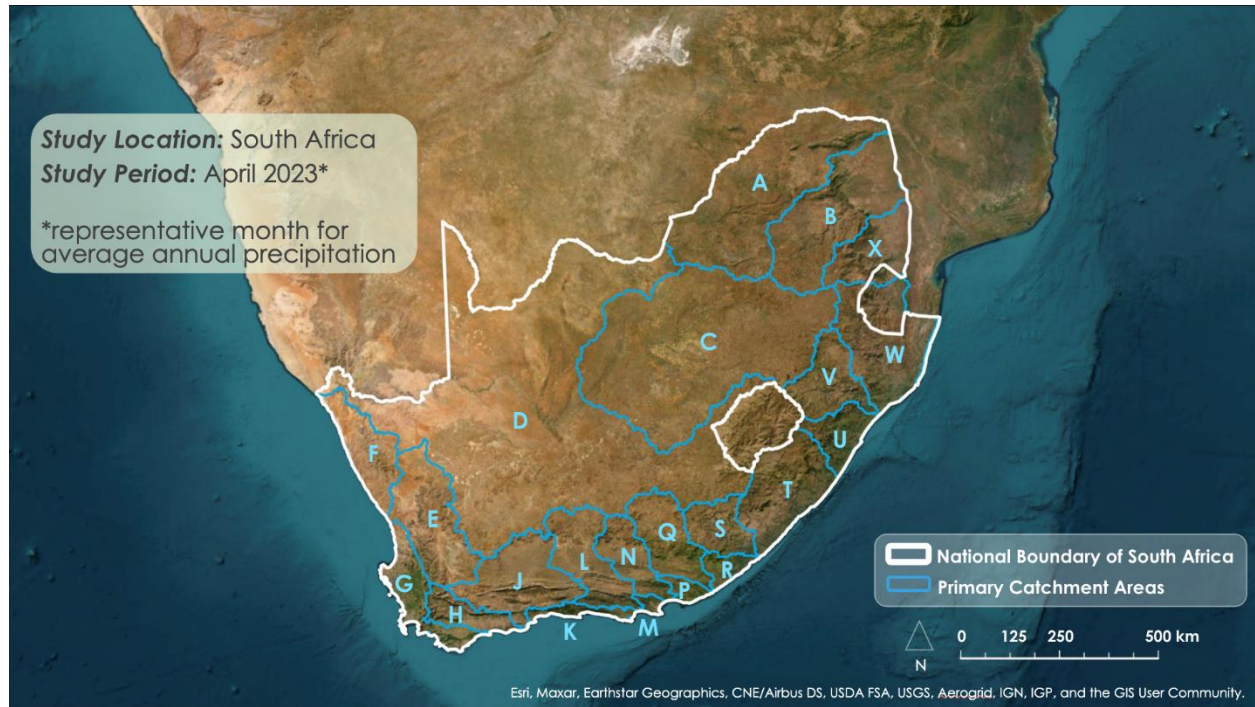


Figure 1. Study area map of South Africa divided into primary catchments.

3. Methodology

3.1 Data Acquisition

3.1.1 Satellite Imagery

To determine potential riparian zones, we first needed to establish the location of river systems across South Africa. SANBI provided us with their National Rivers Dataset (2018 version), which contained all major rivers but lacked smaller tributaries and streams. To address this limitation, we incorporated topographical data from NASA's Shuttle Radar Topography Mission (SRTM) Version 3 Digital Elevation Models (DEMs) to estimate the location of smaller water features (Table 1). We obtained the SRTM Level 2 products through NASA's Earth Data Portal, providing 30-meter resolution elevation data with vertical accuracy of ± 16 meters, collected by the Space Shuttle Endeavour during an 11-day mission in February 2000.

For our classification of woody riparian vegetation, we utilized Harmonized Landsat and Sentinel-2 (HLS) Level-2H surface reflectance products (Table 1). From the HLS 30-meter resolution dataset, we specifically selected bands 3 (green, $0.53\text{--}0.59\ \mu\text{m}$), 5 (NIR, $0.85\text{--}0.88\ \mu\text{m}$), and 7 (SWIR, $2.11\text{--}2.29\ \mu\text{m}$). Through comparative analysis, these bands demonstrated optimal performance in distinguishing woody vegetation within the potential riparian zone layer, outperforming other combinations including true color composites. We acquired the HLS data through NASA's Earth Data Portal, selecting only images with less than 5% cloud cover.

Table 1

Earth observations used in this study

Sensor & Platform	Spectral Bands & Features	Spatial Resolution	Processing Level	Date Range	Use
Harmonized Landsat 8 OLI & Landsat 9 OLI-2 Sentinel-2 MSI	HLS Band 3 (Green): $0.53\text{--}0.59\ \mu\text{m}$, HLS Band 5 (NIR): $0.85\text{--}0.88\ \mu\text{m}$	30 m	Level-2A Surface Reflectance	March – May 2023	Supervised Classification of Woody Riparian Vegetation

	HLS Band 7 (SWIR): 2.11–2.29 μm				
Shuttle Radar Topography Mission (SRTM)	Elevation (DEM)	30 m	Version 3 Vertical Accuracy ± 16 m	February 2000	DEM for Stream Order Analysis

The timing of imagery acquisition proved critical for accurate woody riparian vegetation detection due to South Africa’s pronounced seasonal precipitation patterns. These patterns create significant temporal variability in both vegetation density and surface water presence across the country’s diverse geographic regions. To minimize seasonal bias in our analysis, we selected imagery from April 2023, when the mean monthly precipitation (21.23 mm) closely matched the national annual average (20.6 mm) over the most recent 30-year climate average from 1991 to 2020. When imagery from April with less than 5% cloud coverage was not available, we also obtained tiles from March and May. This strategic timing helped ensure our analysis captured typical climatic conditions rather than seasonal extremes that could confound riparian zone identification and delineation.

3.1.2 Ancillary Data

Additionally, we used non-Earth observing data in our analysis. This included a Major Rivers shapefile provided by SANBI, and a South Africa Land Cover map and Primary Catchments Shapefile both obtained from the South African government, available via internet download (Table 2).

Table 2

Ancillary data used in this project

Name	Product Type	Data Type	Data Provider	Use Cases
The South African National Land Cover map	Land Use Land Cover map	Raster, 20 meters	Republic of South Africa: Department of Forestry, Fisheries, and the Environment	Water feature extraction, agriculture mask, likelihood analysis
Major Rivers’ (NBA2018MajorRivers)	Centerline Shapefile	Vector	SANBI	Used to estimate river systems for potential riparian zone
Primary Catchments	Polygon Shapefile	Vector	Republic of South Africa: Department of Water and Sanitation	Used to clip rivers and elevation data to specific catchments

3.2 Data Processing

3.2.1 DEM Processing and Clipping

We processed the SRTM DEM data using ENVI 6.0 image analysis software to create a seamless country-wide elevation model. The mosaicking process merged individual SRTM tiles while eliminating edge effects. Following the mosaic creation, we clipped both the DEM and the SANBI major river dataset to our study catchment boundaries using ArcGIS Pro 3.4. To prepare the DEM for hydrological analysis, we applied the Fill tool to correct terrain depressions and data voids that could impede accurate flow modeling.

3.2.2 Stream Order Analysis

To estimate and classify woody riparian vegetation across the country, we first had to determine the possible geographic location of riparian zones. Due to their definition as ecotones between riverine and terrestrial ecosystems (Naiman & Décamps, 1997), we determined potential riparian zones by using a combination of

the major rivers' dataset provided by SANBI (Table 2) as well as a stream order raster layer using the SRTM DEM. This stream order analysis estimates the location of smaller streams not included in the major rivers' dataset provided to us by SANBI and classify streams based on expected flow accumulation. Without including this stream order analysis, potential riparian zone wouldn't include many smaller riparian areas that exist along tributaries to larger rivers, leading to an underestimation of woody riparian vegetation.

We obtained flow accumulation and flow direction using the filled and clipped DEM. To exclude streams with low flow accumulation, we reclassified the flow accumulation values, setting a threshold to filter out smaller streams that were unlikely to be considered riparian zones. This step was necessary to prevent the misidentification of small streams that may have existed historically but are no longer present due to landscape alterations. Next, we performed a Strahler stream order analysis, which classifies streams based on their position in the drainage network. In this analysis, streams are assigned an order: the smallest streams are designated as first-order, and streams of higher orders are formed when smaller streams merge. Using the reclassified flow accumulation and direction layers, this analysis produced a raster stream order layer.

3.2.3 Buffer Creation

The stream order raster layer was converted to a vector polygon to enable buffer creation. Buffers were applied to streams based on their order, with larger rivers receiving wider buffers and smaller streams receiving narrower ones. The buffer sizes were determined by stream order, with each increase in stream order corresponding to a proportionally larger buffer. Specifically, the buffer size increased in a linear fashion, with first-order streams receiving a 50-meter buffer, second-order streams receiving a 100-meter buffer, and so on. Additionally, a 30-meter buffer was applied to the major rivers layer to account for their typically broader riparian zones.

To refine the potential riparian zones, the national river layer was processed into a raster using the Euclidean distance function and then merged with the stream order layer. The merged raster, which combined both the Euclidean distance and stream order features, was analyzed using the path distance tool to identify areas within 30 meters of these features. These potential riparian areas were reclassified and converted into a polygon for classification modeling. Finally, the buffer polygons from the stream order layer were merged with the path distance results to ensure comprehensive coverage of riparian zones, incorporating both small streams and major rivers.

3.2.4 Compositing Raster RGB Imagery

To enhance vegetation analysis, we created a composite raster using the green, near infrared, and short-wave infrared bands from the HLS imagery. This combination was selected because SWIR and NIR bands are highly sensitive to moisture and vegetation structure, helping distinguish vegetation types, while the green band enhances visual clarity for training purposes, making it easier to differentiate between woody and non-woody vegetation in the classification (Rahaman et al., 2017). After generating this composite, we clipped it to the potential riparian zone within the catchment area. This clipping step ensures that the analysis focuses solely on riparian areas directly adjacent to water bodies, excluding other regions within the catchment.

3.2.5 Reclassification of the South African National Land Cover map

Using the 2022 South African National Land Cover map (SANLC), we reclassified cultivated agricultural and urbanized areas into a separate layer to remove them from the woody riparian vegetation analysis. By isolating these human-developed land cover types, we avoid misclassification of woody vegetation, which can be misclassified in agricultural areas located adjacent to rivers and streams in particular. This layer is important to ensure that the woody riparian vegetation is identified correctly.

3.2.6 Random Forest Woody Riparian Vegetation Classification

Using the Random Forest classification algorithm in ArcGIS Pro, we applied a machine learning model trained with sample data representing both woody and non-woody areas from within the buffer zones in a smaller catchment area. This classification method segments the data by evaluating multiple decision trees,

enhancing the model's ability to identify woody vegetation, which is often mixed with other vegetation types in riparian zones. We created the training samples from field data collected in collaboration with SANBI, ensuring that the model learned patterns specific to woody vegetation in South Africa's diverse landscapes. This process enabled the model to accurately differentiate woody vegetation from other vegetation types, improving classification accuracy.

We clipped the classified output to the potential riparian zone generated in the previous step, restricting the analysis to areas directly adjacent to water bodies. Next, we excluded agricultural and urbanized areas using the reclassified 2022 SANLC, detailed in Section 3.2.5. This exclusion step is crucial, as agricultural and urban lands may contain vegetation that could mislead the classification, especially in edge areas where woody vegetation might intermingle with non-natural land uses. By isolating natural or semi-natural areas within the potential riparian zone, this step focuses the layer on ecologically significant woody vegetation that is of interest to SANBI.

The final step involved calculating NDVI (Kriegler et al., 1969) using NIR and Red bands (Equation 1) from the HLS imagery and filtering out all values below 0.5 to ensure that only dense vegetation was captured, reducing noise from sparse or low-biomass plants that might otherwise be classified as woody. NDVI values above 0.5 typically indicate healthier, denser vegetation in terms of green leaf area or biomass (Tucker, 1979), which aligns well with the characteristics of woody vegetation in riparian settings. Finally, we combined the Random Forest classification and NDVI results into a single, cohesive layer using the Merge Rasters tool. This integration produced a refined map of woody riparian vegetation by incorporating both structural and spectral vegetation information, yielding a more accurate and ecologically meaningful representation of woody vegetation within riparian zones.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

3.2.7 Simplified Methodology

Due to time and processing power constraints, we simplified the methodology to create the woody riparian vegetation (WRV) map for the entirety of South Africa. This approach, which relied on NDVI thresholds rather than Random Forest classification, allowed for the completion of a nationwide assessment within computational and time constraints. Validation showed that results from this simplified approach closely matched those from the Random Forest classification method.

The simplified methodology filtered NDVI values within the potential riparian zone (PRZ). NDVI thresholds ranging from 0.4 to 0.65 were applied based on regional vegetation characteristics, with higher thresholds used in forested areas to distinguish woody riparian vegetation from other vegetation types. Following the masking of agricultural and urban areas (detailed in section 3.2.5), a likelihood analysis validated the results (section 3.3).

3.2.8 Surface Water Body Classification

In addition to classifying woody riparian vegetation, we classified surface water bodies by calculating the Normalized Difference Water Index (NDWI) of HLS and selecting values that represent water bodies. The NDWI uses the reflectance characteristics of water in the green and near-infrared bands, where water bodies typically have higher reflectance in the green band and lower reflectance in the near-infrared band (Equation 2). By using the full NDWI range from 0 to 1, we captured areas with a high likelihood of representing surface water, helping to differentiate water bodies from surrounding land cover (McFeeters, 1996). This classification will assist SANBI in identifying water features within the riparian zone, which are critical for accurately delineating riparian zones and supporting water-dependent vegetation analyses. In our expanded methodology (section 3.2.6), we used this classification to clip out water bodies. However, we noticed that

our NDVI thresholds were already removing surface water bodies. In order to streamline our methodology for the national dataset (section 3.2.7) we decided to remove this step.

$$NDWI = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}} \quad (2)$$

3.3 Data Analysis

3.3.1 Potential Riparian Zones - Path Distance Analysis

We estimated PRZs using ArcGIS Pro's Path Distance tool, applied to the major rivers' dataset. Prior to this, we ran the Euclidean distance tool on the Major Rivers Map with a 30-meter distance to rasterize it. The Path Distance tool calculates the minimum cumulative travel distance from water sources to each raster cell, considering both horizontal and vertical movement across the landscape. We chose path distance analysis over Euclidean distance because it accounts for topographical variations, which significantly affect the extent of riparian zones. The analysis produced distance values that reflected actual surface traversal costs, allowing for a more accurate, topographically informed delineation of potential riparian zones.

The path distance analysis generated potential riparian zones that adapted to local terrain characteristics: narrower zones in steep valleys and wider zones in flat landscapes. This approach yielded more accurate riparian zone boundaries compared to traditional fixed-width buffers, as it responded to natural topographic constraints. After completing the path distance calculation, we reclassified the output to isolate areas proximate to water features, converted the reclassified raster to a polygon format, and merged it with our buffer polygon to create the final potential riparian zone delineation.

3.3.2 Woody Riparian Vegetation Likelihood Analysis

To refine the accuracy of the woody riparian vegetation layer, we performed a likelihood analysis by intersecting the WRV results with woodland areas identified in the SANLC. This step served to validate and assign confidence levels to areas classified as woody vegetation. Specifically, areas classified as "woody" in both the Random Forest-NDVI layer and the SANLC woodland class were labeled as "High Confidence". This high-confidence category represents regions where multiple datasets consistently identify woody vegetation, making them highly reliable for riparian management and restoration purposes.

Conversely, we classified areas flagged as woody in the Random Forest-NDVI analysis but not identified as woodland in the SANLC as "Low Confidence." These areas might represent transitional or mixed vegetation zones where the presence of woody vegetation is less certain. By distinguishing between high and low-confidence areas, this step improves the layer's accuracy (rather than precision) and allows end users to prioritize high-confidence zones for critical ecological assessments, while also flagging low-confidence areas for further ground validation if needed.

3.3.3 Woody Riparian Zone Analysis

We clipped the NDVI raster layer to the PRZs found in section 3.2.2. NDVI values <0 often associated with non-vegetated surfaces such as water bodies, urban areas, or barren land have mostly been clipped out. Higher NDVI values (>0.5) are generally associated with dense, healthy vegetation, whereas lower values (<0) suggest bare soil, sparse vegetation, or non-vegetated areas. NDVI values <0 often associated with non-vegetated surfaces such as water bodies, urban areas, or barren land have mostly already been clipped out. NDVI thresholds for each province were selected to be above the mean value, taking into account the different distributions of NDVI values for each province and ground truthing using SANLC and aerial imagery. We graphed the distribution of all of the NDVI pixels for validation (section 4.2). Once we had the pixel count of areas with WRV we used equation 3 to find the total amount of land under riparian vegetation (Table B1).

$$\text{Area (km}^2\text{)} = (\text{Number of Pixels}) \times (30)^2 \div (1,000,000) \quad (3)$$

4. Results and Discussion

4.1 Potential Riparian Zone

4.1.1 Analysis of Results

As outlined in our methodology, PRZs are areas adjacent to water bodies that may support riparian ecosystems. These zones were identified using geospatial analysis, considering factors such as proximity to water, stream order, and topographical features. The PRZ layer served as a filter for identifying potential woody riparian vegetation, dramatically narrowing down the area we needed to classify. Our analysis revealed that PRZs cover approximately 120,801 km², or 9.9% of South Africa's total land area (Figure 2). As shown in Tables A1 and A2, PRZs are consistently distributed across hydrological and administrative divisions, with most catchments and provinces showing similar proportional coverage. This suggests a relatively uniform distribution of potentially riparian-bearing rivers and streams across South Africa's diverse landscape.

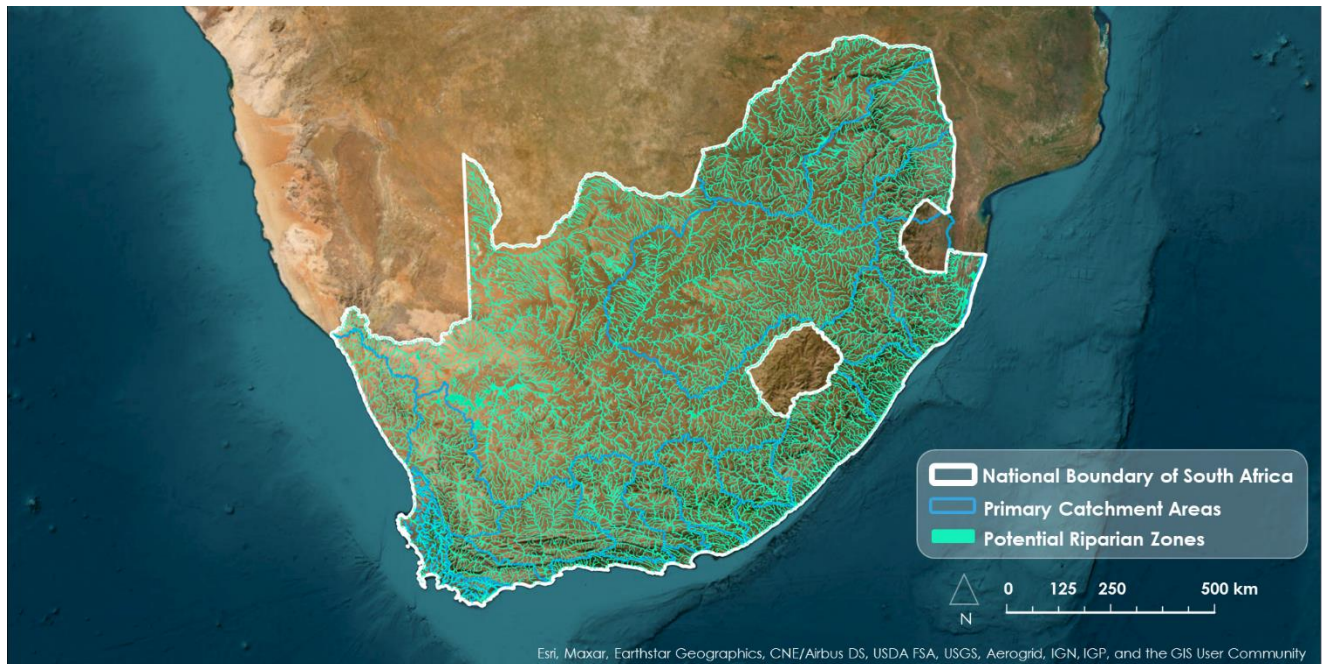


Figure 2. PRZs identified across South Africa

4.1.2 Errors & Uncertainties

The PRZ methodology contains several sources of uncertainty that could have led to over- or under-estimation of woody riparian vegetation later in our methodology. The stream order analysis used for PRZ delineation is based purely on elevation data and may have captured historical or ephemeral streams that no longer support active riparian ecosystems or have been substantially modified by human intervention. Additionally, given the complexity of mapping riparian zones across an entire nation, it was impossible to verify the accuracy of PRZ boundaries at every location, introducing inherent limitations to the comprehensive mapping approach. Accordingly, the methodology could have captured non-riparian vegetation if buffer zones were drawn too broadly, or conversely, excluded small but ecologically significant riparian areas if the zones were too narrow. The complexity of these uncertainties underscores the importance of treating the PRZ mapping as an approximation of geographic location rather than a definitive representation of riparian ecosystems.

4.2 Woody Riparian Vegetation

4.2.1 Analysis of Results

WRV is our estimate of the location of this type of vegetation within the PRZ layer. WRV was identified across the country using the simplified methodology explained in section 3.3.3, yielding 19,733 km², about

1.6% of South Africa's total surface area (Table B1). Out of this 19,733 km², almost 60% was deemed high confidence WRV based on the likelihood analysis performed using SANBI's national land cover map. Table B2 shows the provincial precipitation ranges and the corresponding WRV proportions (SAEON, 2020). Interestingly, although provinces that receive more rainfall on average generally had a greater amount of WRV identified than those that receive less, this was not always the case. For example, on average Mpumalanga receives the 2nd most precipitation annually, but less of the province was classified as WRV than in the case of other, more arid provinces.

For the Western Cape Province, a significant portion of pixels (57%) fall in the range of NDVI values from 0.1 to 0.3, indicating moderate vegetation cover and modified areas impacted by urbanization (Figure 6). Higher NDVI values (0.4–0.9) represent about 31% of the data, showing substantial areas of healthy vegetation. In the North West Province, vegetation showed higher NDVI ranges (0.4–0.7), with about 72% of the pixels in this category (Figure 6). Very high NDVI values (0.7–1) were sparse, suggesting fewer areas with extremely dense vegetation compared to the Western Cape. The Western Cape showed significant vegetation even in very high NDVI ranges (e.g., 0.8–1), which might correspond to riparian zones with robust vegetation. In the North West, extremely high NDVI values (>0.8) were rarely observed, potentially reflecting a less lush riparian environment.

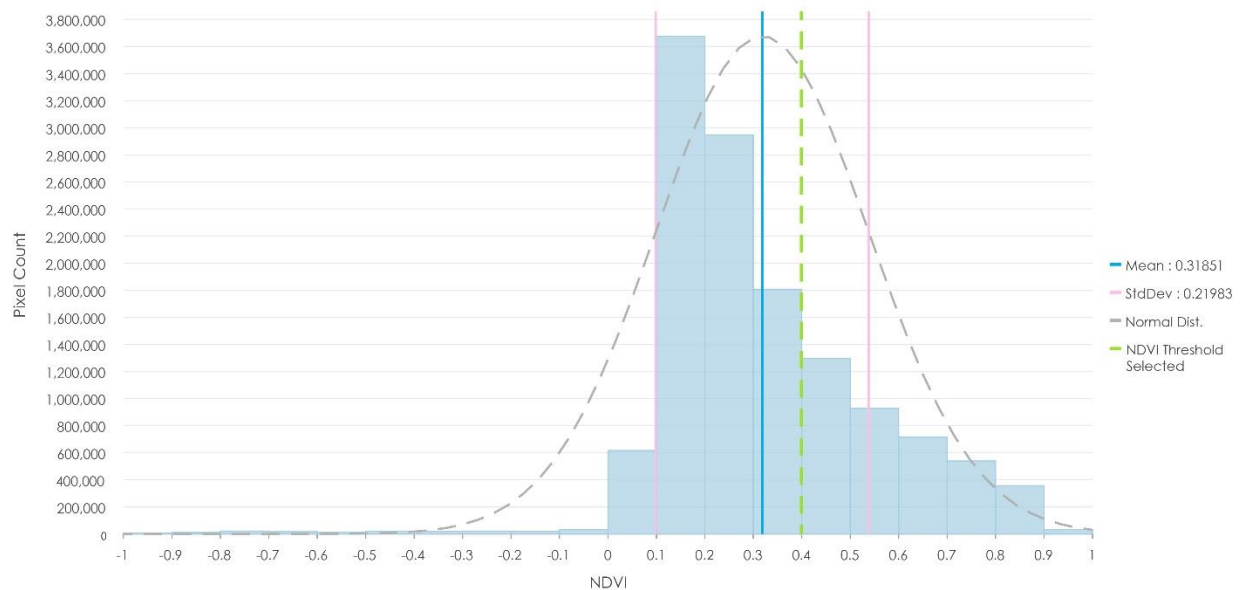


Figure 3. Western Cape Province Potential Riparian Zone Distribution of NDVI Values

The Western Cape's higher NDVI values in the 0.8–1 range could be due to a Mediterranean climate, with more rainfall supporting lush vegetation. The North West is drier and showed most pixels clustering in the 0.4–0.6 range, reflecting vegetation adapted to semi-arid conditions (Figure 7). Areas with lower NDVI values (<0.1) in both provinces, and nationwide, could be targets for riparian restoration efforts. The Western Cape's mix of high and low NDVI values in particular indicates areas with potential for improvement alongside naturally healthy zones. To supplement these findings, field surveys can be used to confirm vegetation health and identify specific stressors in low NDVI areas.

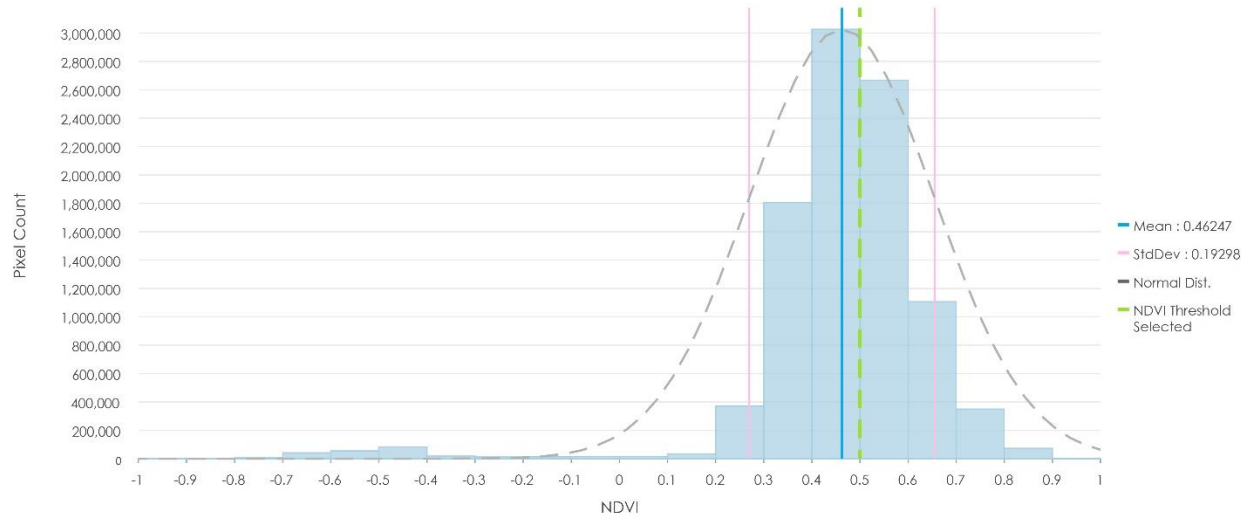


Figure 4. North West Province Potential Riparian Zone Distribution of NDVI Values

4.2.2 Errors & Uncertainties

We encountered several challenges and limitations while using the supervised classification-based methodology for identifying WRV that impacted the accuracy and precision of the results. One of the primary challenges was the moderate 30-meter resolution of HLS imagery, which limited our ability to capture narrow riparian zones. Riparian zones can sometimes be as narrow as just two or three trees wide, and the coarse resolution of the imagery may have missed these smaller riparian areas, especially in fragmented or narrow corridors. While the methodology performed well for larger riparian zones, it sometimes struggled to detect finer-scale features that are crucial for biodiversity preservation and effective conservation management.

Another significant difficulty arose from the training data used for the Random Forest classification. This data, created by our team using satellite imagery, introduced potential uncertainties due to our limited expertise in identifying riparian vegetation. Although we validated the results with the SANLC map, the classification remained approximate. The uncertainty in the training data, compounded by our inexperience with riparian vegetation types in South Africa, led to some misclassification, which may have prevented the model from fully capturing the complexity of riparian vegetation across different regions.

Additional complexities arose from South Africa's diverse precipitation patterns, which have a direct impact on vegetation dynamics and, consequently, the classification accuracy. Due to data constraints, we used imagery from March to May 2023 instead of April 2023, which is most representative of the country's annual average precipitation. This timing introduced variability, as the rainy season differs dramatically across South Africa's climatic regions. In some areas, the imagery corresponded to the start of the rainy season, while in others, it marked the end. This resulted in regions towards the end of the rainy season likely displaying more abundant vegetation, potentially leading to an overestimation of woody riparian vegetation compared to other areas where the rainy season was just beginning. This seasonal effect impacted both the supervised classification and simplified methods and highlights the complexity of using remote sensing data to monitor vegetation in regions with such diverse rainfall patterns.

Finally, the use of NDVI thresholds in the simplified methodology introduced challenges in densely vegetated regions, such as Limpopo. While NDVI is effective for distinguishing vegetated from non-vegetated areas, it faces limitations in regions with dense, homogeneous vegetation. In areas like Limpopo, the uniform NDVI values within PRZs made it difficult to differentiate woody riparian vegetation from the surrounding tree cover. This limitation is particularly evident in forested areas where the NDVI values for both riparian and non-riparian vegetation closely resemble one another. This challenge emphasizes the need for higher-

resolution data or more advanced methods to improve the accuracy of riparian zone delineation in densely vegetated regions.

5. Conclusion

5.1 Interpretation of Results

This study successfully demonstrates the potential of using Earth observation data to map and estimate the distribution of WRV across South Africa. By combining open-source satellite imagery, such as HLS, with sophisticated geospatial analysis techniques, we were able to identify and classify riparian zones at a national scale. The results show a significant extent of potential riparian zones across the country, taking up around 9.9% of the nation's land area. Our simplified methodology, based on NDVI thresholds, proved to be effective for providing nationwide coverage while remaining computationally feasible. However, the methodology encountered challenges such as resolution limitations and variability in vegetation due to seasonal effects. These factors underscore the need for more advanced approaches to improve accuracy in complex, densely vegetated regions.

5.2 Feasibility & Partner Implementation

The methods developed during this project are feasible and effective for use by SANBI and the Western Cape Government in their decision-making processes regarding riparian zone conservation. The PRZ methodology provides a reliable way to estimate the geographic locations of potential riparian zones, and it is reproducible across the entire nation. While the results are not without uncertainty, this project offers a solid foundation for future efforts. The simplified WRV classification using NDVI thresholds proved effective and feasible for use by our partners, although the methodology contains some uncertainty as described above. WRV provides valuable insights for SANBI in monitoring and conserving critical riparian ecosystems across South Africa. By identifying and classifying WRV at a national scale, SANBI can better assess the health of riparian zones and prioritize conservation efforts, contributing to the protection of biodiversity and water resources. Moving forward, SANBI and the Western Cape Government can continue to refine these methods, incorporating more field data, conducting ground validation, and possibly integrating additional temporal datasets to enhance the accuracy and utility of the riparian zone maps.

6. Acknowledgements

We would like to express our sincere thanks to our partner organizations, including (SANBI and the Western Cape Government Department of Environmental Affairs and Developmental Planning, particularly Nancy Job and Anisha Dayaram. Special thanks to our science advisors, Dr. Arnaud Cerbelaud, Dr. Kerry Cawse-Nicholson, and Benjamin M. Holt for their valuable guidance. We also appreciate the leadership and direction provided by our project lead, Michael Pazmino.

Contains public sector information licensed under the Open Government License v3.0.

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

Earth Observation (EO) – The process of collecting information about the Earth's surface using satellite sensors, aircraft, and other technologies. EO data are crucial for environmental monitoring, including land use, vegetation, and climate patterns.

Geospatial Analysis – The process of using geographic data to analyze spatial patterns and relationships between features on the Earth's surface. This can include mapping, modeling, and analyzing data related to land cover, vegetation, and other environmental factors.

Harmonized Landsat Sentinel-2 (HLS) – A dataset that combines Landsat OLI and Sentinel-2 satellite imagery to provide high-quality observations of the Earth's surface. This data is used for monitoring land cover, vegetation, and environmental changes at a resolution of 30 meters.

Multi-Temporal Imagery – Satellite imagery collected at different times to capture changes in the Earth's surface over a period. This is essential for understanding dynamic features like vegetation growth, seasonal changes, and the impact of human activities.

Normalized Difference Vegetation Index (NDVI) – A remote sensing index that measures vegetation health based on the difference between near-infrared and red-light reflectance. NDVI values range from -1 to +1, with higher values indicating healthier and denser vegetation.

Normalized Difference Water Index (NDWI) – An index similar to NDVI but focused on identifying water bodies by measuring the difference in reflectance between the near-infrared and shortwave infrared bands of light. NDWI is commonly used to map water bodies and distinguish them from surrounding land.

Path Distance Analysis – A spatial analysis technique that calculates the minimum travel distance between each raster cell and a specified feature, such as a river. This method accounts for topography and other landscape features, providing more accurate representations of areas like floodplains or riparian zones.

Potential Riparian Zones (PRZ) – Areas identified as likely to support riparian vegetation based on proximity to water bodies. PRZs are delineated using various geospatial analysis techniques to guide conservation efforts and identify regions that may require further ecological management.

Random Forest Classification – A machine learning technique that uses multiple decision trees to classify data. Each tree provides a "vote," and the most common outcome is chosen. This method is often used in remote sensing to classify land cover types like vegetation or water.

Riparian Zones – Transitional areas between riverine and terrestrial ecosystems, typically characterized by distinct vegetation that is adapted to periodic or permanent flooding. Riparian zones are crucial for biodiversity, water purification, and flood regulation.

Satellite Imagery – Images of the Earth's surface captured by satellites equipped with sensors. These images provide a comprehensive view of the landscape and are used for various environmental monitoring applications, including land use, vegetation mapping, and disaster management.

Stream Order – A classification system that ranks streams based on their size and the number of tributaries. First-order streams are the smallest, and higher-order streams are formed when smaller streams merge.

Woody Riparian Vegetation (WRV) – The type of vegetation in riparian zones composed of woody plants, including trees and shrubs, which are important for stabilizing riverbanks, providing wildlife habitat, and filtering water.

8. References

- Capon, S. J., Chambers, L. E., Mac Nally, R., Naiman, R. J., Davies, P., Marshall, N., Pittock, J., Reid, M., Capon, T., Douglas, M., Catford, J., Baldwin, D. S., Stewardson, M., Roberts, J., Parsons, M., & Williams, S. E. (2013). Riparian Ecosystems in the 21st Century: Hotspots for Climate Change Adaptation? *Ecosystems*, 16(3), 359–381. <https://doi.org/10.1007/s10021-013-9656-1>
- Department of Forestry, Fisheries, and the Environment. (2022). *South African National Land Cover (SANLC)*. Republic of South Africa. Retrieved from https://egis.environment.gov.za/gis_data_downloads
- 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. (2000). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2), RG2004. <https://doi.org/10.1029/2005RG000183>
- González, E., Felipe-Lucia, M. R., Bourgeois, B., Boz, B., Nilsson, C., Palmer, G., & Sher, A. A. (2017). Integrative conservation of riparian zones. *Biological Conservation*, 211, 20–29. <https://doi.org/10.1016/j.biocon.2016.10.035>
- Kriegler, F., Malila, W., Nalepka, R., & Richardson, W. (1969). Preprocessing transformations and their effect on multispectral recognition. Proceedings of the 6th International Symposium on Remote Sensing of Environment. Ann Arbor, MI: University of Michigan, 97-131.
- Mcfeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Naiman, R. J., & Décamps, H. (1997). The ecology of interfaces: Riparian zones. *Annual Review of Ecology, Evolution, and Systematics*, 28, 621–658. <https://doi.org/10.1146/annurev.ecolsys.28.1.621>
- National Aeronautics and Space Administration (NASA), United States Geological Survey (USGS), & European Commission. (2023). Harmonized Landsat Sentinel-2 (HLS) Products (Version 2.0). NASA Land Processes Distributed Active Archive Center. <https://doi.org/10.5067/HLS/HLSL30.002>
- National Aeronautics and Space Administration (NASA) & United States Geological Survey (USGS). (2013). NASA SRTM Digital Elevation 30m (Version 3) [Data set]. NASA Earth Data Search. <https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL30.003>
- O'Connor, T. G. (2010). Transformation of riparian forest to woodland in Mapungubwe National Park, South Africa, between 1990 and 2007. *Austral Ecology*, 35(7), 778–786. <https://doi.org/10.1111/j.1442-9993.2009.02084.x>
- Rahaman, K. R., Hassan, Q. K., & Ahmed, M. R. (2017). Pan-Sharpening of Landsat-8 Images and Its Application in Calculating Vegetation Greenness and Canopy Water Contents. *ISPRS International Journal of Geo-information*, 6(6), 168. <https://doi.org/10.3390/ijgi6060168>
- Riis, T., Kelly-Quinn, M., Aguiar, F. C., Manolaki, P., Bruno, D., Bejarano, M. D., Clerici, N., Fernandes, M. R., Franco, J. C., Pettit, N., Portela, A. P., Tammeorg, O., Tammeorg, P., Rodríguez-González, P. M., & Dufour, S. (2020). Global overview of ecosystem services provided by riparian vegetation. *BioScience*, 70(6), 501–514. <https://doi.org/10.1093/biosci/biaa041>

- Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., ... Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154-172. <https://doi.org/10.1016/j.rse.2014.02.001>
- South African Environmental Observation Network (SAEON). (2020). Mean annual precipitation across South African municipalities from 1983-2020. [Dataset]. African GeoPortal. <https://www.africageoportal.com/datasets/NRF-SAEON::mean-annual-precipitation-across-south-african-municipalities-from-1983-2020/explore?location=-29.186131%2C28.019830%2C6.34>
- Sweeney, B. W., Czapka, S. J., & Yerkes, T. (2002). Riparian forest restoration: Increasing success by reducing plant competition and herbivory. *Restoration Ecology*, 10(2), 392–400. <https://doi.org/10.1046/j.1526-100X.2002.02036.x>
- 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>
- Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127-150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0).
- Ward, J. V., Tockner, K., & Schiemer, F. (1999). Biodiversity of floodplain river ecosystems: Ecotones and connectivity. *Regulated Rivers: Research & Management*, 15(1–3), 125–139. [https://doi.org/10.1002/\(SICI\)1099-1646\(199901/06\)15:1/3<125::AID-RRR523>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-1646(199901/06)15:1/3<125::AID-RRR523>3.0.CO;2-E)
- 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>

9. Appendix

Appendix A: Potential Riparian Zone Surface Analysis

Table A1

Potential Riparian Zone Surface Area Analysis (by catchment)

Catchment*	PRZ Area (km²)	Percent of Total PRZ	Catchment Area (km²)	Percent of Catchment
A	9,433	7.81%	109,581	8.61%
B	6,650	5.50%	73,516	9.05%
C	19,105	15.82%	196,299	9.73%
D	38,905	32.21%	409,420	9.50%
E	4,829	4.00%	49,065	9.84%
F	2,095	1.73%	28,586	7.33%
G	2,268	1.88%	25,299	8.96%
H	1,442	1.19%	15,520	9.29%
J	4,234	3.51%	45,134	9.38%
K	530	0.44%	7,216	7.35%
L	4,523	3.74%	34,730	13.02%
M	180	0.15%	2,628	6.86%
N	2,186	1.81%	21,225	10.30%
P	551	0.46%	5,357	10.29%
Q	3,330	2.76%	30,227	11.01%
R	808	0.67%	7,928	10.19%
S	2,317	1.92%	20,481	11.31%
T	5,093	4.22%	46,623	10.92%
U	1,838	1.52%	18,310	10.04%
V	2,791	2.31%	29,040	9.61%
W	5,332	4.41%	59,006	9.04%
X	2,360	1.95%	31,155	7.58%

*Link to the South African Primary Drainage Regions (top-level catchments) dataset provided by the Republic of South Africa Department of Water and Sanitation:

<https://www.dws.gov.za/iwqs/wms/data/000key2data.asp>

Table A2

Potential Riparian Zone Surface Area Analysis (by province)

Province*	PRZ Area (km²)	Percent of Total PRZ	Province Area (km²)	Percent of Province
Eastern Cape	18,255	15.1%	168,966	10.8%
Free State	12,268	10.2%	129,825	9.45%
Gauteng	1,494	1.24%	18,178	8.22%
KwaZulu-Natal	10,220	8.46%	94,361	10.8%
Limpopo	11,205	9.28%	125,754	8.91%
Mpumalanga	6,613	5.47%	76,495	8.65%
North West	9,217	7.63%	104,882	8.79%
Notern Cape	38,863	32.2%	372,889	10.4%
Western Cape	12,664	10.5%	129,462	9.78%
South Africa	120,800.67	100%	1,220,813	9.90%

*Link to province dataset used in this study: <https://geo.btaa.org/catalog/061d4492-56e8-458c-a3fb-e7950991adf0> Provided by Humanitarian Data Exchange under Attribution 3.0 IGO.

Appendix B: Woody Riparian Vegetation Analysis

Table B1

Woody Riparian Vegetation Surface Area Analysis (by province)

Province	WRV Low Confidence Area (km ²)	WRV High Confidence Area (km ²)	WRV Total Area (km ²)	Percent High Confidence	Percent of Province (all WRV)
Eastern Cape	2,214	2,856	5,071	56.3%	3.00%
Free State	450	165	615	26.9%	0.47%
Gauteng	100	89	189	47.1%	1.04%
KwaZulu-Natal	1,030	2,147	3,177	67.6%	3.37%
Limpopo	187	2,066	2,253	91.7%	1.79%
Mpumalanga	154	632	785	80.5%	1.03%
North West	1,167	1,564	2,731	57.3%	2.60%
Nothern Cape	755	922	1,677	55.0%	0.45%
Western Cape	1,817	905	2,723	33.2%	2.10%
South Africa	7,975	11,758	19,733	59.6%	1.62%

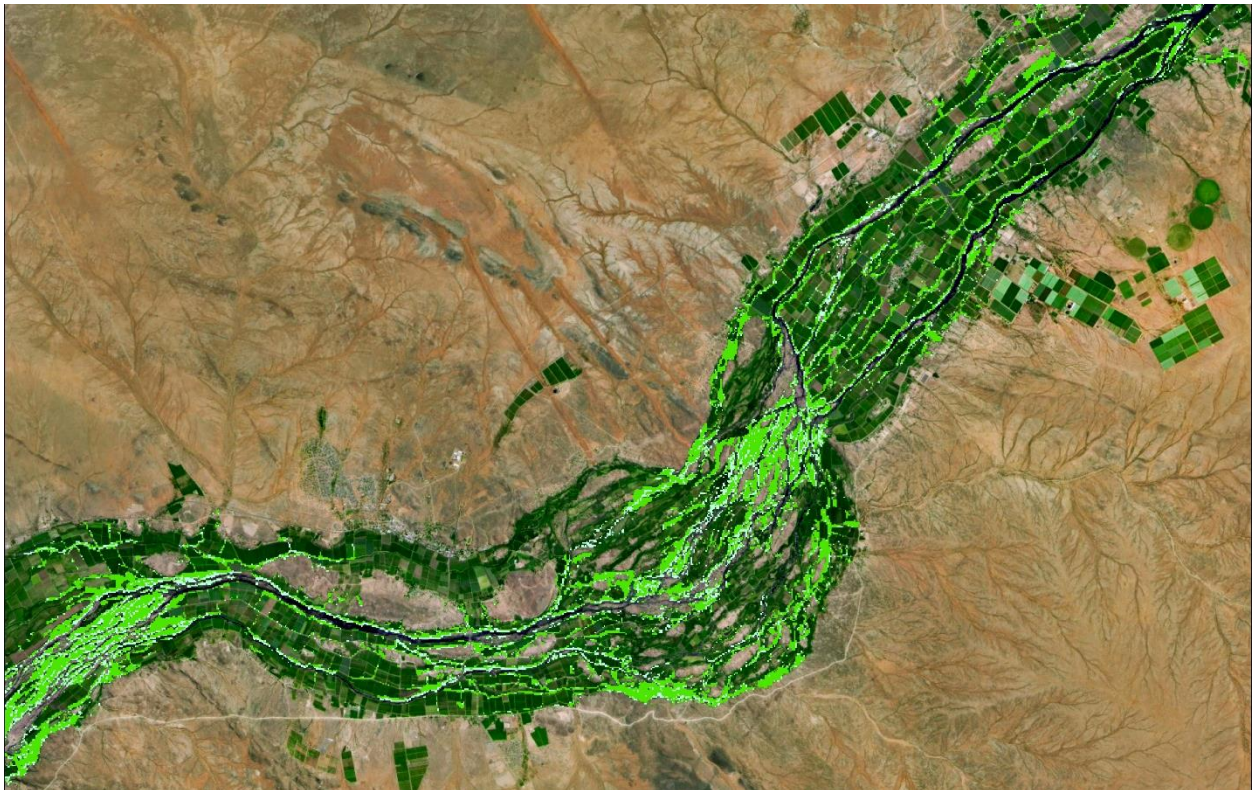


Figure B1. Results of Woody Riparian Vegetation classification, shown in the light green pixels (high likelihood in dark green, low likelihood in light green), over the Orange River riparian zone in Northern Cape Province



Figure B2. Results of Woody Riparian Vegetation classification, shown in the light green pixels (high likelihood in dark green, low likelihood in light green), in agricultural areas of North West Province



Figure B3. Results of Woody Riparian Vegetation classification, shown in the light green pixels (high likelihood in dark green, low likelihood in light green), around Brandvleidam and along Breerivier in Western Cape Province

Table B2

Woody Riparian Vegetation and Precipitation (by province)

Province	Percent of Province Surface Area (all WRV)	Average Precipitation Range (1983-2020)
Eastern Cape	3.00%	400–800 mm
Free State	0.47%	400–600 mm
Gauteng	1.04%	600–800 mm
KwaZulu-Natal	3.37%	800–1,200 mm
Limpopo	1.79%	400–800 mm
Mpumalanga	1.03%	700–1,200 mm
North West	2.60%	300–500 mm
Nothern Cape	0.45%	100–200 mm
Western Cape	2.10%	350–500 mm

Contains public sector information licensed under the Open Government License v3.0. For more information, please see the terms and conditions: <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>